

Able assistants for homeodomain proteins

Recent advances in understanding the regulation of the DNA-binding specificity of gene regulatory proteins may help explain how similar proteins can have diverse effects.

The homeodomain is a 61 residue DNA-binding domain present in a variety of transcription factors [1]. Homeodomain-containing proteins are apparently ubiquitous among eukaryotes, and many have been shown to have important roles in development. For example, those encoded by the well known clustered homeotic genes of the fruit fly *Drosophila melanogaster*, and their homologs in vertebrates and nematodes, specify the identities of different regions along the anterior-posterior axis of the body. In the fly, each homeotic gene is expressed in a specific region of the body, where its protein product is required for the correct development of that region: if the protein is absent, that part of the fly will develop as a copy of another part (a homeotic transformation). Each homeotic gene product directs formation of a characteristic set of structures. For example, *Antennapedia* (*Antp*) directs morphogenetic processes in the thorax of the fly, whereas *Ultrabithorax* (*Ubx*) acts primarily in the anterior abdomen. How do two proteins such as these, with similar homeodomains, have clearly distinct developmental effects? The same question applies to the mammalian homologs of the fly homeotic genes, the *Hox* genes. How does one protein direct rib formation and another not?

In principle, there are several ways in which each protein could have a unique influence on development. Firstly, the different proteins could bind to the same target genes but have different effects once bound. Secondly, they could bind to different sets of target genes. Thirdly, they could bind to both shared and unique targets. In all of these cases, DNA-binding specificity is a crucial determinant of homeotic function. Such specificity could be accomplished with or without the aid of other proteins (Fig. 1). Recent results suggest how the binding specificity can be attained, and in particular how cofactors fine tune the actions of homeodomain proteins. Here we address the question of how homeodomain-containing proteins bind specifically to physiologically relevant target sequences, and how these sequences can be distinct despite the similarity of the different homeodomains.

Both *in vitro* footprinting analyses and transcriptional assays in cotransfected tissue culture cells have suggested that many fly homeodomain proteins can bind to similar or identical DNA sequences [2]. Some studies have shown differences in the DNA sequences preferred by homeodomain proteins [3,4], though these are unlikely to account for the dramatically different developmental pathways the different proteins initiate. For example, although the *Antp* and *Ubx* proteins elicit very different developmental programs, their homeodomains are very similar (54 of 61 residues identical). Moreover, the amino-acid

residues in the recognition helices of these proteins that make specific contacts with the DNA are absolutely conserved, suggesting that binding specificity is only partially achieved by homeodomain-DNA interactions.

How might greater binding specificity be achieved? Possible answers come from studies of bacterial and yeast DNA-binding proteins, the DNA-binding specificities of which are altered by interactions with other proteins. One classic example from bacteria involves the interaction of sigma factors with RNA polymerase. Studies initiated in the 1970s established that different sigma factors can interact with RNA polymerase molecules and that, depending on the type of sigma factor expressed, different sets of promoters would be recognized by the sigma-polymerase holoenzyme [5]. Although RNA polymerase binds tightly to DNA, it does not bind in a sequence-specific fashion. Sigma factors are able to direct RNA polymerase to specific sites, effectively increasing the enzyme's DNA-binding specificity and providing a way of regulating different sets of genes at different times. A second example comes from work on the repression of genes involved in the transport or metabolism of ribonucleosides and deoxyribonucleosides [6]. In the absence of the CytR repressor protein, the cyclic AMP receptor protein (CRP) binds to the operators of such genes and acts as a positive regulator of transcription. When CytR is expressed, however, it also binds these operators but has the opposite effect: transcription is repressed. *In vitro* analyses have shown that the CytR and CRP proteins bind cooperatively to neighboring sites on the DNA. In addition, if cells are deficient in CRP, CytR cannot further repress transcription of its target genes. Thus, CytR appears to require CRP to enable it to bind to its target sites. To what extent do homeodomain proteins use similar mechanisms?

Insights into the ways by which the DNA-binding specificity of homeodomain proteins can be determined in eukaryotic cells have been provided by recent work on mating-type regulation in the yeast *Saccharomyces cerevisiae* [7-9]. Yeast has two haploid cell types, **a** and α , which are determined by the particular allele a cell expresses at the *MAT* locus: **a** cells express the *MAT_a* allele and α cells the *MAT _{α}* allele. In α cells, in the absence of *MAT_a*1 protein, *MAT _{α}* 2 binds to **a**-specific operators, resulting in the repression of **a**-specific genes. In order to bind specifically to the biologically relevant sequences, however, *MAT _{α}* 2 must interact with another DNA-binding protein, *MCM1*, which also binds to **a**-specific operators. The binding sites for *MAT _{α}* 2 and *MCM1* are next to one another, and it has been shown that the proteins bind cooperatively to DNA, forming a complex containing a pair

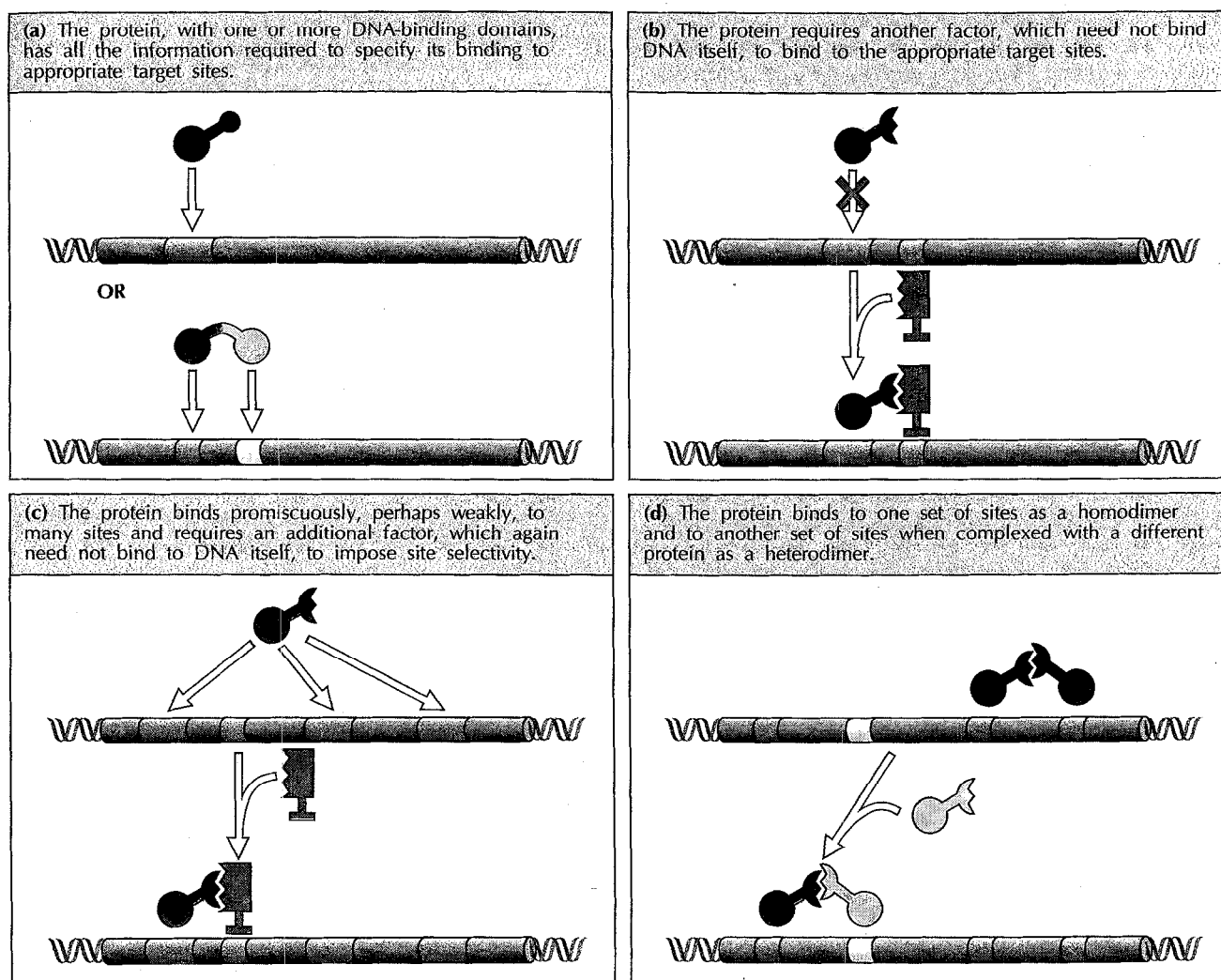


Fig. 1. Models of the different ways in which the DNA-binding specificity of proteins can be determined.

of each of the two proteins [8]. Cooperativity is achieved by protein-protein interactions, and complex formation requires a disordered region of the MAT α 2 protein, which lies immediately amino-terminal to the homeodomain.

Diploid cells express both *MAT α 1* and *MAT α 2* alleles, and in the presence of *MAT α 1* protein, itself a homeodomain protein, the DNA-binding specificity of *MAT α 2* is altered so that it now recognizes and represses the promoters of haploid-specific genes [10]. The alteration of binding specificity results from the formation of heterodimers between *MAT α 2* and *MAT α 1* proteins; the heterodimer has a DNA-binding specificity distinct from that of a *MAT α 2* homodimer [9]. Thus, DNA binding specificity can also be influenced by heterodimer formation, which can allow a particular protein access to a variety of different binding sites, depending on the partner with which it interacts.

Might the homeodomain proteins of animals achieve their specificity in a similar fashion? Several reports have suggested that homeodomain proteins do in fact interact with other proteins, usually resulting in facilitation of binding of the proteins to juxtaposed binding sites. The ubiquitous mammalian homeodomain protein Oct-1 facilitates binding to DNA of the viral transactivator protein VP-16,

which binds DNA rather poorly on its own [11,12]. In a similar fashion, the homeodomain protein PHOX-1 may be involved in facilitating the binding of serum response factor (SRF) to the serum response element of the *c-fos* promoter [13]. Recent work suggests that the pituitary-specific homeodomain protein Pit-1 facilitates binding of the retinoic acid receptor to a sequence in the promoter of the *PIT-1* gene (S Rhodes and MG Rosenfeld, personal communication). Intriguingly, many homeodomain proteins have additional sequence homologies in the region immediately amino-terminal to the homeodomain, equivalent to the region of the yeast α 2 protein required for interaction with MCM1 [7]. These sequences are related to the peptide YPWM, which might be directly involved in interactions with neighboring proteins.

Several different homeodomain proteins have been reported to undergo homodimerization and heterodimerization [14,15]. Most of the work has focused on the 'POU' class of homeodomain proteins, which are involved in both developmental and cell-type specific transcriptional responses. Amino-terminal to their homeodomain, these proteins have an additional region of sequence homology termed the POU-specific domain, which is separated from the homeodomain by a linker region. Studies

with the POU transcription factors Pit-1, Oct-1, Oct-2, Oct-2A and Oct-6 have shown that each can bind cooperatively to DNA as a dimer. Moreover, each protein can form heterodimers, which can also bind cooperatively to DNA, with at least one other POU protein. Generally, both the POU domain and the homeodomain are required for cooperative binding of the heterodimers, although in a couple of cases the homeodomain alone appears able to mediate cooperativity. It is unclear whether heterodimerization affects the DNA-binding specificity of the POU proteins: heterodimers need not have altered DNA-binding affinity or site specificity, but instead may simply be more stable than homodimers *in vivo*. The presence of both components of such a heterodimer would effectively increase the number of active dimers in a cell.

How else could DNA-binding specificity be increased? One way is for a protein to have multiple DNA-binding domains. The POU-specific domain can bind DNA sequence-specifically without a homeodomain, albeit with low affinity [16]. In the absence of the POU-specific domain, the homeodomain of a POU protein can recognize sequences similar to the binding sites of non-POU homeodomain proteins. However, when the POU-specific and homeodomains are linked they recognize different DNA sequences; thus placing the POU-specific domain next to the homeodomain increases the DNA-binding specificity of the POU protein. Similarly, the 'Paired' class homeodomain proteins, known to be important in both fly and mouse development, contain two, separable DNA-binding domains [17]. In addition to a homeodomain, these proteins contain a 128 residue domain known as the Paired domain, which is capable of sequence-specific DNA binding. Taking the idea to extremes, the two *Drosophila* genes *zfb-1* and *zfb-2* encode proteins with multiple copies of the DNA-binding 'zinc finger' motif and one or more homeodomains [18]. These proteins might cross-link several chromosome sites or bind to regions with large arrays of binding sites.

The key to understanding the action of proteins such as MAT α 2 has been to study their interactions with actual target gene sequences. Very few such sequences are known for genes such as *Antp* and *Ubx*, and no binding sites for *Hox* proteins have yet been proven biologically relevant *in vivo*. A clear picture of the mechanisms by which the products of the clustered homeotic genes achieve their specificity will only be obtained when many more genuine *cis*-acting regulatory elements have been characterized. Several recent papers report significant progress in this direction [19–21], and we may hope to learn much more about what distinguishes anterior from posterior in the next few years.

References

- SCOTT MP, TAMKUN JW, HARIZEL JE III: The structure and function of the homeodomain. *Biochim Biophys Acta* 1989, 989:25–48.
- HAYASHI S, SCOTT MP: What determines the specificity of action of *Drosophila* homeodomain proteins? *Cell* 1990, 63:883–894.
- EKKER SC, VON KESSLER DP, BEACHY PA: Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J* 1992, 11:4059–4072.
- DESSAIN S, GROSS CT, KUZIORA MA, MCGINNIS W: Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J* 1992, 11:991–1002.
- GROSS C, LONETTO M, LOSICK R: Bacterial sigma factors. In *Transcriptional Regulation*. Edited by McKnight SL, Yamamoto KR. Cold Spring Harbor: Cold Spring Harbor Press; 1992.
- HOLST B, SOGAARD-ANDERSEN L, PEDERSEN H, VALENTIN-HANSEN P: The cAMP–CRP/CytR nucleoprotein complex in *Escherichia coli*: two pairs of closely linked binding sites for the tAMP–CRP activator complex are involved in combinatorial regulation of the *cdd* promoter. *EMBO J* 1992, 11:3635–3643.
- VERSHON AK, JOHNSON AD: A short, disordered protein region mediates interactions between the homeodomain of the yeast α 2 protein and the MCM1 protein. *Cell* 1993, 72:105–112.
- SMITH DL, JOHNSON AD: A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an α 2 dimcr. *Cell* 1992, 68:133–142.
- DRANGINIS AM: Binding of yeast α 1 and α 2 as a heterodimer to the operator DNA of a haploid-specific gene. *Nature* 1990, 347:682–685.
- GOUTTE C, JOHNSON AD: α 1 protein alters the DNA binding specificity of α 2 repressor. *Cell* 1988, 52:875–882.
- STERN S, HERR W: The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeodomain: evidence for a homeodomain recognition subdomain. *Genes Dev* 1991, 5:2555–2566.
- KRISTIE TM, SHARP PA: Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV α -transactivator protein. *Genes Dev* 1990, 4:2383–2396.
- GRUENBERG DA, NATESAN S, ALEXANDRE C, GILMAN MZ: Human and *Drosophila* homeodomain proteins that enhance the DNA-binding activity of serum response factor. *Science* 1992, 257:1089–1095.
- ROSENFELD MG: POU-domain transcription factors: pou-er-ful developmental regulators. *Genes Dev* 1991, 5:897–907.
- VERRIJZER CP, VAN OOSTERHOUT JAWM, VAN DER VLIET PC: The Oct-1 POU domain mediates interactions between Oct-1 and other POU proteins. *Mol Cell Biol* 1992, 12:542–551.
- VERRIJZER CP, ALKEMA MJ, VAN WEPEREN WW, VAN LEEUWEN HC, STRATING MIJ, VAN DER VLIET PC: The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J* 1992, 11:4993–5003.
- TREISMAN J, HARRIS E, DESPLAN C: The paired box encodes a second DNA-binding domain in the Paired homeodomain protein. *Genes Dev* 1991, 5:594–604.
- FORTINI ME, LAI Z, RUBIN GM: The *Drosophila zfb-1* and *zfb-2* genes encode novel proteins containing both zinc-finger and homeodomain motifs. *Mech Dev* 1991, 34:113–122.
- GOULD AP, BROOKMAN JJ, STRUTT DI, WHITE RAH: Targets of homeotic gene control in *Drosophila*. *Nature* 1990, 348:308–312.
- GRABA Y, ARAGNOL D, LAURENTI P, GARZINO V, CHARMOT D, BERENGER H, PRADEL J: Homeotic control in *Drosophila*: the *scabrous* gene is an *in vivo* target of *Ultrabithorax* proteins. *EMBO J* 1992, 11:3375–3384.
- VACHON G, COHEN B, PFEIFLE C, MCGUFFIN ME, BOTAS J, COHEN SM: Homeotic genes of the bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. *Cell* 1992, 71:437–450.

J. Robert Manak and Matthew P. Scott, Departments of Developmental Biology and Genetics, Stanford University School of Medicine, Stanford, California 94305–5427, USA.