Am. J. Hum. Genet. 63:1573–1577, 1998

PROTEIN COMPLEXES '98
Flexing DNA: HMG-Box Proteins and Their Partners

Marco E. Bianchi and Monica Beltrame

Dipartimento di Genetica e di Biologia dei Microrganismi, Universitá di Milano, and San Raffaele Scientific Institute, Milan

The initiation of transcription by RNA polymerase II is controlled by the assembly of the basal transcription machinery, which integrates inputs from transcription factors bound to promoters and enhancers. Overall, a typical gene may depend on the binding of tens of transcription factors to its controlling-sequence elements. Combinatorial assembly ensures that each gene can exhibit a distinctive and sometimes variable pattern of expression, despite the fact that the number of transcription factors is much lower than the number of genes.

DNA is usually represented as a straight thread over which transcription factors perch like birds on a telephone line. During the past few years, however, it has been realized that DNA and transcription factors form a tightly knit three-dimensional structure, in which the DNA is bent and twisted and in which the proteins establish a vast number of contacts to each other (Tjian and Maniatis 1994). This newer model neatly explains how each transcription factor can establish contacts with several others, which are not necessarily bound to contiguous sites on the DNA. Hence, the model accommodates a large number of otherwise puzzling biochemical data, but it raises a distinct biophysical problem: DNA is intrinsically a stiff molecule, with very little inclination to bend and twist. In dilute solutions, DNA can be modeled as if it consisted of freely rotating rigid rods, 300 bp in length; only beyond the kilobase scale does DNA appear serpentine (Cantor and Schimmel 1980). How, then, is DNA persuaded to follow tightly curved paths? Perhaps not surprisingly, specialized proteins are responsible.

SRY and the Genetics of Sex Determination

The most famous of DNA-flexing proteins is SRY, the factor that determines the male sex in mammals (Goodfellow and Lovell-Badge 1993, and references therein). SRY is coded by an intronless gene on the Y chromosome, which redirects the undifferentiated gonad of mammalian embryos from the default developmental fate (ovaries and female sexual organs), promoting instead the development of testes and male sexual organs. In the mouse, the *Sry* gene is expressed in pre-Sertoli cells at ~10.5–12.5 d postcoitus. Sry mRNA expression resumes in the adult testis, but no Sry protein is formed in this tissue, because the transcript is circularized, leaving no entry point for ribosomes. Approximately 30% of XX mice transgenic for the *Sry* gene become infertile but otherwise apparently normal males. In humans, as well, the translocation of the genome segment encompassing *SRY* to other chromosomes causes the switch of XX individuals to male gonads and habitus. Conversely, mutations of the *SRY* gene in XY individuals lead to female phenotypes with variable penetrance.

SRY protein consists of a DNA-binding domain of the HMG-box class, flanked by protein sequences with little similarity to other transcription factors. Surprisingly, the flanking sequences are not homologous in SRY protein from closely related species (Tucker and Lundrigan 1993; Whitfield et al. 1993). Moreover, all but one of the known sex-reversing mutations in human SRY protein encompass the DNA-binding domain, pointing to a critical role for this segment and little or no role for the rest of the protein. In particular, no transactivation domain has been identified. How such a minimal protein can control gene expression and cell differentiation cannot directly be tested, however, because, so far, no target gene for SRY has been identified with certainty. Detailed information is nonetheless available on the mode of interaction between the HMG box of SRY and DNA.

The HMG-box domain was first recognized as duplicated 80-amino-acid regions present in the abundant nonhistone chromatin protein HMG1. Similar sequences are also present in a variety of transcription factors, and there are currently >100 examples known (reviewed in Bianchi 1995). All of these proteins share a marked effect on DNA structure and an affinity for distorted DNA structures such as four-way junctions (Bianchi et al. 1989) and bent cisplatin adducts (Pil and Lippard 1992). In addition, some of these proteins, including SRY, recognize specific sequences in linear DNA, with high af-

Received October 8, 1998; accepted for publication October 15, 1998; electronically published November 10, 1998.

Address for correspondence and reprints: Dr. Marco E. Bianchi, Dipartimento di Genetica e di Biologia dei Microrganismi, Universitá di Milano, via Celoria 26, 20133 Milano, Italy. E-mail: marco.bianchi@unimi.it

0002-9297/98/6306-0002$02.00

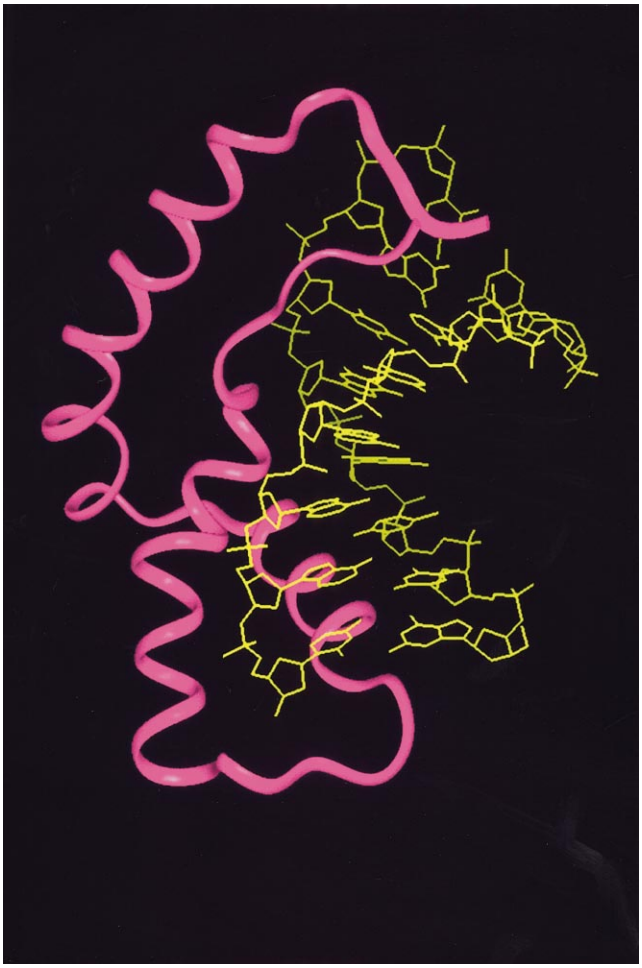

Figure 1 Structure of the complex between the HMG box of human SRY and the octanucleotide GCACAAAC. HMG boxes are composed of three α-helices and an extended amino acid stretch at their N-terminus. The Cα backbone of the SRY HMG-box (*magenta*) is almost exactly superimposable upon that of HMG boxes from mammalian HMG1 and *Drosophila* HMG-D, which share only ~20% identity in primary sequence. The DNA (*yellow*) is significantly distorted: the minor groove is widened considerably to accommodate the extended stretch of the HMG box; the planes of the bases are tilted; and the double helix is unwound and bent. Very similar features are also present in DNA complexed to TBP, which also binds via the minor groove but which has a completely different protein fold. The contact between the HMG box and the DNA occurs over a surprisingly large area and with a very tight fit; thus, the energetic cost of distorting the DNA is more than compensated by the chemical energy freed by the protein-DNA interaction. This model is derived from the atomic coordinates provided by Werner et al. (1995).

finity and moderate selectivity. The molecular basis of these unusual binding specificities is revealed by the structure of a complex between a DNA octamer and the DNA-binding domain of SRY (fig. 1; Werner et al. 1995). The HMG box of SRY has an L-shaped structure formed by an extended segment and three α-helices, stabilized by hydrophobic contacts. The concave side of the protein contacts the minor groove of the DNA, over a surprisingly large surface area. The minor groove is extensively widened to accommodate the extended segment of the protein domain, and the double helix is unwound by a succession of positive roll angles and is bent back by 70°–80°.

The observations that SRY consists essentially of the DNA-binding domain and that it bends DNA have led to the proposal that DNA flexing might be the main activity of SRY. The local structural deformation that SRY induces in the double helix might mediate effects at a distance, via the mechanical displacement of the DNA segments (and associated factors) at either side of the point of flexure (fig. 2). In accordance with this idea, a human SRY mutation that causes a very moderate reduction of the affinity of protein binding to target sites—but a marked variation in the angle of flexure of the DNA—was found to cause sex reversal with full penetrance, as mutations that completely abolish DNA binding (Pontiggia et al. 1994). A critical prediction in the DNA-flexing hypothesis, however, is that SRY should be a component of a multiprotein complex, whose stability it promotes. Again, this prediction cannot be tested, for lack of a known target gene for SRY. Interesting clues are nonetheless offered by the biology of other HMG-box proteins.

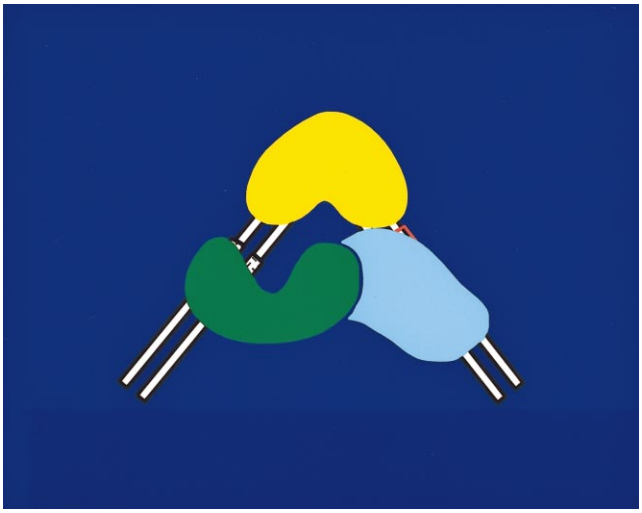

Figure 2 Model for the action of SRY and other HMG-box proteins. An HMG-box protein (*orange*) such as SRY binds to target DNA sequences, which it bends considerably. The angle in the DNA depicted here approximates that caused by SRY. The flexure of the DNA brings proteins (*blue* and *green*) bound to each side of the site into proximity, thus promoting their interaction. The presence of contact surfaces between SRY and one or both neighboring proteins may stabilize the whole complex and increase its target specificity. The same model can apply to HMG1. HMG1 binds to DNA regardless of sequence, so protein-protein interactions with partners are essential to direct it to the DNA sites where flexing is to occur.

HMG1 and Its Relatives

HMG1, which gives the name to the whole family, is a very abundant nuclear protein in all mammalian cells: >1 million HMG1 molecules may be found in a single nucleus (reviewed by Bustin and Reeves 1996). It has two close relatives, HMG2 and the recently discovered HMG4 (Vaccari et al. 1998), which are expressed in embryos and in a limited number of adult tissues. HMG1 and its relatives are composed of two HMG boxes and a long acidic tail, separated by short linker sequences. The tail is longest in HMG1 (30 aspartic and glutamic acid residues) and shortest in HMG4; no specific function has been attributed to it so far, and certainly it has no transactivation potential (Landsman and Bustin 1991). Thus, mammalian HMG1-like proteins consist almost entirely of their DNA-binding domains. Their primary sequence is extremely conserved in all vertebrates and down to echinoderms; distantly related proteins are also present in insects, plants, and fungi.

HMG1 shares very similar domain structures and DNA-binding features with SRY. It binds avidly to four-way junctions and other DNA molecules displaying a widened major groove. Ring-closure assays show, conversely, that, when present at high concentrations, it bends short linear-DNA fragments on binding. Unlike SRY, however, HMG1 has no sequence specificity and binds very inefficiently to linear DNA. This lack of specificity in DNA target sites is offset, in the case of HMG1, by several specific interactions with such DNA-binding proteins as Oct and Hox proteins (Zwilling et al. 1995; Zappavigna et al. 1996), nuclear hormone receptors (Boonyaratanakornkit et al. 1998), p53 (Jayaraman et al. 1998), RAG1 (V. Aidinis, T. Bonaldi, M. Beltrame, S. Santagata, M. E. Bianchi, and E. Spanopoulou, unpublished data) and some components of the basal transcription machinery (Sutrias-Grau et al., in press, and references therein). Strikingly, all such partners interact directly with HMG boxes, via their DNA-binding domain. In doing so, these proteins increase the protein surface contacting the DNA, from both the major and minor groove, to ultimately achieve high-affinity interaction with their cognate DNA site, without affecting sequence specificity. How these diverse protein folds may all interact with HMG boxes is unclear at this time. Protein-protein interactions also occur in the absence of DNA but are rather weak and dynamic. They can easily be verified by copurification of the components, but yeast two-hybrid screens have proved disappointing for this purpose. Within the cell nucleus, the large pool of HMG1 probably interacts with a variety of macromolecules, including nucleosomes (Falciola et al. 1997, and references therein). In vitro, HMG1 forms stable complexes with reconstituted nucleosomes. It is not a component of condensed metaphase chromosomes, and it readily diffuses away from the chromatin of detergent-permeabilized interphase cells. The most telling observation is that, abundant as it is, HMG1 is limiting within cells: transient overexpression of HMG1 enhances the transcriptional activity of Hox protein and steroid hormone receptors, as well as the yield of V(D)J recombination products, in transfection assays (Zappavigna et al. 1996; Boonyaratanakornkit et al. 1998) (V. Aidinis, T. Bonaldi, M. Beltrame, S. Santagata, M. E. Bianchi, and E. Spanopoulou, unpublished data). The emerging picture, thus, is that HMG1 serves as an all-purpose DNA-bending, -wrapping, and -looping factor that can be recruited for transcription, DNA repair, and recombination. It is therefore not surprising that no human mutation in *HMG1* has been reported so far. Genetic information is expected from the targeted deletion of the mouse *Hmg1* gene, currently underway in our lab.

Building up Targeting Specificity

This brief review of HMG1 molecular and cellular biology suggests that a ubiquitous and abundant protein with no sequence specificity can be directed where DNA flexing is required by interactions with its partners. Such protein-protein interactions may well help localize sequence-specific DNA-flexing proteins, too. More than 400,000 sites with sufficient affinity for SRY are expected in the human haploid genome, a problem that is exacerbated by the existence of multiple proteins with DNA-binding domains very similar to SRY, conferring exactly the same binding specificity. There are>20 *SOX* (Sry HMG box) genes in mammals, although some have been identified only via the PCR-derived sequence coding for their conserved HMG box (Pevny and Lovell-Badge 1997). Most Sox proteins are expressed in the developing nervous system and are involved in neural determination, but a few are also expressed in other embryonic and adult tissues.

Target genes for some SOX proteins have been identified: Sox-2 binds the *Fgf-4* enhancer (Ambrosetti et al. 1997), whereas Sox-9 binds the enhancer of the collagen gene *Col2a1* (Bell et al. 1997). In vitro, the binding specificity of Sox-2, SRY, and Sox-9 HMG boxes is indistinguishable, as is the degree of bending that they impose on DNA (M. E. Bianchi, unpublished data). In vivo, some target specificity may derive from the pattern of expression of different SOXs, but a lot certainly derives from interaction with partners. Sox-2 interacts with Oct-3 (Ambrosetti et al. 1997), Sox-10 with Oct-6 (Kuhlbrodt et al. 1998*b*), and Sox-11 with Brn-1 (Kuhlbrodt et al. 1998*a*). Where tested, interactions were found to occur between the HMG box and the POU domain in the partner, suggesting that a SOX-POU code may exist. Moreover, in all cases, SOX domains and POU domains

bind to contiguous sites on the DNA, and the partners interact synergistically in driving the expression of reporter genes. As predicted by models that implicate DNA flexing, the spacing of the binding sites of the protein partners is essential for proper regulation of target genes (Ambrosetti et al. 1997; Kuhlbrodt et al. 1998*a*), and mutations that alter the DNA-bending ability of SOXs affect the transcription of target genes (P. Scaffidi and M. E. Bianchi, unpublished data).

On the basis of the features of other HMG-box proteins, SRY is strongly predicted to have a partner, which is expected to interact with it via its HMG box. Two-hybrid screens for SRY interactors have been performed in several laboratories, and, indeed, a single human 34-kD protein called "SIP-1" (SRY interacting protein) has been identified (Poulat et al. 1997). This protein is ubiquitous, nuclear, and interacts, via its two PDZ domains, with the most C-terminal seven amino acids in human SRY. No such interaction is seen with rodent Sry; because the non-HMG box segments of the Sry protein diverge from the human SRY, it is possible that SIP-1 interacts with SRY in humans only and that the mouse Sry partner might be an entirely different protein. A candidate-protein approach might work best to uncover more SRY partners. Good candidates include the transcription factors of the POU family, whose interactions with Sox proteins are well documented but are not evident by the two-hybrid method. Several nuclear hormone receptors and homeodomain proteins have been shown to interact with HMG1 and, therefore, should be considered as potential SRY partners. At a minimum, it is clear that the molecular and cellular biology of HMG1 and other DNA-flexing proteins have provided a powerful inspiration to understand the three-dimensional aspects of gene regulation.

Acknowledgments

The authors are supported by grants from Fondazione Telethon, Associazione Italiana Ricerca sul Cancro, and the TMR Program of the European Union.

References

Ambrosetti D-C, Basilico C, Dailey L (1997) Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol 17:6321–6329

Bell DM, Leung KKH, Wheatley SC, Ng LJ, Zhou S, Ling KW, Sham MH, et al (1997) SOX9 directly regulates the type-II collagen gene. Nat Genet 16:174–178

Bianchi ME (1995) The HMG-box domain. In: Lilley DMJ (ed) DNA-protein: structural interactions, Oxford University Press, Oxford, pp 177–200

Bianchi ME, Beltrame M, Paonessa G (1989) Specific recognition of cruciform DNA by nuclear protein HMG1. Science 243:1056–1059

Boonyaratanakornkit V, Melvin V, Prendergast P, Altmann M, Ronfani L, Bianchi ME, Tarasaviciene A, et al (1998) High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. Mol Cell Biol 18:4471–4487

Bustin M, Reeves R (1996) High mobility group chromosomal proteins: architectural components that facilitate chromatin function. Prog Nucleic Acid Res Mol Biol 54:35–100

Cantor CR, Schimmel PR (1980) Biophysical chemistry. WH Freeman, San Francisco

Falciola L, Spada F, Calogero S, Längst G, Voit R, Grummt I, Bianchi ME (1997) High mobility group 1 (HMG1) protein is not stably associated with the chromosomes of somatic cells. J Cell Biol 137:19–26

Goodfellow PN, Lovell-Badge R (1993) SRY and sex determination in mammals. Annu Rev Genet 27:71–92

Jayaraman L, Moorthy NC, Murthy KGK, Manley JL, Bustin M, Prives C (1998) High mobility group protein-1 (HMG1) is a unique activator of p53. Genes Dev 12:462–472

Kuhlbrodt K, Herbarth B, Sock E, Enderich J, Hermans-Borgmeyer I, Wegner M (1998*a*) Cooperative function of POU proteins and SOX proteins in glial cells. J Biol Chem 273: 16050–16057

Kuhlbrodt K, Herbarth B, Sock E, Hermans-Borgmeyer I, Wegner M (1998*b*) Sox10, a novel transcriptional regulator in glial cells. J Neurosci 18:237–250

Landsman D, Bustin M (1991) Assessment of the transcriptional activation potential of the HMG chromosomal proteins. Mol Cell Biol 11:4483–4489

Pevny LH, Lovell-Badge R (1997) Sox genes find their feet. Curr Opin Genet Dev 7:338–344

Pil PM, Lippard SJ (1992) Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin. Science 256:234–236

Pontiggia A, Rimini R, Harley VR, Goodfellow PN, Lovell-Badge R, Bianchi ME (1994) Sex-reversing mutations affect the architecture of SRY-DNA complexes. EMBO J 13: 6115–6124

Poulat F, Barbara PS, Desclozeaux M, Soullier S, Moniot B, Bonneaud N, Boizet B, et al (1997) The human testis determining factor SRY binds a nuclear factor containing PDZ protein interaction domains. J Biol Chem 272:7167–7172

Sutrias-Grau M, Bianchi ME, Bernués J. HMG1 interacts with the core domain of human TBP and interferes with TFIIB within the pre-initiation complex. J Biol Chem (in press)

Tjian R, Maniatis T (1994) Transcriptional activation: a complex puzzle with few easy pieces. Cell 77:5–8

Tucker PK, Lundrigan BL (1993) Rapid evolution of the sex determining locus in Old World mice and rats. Nature 364: 715–717

Vaccari T, Beltrame M, Ferrari S, Bianchi ME (1998) *Hmg4*, a new member of the *Hmg1/2* gene family. Genomics 49: 247–252

Werner MH, Huth JR, Gronenborn A, Clore GM (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell 81:705–714

Whitfield LS, Lovell-Badge R, Goodfellow PN (1993) Rapid sequence evolution of the mammalian sex-determining gene *SRY*. Nature 364:713–715

Zappavigna V, Falciola L, Helmer Citterich M, Mavilio F, Bianchi ME (1996) HMG1 cooperates with HOX proteins in DNA binding and transcriptional activation. EMBO J 15: 4981–4991

Zwilling S, König H, Wirth T (1995) High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. EMBO J 14:1198–1208