Protein–DNA Interactions

Marianne Rooman, *Free University of Brussels, Brussels, Belgium*

René Wintjens, *Free University of Brussels, Brussels, Belgium*

The control of the information stored in the genome is managed by DNA-binding proteins, which are therefore of fundamental importance to cellular life. A wide diversity of such proteins, exhibiting various DNA-binding modes, specificities and functions, is observed in all living organisms.

Introductory article

Article Contents

doi: 10.1038/npg.els.0003906

Introduction

In recent years, significant progress has been made towards our understanding of the interactions between proteins and DNA, mainly due to advances in X-ray crystallography and nuclear magnetic resonance, which have allowed the determination and subsequent analysis of the three-dimensional structure of many protein–DNA complexes. These advances have led, for instance, to the observation that some proteins are able to find a single DNA site of 10–20 base pairs (bp) in a background of 10^6–10^9 bp. It has also been realized that proteins that rule gene expression by their binding to DNA are themselves coded for in the genome and must first be expressed before being able to regulate other genes. However, although the different mechanisms by which proteins recognize DNA are beginning to be elucidated, we are far from understanding the underlying principles that rule the complex biological processes of cellular life, in which protein–DNA interactions appear to be just a remarkable tool.

Roles of DNA-binding Proteins

DNA-binding proteins are involved in many fund amental cellular processes, such as control of gene expression, response to extracellular signals, homeostasis, cell growth and development, as well as cell division and differentiation. Some of them possess well-defined enzymatic activities. For example, RNA and DNA polymerases catalyse the addition of ribo- and deoxyribonucleotides to a polynucleotide chain during DNA transcription to form messenger RNA and during DNA replication, respectively; endonucleases cleave DNA or RNA chains often at very specific sites; recombinases are crucial to the recombination of single-stranded into double-stranded DNA.

The majority of DNA-binding proteins, however, are involved in the regulation of gene expression. The fraction of such proteins encoded by the human genome is estimated to be of the order of 5–10% of the total open reading frames. Regulatory proteins are extremely important as they are able to either block or favour the transcription of certain genes in some cell environments, at certain stages of development and under some external conditions. Their malfunctioning can cause many genetic disorders. A well-known example of transcriptional deregulation in *Drosophila* is caused by mutation in a DNA-binding protein called antennapedia homeodomain. Under normal circumstances this protein is expressed during the stage of development and ensures the correct leg formation. Following a mutation, however, it is expressed in antenna cells which are, as a consequence, transformed into legs. Other mutations cause the loss of the protein's activity and lead to the formation of antennae instead of legs.

Other DNA-binding proteins are only indirectly related to transcriptional regulation. Histones, DNA benders and some high mobility group proteins, for example, fulfil the function of packaging long DNA molecules into a compact form. However, a gene can be decoded only if it is accessible to the transcription proteins. The packaging DNA-binding proteins thus exert a negative control on transcription by not packaging the genes to be expressed. They also exert a positive control by locally bending and compacting the stretch of DNA situated between the promoter sequence of a given gene, which is bound by transcription factors, and sometimes distant upstream or downstream regions bound by activator proteins; this allows the activator and transcription proteins to interact, thereby favouring the initiation of transcription.

Note that some proteins that do not bind DNA yet influence transcriptional regulation, in a doubly indirect manner. Histone chaperones, for example, are acidic proteins that bind to free histones and hence prevent their improper or premature interaction with DNA.

Most of the known DNA-binding proteins bind mainly or exclusively to double-stranded DNA, and we will focus upon them. Note, however, that some proteins bind to single-stranded DNA and play important roles as accessory proteins in DNA replication, recombination and repair. For instance, specialized packaging proteins are used to stabilize single-stranded DNA before its recombination into double-stranded DNA.

Nature of Protein–DNA Interactions

The intermolecular forces that determine how proteins interact with DNA involve electrostatic interactions, hydrogen bonds sometimes mediated by water molecules, van der Waals interactions, hydrophobic forces, cation–π interactions between positively charged amino acids and nucleic acid bases, and π–π stacking between aromatic amino acids and nucleic acid bases. Electrostatic forces primarily rule the attraction between the positively charged protein surface and the negatively charged DNA phosphate backbone. Once protein and DNA have been brought together, the other forces, which are shorter range, become effective. These, and predominantly hydrogen bonds and cation–π interactions between amino acid side-chains and nucleic acid bases, determine the protein–DNA binding specificity. Note that these two types of interactions often occur concomitantly, with a positively charged amino acid side-chain linked through a hydrogen bond and a cation–π interaction with two successive nucleobases along the DNA stack. The stacking interactions between aromatic side-chains intercalated between successive bases are particularly important in proteins that bind single-stranded DNA, in which the nucleic acid bases are highly exposed to the solvent. In proteins binding duplex DNA, intercalation provokes DNA bending.

These forces all contribute to the overall free energy of protein–DNA association and determine the affinity of a protein for its DNA target. High affinity requires the exact structural and energetic complementarity of the protein and DNA contact surfaces. An exact fitting of these surfaces is often ensured by a residual flexibility in the DNA-binding protein domain or the DNA conformation. **Figure 1** shows, as an example, the complementarity of the contact surfaces of the O_r1 DNA operator and 434 Cro.

A quantitative measure of the affinity is given by the equilibrium binding constant, which corresponds to the ratio of the concentration of the protein–DNA complex to that of the free protein and of free DNA. The affinity of single DNA-binding domains for their DNA target sites is rather weak, with typical binding constant values in the 10^{-7}–10^{-9} mol L^{-1} range. In general, however, this relatively low affinity is increased *in vivo* by interactions between several domains and their cooperative binding to DNA.

The affinity of a protein for DNA is sometimes influenced by its interaction with specific ligands. For example, Lac repressor binds DNA in the absence of ligand and dissociates from DNA in the presence of ligand. In contrast, purine repressor is unable to bind DNA in unliganded form.

DNA-binding proteins usually bind not only their target DNA sites, but also all other sites, with lower affinities. They can thus be found in two types of complexes: one that shows tight, sequence-specific binding and another that involves looser, nonspecific binding with a nontarget site. This suggests that proteins first bind transiently to nontarget DNA sites and diffuse along the DNA in a one-dimensional random walk until they find a target site. Their association with other proteins and their cooperative binding to DNA can also help to recognize target sites. When a protein remains bound for a sufficiently long time to a given site, it can fulfil its biological role.

Families of DNA-binding Proteins

Many DNA-binding protein structures can be separated into two domains: one that interacts with DNA and another that possesses the biological activity or interacts with other protein partners. This separation means that similar DNA-binding motifs can be associated with different roles in various organisms. The DNA-binding domain usually interacts with DNA via a characteristic structural motif. Many motifs include a helix that enters into the major groove of DNA, but others contain a pleated β sheet or an extended stretch that binds to either the major or minor grooves. We briefly describe below the five major families of structural DNA-binding domains: helix-turn-helix, zinc finger, leucine zippers, β-ribbon domains and high mobility group (**Figure 2**). A sixth group mixes miscellaneous motifs, with atypical binding motifs, for which too few examples are known to define a family.

Other DNA-binding proteins cannot be classified on a conformational basis, because the functional and DNA-binding domains are intertwined, or perhaps because their characteristic structural motif has not yet been identified. This is the case for several enzymes, which we therefore group under the name DNA-binding enzymes.

Helix–turn–helix (HTH) domains

The HTH structure was the first DNA-binding motif to be discovered and is certainly the most thoroughly studied. It is found within protein domains of varied structure, origin and function, such as prokaryotic transcription regulatory proteins, eukaryotic homeodomains involved in cell differentiation during cell development, and histones whose role consists of DNA packaging. Notwithstanding this variability, the HTH motif has a very characteristic shape. It consists of two α helices that cross at an angle of about 120°, often connected by a specific turn of three amino acids with a glycine preferred in the first position. The second helix in the motif is the 'recognition' helix, which protrudes from the protein surface and penetrates into the major DNA groove. In addition to this canonical motif, the HTH domains contain a third helix of variable orientation, with sometimes a fourth helix, a small β sheet (in which case they are referred to as winged HTHs) or a flexible N- or C-terminal region. These additional elements confer to the HTH domains their structural and

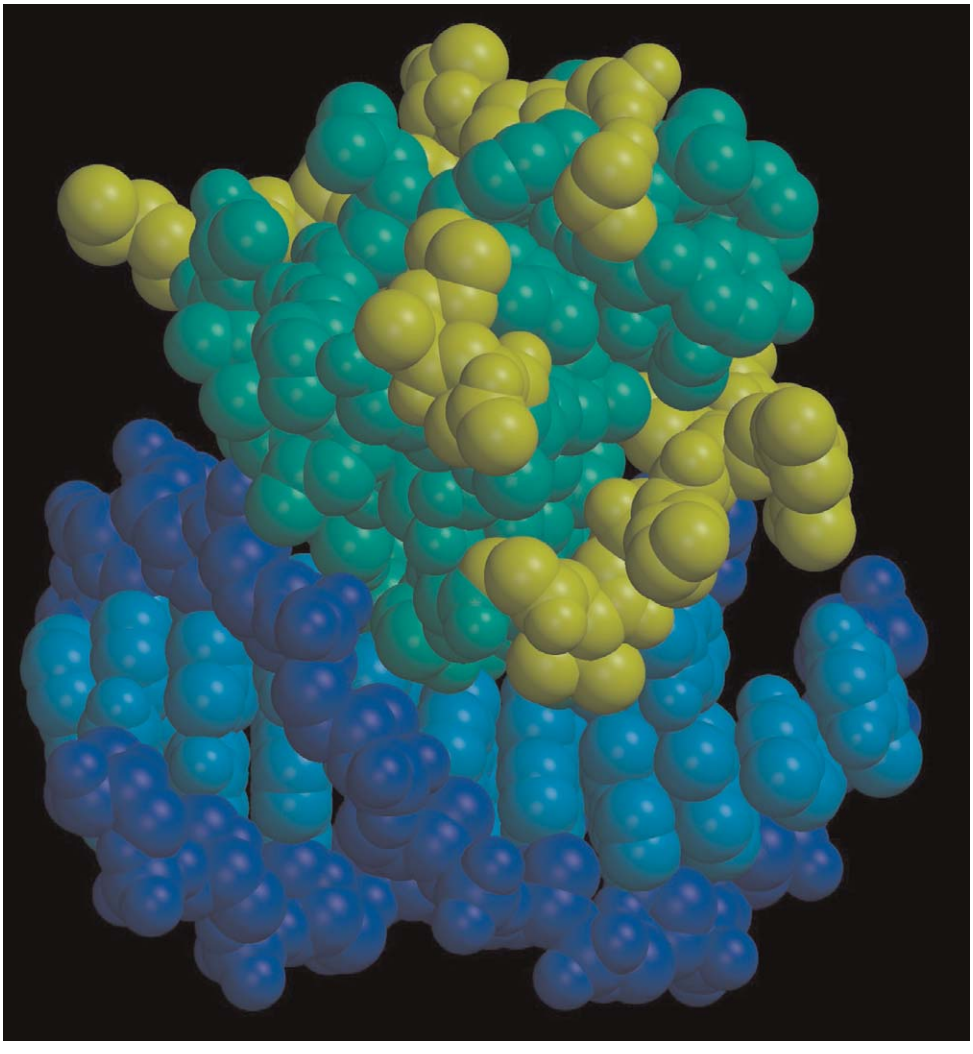

Figure 1 Space-filling model of 434 Cro in complex with the O_r1 DNA operator. The helices of Cro are in green and the coil regions in yellow-green; the DNA phosphate backbone is in dark blue and the base pairs in light blue.

DNA-binding specificity, as they often contact the border of the major DNA groove penetrated by the recognition helix, or the adjacent minor groove.

Zinc fingers

Zinc fingers contain zinc atoms as a structural element. Apart from this, they show a wide diversity in folds, metal-binding strategy and DNA-binding modes. They can be classified into three classes. Class 1, denoted Cys_2His_2-Zn_1, comprises domains of about 30 residues formed of an α helix and a β hairpin, sandwiched around a zinc ion liganded by two cysteines and two histidines. Usually, such domains are composed of three or more fingers in direct succession along the polypeptide chain. These consecutive fingers wrap around the DNA while their helices bind to the major groove in an essentially identical manner, at intervals of 3 bp. This type of zinc finger is involved in many aspects of eukaryotic gene regulation.

Class 2, referred to as Cys_8-Zn_2, comprises domains of about 70 residues and two zinc ions, each bound by four cysteines. Their structure consists of two loop–helix modules, with the two helices packed against each other at nearly right-angles and a zinc ion between each helix and loop. These zinc fingers bind DNA as homo- or heterodimers, with the first helix of each subunit fitting into the major groove. This class of zinc fingers is often encountered in receptors for steroids and related hormone-like molecules.

Class 3 zinc fingers, called Cys_6-Zn_2, are mainly found in yeast transcriptional activators. They have two zinc ions close in space, sharing six cysteines. The DNA-binding domain contains about 70 residues wrapped around the two zinc ions, among which is a helix that enters into the

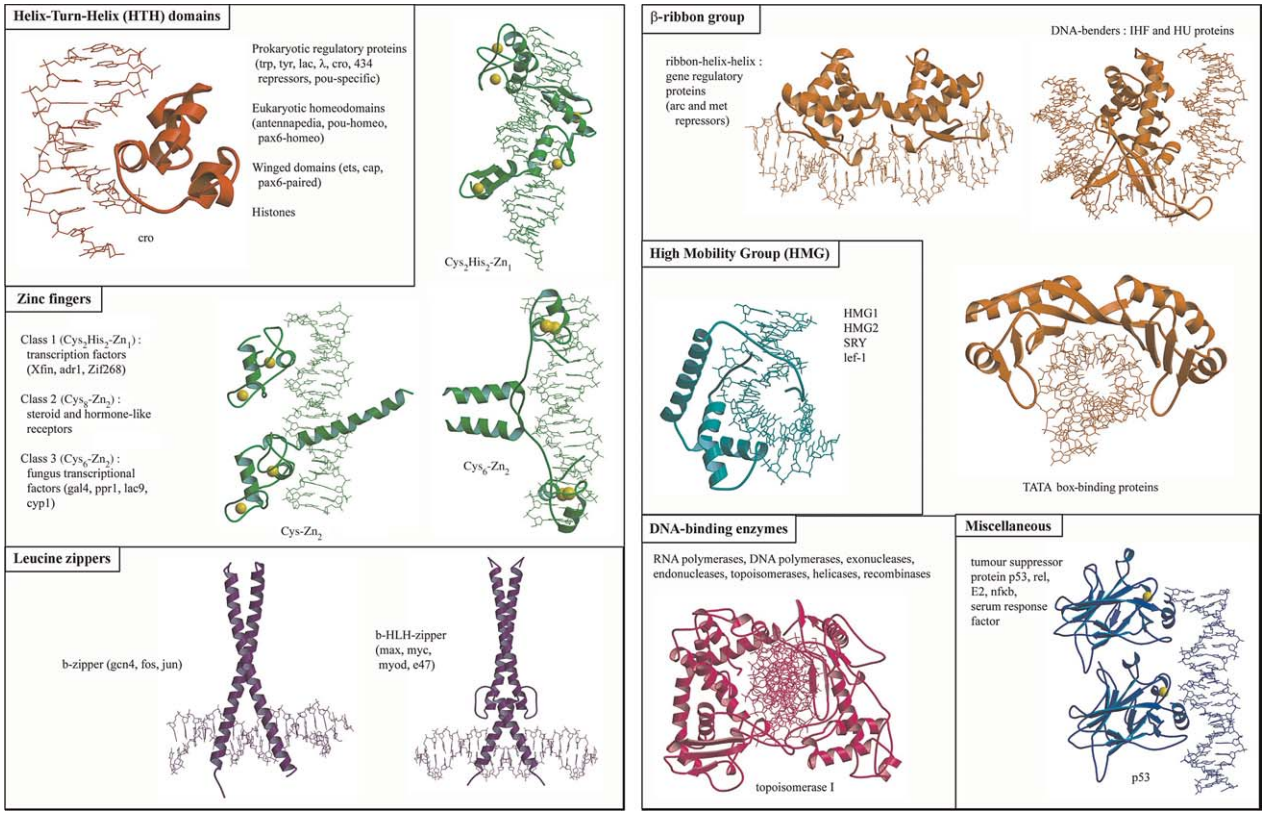

Figure 2 Illustration of the families of DNA-binding domains (colour ribbons) in complex with DNA (dark grey sticks). Zinc ions are depicted as yellow spheres. For each family, several members are listed and a characteristic structure is represented. For the DNA-binding enzymes and miscellaneous families, the member chosen to be depicted is indicated.

major DNA groove. These domains homodimerize by means of the C-terminal regions, which form a short α-helical coiled coil.

Leucine zippers

Proteins belonging to this family can easily be identified on the basis of their particular dimer conformation, called a leucine zipper: the C-terminal α-helical elements of each monomer are arranged in parallel and adopt a coiled-coil conformation. The sequences of the helices are characterized by a heptad repeat with predominately nonpolar residues in the first positions and leucines in the fourth positions. The N-termini are also characteristic. They consist of a positively charged basic region, which becomes helical upon contact with DNA. The helices of the two monomers enter the major groove on opposite sides of the DNA double helix, and hold it like a pincer.

Leucine zipper proteins subdivide into two structurally distinct classes, according to the conformation of the region connecting the leucine coiled-coil (zipper) and basic (b) regions. In the first subfamily, called b-zipper, there is no connecting region: the domain consists of two almost parallel helices, which have a coiled-coil arrangement at the C-terminus, and hold the DNA target by the N-terminus. The second subfamily possesses a connecting domain formed of a helix-loop-helix (HLH) motif, which, together with the HLH motif of the second monomer, form a parallel four-helix bundle. This subfamily is called b-HLH-zipper. A variant of it has the leucine zipper region suppressed and consists only of a b-HLH motif. Proteins of this family have important roles in cell differentiation and development.

β-Ribbon group

The defining feature of these proteins is their DNA binding through *β* structures. The interaction 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 gene regulatory 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.

The DNA benders, including the HU protein and integration host factor (IHF), introduce sharp bends in DNA and compact it. They are sometimes referred to as histone-like proteins, because of their presumed similar role. They are dimeric and composed of a body of α helices, with two protruding β-ribbon 'arms', which are flexible in the absence of DNA. The arms penetrate the minor groove on either side of the DNA and induce its bending by partial intercalation of hydrophobic residues. As a result, the DNA is tightly wrapped around the protein, and the β-ribbon arms in turn wrap around the DNA.

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. The DNA binding is mediated by a saddle-shaped, eight-stranded antiparallel β sheet, which forms a large concave surface for DNA interactions. Upon DNA recognition, hydrophobic side-chains are partially intercalated into the minor groove, and DNA is kinked and locally unwound so that the minor groove edges can make contact with the protein.

High mobility group (HMG)

This group is composed of several subgroups, with only similar extraction and solubility properties in common. The best known subgroup, HMG1/2, includes HMG1 and HMG2 that bind DNA with low specificity, and the sex-determining factor SRY that binds DNA with high specificity. The characteristic HMG1/2 motif looks like a V-shaped arrowhead, composed of an extended structure followed by three helices. The first two helices form one arm of the V, the third helix and the extended N-terminal region the other arm. The concave surface of the V binds to the DNA minor groove with partial insertion of hydrophobic amino acids into the DNA base stack. These contacts widen the minor groove and severely bend the DNA towards the major groove. As in the case of histones and DNA benders, the induced DNA bending seems to play an architectural role in transcription and replication, by juxtaposing non-adjacent factor-binding sites and allowing distant regulatory proteins to interact together and with the transcription or replication apparatus. Some HMG1/2 proteins seem also to have a role in DNA recombination and repair, as they bind to double- and single-stranded DNA, four-way junctions and other irregular DNA structures.

DNA-binding enzymes

These enzymes are usually larger than the other DNA-binding domains and present a wide diversity of structures and functions. We describe a limited number of examples.

DNA polymerases catalyse DNA replication by adding monodeoxyribonucleotides to a polydeoxyribonucleotide chain, and are thus responsible for copying genetic information in living systems. Similarly, RNA polymerases catalyse the transcription of DNA to form messenger RNA by adding monoribonucleotides to a polyribonucleotide chain, and are at the heart of gene expression. The faithful transmission of the genetic material, in other words the ability of the polymerases to choose at each step the correct nucleotide among the four possible ones, is obviously extremely important. Errors in replication or transcription may have disastrous consequences. To limit such errors, DNA polymerases use an error correction scheme, in which an exonuclease removes misincorporated nucleotides from the nascent chain; RNA polymerases have recently been shown to have a similar correction mechanism. The error rate attained by DNA polymerases is of the order of 10^{-7}; this is among the lowest error rates of all enzymes.

Other enzymes, grouped under the name DNA-repair enzymes, tend to preserve the integrity of the genetic information. They act to correct mutagenic, oxidative and toxic DNA damage produced by endogenous intracellular sources or extracellular agents, which otherwise might cause deficiencies such as cancer, ageing and death. Basically, they proceed by detecting the improper nucleotides, removing them and replacing them with the correct ones. There are a variety of enzymes participating in DNA repair in each organism, including endonucleases that detect and cleave improper nucleotides, certain kinds of DNA polymerases that add proper nucleotides, and recombinases that promote strand-exchange reactions. The latter, which also play a central role in recombination, typically act by binding to single-stranded DNA, then by interacting with duplex DNA and searching for homology between the single strand and the duplex. When homology is found, triple-stranded DNA is formed, the duplex is unwound and the original single strand is intertwined with the complementary strand from the duplex.

Yet other enzymes modify DNA conformations. Topoisomerases, for example, catalyse DNA rearrangements such as relaxation of supercoiled DNA and catenation and decatenation of DNA rings. They function by forming transient bonds with the DNA backbone, thereby allowing the passage of single- or double-stranded DNA through the broken backbone. They have characteristic structures, with four domains adopting a toroidal arrangement around a central hole, which is large enough to accommodate duplex DNA, and whose internal surface has an overall positive electrostatic charge. Helicases are involved in the unwinding and melting of DNA or RNA duplexes, and can bind to single- and double-stranded DNA. There are many different types of helicases, with various sequences and specificities, but all seem to share sequence and structural motifs and common mechanisms.

Methylation is a reversible, epigenetic modification of the mammalian genome implicated in parental imprinting,

X-chromosome inactivation, differential gene expression and carcinogenesis. Several DNA-binding enzymes are involved in methylation and demethylation events, and cooperate to fashion specific patterns of methylation. For example, methyltransferase catalyses the transfer of a methyl group to a cytosine in a DNA molecule. It proceeds by pushing the cytosine out of the DNA stack and replacing it by an amino acid side-chain. The expulsed cytosine is placed in the catalytic pocket of the protein, where it is methylated.

In several of the DNA-binding enzymes, such as some polymerases and endonucleases, a recurrent DNA-binding motif has recently been identified which is totally aspecific and differs both in structure and in DNA-binding mode from all other known motifs. It is called the helix-hairpin-helix motif and is formed by a pair of antipa rallel helices separated by a hairpin-like turn. Nonspecific interactions with DNA are mediated through the formation of hydrogen bonds between the DNA phosphate groups and the protein backbone situated in the hairpin region.

Miscellaneous

Many DNA-binding proteins could be categorized in this group, either because they really present atypical binding motifs, or because they are the first known members of a new family. Among these are the papillomavirus proteins E2 that play central roles in regulating transcription and viral DNA replication, the transcription regulatory proteins rel/nfκb and NFAT1 (nuclear factor of activated T cells 1), the serum response factor that regulates cell proliferation and differentiation, and finally p53 and PCNA which are described below.

p53 is a transcription regulatory protein that is very important in terms of biological relevance, as it has been implicated in about half of all cases of human cancer. Mutations affecting p53, most of them situated in the DNA-binding domain, have been shown to provoke the formation of tumours. This protein is therefore called a tumour suppressor. Its structure reveals a nonstandard DNA-binding motif, which differs from zinc fingers although it contains a zinc atom. The binding to DNA is mediated through contacts in the major groove by the α helix and a loop of a helix-sheet-loop motif, as well as by a contact in the minor groove from a residue in another loop.

Proliferating cell nuclear antigen (PCNA) is an unusual type of protein, of toroidal shape, which encircles DNA and can slide bidirectionally along the duplex. It interacts with a large number of DNA-binding proteins, such as cell cycle regulators and enzymes involved in DNA replication and repair. For example, PCNA tethers the catalytic unit of DNA polymerase to the DNA template for rapid and processive DNA synthesis.

Specificity of Protein–DNA Interactions

The specificity of a DNA-binding protein is determined by its ability to distinguish its DNA target site(s) from others. It is measured by the ratio of the protein–DNA equilibrium binding constants for target and nontarget DNA sites.

Many DNA-binding proteins can recognize specific DNA sequences among thousands of others. Transcription factors are examples of sequence-specific binding proteins that modulate the expression of genes. Lac repressor, for example, has a specificity of the order of 10^6. TATA box-binding proteins are also highly specific: they specifically recognize TATA promoter sequences and thereby initiate transcription. Other proteins specifically recognize methylated DNA sites.

At the other extreme, some proteins are essentially sequence nonspecific, but rather show specificity for certain DNA structures and topologies. Several repair enzymes recognize mispaired or damaged DNA: topoisomerases detect supercoiled or circular DNA, and HMG1/2-type proteins bind to four-way junctions. Histones and DNA benders, whose primary role consists of packaging of DNA, also have relaxed sequence specificity. For example, the specificity with respect to random DNA sequences of the DNA bender IHF is about 10^3–10^4.

Different mechanisms can render a DNA-binding protein specific for a given DNA target. The intrinsic propensity of the motif to recognize a DNA sequence constitutes the first level of specificity. It is intimately related to the local bending propensities of the DNA target, which are basically determined by the DNA sequence. Specificity can be increased by the assembly of different DNA-binding motifs and can also be acquired indirectly, when nonspecific proteins bind to specific ones.

Intrinsic specificity of the DNA-binding motif

Sequence-specific binding is achieved essentially through hydrogen bonds, cation–π and π–π interactions as well as favourable van der Waals interactions between amino acid side-chains, typically located in the recognition helix or β structure that interacts with DNA, and nucleic acid bases, situated in the major or minor groove penetrated by the protein motif. The most marked hydrogen bond preferences involve Arg-Gua, Asn-Ade and Gln-Ade in the major groove.

Protein motifs bound to their target groove sometimes display additional contacts in the adjacent grooves. These additional contacts are important as they increase binding affinity and specificity. In the HTH family for example, the N-terminal arms of the homeodomain motifs, which are unstructured in the absence of DNA, and the 'winged' β-hairpin structures of the Ets domains, make particular contacts in the minor groove adjacent to the target major groove

where the recognition helix lies. In this way, these proteins increase their DNA target region, including additional DNA base pairs that flank the central major groove core target, thereby enabling the recognition of target sequences that are long enough to ensure some specificity.

Several authors were tempted to reduce the general mechanism of sequence-specific DNA recognition to the existence of a 'DNA recognition code', i.e. a code whereby certain DNA base pairs are recognized by certain amino acids. However, though some amino acid–nucleobase pairing preferences are observed, as well as certain rules within families such as the Cys_2His_2-Zn_1 zinc fingers, it is now clear that no simple code describing all DNA–protein interactions does exist. Rather, specificity is achieved through the combination of several mechanisms, including the intrinsic specificity of the motifs but not reduced to it.

Association of DNA-binding motifs

Another way of increasing the DNA target region is through the association of two or more DNA-binding domains. These domains may be identical or distinct, and may belong to the same polypeptide chain or be formed by dimers or multimers.

In the case of identical domains, the DNA-recognition site comprises two 'half-sites' that are either direct or inverse (palindromic) repeats of each other. The base pair separation between the two half-sites is specific to each protein; it depends upon the linker region between the motifs when these belong to the same sequence, and on the protein–protein dimerization interface otherwise. The HTH subfamily of prokaryotic repressors, such as cro and λ, bind DNA as homodimers and hence use this scheme to acquire specificity.

Leucine zippers exemplify a different kind of association. They bind DNA only as dimers, with two helices wrapped in coiled-coil conformation. Though they can bind as homodimers, they usually prefer to assemble as heterodimers. This modulates their *in vivo* activity, where enhanced DNA binding may be the result of higher dimer stability or more favourable DNA contacts. For example, the b-zipper domains Fos and Jun and the b-HLH-zipper domains Myc and Max preferentially form the heterodimeric complexes Fos/Jun and Max/Myc.

When several proteins bind cooperatively to DNA, specificity is augmented through the interaction of protein partners which themselves already possess some specificity for DNA sites. For example, the heterodimeric Fos/Jun leucine zipper protein interacts with an Ets protein, exhibiting a winged HTH motif, or with the nuclear factor NFAT1. The heteromultimeric NFAT1–Fos–Jun–DNA complex is shown in **Figure 3**.

A rather large and complicated association of DNA-binding proteins is the pre-initiation complex (PIC), which initiates transcription in eukaryotes. It contains the TATA box-binding protein (TBP), which specifically recognizes TATA promoter sequences situated about 25 bp upstream of the transcription start site. The binding of TBP to DNA produces a strong DNA bend and distortion, which partially unwinds DNA, thereby bringing the DNA sequences at both sides of TBP in close apposition. This DNA deformation favours the assembly of the PIC complex, which encompasses RNA polymerase II, TBP and at least six other polypeptide subunits. In the presence of adenosine triphosphate (ATP), the assembly of the complete PIC is followed rapidly by DNA melting and transcription initiation. *In vivo*, transcription also involves a number of regulatory proteins which, by interacting with the PIC assembly, repress or enhance transcription.

DNA bending propensities

Specificity can also be modulated by the conformation of the DNA chain itself and by its capacity to accommodate structural modifications. Several proteins bind to DNA sites that have a specific local conformation and flexibility, which depend on the base pair composition. B-form DNA, for example, where the major groove is much wider than the minor groove, is favoured by AT-rich sequences and is much more flexible than A-form DNA. In particular, the intrinsic bendability of the TATA box regions facilitates recognition by the TATA box-binding proteins and thereby helps to promote transcription initiation and elongation. Upon binding, these proteins increase the bending of DNA that has already a natural propensity to bend. Another well-documented example is the homodimeric HTH-type proteins Cro and catabolite activator (Cap) from bacteriophage λ, in which the recognition helix of each monomer penetrates the major groove and compresses the minor groove between the two insertion sites. This leads to a bending of DNA of about 90° in the case of Cap and of about 40° in the case of Cro. The affinity of Cro and Cap proteins for their operator is therefore strongly dependent upon the sequence composition of the central region of the target DNA. More recently, several structures of EcoRV endonuclease resolved with target and nontarget DNA sequences reveal that the deformation of the DNA structure is needed to allow specific nucleobase contacts to be formed.

Summary

Analysis of the increasing number of experimentally determined protein–DNA structures has revealed the existence of several protein–DNA interaction modes, mediated through distinct structural motifs. These motifs are occasionally associated with a specific function. Usually, however, they are found in proteins with various functions and coming from different organisms. The affinity and specificity of the DNA-binding domains for their DNA

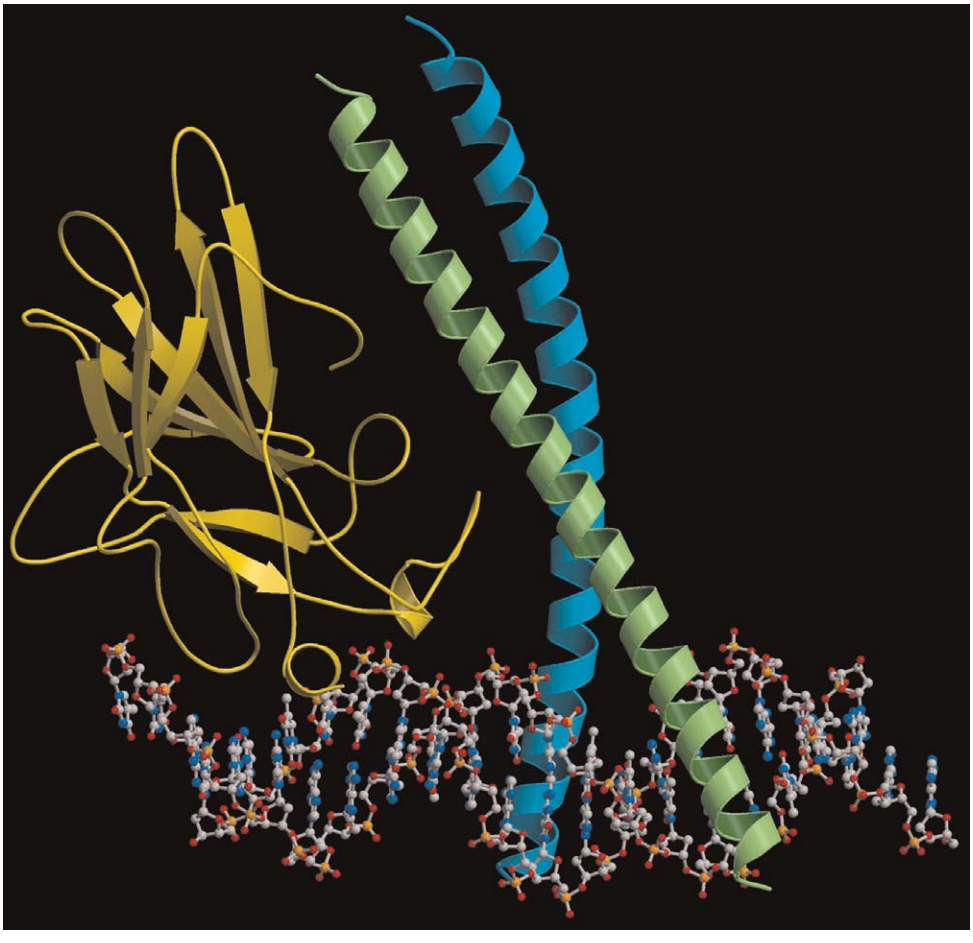

Figure 3 Example of heteromultimeric protein–DNA complex, consisting of the DNA-binding domain of NFAT1 (nuclear factor of activated T cells 1), Fos and Jun bound to DNA. Proteins are depicted in yellow (NFAT1), green (Fos) and blue (Jun) ribbons. The DNA molecule is represented by ball and sticks; carbon atoms are in grey, oxygen in red, nitrogen in blue, and phosphorus in orange.

target are determined by their amino acid sequence, but also by the association of several domains which cooperatively bind to DNA, and by the local conformational properties of the DNA target itself.

Revised and updated: December 2003

Further Reading

Banaszak LJ (2000) *Foundations of Structural Biology*. London: Academic Press.

Biomolecular Structure and Modelling group (2000) Summary of DNA-binding protein structural families, grouped by DNA recognition motif. A structural classification of all protein– DNA complexes solved by X-ray crystallography to a resolution of 3.0 angstroms or better. University College London. http://www.biochem.ucl.ac.uk/bsm/prot_dna/prot_dna_cover.html

Branden C and Tooze J (1998) *Introduction to Protein Structure*, 2nd edn. New York: Garland Publishing.

Goldman A and Ollis DL (1990) Interaction on nucleic acids with proteins. In: Blackburn GM and Gait MJ (eds) *Nucleic Acids in Chemistry and Biology*, pp. 337–381. Oxford: IRL Press.

Kyushu Institute of Technology (2003) The Biomolecules Gallery. An image database providing a wide spectrum of high quality pictures of biomolecules, such as nucleic acids, proteins and their complexes. Kyushu Institute of Technology, Japan. http://gibk26.bse.kyutech.ac.jp/jouhou/image/gallery.html

Travers A and Buckle M (2000) *DNA–Protein Interactions. A practical approach* Oxford: Oxford University Press.