

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Characterization and mutations in the human sex determining factor SRY

### Thesis

How to cite:

Pontiggia, Andrea (1998). Characterization and mutations in the human sex determining factor SRY. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1998 The Author

Version: Version of Record

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

31 0219257 2



UNRESTRICTED

CHARACTERIZATION AND MUTATIONS IN THE HUMAN  
SEX DETERMINING FACTOR SRY

ANDREA PONTIGGIA

A thesis submitted in partial fulfilment of the requirements of the  
Open University  
for the Degree of Doctor of Philosophy

June 1998

Author number: P9275825  
Date of award: 13 October 1998

DIBIT  
Department of Biological and Technical Research  
Milan, Italy

**BEST COPY**

**AVAILABLE**

If binding and bending are really important for the biological function of SRY proteins, these features will be positively selected and conserved during evolution. The second part of this work demonstrates that the architecture and stability of nucleoprotein complexes formed by the DNA-binding domains of SRY proteins is indeed of primary importance. It is shown that these parameters are conserved amongst primates, irrespective of the time since divergence from a common ancestor or the number of residues which have changed during evolution.



## **Acknowledgments**

I wish here to thank all the members of the DNA-proteins interactions laboratory of DIBIT, San Raffaele Scientific Institute, Milan for the help they gave me in so many different ways.

In particular I owe a lot to Marco Bianchi, for his patience and unfailing guidance and support he gave me throughout the duration of this project. I am also very grateful to all the people I turned for help and advice, among them Vince Harley, Simon Whitfield and Peter Goodfellow.

I am very grateful to the people of Glaxo Wellcome Research Center, and among them to Emiliangelo Ratti and Harald Mottl who encouraged me to finish this thesis and supported me with all their patience and help.

Immense thanks also to Daniel Simon and Alessandra Polissi whom, with their friendship and lighthearted helping made more comprehensible my poor English.

Finally, I owe a great deal to Anna, not only with the proofreading of this thesis, but for the boundless support and cooperation she gave me at all times.

## **Declaration**

I hereby declare that I alone have written this thesis and that, except where stated, the work described is my own.

**Andrea Pontiggia**

# CONTENTS

|   |               |
|---|---------------|
| <b>CHAPTER 1: INTRODUCTION</b>  | <b>PAG.12</b> |
| 1.1 SRY: A SEXY STORY   | PAG.14        |
| 1.2 SEXUAL COMMITMENT AND SRY TRANSCRIPTION   | PAG.16        |
| 1.3 PATHWAYS AND GENES INVOLVED IN SEX DIFFERENTIATION  | PAG.19        |
| 1.4 SRY PROTEIN AND HMG BOX DOMAIN: A GENERAL OVERVIEW  | PAG.23        |
| 1.5 DNA BINDING   | PAG.25        |
| 1.6 BINDING, BENDING AND BIOLOGICAL EFFECT OF HMGBOX.   | PAG.27        |
| <br>  |               |
| <b>CHAPTER 2: RESULTS</b>   | <b>PAG.30</b> |
| 2.1 PRODUCTION OF WILD TYPE AND MUTANT SRY DNA BINDING DOMAINS                                    | PAG.31        |
| 2.2 THE AFFINITY OF WT AND MUTANT HMG-BOXES FOR BENT DNAS   | PAG.34        |
| 2.3 THE DNA-BENDING ACTIVITY OF NORMAL AND MUTANT SRY BOXES                                       | PAG.39        |
| 2.4 WILD TYPE SRY AND THE M64I MUTANT RECOGNISE THE SAME<br>REPERTOIRE OF SEQUENCES IN LINEAR DNA | PAG.44        |
| 2.5 DNA SEQUENCE AFFECTS DNA BENDING BY SRY   | PAG.47        |
| 2.6 THE POSITIVE CONTROL: THE EVOLUTIONARY TREE OF PRIMATES                                       | PAG.49        |
| 2.7 PRODUCTION OF SRY DNA BINDING DOMAINS OF DIFFERENT PRIMATES                                   | PAG.50        |
| 2.8 ARCHITECTURE OF THE DNA COMPLEXES INDUCED BY PRIMATE SRY<br>PROTEINS                          | PAG.52        |
| <br>  |               |
| <b>CHAPTER 3: DISCUSSION</b>  | <b>PAG.55</b> |
| 3.1 HOW DOES SRY WORK?  | PAG.56        |
| 3.2 DISSECTION OF THE DNA BINDING AND DNA BENDING ACTIVITIES OF<br>SRY PROTEIN                    | PAG.58        |

|      |   |        |
|------|---|--------|
| 3.3  | THE CONTRIBUTION OF THE DNA SEQUENCE TO THE ORGANISATION OF SRY-INDUCED NUCLEOPROTEIN COMPLEXES.                                | PAG.61 |
| 3.4  | CORRELATION BETWEEN THE DNA BINDING PROPERTIES OF THE SRY BOX MUTANTS AND THE PHENOTYPE OF THE AFFECTED PATIENTS                | PAG.62 |
| 3.5  | MOLECULAR BASIS OF HUMAN 46X,Y SEX REVERSAL REVEALED FROM THE THREE DIMENSIONAL SOLUTION STRUCTURE OF THE HUMAN SRY-DNA COMPLEX | PAG.65 |
|      | <b>CHAPTER 4: MATERIALS AND METHODS</b>   | PAG.67 |
| 4.1  | ALKALINE LYSIS MINIPREP   | PAG.68 |
| 4.2  | PREPARATION OF CRUDE LYSATE BY ALKALINE LYSIS: MAXI PREP  | PAG.70 |
| 4.3  | TRANSFORMATION USING CALCIUM CHLORIDE   | PAG.76 |
| 4.4  | PREPARE COMPETENT CELLS   | PAG.77 |
| 4.5  | AGAROSE GEL ELECTROPHORESIS   | PAG.81 |
| 4.6  | RESOLUTION OF DNA FRAGMENTS ON STANDARD AGAROSE GELS  | PAG.81 |
| 4.7  | DIGESTING A SINGLE DNA SAMPLE WITH RESTRICTION ENDONUCLEASES  | PAG.86 |
| 4.8  | DNA-DEPENDENT DNA POLYMERASES   | PAG.87 |
| 4.9  | ENZYME KLENOW FRAGMENT OF ESCHERICHIA COLI DNA POLYMERASE   | PAG.90 |
| 4.10 | LABELING CIRCULAR PERMUTE PROBS AT 3' ENDS OF DNA.  | PAG.90 |
| 4.11 | REPAIRING 3' OR 5' OVERHANGING ENDS TO GENERATE BLUNT ENDS.   | PAG.92 |
| 4.12 | ENZYME T4 DNA POLYMERASE  | PAG.93 |
| 4.13 | ENZYME ALKALINE PHOSPHATASES: BACTERIAL ALKALINE PHOSPHATASE  | PAG.95 |
| 4.14 | CIP REACTION. CONDITION FOR DEPHOSPHORYLATION OF DNA  | PAG.96 |
| 4.15 | ENZYME T4 POLYNUCLEOTIDE KINASE   | PAG.97 |



|   |         |
|---|---------|
| 4.16 PHOSPHORYLATING SYNTHETIC OLIGONUCLEOTIDES<br>BY THE FORWARD REACTION                  | PAG.97  |
| 4.17 LABELING 5' TERMINI BY THE EXCHANGE REACTION   | PAG.99  |
| 4.18 ENZYME T4 DNA LIGASE   | PAG.100 |
| 4.19 SUBCLONING OF DNA FRAGMENTS  | PAG.101 |
| 4.20 ALTERNATIVE PROTOCOL: LIGATION OF DNA FRAGMENTS IN GEL SLICES                          | PAG.105 |
| 4.21 CONSTRUCTION OF RECOMBINANT DNA FRAGMENT OF SRY BOXES BY<br>POLYMERASE CHAIN REACTION. | PAG.107 |
| 4.22 SUBCLONING DNA FRAGMENTS   | PAG.108 |
| 4.23 DNA SEQUENCING STRATEGIES  | PAG.114 |
| 4.24 ALKALI DENATURATION OF DOUBLE-STRANDED<br>PLASMID DNA FOR DIDEOXY SEQUENCING           | PAG.114 |
| 4.25 LABELING/TERMINATION SEQUENCING REACTIONS<br>USING SEQUENASE                           | PAG.116 |
| 4.26 COLORIMETRIC METHODS FOR QUANTIFYING PROTEINS  | PAG.121 |
| 4.27 BRADFORD METHOD  | PAG.121 |
| 4.28 GEL ELECTROPHORESIS  | PAG.123 |
| 4.29 DENATURING (SDS) DISCONTINUOUS<br>GELELECTROPHORESIS: LAEMMLI GEL METHOD               | PAG.123 |
| 4.30 ALTERNATIVE PROTOCOL 1: ELECTROPHORESIS IN TRIS-TRICINE BUFFER<br>SYSTEMS              | PAG.133 |
| 4.31 ALTERNATIVE PROTOCOL: RAPID COOMASSIE BLUE STAINING                                    | PAG.137 |
| 4.32 SILVER STAINING  | PAG.139 |
| 4.33 PURIFICATION OF SRY BOXES  | PAG.142 |
| 4.34 MOBILITY SHIFT DNA-BINDING ASSAY USING GEL ELECTROPHORESIS                             | PAG.145 |
| 4.35 MAKING THE LABELED CROSS   | PAG.150 |
| 4.36. CIRCULAR PERMUTATION ASSAY  | PAG.152 |
| 4.37 CALCULATION OF DNA BEND PARAMETERS   | PAG.153 |
| 4.38 DETERMINATION OF COMPLEX DISSOCIATION CONSTANTS  | PAG.156 |

|   |                |
|---|----------------|
| <b>4.39 DETERMINATION OF PROTEIN-DNA SEQUENCE</b>                 |                |
| <b>SPECIFICITY BY PCR-ASSISTED BINDING-SITE SELECTION</b>         | <b>PAG.156</b> |
| <b>4.40 ISOLATION AND ANALYSIS OF BOUND OLIGONUCLEOTIDES FROM</b> |                |
| <b>MOBILITY SHIFT GELS</b>  | <b>PAG.164</b> |
| <b>REFERENCES</b>   | <b>PAG.165</b> |

## FIGURES

- FIG. 1 ALIGNMENT OF HMG BOXES. PAG. 32 B
- FIG. 2 PURIFICATION OF THE E.COLI HMG BOXES. PAG. 33 B
- FIG. 3 DESIGN OF AN ARTIFICIAL CRUCIFORM DNA. PAG. 35 B
- FIG. 4 PEPTIDE HSRYBOX BOTH RECOGNIZE FOUR-WAY JUNCTION. PAG. 35 C
- FIG. 5 DESIGN OF AN ARTIFICIAL CRUCIFORM DNA PAG. 36 B
- FIG. 6 PEPTIDE HSRYBOX DOES NOT RECOGNIZE LINEAR CONTROL  
DNAS AZ AND BZ AT THE CONCENTRATIONS OPTIMAL FOR  
COMPLEX FORMATION WITH JUNCTION Z. PAG. 36 C
- FIG. 7 DNA BINDING PROPERTIES OF SOME HSRYBOXES. PAG. 38 B
- FIG. 8 CIRCULAR PERMUTED PROBES. PAG. 40 B
- FIG. 9 CIRCULAR PERMUTATION ANALYSIS OF DNA FLEXURE  
INDUCED BY BINDING OF HUMAN SRY TO  
THE AACAAAG SEQUENCE. PAG. 40 C
- FIG. 10 A AND B. DNA BENDING ACTIVITY OF MUTANT SRY BOXES  
ON THE CD3E SEQUENCE. PAG. 41 C
- FIG. 10 A AND C. DNA BENDING ACTIVITY OF MUTANT SRY BOXES  
ON THE CD3E SEQUENCE. PAG. 41 D
- FIG. 10 A AND D. DNA BENDING ACTIVITY OF MUTANT SRY BOXES  
ON THE CD3E SRY PROTEINS. PAG. 41 E
- FIG. 11 THE HMG BOXES OF WT SRY AND MUTANT M64I RECOGNISE  
THE SAME REPERTOIRE OF SITES. PAG. 45 B
- FIG. 12 WT SRY BOX AND MUTANT M64 RECOGNISE  
THE SAME REPERTOIRE OF SITES. PAG. 45 C
- FIG. 13 DNA BINDING SITES WITH DIFFERENT SEQUENCES ARE  
BENT DIFFERENTLY BY SRY BOXES. PAG. 48 C
- FIG. 14 ALIGNMENT OF HMG-BOX SEQUENCES FROM SRY  
PROTEINS OF PRIMATES. PAG. 50 B

|         |   |           |
|---------|---|-----------|
| FIG.15  | DNA BINDING AND BENDING PROPERTIES OF HMG-BOXES<br>OF PRIMATE SRY PROTEINS.                             | PAG. 52 B |
| FIG. 16 | CIRCULAR PERMUTATION ANALYSIS OF THE DNA FLEXURE<br>INDUCED SRY-BOXES OF MARMOSET AND PYGMY CHIMPANZEE. | PAG. 53 B |
| FIG. 17 | A MODEL FOR THE ACTION OF SRY PROTEIN.  | PAG. 56 B |
| FIG. 18 | THERMODYNAMICS OF THE BINDING OF SRY BOXES TO DNA.  | PAG. 58 B |
| FIG.19  | A SCHEMATIC MODEL FOR THE INTERACTION OF SRY PROTEINS<br>WITH LINEAR AND FOUR-WAY JUNCTION DNAS.        | PAG. 59 B |
| FIG. 20 | THE INTERACTION OF HSRYBOX WITH DNA   | PAG. 65 B |
| FIG. 21 | RIBBON DRAWING OF HSRYBOX   | PAG. 66 B |
| FIG. 22 | SUMMARY OF THE CONTACTS BETWEEN HSRYBOX AND DNA   | PAG. 66 C |



## **TABLES**

|                |  |                  |
|----------------|--|------------------|
| <b>TABLE 1</b> | <b>SRY MUTATIONS CHARACTERISTICS AND PROPERTIES OF<br/>CRUCIFORM DNA-SRY BOXES INTERACTION.</b>  | <b>PAG. 38 C</b> |
| <b>TABLE 2</b> | <b>SUMMARY OF SRY MUTATION CHARACTERISTICS AND<br/>PROPERTIES OF THE SRY BOXES.</b>  | <b>PAG. 41 B</b> |
| <b>TABLE 3</b> | <b>INFLUENCE OF THE DNA SEQUENCE ON THE AMPLITUDE OF<br/>THE ANGLE INDUCED BY SRY PROTEINS.</b>  | <b>PAG. 48 B</b> |
| <b>TABLE 4</b> | <b>NO CORRELATION BETWEEN THE TIME OF DIVERGENCE<br/>FROM THE HUMAN LINE OF DESCENT AND THE AMPLITUDE<br/>OF THE DEFLECTION IN DNA INDUCED BY SRY.</b> | <b>PAG. 53C</b>  |

# **INTRODUCTION**

Since 1953 (Jost, 1953) it was known that the presence of testes during mammalian embryogenesis resulted in male differentiation of the internal reproductive organs and external genitalia, while the presence of ovaries or the complete absence of gonads resulted in female differentiation. He deduced that the secretion of testicular hormones during a critical period in embryogenesis redirected differentiation of the reproductive tract from a female to a male pathway. Subsequently, the presence or absence of a Y chromosome was shown to correlate with male or female phenotype: XY and XXY individuals are male, while XO and XX individuals are female. Yet the studies by Jost (1953), made clear that the Y chromosome-encoded mediator of sex determination needed only specific male gonadal development, and it was therefore called testis determining factor (TDF in humans, Tdy in mice). It is now accepted that, apart from the gonads, all morphological differences between male and female eutherian mammals result from the action of two testicular hormones, Mullerian inhibiting substance (MIS or AMH) and testosterone. MIS is a member of the TGF beta family of growth factor, is secreted by Sertoli cells and causes regression of the Mullerian ducts, which would otherwise develop to form the uterus, cervix, fallopian tubes, and part of the vagina. Testosterone, secreted by Leydig cells, induces the development of male structures, derived from the Wolffian duct, including the epididymis, vas deferens, and seminal vesicles. The formation of male rather than female external genitalia requires the conversion of testosterone into dihydrotestosterone by steroid alpha reductase in the target tissue. The embryogenesis of mammalian sexual dimorphism has thus been divided into primary and secondary sex differentiation, the former referring to the development of the bipotential gonads into testes or ovaries, and the latter evocative of subsequent hormonal effects.

## **1.1 SRY: A SEXY STORY**

In a set of experiments, Koopman et al, (1991) demonstrated that XX mice differing from their sisters by addition of a single 14 Kb DNA fragment from the mouse Y chromosome including the *SRY* gene, can develop as phenotypic male. Though devoid of spermatogonia, the testes of these transgenic XX male mice appear to have a normal hormonal function; the mice display normal male secondary sexual differentiation and mating behavior. In contrast, XY mice differing from normal XY males only by the deletion of a Y DNA fragment, 11 Kb including the *SRY* gene, develop as phenotypic females. These XY female mice are fertile and can transmit the mutant Y chromosome to their offspring (Lovell-Badge and Robertson, 1990; Gubbay et al., 1992). Thus, as judged by gonadal histology and hormone production, the presence and absence of SRY is sex determining in mice. Interpretation of the data from humans is less straightforward, though rare sex-reversed individuals provide counterparts to both XX male and XY female mice. Crossing over normally occurs near the short arm telomers of the human X and Y chromosomes during male meiosis, and aberrant recombination occurring more proximal can transfer fragments of Y-specific DNA (including SRY) into the genome of an otherwise XX individual. Most such translocations are cytologically undetectable, and the resulting XX males do not produce spermatozoa but have normal male internal and external genitalia. When small (35 Kb) fragments of Y-specific DNA are translocated, the resulting XX+SRY individuals invariably have genital abnormalities, sometimes despite apparently normal testicular histology (Palmer et. al., 1989). These abnormalities may include "sexually ambiguous" external genitalia (incomplete fusion of the labioscrotal folds with a resultant bifid scrotum or displaced urethra opening), undescended testes, or



ovotestes (gonads containing areas of both testicular and ovarian histology) with persistence of female structures derived from the Mullerian duct. There are several possible explanations for the phenotypic differences between these  $XX+SRY$  humans and  $XX+SRY$  transgenic male mice. Perhaps the most likely scenario is that *SRY* transcription is reduced in  $XX+SRY$  humans (Palmer et. al., 1989). This could result from the spread of X inactivation across a nearby X;Y translocation breakpoint in humans. The murine *SRY* transgene would not be subject to this effect if it inserted at an autosomal locus. Alternatively, loss of regulatory sequences 5' of the human *SRY* gene might result from X;Y translocation involving only small amounts of Y-specific DNA. It is unclear whether reduced expression of *SRY* could explain the phenotype; presumably decreased or delayed hormonal action would have to result. It remains a formal possibility that, in humans, but not in mice, full development of the male phenotype requires expression of a second Y-linked gene, located near *SRY* and perhaps expressed outside the testis. In contrast with XY female mice, XY female humans are sterile; some of these individuals have mutations affecting the *SRY* open reading frame (Berta et. al., 1990). Such XY individuals fail to develop mature testes or ovaries and instead form poorly differentiated gonads without clear male or female histology. Because the gonads lack male hormonal function, Mullerian structures persist and female secondary sexual differentiation follows. Infertility may result in part from X chromosome monosomy: unlike X0 mice, X0 humans are sterile.

## **1.2 SEXUAL COMMITMENT AND *SRY* TRANSCRIPTION.**

The mammalian gonad is composed of germ cells and three types of somatic cells. Primordial germ cells, which are first observed in the extraembryonic mesoderm (Ginsburg et al., 1990), migrate into the primitive gonad at 10.5-12 days post coitum (dpc) in the mouse. The somatic portion of the gonad is composed of supporting cells (Sertoli cells in the male and follicle cells in the female, thought to derive from common progenitors), steroidogenic cells (Leydig cells in the male and theca cells in the female), and connective tissue cells. When is gonadal sex determined? Though the sexual fate of an embryo is set at fertilisation, it is not known when gonadal cells become committed to male or female development. XX and XY mouse embryos are morphologically indistinguishable until 11.5-12.5 dpc, when pre Sertoli cells align to form testis cords in the XY gonad. By this time a male or female developmental program has clearly been initiated. Yet it is not obvious that sex determination (commitment) is coincident with the onset of sex differentiation (e.g. histologic change). When does the presence or absence of *SRY* protein result in irreversible commitment to male or female differentiation? Barring pleiotropic effects of *SRY*, it is reasonable to suppose that sex differentiation begins when *SRY* is first expressed. Initial experiments suggested that this is from 10.5 to 12.5 dpc in the murine gonads. *Sry* was previously shown to be expressed by germ cells in adult mouse testis surprisingly as a circular transcript that seems not to be translated (Capel et al., 1993). The embryonic promoter is turned on specifically in the genital ridge at about 10.5-11 dpc, just before the morphological differentiation between sexes is observed and is turned off at about 12.5 dpc. *Sry* expression in the preimplantation mouse embryo has been reported (Zwingman et al., 1993). Although there is still doubt about these



findings, the transcript present in pre implantation embryos may have no function and they could all be circular. Moreover there is no evidence for any function for *Sry* at this stage and a role in sex determination seem highly unlikely.

If it is not completely clear when sex determination begins, it is equally unclear when it finishes. By 13.5 dpc, *Sry* transcripts are no longer detectable by PCR of gonadal RNA (Koopman et al., 1990). While it is not known how long SRY protein persists, the function of the Y chromosome in determining the sex of the embryo appears to be complete by this time. Given that other proteins must be involved, sexual commitments could still be in flux at the time of the disappearance of SRY.

In which cells do the initial steps in gonadal sex determination occur? Not in the germ cells: in mouse embryos that are defective in migration of primordial germ cells to the gonad, the process of gonadal differentiation is otherwise undisturbed (reviewed by McLaren, 1991). XY animals still develop testes, and XX animals still begin to develop ovaries (although follicles will not form). Sex determination must unfold in the somatic cell lineage of the gonad.

In which of the somatic lineage, then, does gonadal sex determination occur? Experiments involving XX -- XY chimeric mice, reminiscent of mosaic analysis in invertebrates, have sought to identify gonadal cell lineages that become committed to male or female development as a direct result of the presence or absence of the Y chromosome. These should be the cell lineages in which sex determination initiates. If sex determination occurs autonomously in each gonadal cell, then all of XY cells in a chimeric gonad should be male and all of XX cells should be female. If, on the other hand, the sexual phenotype of a gonadal cell is determined in consultation with its neighbors, then the strict correlation of chromosome constitution with cellular phenotype should break down. In chimeric mice with testes, the proportion of XX and

XY cells in the Leydig population are similar to those seen outside the gonad, implying that sex determination does not occur autonomously in Leydig cell precursors. In contrast, most but not all Sertoli cells are XY, indicating some direct action by the Y chromosome, presumably by *Sry* (Patek et al, 1991; Palmer and Burgoine, 1991). The presence of a few XX Sertoli cells in XX --XY embryos indicates that cells lacking *Sry* can be recruited into the Sertoli population. The mechanism by which such recruitment occurs is unknown. Complementary studies in XX-XY chimeras with ovaries show that XY precursors can become follicle cells, the female counterparts of Sertoli cells. The major conclusion that can be drawn from these experiments is that the initial steps that commit the gonad to male or female differentiation likely occur in the supporting cell lineage (i.e., in pre Sertoli or prefollicle cells). Since *Sry* encodes a nuclear factor and must be expressed and function within the cells that carry out these steps, it is presumably expressed in pre-Sertoli cells.



### **1.3 PATHWAYS AND GENES INVOLVED IN SEX DIFFERENTIATION**

Apart from what was reviewed above, rather little is known about the molecular nature of gonadal differentiation in mammals, although there is a substantial knowledge of the cellular processes that occur during early testis and ovary development. The first cells to be differentiated by the action of *SRY* in the developing genital ridge are the Sertoli cells, and they in turn direct the differentiation of the other cell type in the testis (Burgoyne et al., 1988). The timing of *SRY* action during development is critical, the *SRY* gene is normally necessary and sufficient to determine testis differentiation, but nothing is known about the genes it regulates to produce this effect.

Sex reversed patients have been the main source of information on candidate sex determination genes. The establishment of epistatic relationship between these genes and *SRY*, however, must await their analysis in the mouse. The gene for a severe dwarfism syndrome, campomelic dysplasia, often associated with XY female sex reversal, has been cloned (Wagner et al., 1994; Wright et al. 1995; Foster et al. 1994). The gene encodes a protein related to *SRY*, and has been named Sox9. The phenotype of these patients is associated with inactivating mutations within the open reading frame of Sox 9 or with translocations at some distance upstream of the gene. These mutations are present in only one allele of the gene, suggesting that sex reversal and campomelic dysplasia in these patients could be due to haploinsufficiency of the Sox 9 gene product. Sox9 expression was found to closely follows differentiation of Sertoli cells in the mouse testis and the timing and the cell type specificity of Sox9 expression suggests that Sox9 may be directly regulated by *SRY* (Da Silva et al., 1996; Kent et al., 1996). Another case of dosage sensitive sex reversal in humans is

found in XY females that have duplications in a region of the short arm of the X chromosome. One model for this effect assumes that this region contains a gene, termed *DSS* for dosage-sensitive sex reversal, which encodes a female specific function. Perhaps this is negatively regulated by *SRY* in males, but this repression fails when the gene is present in a double dose or not X inactivated. A gene that encodes a member of the nuclear hormone receptor superfamily has been cloned from this region. This gene is responsible for the condition known as adrenal hypoplasia congenital (AHC) and has therefore now been named *DAX-1* for DSS-AHC critical region on the X chromosome (Zanaria et al., 1994; Bardoni et al., 1994). *Dax-1* is expressed in the genital ridge in mice at 11.5 dpc, consistent with a role in sex determination, but apparently at similar levels in both female and males. By 12.5 dpc, *Dax-1* is switched off in the male gonad, but remains on in females (Swain et al 1996). If it is a candidate for a female specific function that is negatively regulated in males, any repression will be post transcriptional. Studies in mice on another member of the nuclear hormone receptor superfamily, the steroidogenic factor 1 (*SFI*) gene, has shown that it is involved in gonad development and possibly in sex determination. The *SFI* gene was first identified as a factor that regulates steroidogenic enzymes in the adrenal cortex and gonads (Rice et al., 1991; Lala et al., 1992; Honda et al., 1993). However, *SFI*-deficient mice showed a complete absence of gonads as well as adrenals, implying that the gene has a function in the initial development of these organs before *SRY* is turned on (Luo et al., 1994). *SFI* has also been implicated in sex determination because of its sexually dimorphic expression in the gonad. In fact during the sexually undifferentiated stage of gonadal development (9-12 dpc), all embryos express *SFI* in the genital ridge. As testicular cords form in males, *SFI* transcripts are diffusely expressed throughout the testes. By Northern blotting and in

situ hybridisation *SF1* expression is shown in Sertoli cells (Ikeda et al., 1994). In contrast to its persistent expression in the embryonic testes, *SF1* transcripts disappear from the ovary between 13-16.5 dpc, reappearing only during late gestation (Ikeda et al., 1994). Coupled with the demonstration of *SF1* mRNA in Sertoli cells, these data suggest that *SF1* plays a role in gonadal development. Indeed male and female *SF1* null mice had female internal genitalia, despite complete gonadal dysgenesis. These studies establish that the gene coding SF1 is essential for sexual differentiation and formation of the primary steroidogenic tissues (Luo et al., 1994). Furthermore, in primary Sertoli cells, SF1 regulates the *MIS* gene by binding to a conserved upstream regulatory element and can activate its transcription in cotransfection assays. In heterologous HeLa cells, *MIS* gene activation by SF1 requires removal of the SF1 ligand-binding domain, implicating a Sertoli cell-specific ligand or cofactor (Shen et al., 1994). Since sexually dimorphic expression of SF1 during development coincides with *MIS* expression and Mullerian duct regression, SF1 is proposed to regulate *MIS* in vivo and participate directly in the process of mammalian sex differentiation (Shen et al., 1994). However, because it is expressed earlier and in both sexes, SF1 alone cannot account for the testis-specific activation of *MIS* expression at 11.5 dpc. Several studies have implied that *SRY* does not directly activate *MIS* but acts through an unidentified factor. Differential patterns of expression in the male mouse gonad at the time where *SRY* is thought to act are found for *MIS* and for desert hedgehog, a member of a family of proteins that are known to be important in cell signalling. However, mice deficient in either of these genes do not show a phenotype consistent with a role in sex determination (Behringer et al., 1994). Recent results show that *AMH* mutations in mice modify the regression of Mullerian ducts and suggested that it might be involved in the control of proliferation of steroidogenic cells in the testis.

**From all these results sex determination in mammals appears to be a puzzle of several genes whose function is far from completely understood.**



#### **1.4 SRY PROTEIN AND HMG BOX DOMAIN: A GENERAL OVERVIEW.**

The human *SRY* gene encodes a protein with a central high mobility group domain (HMG box) of about 78 amino acids. This box is also present and conserved in several eukaryotic proteins. It is a DNA binding domain recently discovered and first recognised by sequence alignments of hUBF, a transcription factor for human RNA polymerase I, with HMG1, an abundant and strongly conserved component of mammalian chromatin (Jantzen et al., 1990). Several additional members of the HMG family have since been described. Sequence analysis of this family indicates that the HMG box is a minimally stretch of about 70 amino acids, with a net positive charge and an abundance of aromatic residues and prolines. Among the members of this family the similarity of the primary sequences of HMG boxes is modest: no residue is absolutely conserved, and just three residues show only conservative substitutions in all known HMG boxes (Fig. 1). The sequence variation between HMG boxes is paralleled by the diversity of their presumed biochemical functions. One subgroup of proteins is clearly related to HMG1 and comprises structural components of eukaryotic chromatin. HMG1-like proteins are present in all eukaryotes and in all tissues of higher organisms. However their physiological function remains elusive; roles have been suggested in DNA replication, in nucleosome assembly and in transcription (reviewed by Bustin et al., 1990, and Bianchi et al., 1992a). Another subgroup of proteins comprises general transcription factors of RNA polymerase I and mitochondrial RNA polymerase; some of these proteins, such as hUBF (Jantzen et al., 1990), contain HMG boxes as moderately repeated elements. A third subgroup has recently attracted much interest: it comprises a number of fungal proteins involved in mating-type expression (Staben and Yanofsky, 1990; Sugimoto et al., 1991), the

mammalian testis determining factor SRY (Gubbay et al., 1990; Sinclair et al., 1990), the protein product of Sox genes and a set of lymphoid-specific enhancer binding factors (Travis et al., 1991; van de Wetering et al., 1991; Waterman et al., 1991). These proteins are most likely transcriptional regulators: protein LEF-1, for example, is expressed specifically in pre-B and T lymphocytes and is involved in the transcription of the gene for the T cell receptor alpha chain (Travis et al., 1991; Waterman et al., 1991).

Up to now there is no clear evidence of the function of the portion of SRY proteins outside the HMG box. Also within mammals, the comparison of SRY proteins during evolution shows that protein segments flanking the HMG-box are highly variable in sequence and length and cannot be aligned (Whitfield et al., 1993; Tucker et al., 1993). This feature is also shared within primates, where the C-terminal domain varies extensively. All these data may be explained by if all the functions of SRY reside in the highly conserved HMG-box. Another possibility is that natural selection favors the fast adaptive divergence of the non-box regions. In accordance with the first hypothesis, the importance of the HMG box is underlined by the analysis of human sex reversed XY females. All the single point mutations of the *hSRY* gene found in patients associated with sex reversal phenotypes fall within the HMG box.

Both NLS motifs are conserved in the HMG-boxes of other transcription factors, and are supposed to act as a bipartite basic motif, both able to independently direct the protein into the nucleus.

## **1.5 DNA BINDING**

HMG boxes from all three subgroups have been shown to be necessary and sufficient to bind DNA: the two HMG boxes of HMG1 have been produced in *E. coli* as separate polypeptides, which bind to DNA with about the same affinity and specificity as the full-length protein (Bianchi et al., 1992b); *Xenopus* UBF binds to DNA via its boxes (McStay et al., 1991); some mutations in the HMG box of human SRY can abolish its capacity to bind to DNA as well as its biological activity (Berta et al., 1990; Harley et al., 1992), and the isolated HMG box of LEF-1 retains the DNA binding properties of the whole protein (Giese et al., 1992). HMG boxes are therefore authentic DNA binding domains, which can fold independently from the rest of the polypeptide. The nature of the DNA targets recognised by HMG boxes, however, is not obvious. A common feature of all HMG boxes is that they have considerable affinity and specificity towards bent DNA, e.g. as cruciform structures, irrespective of their sequences. The HMG box chromatin proteins (subgroup 1) seem to be indifferent to DNA sequence information. They only share with the HMG box family the considerable affinity and specificity towards four-way DNA junctions and cruciform structures, such as those extruded from inverted repeat sequences under the effect of supercoiling (Bianchi, 1994; Lilley, 1992). The HMG-box proteins of subgroup 2 (the nucleolar and mitochondrial transcription factors), while still endowed with sufficient sequence discrimination to produce specific footprints, do not bind to DNA sites with a recognisable consensus sequence (Pikaard et al., 1990 a,b; Parisi and Clayton, 1991). Therefore the binding to DNA does not seem to depend entirely on sequence recognition. The transcriptional regulators (subgroup 3) produce specific footprints on DNA, spanning sequences with a recognisable consensus



(Waterman et al., 1991; Nasrin et al., 1991; Travis et al., 1991). Human SRY can recognize synthetic duplex DNA fragments of the sequence AACAAAG (Harley et al., 1992). Methylation interference and base substitution experiments show that LEF-1 and the related TCF-1 proteins recognize the AACAAAG motif predominantly through minor groove contacts (Giese et al., 1991, 1992; van de Wetering and Clevers, 1992), where the recognition of bases is therefore restricted. This limits their binding specificity; for example numerous substitutions in the target site of SRY are allowed (Harley et al., 1992) and specific and non-specific binding affinities differ by a factor of only 20 to 50 (Ferrari et al., 1992). Upon binding, these proteins distort the DNA to dramatic extents: SRY produces a bend of about 80 degrees centered at the GAACAAAG sequence in the CD3e enhancer (Ferrari et al., 1992). Genetic experiments have identified the weakly related sequences RAACAAAGAA and GAGAACAATRR as the consensus target sites of the fungal protein Ste11 (Sugimoto et al., 1991). In general, all the binding sites for the HMG-box transcriptional regulators are AT-rich, and the same sequences are recognised by several proteins of this group. HMG box transcription regulators of this family upon binding induce a sharp angle in their DNA consensus sequence. Although fairly low, the sequence specificity of each individual protein is compatible with a function in gene-specific transcription regulation. The role of the DNA binding and bending ability of the HMG box is so far elusive, at least for SRY.



## **1.6 BINDING, BENDING AND BIOLOGICAL EFFECT OF HMG BOX.**

HMG1 has been recently purified as the factor able to increase the binding between human progesterone receptor (PR) and its DNA binding site, the progesterone response element (PRE). This effect appears to be highly selective for HMG1, and the increase in binding affinity for a partial palindromic PRE present in natural targets genes was greater than 10 fold.

Because HMG boxes of HMG1 stimulate the binding and can flex DNA it could be possible that HMG1 facilitates the binding of PR inducing a structural change in the target DNA (Onate et al., 1994). A similar role could be played by the protein HMG2 in another context. This protein increases the DNA binding activity of octamer transcription factors (OCT proteins), supporting these factors in their role as transcriptional activators (Zwilling et al., 1995).

UBF, the upstream binding factor of RNA polymerase I, can also bend linear DNA inducing positive supercoiling, suggesting that UBF wraps the DNA. Additional studies provide evidence that wrapping DNA is produced in a right handed direction, opposite to the nucleosome's one.

UBF-induced DNA wrapping could be a mechanism by which UBF counteracts histone mediated gene repression.

Protein LEF1 contains, like SRY, one HMG box domain that induces a sharp angle in the DNA. This protein recognises a DNA consensus sequence on the T cell receptor alpha gene enhancer; the minimal enhancer element was shown to contain binding sites for at least others two distinct factors, flanking the binding site of LEF1. All three sites are recognised for enhancer function. A series of experiments revealed that mutations that alter the relative positions of the three sites inactivate the enhancer.

Moreover, it appeared that LEF 1 protein cannot activate transcription on its own, but must act in concert with factors that bind to the other two sites of the enhancer.

By using LEF1 in a different context it is also possible to study the connection of its HMG box and the in vivo topology of DNA . This protein is in fact able to interact with its DNA binding site in the human HIV1 enhancer. By using a nucleosome-assembly system derived from *Drosophila* embryos, it was found that the packaging of DNA into chromatin, in vitro, strongly represses HIV1 transcription and that repression can be counteracted efficiently by preincubation of DNA with LEF1 supplemented with a fraction containing the promoter binding protein Sp1 (Sheridan et al., 1995). This observation and additional studies of LEF1 ability to bend DNA led to the proposal that LEF1 acts as an architectural component in the assembly of the T cell enhancer complex. According to this model, LEF 1 induces a bend in DNA so that transcription factors bound to recognition sequences flanking the LEF1 binding site can interact with each other.

SRY action could be explained in the same way. Proteins bound at either side of the SRY binding site may then be brought into close contact and form a stable complex. In turn, the complex would freeze the deformation of the DNA molecule, and indirectly stabilise the binding of SRY. In fact, it was shown previously that SRY binds with high affinity to distorted DNA molecules, like four-way junctions, irrespective of their sequence (Ferrari et al., 1992). This possible role is supported by two pieces of evidence:

- 1) among different species, the only part of the protein conserved is the HMG box domain.
- 2) the only mutations in SRY known to cause sex-reversal in humans fall within the HMG-box domain (Berta et al., 1990, Hawkins et al., 1992, Hawkins, 1994 ).

To find evidence for a possible role of the DNA bending activity of the HMG box of SRY in gene expression and differentiation, and a possible correlation between an anomalous geometry of the DNA-protein complex and a biological effect of the mutation, we analysed SRY mutants associated with sex-reversal in humans. Five mutations with varying degrees of severity were considered. Moreover, we considered also the behavior of seven SRYs from different primates that, irrespective of substitutions occurring within the HMG box, have biological functions selected by evolution.

Five mutations with varying degrees of severity were considered. Mutations F109S and I90M were associated with complete gonadal dysgenesis in the patients where they were first identified, but were also present in normal male relatives of the patients, including the father (Hawkins et al., 1992b; Jager et al., 1992). These mutations therefore have low penetrance, are transmissible, and may cause a differentiation defect only in specific environments or in association with specific genetic backgrounds. Mutations G95R and M64I were identified only in patients with complete gonadal dysgenesis but not in their relatives, and have arisen de novo (Berta et al., 1990; Hawkins et al., 1992a). Mutation K106I was also associated with complete gonadal dysgenesis, but no male relatives of the patient were available for testing (Hawkins et al., 1992b).

# RESULTS



## **2.1 PRODUCTION OF WILD TYPE AND MUTANT SRY DNA BINDING DOMAINS**

We first chose to compare HMG box A of rat HMG1 (HMG1bA, amino acids -8 to 81 in the numeration of figure 1) with the HMG box of human SRYs, identified on the basis of its sequence similarity to other HMG boxes (Sinclair et al., 1990; figure 1). To express and study protein mutations of HMG boxes, it was first necessary to subclone their DNA coding sequences. Since the *SRY* gene does not contain introns, DNA fragments coding for the HMG box of normal and mutant human SRY were obtained by polymerase chain reaction (PCR) of total genomic DNA with specific oligonucleotides.

PCR is a method for amplifying DNA enzymatically. The first step simply entails mixing template DNA, two appropriate oligonucleotide primers, Taq DNA polymerase, deoxyribonucleoside triphosphates (dNTPs), and a buffer. The mixture is then cycled many times (in this case 25) through temperatures that permit denaturation, annealing, and synthesis in order to exponentially amplify a product of specific size and sequence. The PCR products are then displayed on an appropriate gel and examined for yield and specificity. The oligonucleotides used in our experiments contain two restriction enzymes sequences (*NdeI-HindIII*) in order to subclone the PCR fragments in the expression vector pT7-7. This vector (Studier et al, 1991) is a T7 expression vector containing the T7 promoter upstream of the gene to be expressed. The vector is designed for the exclusive expression of the cloned gene. This vector can then be used for production of intact native proteins by fusing coding sequences at the *NdeI* site immediately preceding the ATG. In addition to the promoter, this vector contains also a ribosome binding site. In *E.coli* the ribosome

binding site includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-9 nucleotides upstream of the initiation codon. This sequence, which is called the Shine Delgarno (SD) sequence, is complementary to the 3' terminus of E coli 16S rRNA. Binding of the ribosome to mRNA is thought to be promoted by base pairing between the SD sequence in the mRNA and the sequence at the 3' terminus of the 16S rRNA.

Another factor required to express cloned DNA with this system is bacteriophage T7 RNA polymerase. This polymerase recognises solely bacteriophage T7 promoters; it is a very processive RNA polymerase, and may transcribe sequences that are not efficiently transcribed by E. coli RNA polymerase; this system allows high levels of expression of some genes that are not expressed efficiently in other systems. T7 polymerase is the product of bacteriophage T7 gene 1 and can be provided on an infecting bacteriophage lambda vector or produced from a gene copy inserted into E coli chromosomes. If expression of the cloned gene is toxic, then the level of bacteriophage T7 RNA polymerase must be kept low during cell growth. One way to accomplish this is to use the lysogen BL21(DE3), in which bacteriophage T7 gene 1 is expressed from the LacUV5 promoter, such that it is inducible by IPTG.

To allow the synthesis of the HMG box in E. coli, these primers were also designed in order to change valine 4 (see figure 1) to methionine, to provide a translation start site. Also lysine 77 was changed to a stop codon. DNAs from PCRs were then digested with restriction endonucleases *NdeI* and *HindIII*, loaded on an agarose gel, gel purified and cloned between the *NdeI* and *HindIII* sites of plasmid pT7-7. The resulting plasmids were then introduced in strain BL21(DE3), a host strain containing an IPTG inducible gene for T7 RNA Polymerase. Plasmid DNA from positive colonies was extracted and checked by sequencing with T7 DNA polymerase (see

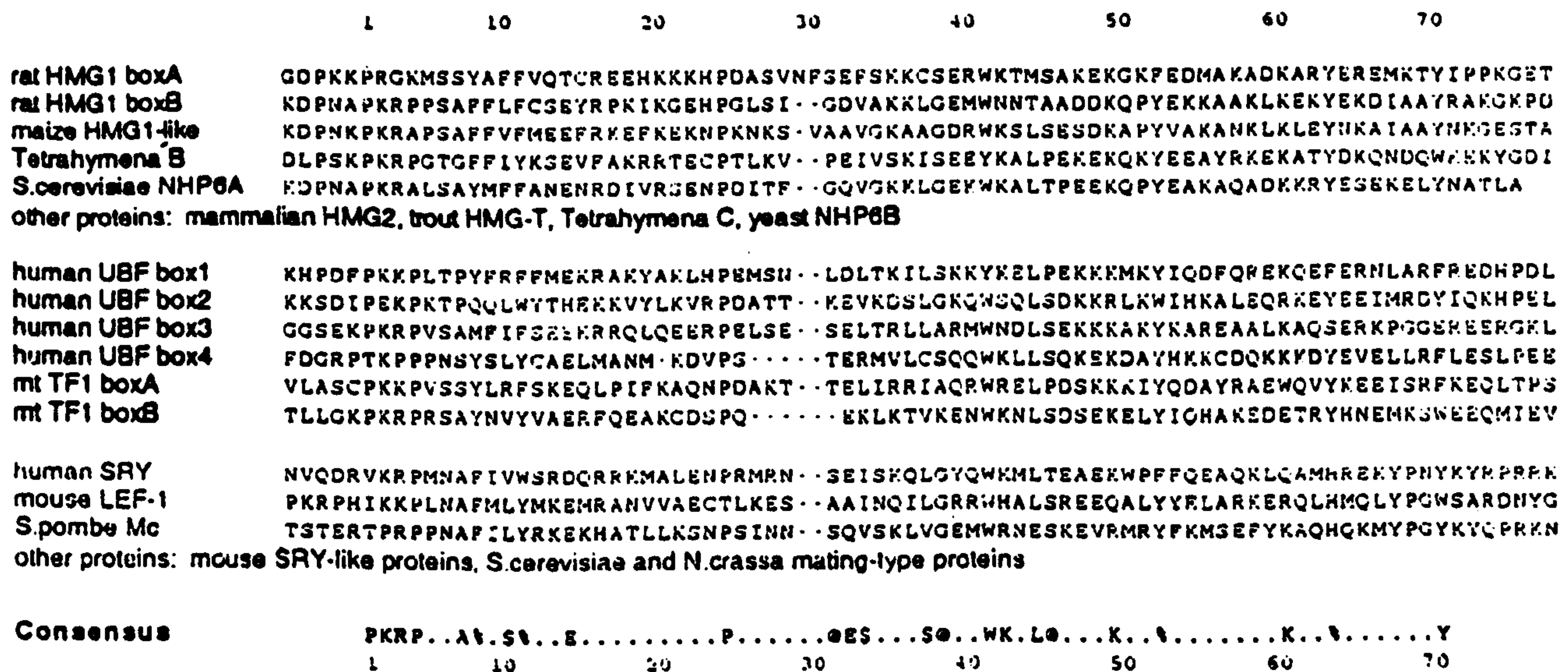


Fig. 1. Alignment of HMG boxes. Only proteins mentioned in the main text are shown; a more exhaustive compilation may be found in Bianchi et al. (1992a). The first group of HMG boxes are from chromatin proteins. HMG1 proteins of mammals have two HMG boxes and are identical save conservative substitutions in the C-terminal acidic stretch (not shown here); plant, protozoan and yeast HMG1-like proteins contain a single HMG box. The second group of HMG boxes are from general transcription factors for RNA polymerase I (human UBF, with four boxes) and mitochondrial RNA polymerases (human mtTF1, with two boxes). The third group of HMG boxes are from transcriptional regulators. Dashes indicate gaps in the alignment; Z indicates a stop codon in the gene. The consensus sequence for the HMG box motif was obtained from 21 protein sequences: one-letter symbols indicate amino acids present in 50-90% of the sequences; conservative substitutions (at least 75% of the occurrences at a particular position) are indicated as follows: @ for proline, alanine, glycine, serine and threonine; % for tryptophan, phenylalanine and tyrosine; \$ for methionine, valine, leucine and isoleucine. The numbering system starts from the first conserved proline and is based on HMG1 box A.



materials and methods). Bacterial cells containing the right constructs were grown up to O.D.<sub>600</sub>=0.6 and then T7 protein expression was induced by IPTG for two hours. Part of the *E. coli* extract was boiled in a denaturing loading buffer (Laemmli buffer) and separated on SDS PAGE to verify if corresponding polypeptides were expressed at high levels in *E.coli*. The majority of the bacterial pellet was resuspended in a sonication buffer, sonicated until clarification, treated with DEAE cellulose to remove nucleic acids and then centrifuged. The supernatant containing the soluble recombinant protein was filtered through a cache of glass wool, applied to a Mono S FPLC column. This column is a strong cation exchanger with charged sulfonic groups which remain negative over the pH range 2-12; it was used because wt and mutated hSRY boxes have a high content of basic residues. Fractions were then analysed on SDS PAGE, and also monitored with absorbance. Purified proteins were eluted in fractions containing 1M NaCl (see Materials and Methods). Proteins were 99% pure (figure 2).



12 11 10 9 8 7 6 5 4 3 2 1 M

**Fig.2. Purification of E. coli HMGboxes.**

Proteins were purified by removing most E. coli proteins by precipitation with ammonium sulphate, and fractionation of the resulting supernatant by chromatography (see Materials and Methods).

Aliquots from the mono S fractions were mixed with 1  $\mu$ g of BSA, to serve as an internal quantity standard. When electrophoresed on SDS page, a single band with an apparent molecular mass of about 9 Kd appears in fraction 8 and 9. Molecular mass standards are shown in lane M.

## **2.2 THE AFFINITY OF WT AND MUTANT BOXES FOR BENT DNAS**

The HMG box domains of HMG1 (an abundant chromatin protein) and UBF (a general transcription factor for RNA polymerase I) all recognise four-way junctions structure-specifically and sequence-independently (Bianchi et al., 1992a, 1992b; Kuhn et al., 1994). The recognition of this unusual DNA structure is related to the functional properties of this class of proteins (Lilley, 1992), and depends on the structural integrity and correct folding of the HMG box domain (Falciola et al., 1994). To find evidence for a possible role of the stability of DNA-protein complex of the HMG boxes of SRY in gene expression and sex determination, we tried to compare their affinities for DNA.

Binding specificity and affinity of the normal and the five purified mutant HMG boxes (hSRY boxes) for a bent DNA fragment were determined in band shift assays using a limiting amount of labelled DNA and titrating the polypeptides (see below and Materials and Methods).

The gel shift assay is one of the most powerful methods for the analysis of DNA-protein interactions. The assay itself is simple; DNA and protein are mixed together, the solution subjected to vertical electrophoresis through polyacrylamide gel, and the gel is then analysed for DNA, usually by autoradiography of radiolabeled DNA. Binding of the protein to the DNA can result in a complex that has a different electrophoretic mobility from the free DNA. In general, the mobility of the complex is retarded relative to the unbound DNA and thus the assay is often called gel retardation. The separation of the complex from the free DNA, and therefore the detection of the complex, is dependent on a variety of factors. Factors that influence the electrophoretic mobility of DNA-protein complexes include the molecular weight

of the protein and the DNA, the ionic strength and the pH of electrophoresis buffer, the concentration of the gel matrix and the temperature. The principle of the gel shift assay is that the entry of the mixture of free DNA and DNA-protein complex into the gel matrix results in the physical separation of the two species. In the subsequent electrophoresis, the protein generally makes no difference to the mobility of the free DNA, although it has been shown that complexes reversibly dissociate and reassociate within the gel. However, even if the bound DNA dissociates from the protein during electrophoresis, it can never "catch up" with the DNA that was free at the start of the run. The concentration of each species can then be determined. Assays of this type can yield the equilibrium constant for the binding of the protein to its DNA ligand and also the kinetics of the interaction.

With this technique, mutated and wt hSRY proteins were demonstrated to recognize with high affinity bent DNA, irrespective of its sequence. A probe with bent DNA structure was obtained by annealing four appropriately chosen oligonucleotide sequences. In this way molecules were generated with the shape of cruciform (cruciform c), similar to those of Holiday junctions (Fig. 3). Cruciform DNA was labelled with  $^{32}\text{P}$  and gel purified; wt and mutated SRY boxes were then incubated with labelled junction probe c and samples were applied to vertical polyacrylamide gels. Gels were electrophoresed, dried and autoradiographed for 16 hours.

All hSRY boxes formed well defined complexes with the four-way junction probe c. Figure 4 shows the band shift experiment obtained with wt hSRY box. About 0.5 ng of peptide hSRY box (lanes 1-4) or control buffer with no protein (lane 5) were mixed in standard binding buffer with various concentrations of four-way junction DNA c, and assayed by gel electrophoresis as described before. Wt hSRY box forms well defined complexes with four-way junction probe c; the slower moving complexes



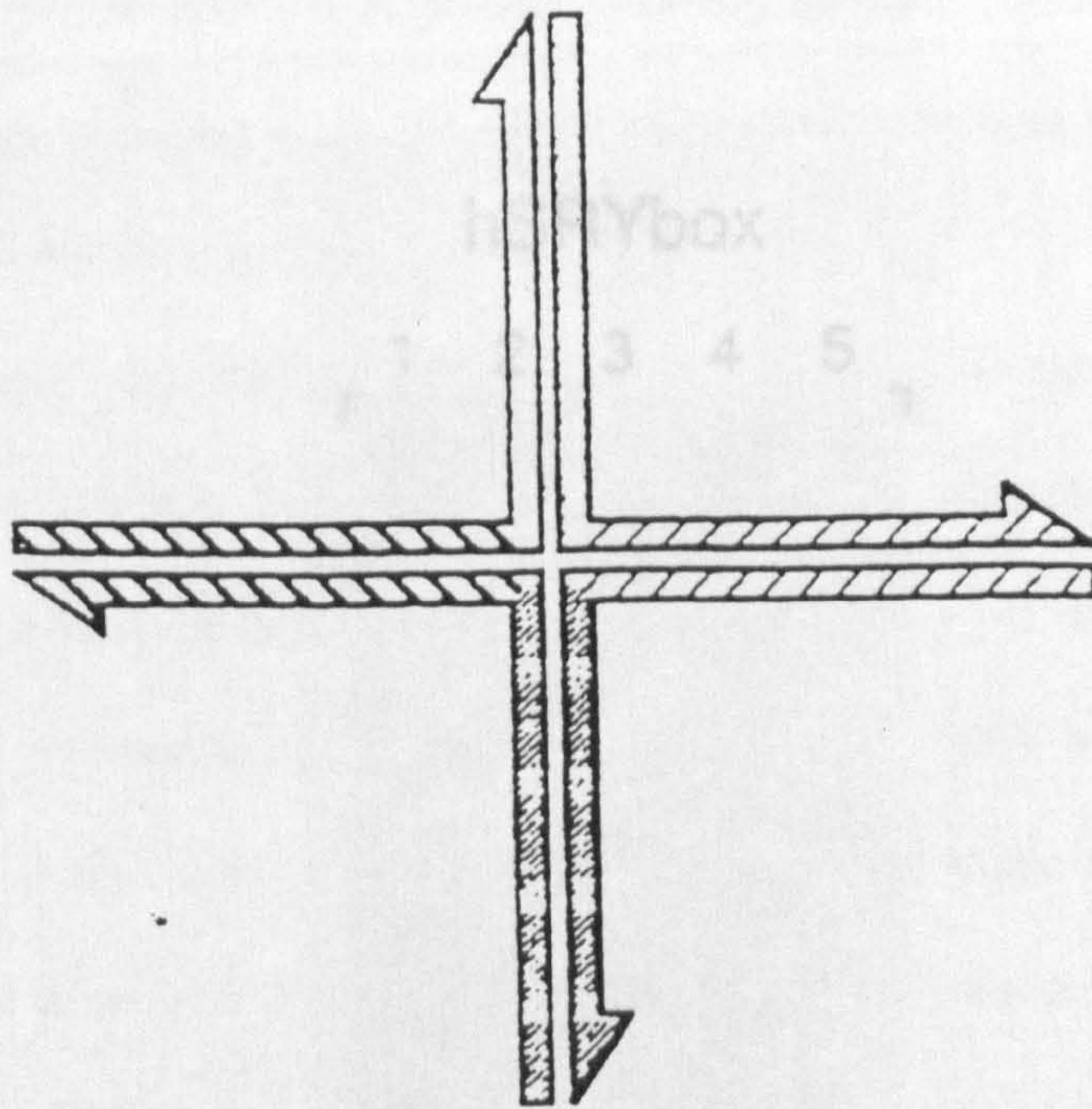


FIG.3 . Design of an artificial cruciform DNA.

The molecule c, unlike natural cruciform DNA, has no sequence symmetry and cannot dissociate through branch migration. Cruciform DNA molecules were produced by annealing chemically synthesised oligonucleotides 1, 2, 3, and 4.

1: CCCTATAACCCCTGCATTGAATTCCAGTCTGATAA

2: GTAGTCGTGATAGGTGCAGGGGTTATAGGG

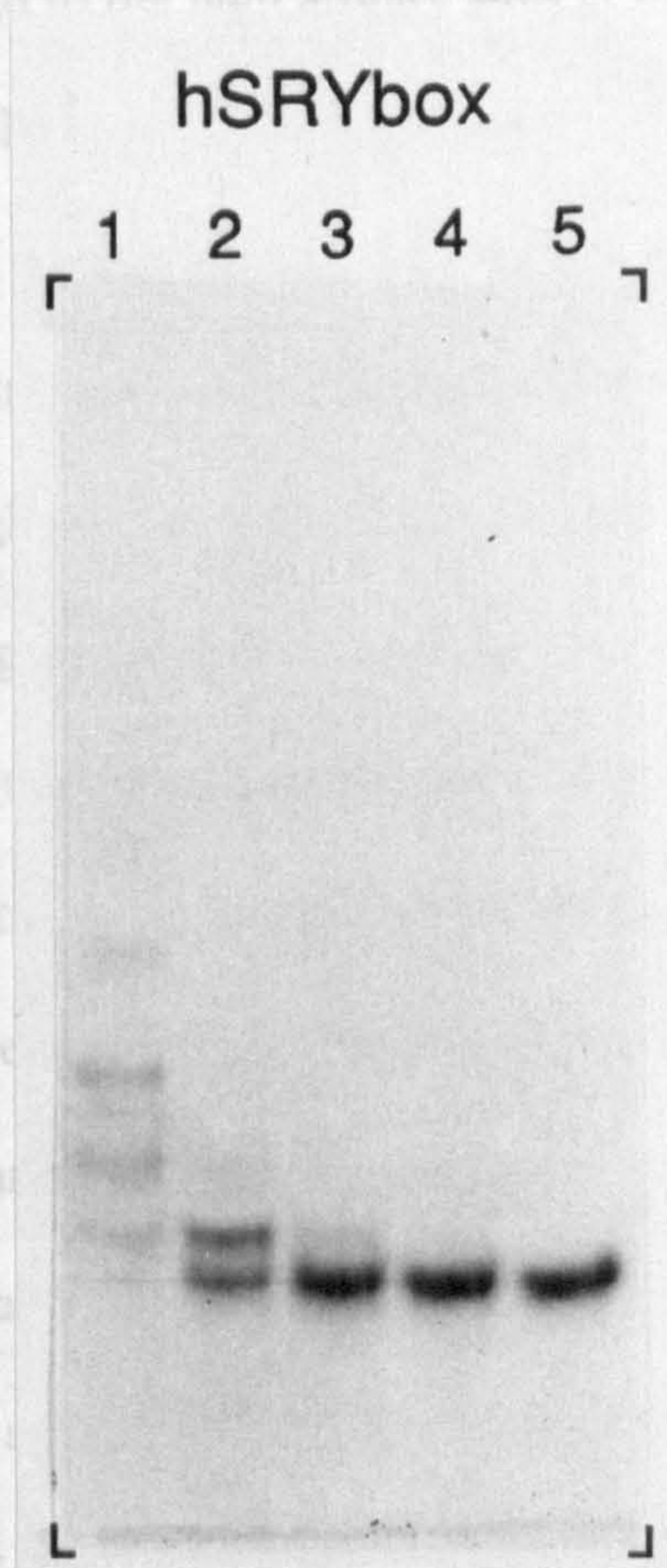
3: AACAGTTAGCTCTTATTCGAGCTCGCGCCCTATCACGACTA

4: TTTATCAGACTGGAATTCAAGCGCGAGCTCGAATAAGAGCTACTGT



formed by wt hSRY box probably contain multiple copies of polypeptide per DNA molecule, some bound to the high affinity sites at the base of the junction and some to low affinity sites on the arms.

In similar band shift assays, the four-way junction probe c with wt hSRY box was retarded by 100-fold at high concentrations of protein. To obtain convincing evidence of the shape of four-way junction DNA present in the junction, four-way junction DNAs were constructed. In this experiment a linear duplex DNA (az and bz) which contains the same sequence as cruciform DNA c but with different shapes were used as controls. Control DNAs were



are shown to recognise four-way junction DNA c with wt hSRY protein and the same multiple bands were observed.

wt hSRY boxes recognise the peculiar shape of four-way junction DNA c. Specific binding sites adventitiously present in the junction DNA were identified by additional band shift assays.

Linear duplex DNA (az and bz) which contains sequences identical to those of cruciform DNA c (Harley et al., 1992). Cruciform DNA c has a different structure.

Linear duplex DNAs containing the same sequence of the arms of cruciform z (az and bz) but they will have a different shape (figure 5). These molecules do not have angles distorting the axis of the DNA.

Although the artificial cruciform z has no obvious sequence similarity with cruciform

Fig. 4. Peptide hSRY box recognises the four-way junction DNA c. About 0.5 nanograms of peptide hSRYbox (lanes 1 to 4) or control buffer with no protein (lanes 5) were mixed in standard binding buffer with various concentrations of four-way junction DNA c, and assayed by gel electrophoresis as described in Materials and methods. The concentration of DNA was as follows: lanes 1, 1.5 nM; lanes 2, 8 nM; lanes 3, 40 nM; lanes 4 and 5, 200 nM.

peptide. Samples were applied to a vertical 6.5 polyacrylamide gel and electrophoresed as described in materials and methods.



formed by wt hSRY box probably contain multiple copies of polypeptide per DNA molecule, some bound to the high affinity sites at the base of the junction and some to low affinity sites on the arms of the junctions.

In similar band shift assays mutant SRY boxes were shown to recognise four-way junction probe c with the same affinity as the wt protein and the same multiple retarded bands at high protein to DNA ratios were observed.

To obtain convincing proof that wt and mutated hSRY boxes recognise the peculiar shape of four-way junctions, and not sequence-specific binding sites adventitiously present in the junction, binding affinities of wt and mutant SRY boxes for linear and junction DNA were examined in more detail by means of additional band shift assays.

In this experiment another type of junction (z) was used, which contains sequences deliberately chosen as poor binding sites for SRY (Harley et al., 1992). Cruciforms c and z have different sequences but share the same bent structure.

Control DNAs were obtained by synthesising linear duplex DNAs containing the same sequence of the arms of cruciform z (az and bz) but they will have a different shape (figure 5). These molecules do not have angles distorting the axis of the DNA.

Although the artificial cruciform z has no obvious sequence similarity with cruciform c, it is recognised in the same way by wt and mutated SRY boxes. Control linear duplex DNA (az and bz) which contains the same sequence as cruciform DNA but a different structure are not recognised with high affinity by these proteins. Figure 6 shows the band shift assay performed with wt hSRY box and probe junction z, versus linear duplex arms az and bz. Labelled junction z, duplex az and bz were mixed in 10 microliters of standard binding buffer with the identical amounts of wt hSRY box peptide. Samples were applied to a vertical 6.5 polyacrylamide gel and electrophoresed as described in materials and methods.

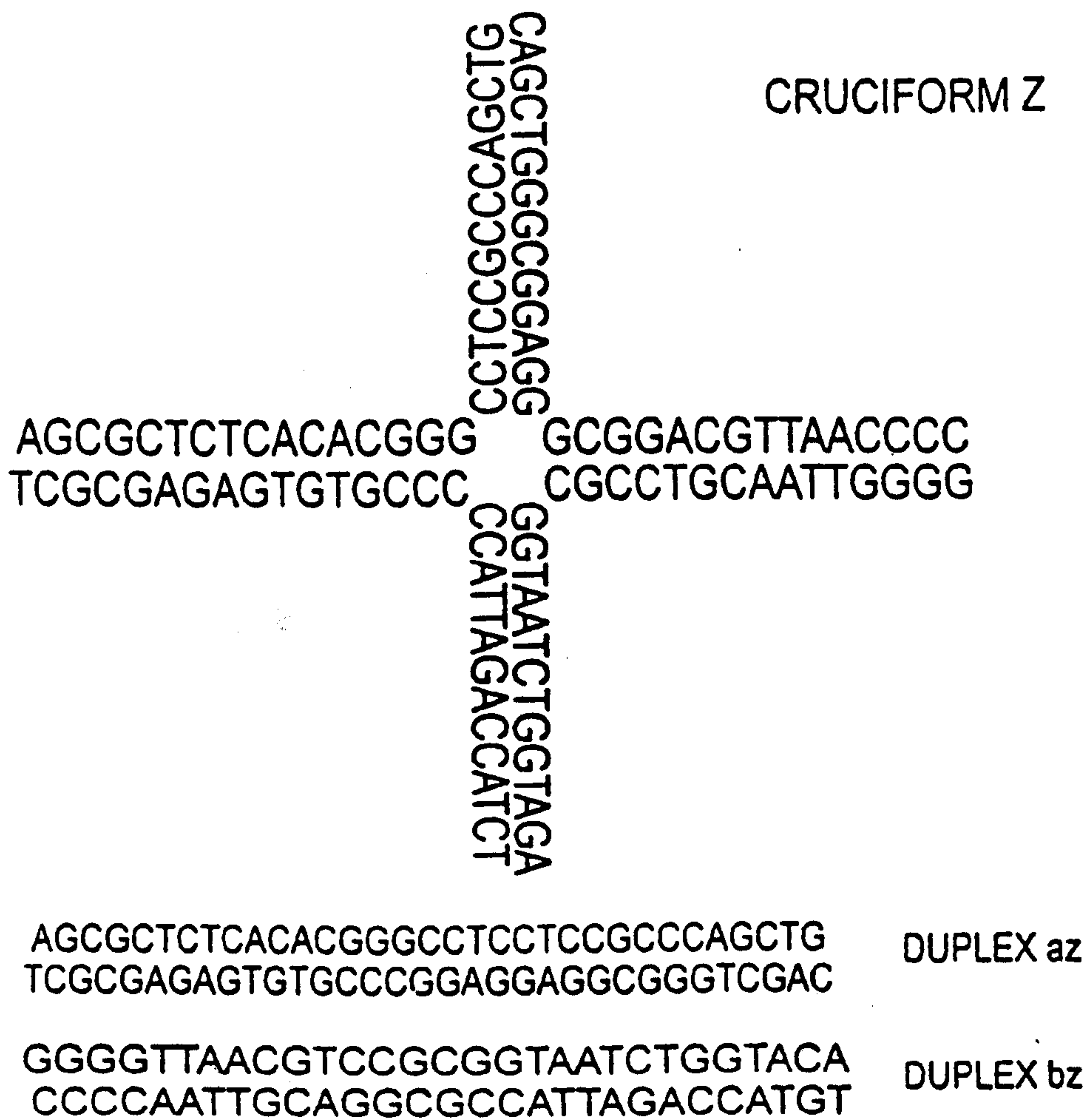


Fig. 5 Design of an artificial cruciform DNA.

The molecule z, unlike natural cruciform DNA, has no sequence symmetry and cannot dissociate through branch migration. Cruciforms z and c have no obvious sequence similarity, but are both recognised by SRY box. Cruciform and linear duplex DNA molecules were produced by annealing chemically synthesised oligonucleotides.



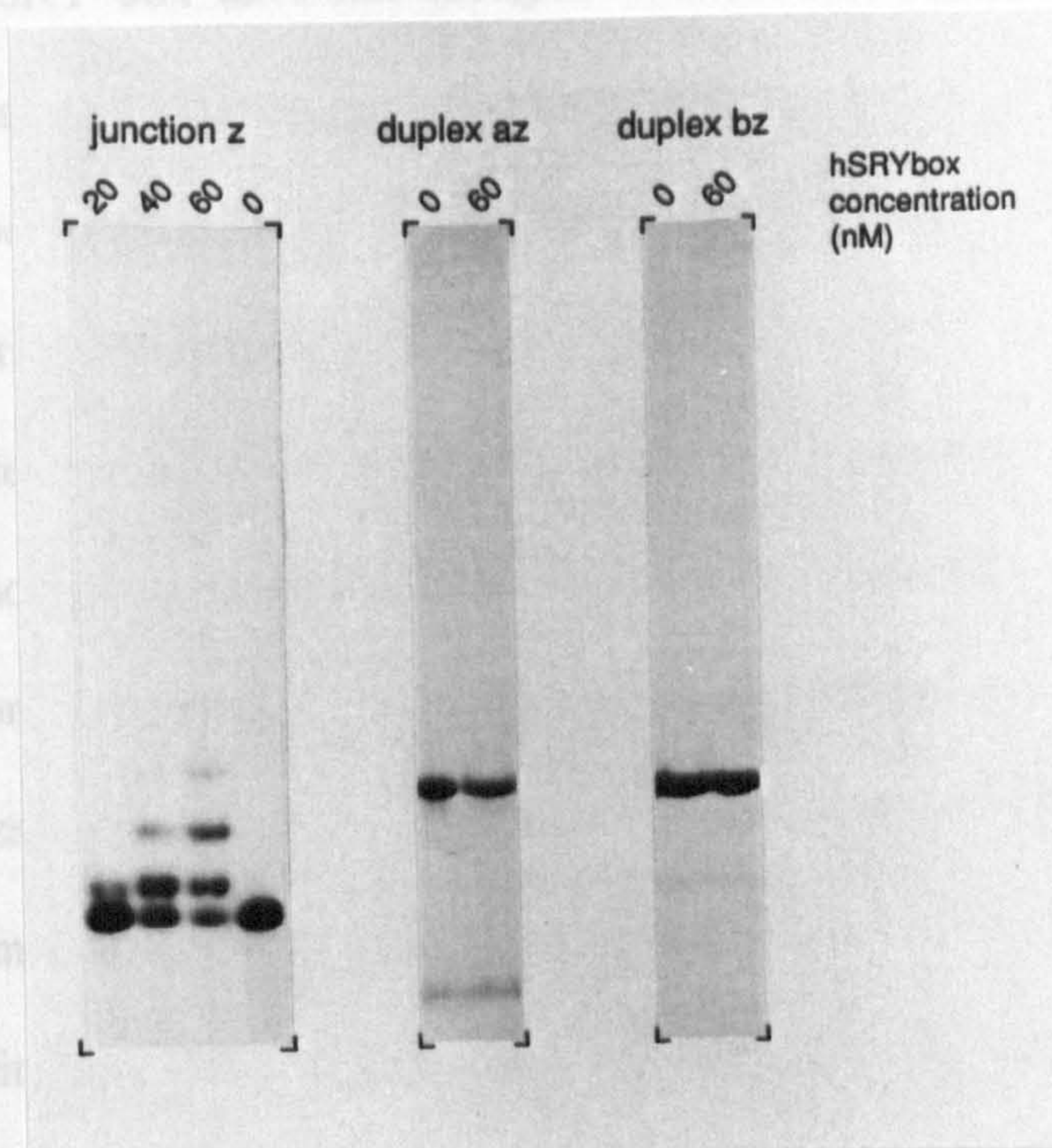


Fig.6. Peptide hSRYbox does not recognize linear control DNAs az and bz at the concentrations optimal for complex formation with junction z. Labeled junction z, duplex az and bz (0.2 nM) were mixed in 10  $\mu$ l of standard binding buffer with the indicated amounts of hSRYbox peptide. Electrophoresis and autoradiography were done as described in materials and methods.

[D] = concentration of free DNA

[PD] = concentration of protein-DNA-DNA complex

$K_{eq}$  = equilibrium dissociation constant (in this case referred to the dissociation constant,  $K_d$ ).

The  $K_d$  of a given protein-DNA interaction is a measure of the affinity of the protein for that particular piece of DNA. The apparent  $K_d$  can be determined by a band shift assay in which protein is titrated against a known amount of DNA containing the



Peptide wt hSRY box does not recognise linear control DNAs az and bz at the concentrations optimal for complex formation with junction z. The identical result was obtained by performing the same experiment with all mutated SRY boxes associated with sex reversal phenotypes.

To investigate in more detail if various single point mutations associated with sex reversal phenotype could slightly affect the binding affinity of a bent DNA, constants of dissociation of the DNA-SRYs complexes were measured and compared.

Band shift assays are ideal for equilibrium studies, because bound DNA can rapidly separate from free DNA. DNA retained by the protein can be easily quantified and compared with free DNA. The protein DNA binding interaction can be looked at as an equilibrium represented by the equation:

$$[P] + [D] = [PD],$$

$$\text{so that } K_{eq} = \frac{[P][D]}{[PD]}$$

where  $[P]$  = concentration of total protein. It corresponds to the concentration of free protein if the concentration of protein bound to DNA is very low; this is possible when the concentration of total DNA is  $\ll [P]$ .

$[D]$  = concentration of free DNA

$[PD]$  = concentration of protein DNA-DNA complex.

$K_{eq}$  = equilibrium dissociation constant (in this case referred as the dissociation constant,  $K_d$ ).

The  $K_d$  of a given protein-DNA interaction is a measure of the affinity of the protein for that particular piece of DNA. The apparent  $K_d$  can be determined by a band shift assay in which protein is titrated against a known amount of DNA containing the



binding structure or site being tested. The half -maximal point of the resulting curve is equal to the  $K_d$ , since at the 50% saturation of the DNA,  $[PD]=[D]$ .

Therefore, the equation

$K_{eq}=K_d= [P][D]/[PD]$  is reduced to  $K_d=[P]$ .

This simplification is true only if  $[D]\ll K_d$ .

Therefore, as in our case, the amount of DNA used in these experiments is usually very small.

This basic algebra means that in a band shift assay, under conditions of protein excess, the dissociation constant is equivalent to the concentration of polypeptide when half of the input DNA is taken up in the complexes and half is free.

All mutant SRY boxes were then individually titrated and analysed in a band shift assays, with four way junction probe z.

The affinities of the mutant SRY DNA binding domains towards junction z were indistinguishable from that of the wild type domain (Table 1). From these observations we can conclude that the various proteins were all equally active towards these DNA molecules, indicating that the amount of inactive polypeptide chains was irrelevant and that the amino acid substitutions did not cause gross misfoldings of the SRY proteins.



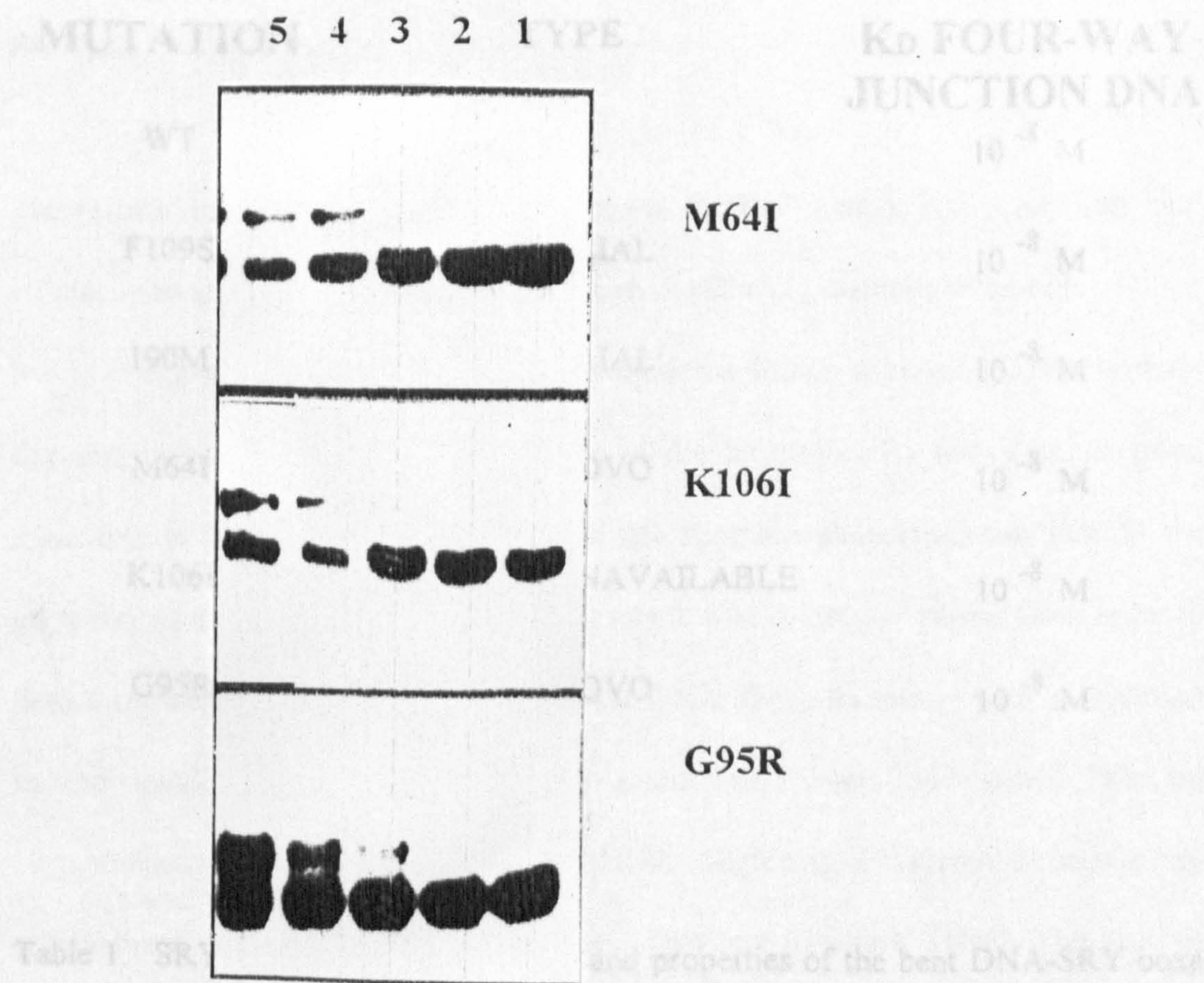


Table 1. SRY box DNA binding properties of the best DNA-SRY boxes interaction. The mutations in SRY box have considered were described by Berta et al. (1990), Hawkins et al. (1993a,b) and Jager et al. (1992). The dissociation constants for complexes between the various SRY boxes and DNA were calculated by titrating the SRY box polypeptides against 0.25 nM DNA and leaving the formation of complexes by electrophoretic mobility shift assays. "Four-way junction" refers to

Fig. 7 DNA binding properties of some hSRYboxes. DNA binding mixtures were set up to contain 1.5 nM labelled four way junction DNA. Each panel contains the following amounts of the indicated peptides: lanes 1, no protein; lanes 2, 5 nM; lanes 3, 10 nM; lanes 4, 20 nM. Samples were assayed by electrophoresis as described in materials and methods. The faster migrating band is free DNA, the slower is the SRY box -junction complex.



| MUTATION | TYPE                  | K <sub>D</sub> FOUR-WAY<br>JUNCTION DNA |
|----------|-----------------------|---|
| WT       |                       | 10 <sup>-8</sup> M                      |
| F109S    | FAMILIAL              | 10 <sup>-8</sup> M                      |
| I90M     | FAMILIAL              | 10 <sup>-8</sup> M                      |
| M64I     | DE NOVO               | 10 <sup>-8</sup> M                      |
| K106I    | RELATIVES UNAVAILABLE | 10 <sup>-8</sup> M                      |
| G95R     | DE NOVO               | 10 <sup>-8</sup> M                      |

Table 1. SRY mutations characteristics and properties of the bent DNA-SRY boxes interaction. The mutations in SRY we have considered were described by Berta et al. (1990), Hawkins et al. (1992a,b) and Jager et al. (1992). The dissociation constants for complexes between the various SRY boxes and DNA were calculated by titrating the SRY box polypeptides against 0.25 nM DNA and assaying the formation of complexes by electrophoretic mobility shift assays. "Four-way junction" refers to junction molecule z, which was chosen because the sequences it contains cannot be recognised sequence-specifically by SRY protein (Ferrari et al., 1992). The reported values are the mean of at least two assays.

### **2.3 THE DNA-BENDING ACTIVITY OF NORMAL AND MUTANT SRY BOXES**

The results described in the previous sections establish that hSRY boxes can bind efficiently to structurally similar targets where the DNA is strongly distorted.

It was previously shown that wt hSRY recognises a double stranded DNA consensus site and induces a distortion in the axis of the molecule. To test if single point mutations in hSRY box, associated with sex reversed phenotype, can modify the geometry of the DNA-protein complex, we performed a circular permutation assay to detect any large distortions induced in linear DNA. DNA fragments with a distortion in the middle of the molecule have a different shape, and hence different electrophoretic mobility, compared to DNA fragments of identical length and composition with a distortion near one end (Wu and Crothers, 1984). The bending angle  $\alpha$  is defined as the angle by which a segment of the rod like DNA duplex departs from linearity. It is possible to estimate  $\alpha$  by measuring the mobility of the complex with the protein bound at the middle of the fragment and the mobility of the complex with the protein bound near the end of the DNA fragment using the relationship reported in Materials and Methods.

Although the relationship between electrophoretic mobility and conformation is complex (Levene and Zimm, 1989), the algebra allows one to map the locus of protein-DNA interaction and to estimate the amount of distortion introduced in DNA simply by measuring the rate of migration in polyacrylamide gels of complexes of protein with DNA fragments of circularity permuted sequence (Liu-Johnson et al., 1986; Thompson and Landy, 1988; also see Materials and Methods).



To generate the probes for the permutation assay, a short sequence from the enhancer of the CD3e gene, containing the AACAAAG sequence and previously shown to be a good binding site for human SRY protein (Harley et al., 1992), was cloned between two directly repeated sequences in plasmid pBend2. These two duplicated circular permuted sequences contain the same set of 17 cloning sites for insertion of protein binding sequences (figure 8). The duplicated sites can be used to generate DNA fragments of identical length, but in which the protein binding sequence is shifted. Sites *XbaI* and *SalI* are unique, in the middle of the tandemly repeated sequences and suitable for the cloning of the protein binding sequence.

Briefly, to subclone the CD3 epsilon site in the vector, synthetic oligonucleotides CTAGAGAGCGCTTTGTTCTCAG and TCGACTGAGAACAAAGCGCTCT were annealed, generating a double stranded molecule containing sticky ends complementary to the overhangs produced by cutting with restriction endonucleases *XbaI* and *SalI*.

This molecule was subcloned within the *XbaI* and *SalI* sites of the vector p Bend 2; the recombinant vector was checked by sequencing and digested with the set of restriction enzymes reported in figure (8). Cleavage with restriction endonucleases indicated in the map of fig. 8 yielded a set of fragments of DNA (designated A-G) containing circular permutations of the same sequence of 141 b.p. These molecules therefore have identical lengths, the same sequences but each has a different distance from the binding site and the ends of the molecule (circular permuted sequences). Circular permuted fragments were gel purified and the same amount of DNA was labelled with <sup>32</sup>P (see material and methods) and incubated with purified wt and mutated SRY boxes. Samples were applied to a polyacrylamide gel and the electrophoretic mobility of the resulting complexes were analysed. Figure 9 shows the electrophoretic mobility of the circular permuted DNA fragments complexed to the

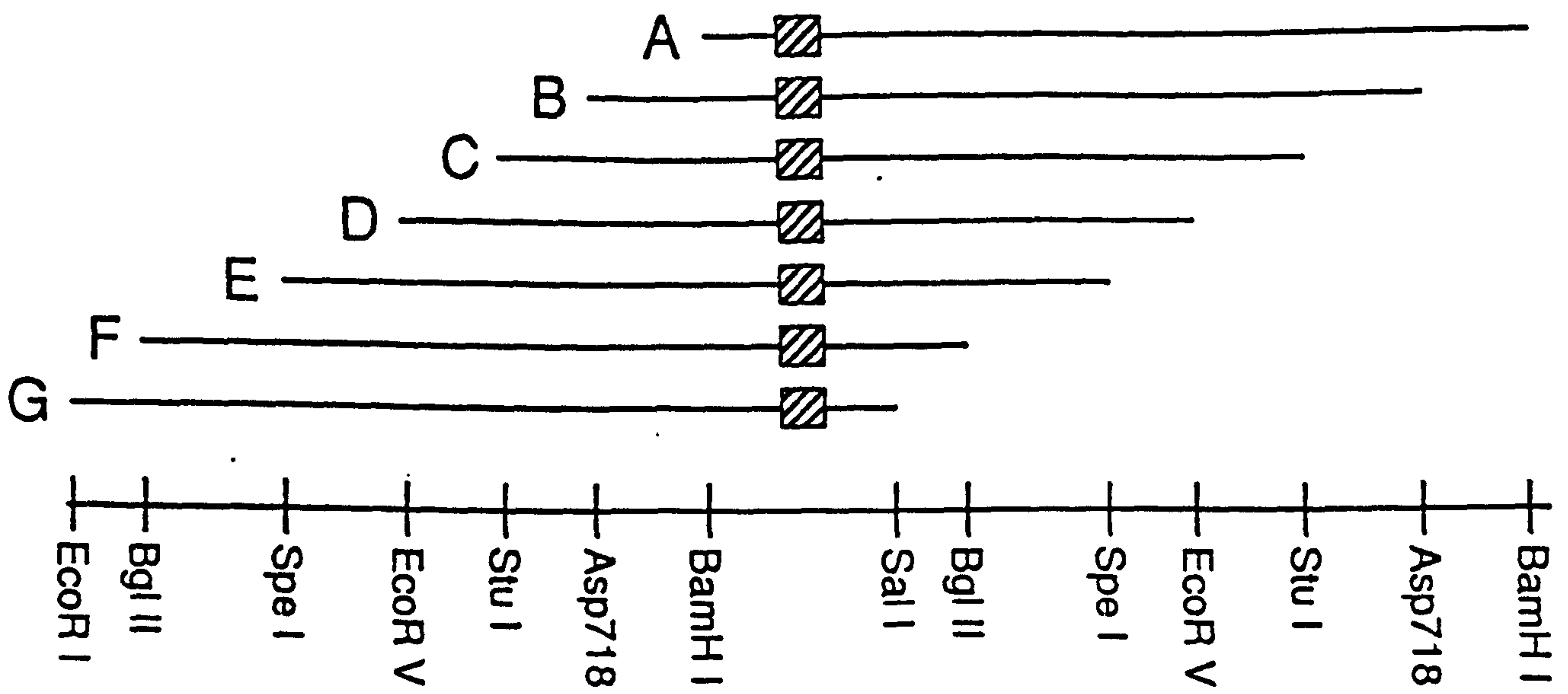


Fig.8 Circular permuted probes. Plasmid pB2CD3e, containing the CD3e site (stippled box) flanked by tandemly duplicated DNA sequences, was cleaved at the restriction sites indicated in the map. The DNA fragments obtained in this way (designated A to G) all contain circular permutations of the same sequence of 141 bp



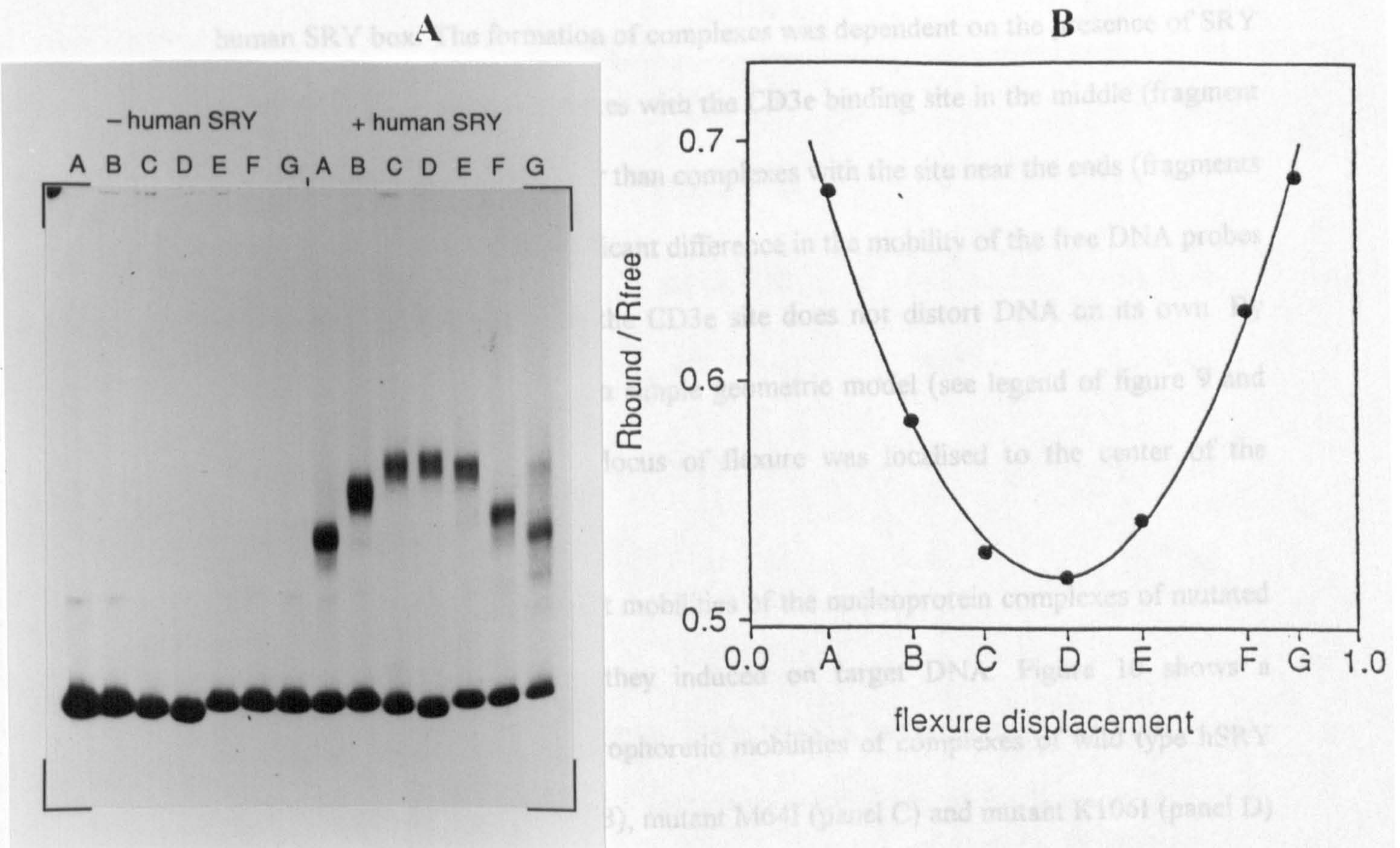


Fig.9. Circular permutation analysis of DNA flexure induced by binding of human SRY to the AACAAAG sequence. A. Electrophoretic mobility of the circularly permuted DNA fragments complexed to the human SRY protein. DNA fragments A to G (Fig 6, 8 femtomoles) were mixed with sonicated salmon sperm DNA (50 ng) in 9  $\mu$ l of standard DNA binding buffer (see Materials and methods). To the various mixtures we added 1  $\mu$ l of purified fraction (about 5 ng of protein) containing human SRY, or the same volume of water. Electrophoresis and autoradiography were performed as indicated in materials and methods. B. Mapping of the locus of flexure and analysis of the bending parameters. The mobilities of the protein-DNA complexes ( $R_{bound}$ ) were normalized to the mobility of the corresponding free DNA ( $R_{free}$ ). The distance of the center of the CD3e site from the 5' end of the probe was divided by the total length of the probe (flexure displacement). The plotted points were interpolated with a quadratic function as described in materials and methods. The fitting second-order equation was  $y = 1.117x^2 - 1.131x + 0.804$  ( $R^2 = 0.997$ ). The first and second-order parameters of the equation are in close agreement and yield an estimate of deviation from linearity of about  $83^\circ$ . The locus of flexure was localized to the center of the AACAAAG sequence,  $\pm 2$  bp.



human SRY box. The formation of complexes was dependent on the presence of SRY proteins. DNA-protein complexes with the CD3e binding site in the middle (fragment D) migrated significantly slower than complexes with the site near the ends (fragments A and G). No statistically significant difference in the mobility of the free DNA probes was observed, indicating that the CD3e site does not distort DNA on its own. By analysing the data in terms of a simple geometric model (see legend of figure 9 and Materials and Methods), the locus of flexure was localised to the center of the AACAAAG site.

We then compared the different mobilities of the nucleoprotein complexes of mutated SRY boxes and the angles they induced on target DNA. Figure 10 shows a comparison between the electrophoretic mobilities of complexes of wild type hSRY box and mutant F109S (panel B), mutant M64I (panel C) and mutant K106I (panel D) with probes A-G containing the CD3e site GAACAAAG. Briefly, 3 fmol of labelled DNA probes were mixed in 9  $\mu$ l of standard DNA binding buffer (see Materials and Methods). To the various mixtures 1  $\mu$ l of purified protein was added (2 ng of hSRYbox wild type, 2 ng of hSRYbox F109S, 10 ng of hSRYbox M64I, 25 ng of hSRYbox K106I). Electrophoresis and autoradiography were performed as indicated in Materials and Methods; however, the autoradiography in panel D was overexposed to detect the faint bands of DNA complexed to hSRYbox K106I. The retarded bands running behind the principal protein-DNA complex band are complexes containing more than one polypeptide molecule per DNA molecule. No significant difference in the mobility of the free DNA probes can be observed, indicating that the CD3e site does not distort DNA on its own. By analysing the data in terms of a simple geometric model (see Materials and Methods), it appears that the mutant SRY boxes also flex

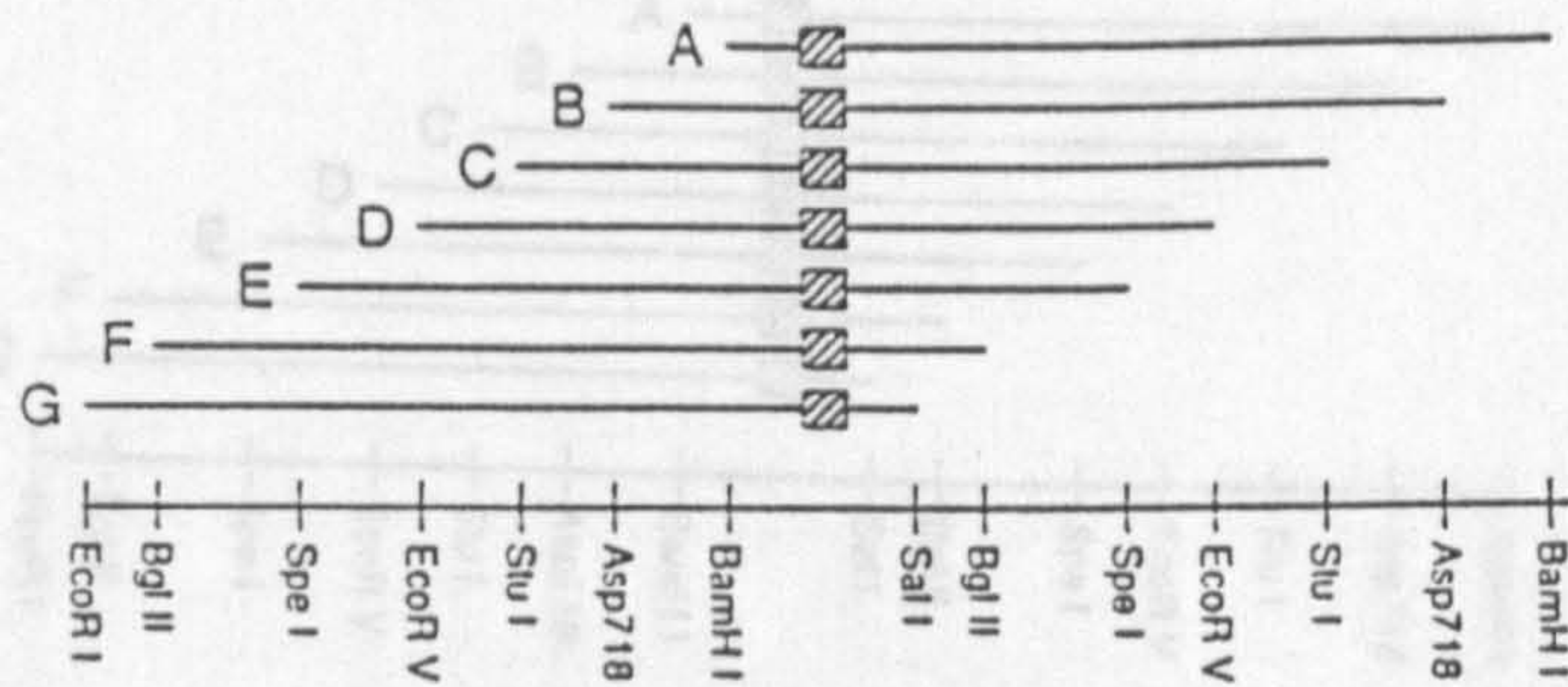


| Mutation | Type                     | KD<br>linear DNA       | Angle      | Kd four way<br>junc. DNA |
|----------|--------------------------|------------------------|------------|--------------------------|
| WT       |                          | $2 \times 10^{-8}$ M   | $76^\circ$ | $10^{-8}$ M              |
| F109S    | FAMILIAL                 | $2.5 \times 10^{-8}$ M | $76^\circ$ | $10^{-8}$ M              |
| I90M     | FAMILIAL                 | $4 \times 10^{-8}$ M   | $75^\circ$ | $10^{-8}$ M              |
| M64I     | DE NOVO                  | $6.5 \times 10^{-8}$ M | $56^\circ$ | $10^{-8}$ M              |
| K106I    | RELATIVES<br>UNAVAILABLE | $2 \times 10^{-6}$ M   | $73^\circ$ | $10^{-8}$ M              |
| G95R     | DE NOVO                  | $> 10^{-5}$ M          |            | $10^{-8}$ M              |

Table 2. Summary of SRY mutation characteristics and properties of the SRY boxes. The mutations in SRY we have considered were described by Berta et al. (1990), Hawkins et al. (1992a,b) and Jager et al. (1992). The dissociation constants for complexes between the various SRY boxes and DNA were calculated by titrating the SRY box polypeptides against 0.25 nM DNA and assaying the formation of complexes by electrophoretic mobility shift assays (see materials and methods). "Linear" refers to the D probe bearing the CD3e site, "four-way junction" refers to junction molecule z, which was chosen because the sequences it contains cannot be recognised sequence-specifically by SRY protein (Ferrari et al., 1992). "Angle" refers to the angle calculated by the algorithm of Ferrari et al. (1992) from the circular permutation assays on the set of probes bearing the CD3e site. The reported values are the mean of at least two assays.



A



B

probe CD3ε

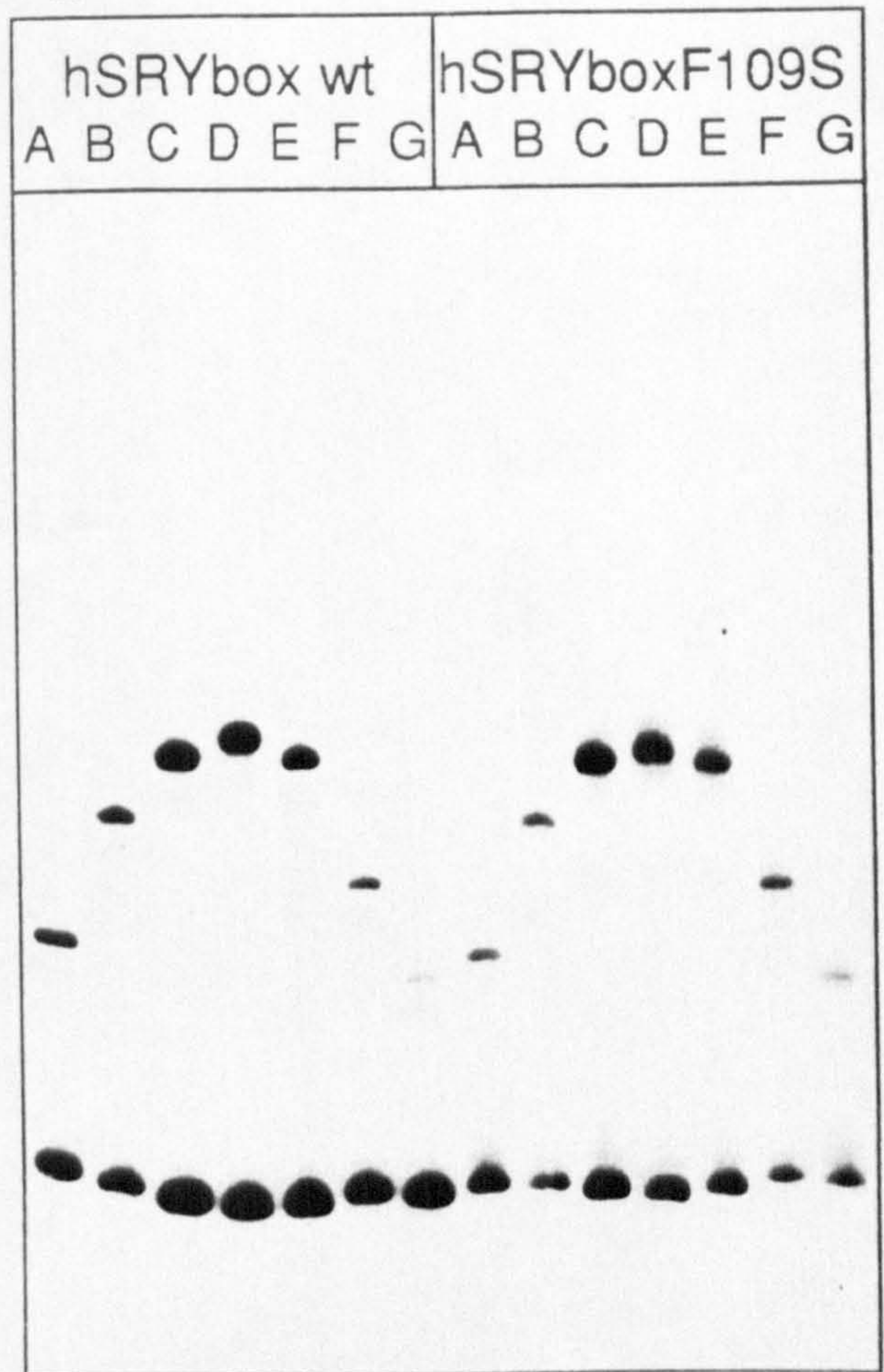
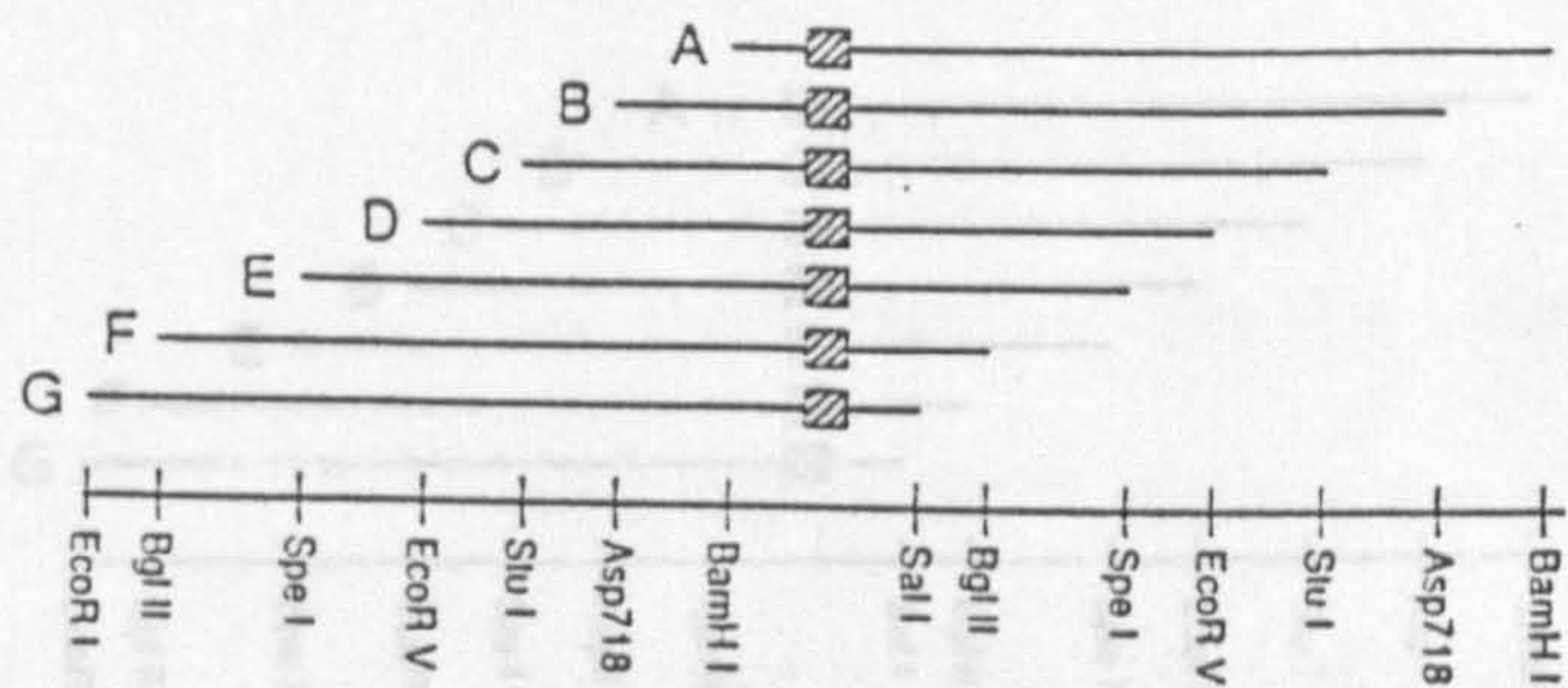


Fig. 10 A and B. DNA bending activity of mutant SRY HMG-boxes on the CD3ε sequence. (A) DNA probes used for the circular permutation analysis. Plasmid pBend2CD3ε, containing the CD3ε site (hatched box) flanked by tandemly repeated DNA sequences, was cleaved at the restriction sites indicated in the map. The DNA fragments obtained in this way (designated A-G) all contain circular permutations of the same sequence of 141 bp. DNA fragments containing CD3ε sequence, designated MUT11, were generated in the same way. (B) Circular permutation analysis of DNA bending induced by mutant hSRY box F109S. This panel shows a comparison between the electrophoretic mobilities of complexes of wild type hSRYbox and mutant F109S with probes A-G containing the CD3ε site GAACAAAG. Three fmol of labelled DNA probes were mixed in 9 μl of standard DNA binding buffer (see Experimental Procedures). To the various mixtures 1 μl of purified protein was added (2 ng of hSRYbox wild type and 2 ng of hSRYbox F109S). Electrophoresis and autoradiography were performed as indicated in material and methods. The retarded bands running behind the principal protein-DNA complex band are complexes containing more than one polypeptide molecule per DNA molecule. No significant difference in the mobility of the free DNA probes can be observed, indicating that the CD3ε site does not distort DNA on its own.



A



C

probe CD3ε

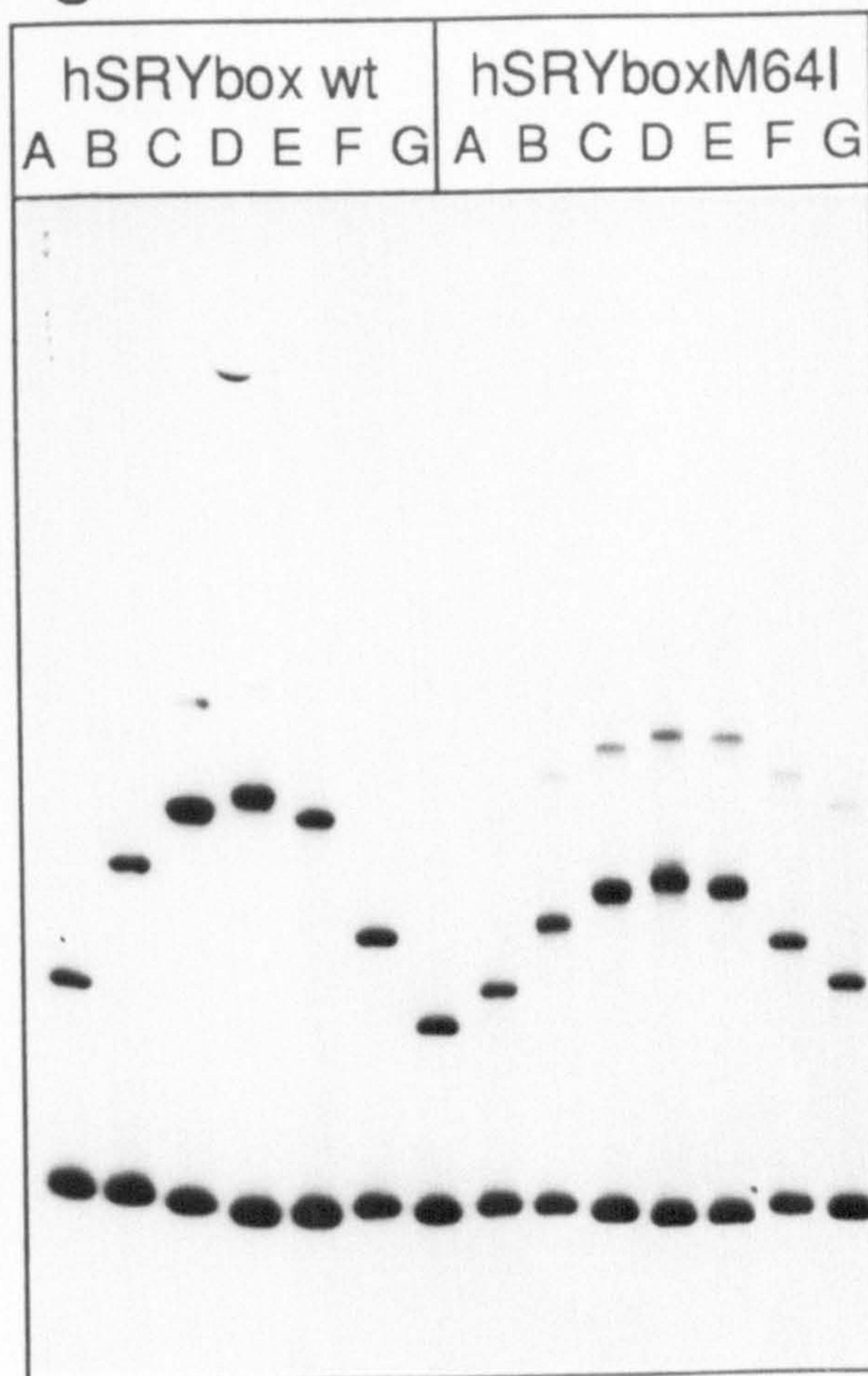
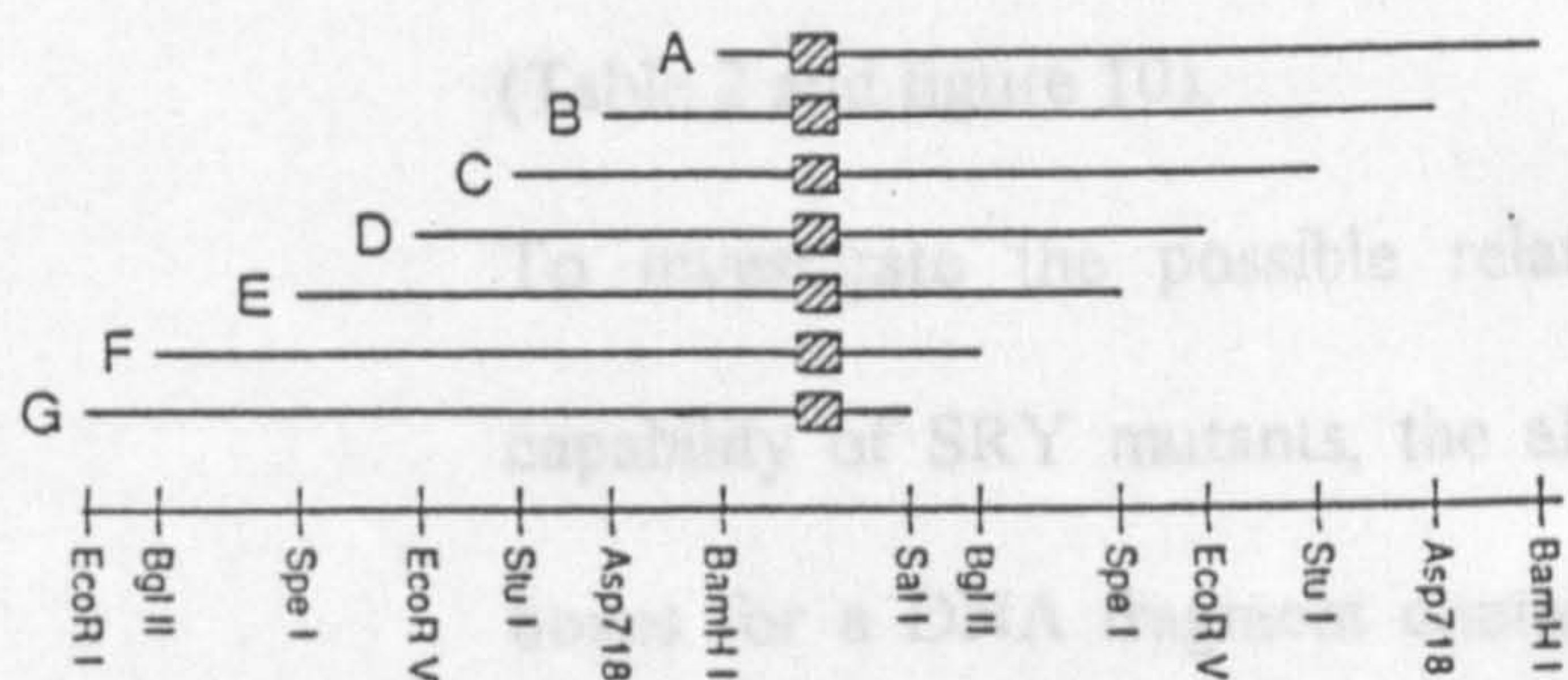


Fig. 10 A and C. DNA bending activity of mutant SRY HMG-boxes on the CD3ε sequence. (A) DNA probes used for the circular permutation analysis. Plasmid

Fig.10 A and C. DNA bending activity of mutant SRY HMG-boxes on the CD3ε sequence. (A) DNA probes used for the circular permutation analysis. Plasmid pBend2CD3ε, containing the CD3ε site (hatched box) flanked by tandemly repeated DNA sequences, was cleaved at the restriction sites indicated in the map. The DNA fragments obtained in this way (designated A-G) all contain circular permutations of the same sequence of 141 bp. DNA fragments containing CD3ε sequence, designated MUT11, were generated in the same way. (C) Circular permutation analysis of DNA bending induced by mutant hSRY box M64I. This panel shows a comparison between the electrophoretic mobilities of complexes of wild type hSRYbox and mutant M64I with probes A-G containing the CD3ε site GAACAAAG. Three fmol of labelled DNA probes were mixed in 9 μl of standard DNA binding buffer (see Experimental Procedures). To the various mixtures 1 μl of purified protein was added (2 ng of hSRY box wild type and 10 ng of hSRY box M64I). Electrophoresis and autoradiography were performed as indicated in materials and methods. The retarded bands running behind the principal protein-DNA complex band are complexes containing more than one polypeptide molecule per DNA molecule. No significant difference in the mobility of the free DNA probes can be observed, indicating that the CD3ε site does not distort DNA on its own.



A



D

probe CD3ε

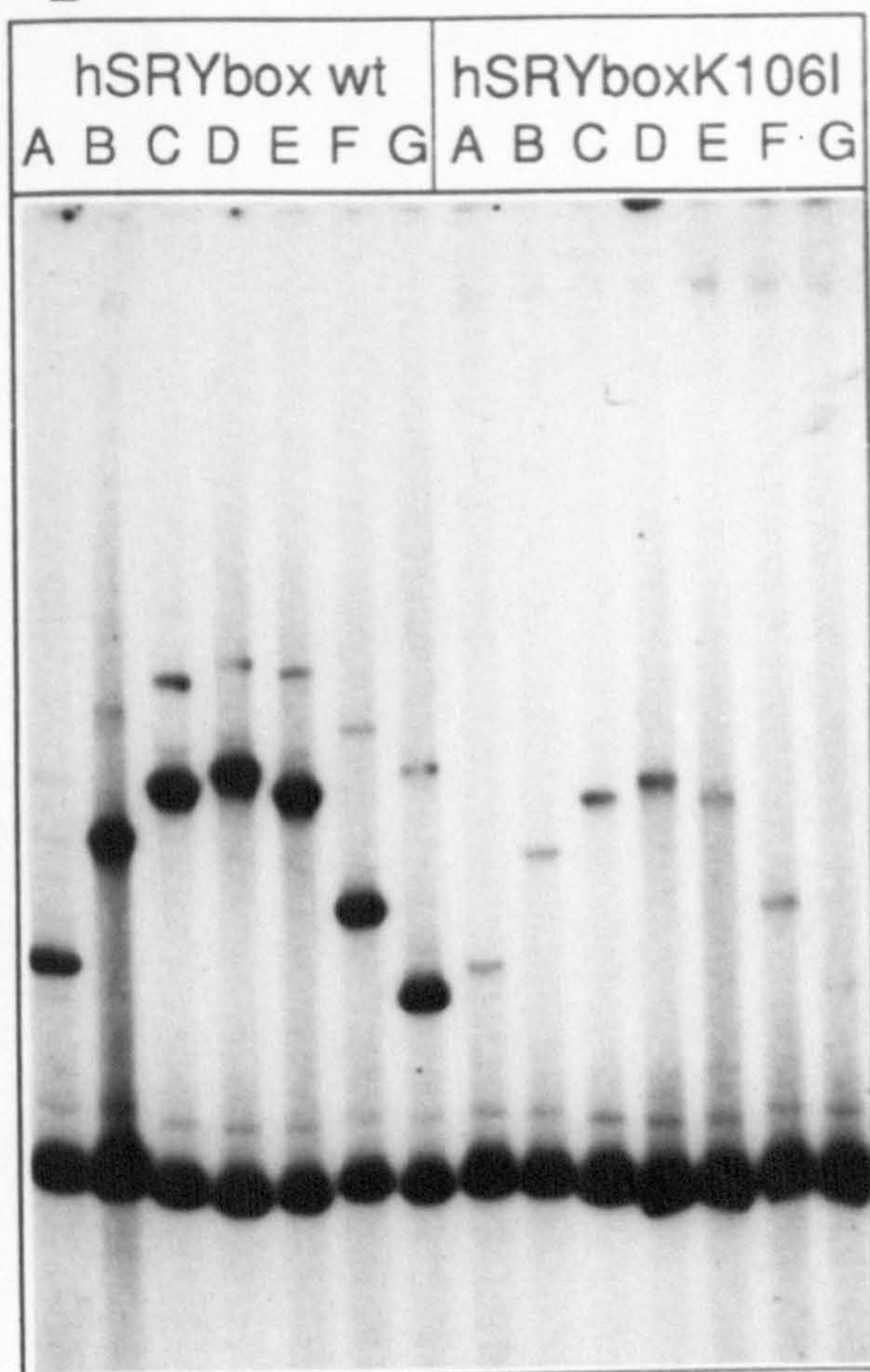


Fig. 10 A and D. DNA bending activity of mutant SRY HMG-boxes on the CD3ε sequence. (A) DNA probes used for the circular permutation analysis. Plasmid pBend2CD3ε, containing the CD3ε site (hatched box) flanked by tandemly repeated DNA sequences, was cleaved at the restriction sites indicated in the map. The DNA fragments obtained in this way (designated A-G) all contain circular permutations of the same sequence of 141 bp. DNA fragments containing CD3ε sequence, designated MUT11, were generated in the same way. (D) Circular permutation analysis of DNA bending induced by mutant hSRY box K106I. This panel shows a comparison between the electrophoretic mobilities of complexes of wild type hSRYbox and mutant K106I with probes A-G containing the CD3ε site GAACAAAG. Three fmol of labelled DNA probes were mixed in 9 μl of standard DNA binding buffer (see Experimental Procedures). To the various mixtures 1 μl of purified protein was added (2 ng of hSRY box wild type and 25 ng of hSRYbox K106I). Electrophoresis and autoradiography were performed as indicated in Experimental Procedures; however, the autoradiography in panel D was overexposed to detect the faint bands of DNA complexed to hSRY box K106I. The retarded bands running behind the principal protein-DNA complex band are complexes containing more than one polypeptide molecule per DNA molecule. No significant difference in the mobility of the free DNA probes can be observed, indicating that the CD3ε site does not distort DNA on its own.



the DNA in the centre of the AACAAAG site. Table 2 summarises results from circular permutation assays obtained with the complete set of mutated SRY boxes.

Wild type SRY box bends the DNA at CD3e site by an angle of 75 degrees, while mutant M64I induces a significantly smaller angle, of about 56°. All other mutants, including K106I, bend the DNA by an angle extremely close to the wild type protein (Table 2 and figure 10).

To investigate the possible relationships between binding affinity and bending capability of SRY mutants, the affinities of the normal and the five mutant HMG boxes for a DNA fragment containing the CD3e enhancer site were determined in band shift assays using a limiting amount of labelled DNA and titrating the polypeptides. As described before, under conditions of protein excess, the dissociation constant is equivalent to the concentration of polypeptide where half of the input DNA is taken up in the complexes and half is free.

Mutant HMG boxes varied extensively in their DNA binding ability, in accordance with a previous report (Harley et al., 1992). The affinity for DNA of mutant F109S was indistinguishable from that of the wild type, as already noted by Jager et al. (1992). The affinity of mutants I90M and M64I was moderately reduced in comparison to the wild type, whereas the affinity of K106I was reduced by more than two orders of magnitude, and G95R did not form any complexes under the conditions used (Table 2). Variations of activity in a bandshift assay can be caused by a variation in the intrinsic affinity of the proteins for DNA, or by the presence of a variable proportion of denatured or inactive protein molecules in the preparation. To rule out the latter alternative, we tested the DNA binding activity of various preparations of the same polypeptide, and found them identical.

Our results indicate clearly that mutant M64I bends the DNA much less than the wild type, but its binding affinity is reduced only mildly, whereas mutant K106I bends the DNA almost as much as the wild type, although its DNA binding affinity is reduced by more than two orders of magnitude. Thus, contrary to a widely held preconception, affinity and DNA bending are not correlated, and a weakly binding protein can produce a large DNA distortion.

## **2.4 WILD TYPE SRY AND THE M64I MUTANT RECOGNISE THE SAME REPERTOIRE OF SEQUENCES IN LINEAR DNA**

The difference in the angle induced by mutant M64I, viewed against the relative uniformity of the other mutants, appears as quite large. In comparison, the DNA bending properties of SRY proteins from mutations associated with a mild biological effect are strictly conserved. Thus, it is tempting to consider the difference in bending ability of the M64I mutant as a key to the biological consequences of the mutation. A more trivial explanation is possible, however: mutation M64I might modify the sequence specificity of SRY protein. In order to test this hypothesis, we performed an experiment of PCR-assisted binding site selection. Binding-site selection is used to determine the target specificity of a sequence-specific DNA-binding protein. The technique has a number of applications, for example to identify DNA target sequences for proteins with unknown DNA-binding specificities to or to provide additional information on the protein-DNA interactions of previously characterised DNA binding domains.

As indicated in Materials and Methods, a pool of random-sequence oligonucleotides is used as the source of potential binding sites. The oligonucleotide pool is made double stranded and labelled with  $^{32}\text{P}$  dCTP. Several cycles of selection and amplification of a random population of double-stranded DNA are performed. The starting population consists of a family of double stranded 60 bp DNA molecules initially displaying all possible sequence variations at the central 10 nucleotide residues. Flanking regions are 25 base pairs long and constant in sequence. Two *EcoRI* restriction sites were present in the two constant regions flanking the 10 bp central random sequence. The starting population of molecules was labelled with  $^{32}\text{P}$  and then incubated with different



amounts of M64I hSRY box protein. Samples were applied on a vertical polyacrilamide gel and electrophoresed at room temperature. The shifted band, containing a subset of the original population of labelled molecules, was recovered by cutting out a gel slice, crushing it and eluting the DNA. The eluted DNA was amplified by PCR, using dCTP labelled with  $^{32}\text{P}$ . The radioactive product was gel purified, incubated again with M64I hSRY box and electrophoresed as described. Shifted band was recovered, eluted, amplified and incubated again as described before. Five rounds of selection/amplification were sufficient to select a population with maximal affinity for the protein. Additional cycles of selection/amplification did not increase the average affinity of the population for the protein. Selected molecules were digested with *EcoRI*, gel purified and subcloned in the *EcoRI* site of the vector pUC19. Bacteria carrying recombinant plasmids give rise to white colonies. White colonies were then picked at random; plasmids were extracted and selected molecules were sequenced (figure 11). These sequences can be aligned, are rich in A's and T's and the distribution of bases at seven individual positions is clearly non random. The consensus that can be derived matches exactly the consensus determined for molecules selected in selection/amplification protocol by wild type SRY protein, A/T A A C A A A/T (Harley et al., 1994).

We also compared the binding preferences of both mutant M64I and wild type SRY boxes with respect to the entire repertoires of sequences, rather than individual sequences. A competition between the initial (non-specific) and final (specific) population of DNA molecules was performed (figure 12). Labelled templates selected by M64I hSRY protein were purified and amplified after the first round of selection (first round) and after the fifth round (last round). These two families of probes were mixed in different ratios. Mixing the initial and final populations in different

```

1  cggCATTACAAACacg
2  ggCGATAACTCAccga
3  gGGTCAAACAAccacg
4  gAGAGAAACAAccacg
5  ggCCCAATGAACccac
6  gGACAAATCACccacg
7  gt cggAAACACAGTGc
8  gt ggTAAACTTACCcc
9  accCAAAGGACAAggc
10 ggTTGAAACAAAccac
11 ggTCCTATCAGTccac
12 CACCGAAAGGAggtgc
13 cggTAAATCTAGTcca
14 ccCCCAAAGAGAggtg
15 tggACAAACTCCAaccg
16 ggCATAAAGTTCccga
17 gt cggAAAGAACTTTc
18 gt cggAAACAAAGCTc
19 t cggGAAACTAATTcc
20 ggAGTAAACATCccac
21 cggCGAATCTAATcca
22 ccCTTAATCTAAGggt
23 gt cggTAACAAACAAc

```

```

A    19 22 16 - 14 13 12
C    - - - 17 - 5 8
G    - - 1 6 1 2 2
T    4 1 6 - 8 3 1

```

consensus A/T A A C A A A/T

Fig 11 The HMG boxes of wt SRY and mutant M64I recognise the same repertoire of sites. Sequence of some DNA molecules selected after the fifth selection/amplification cycle with SRY HMG-box M64I. Sequences were aligned following the algorithm of the program "Align". The lower-case bases belong to the polylinker of plasmid. The frequency of each base at each position is indicated below the sample set of sequences. This consensus that can be derived matches the consensus for wild type SRY (Harley et al., 1994).



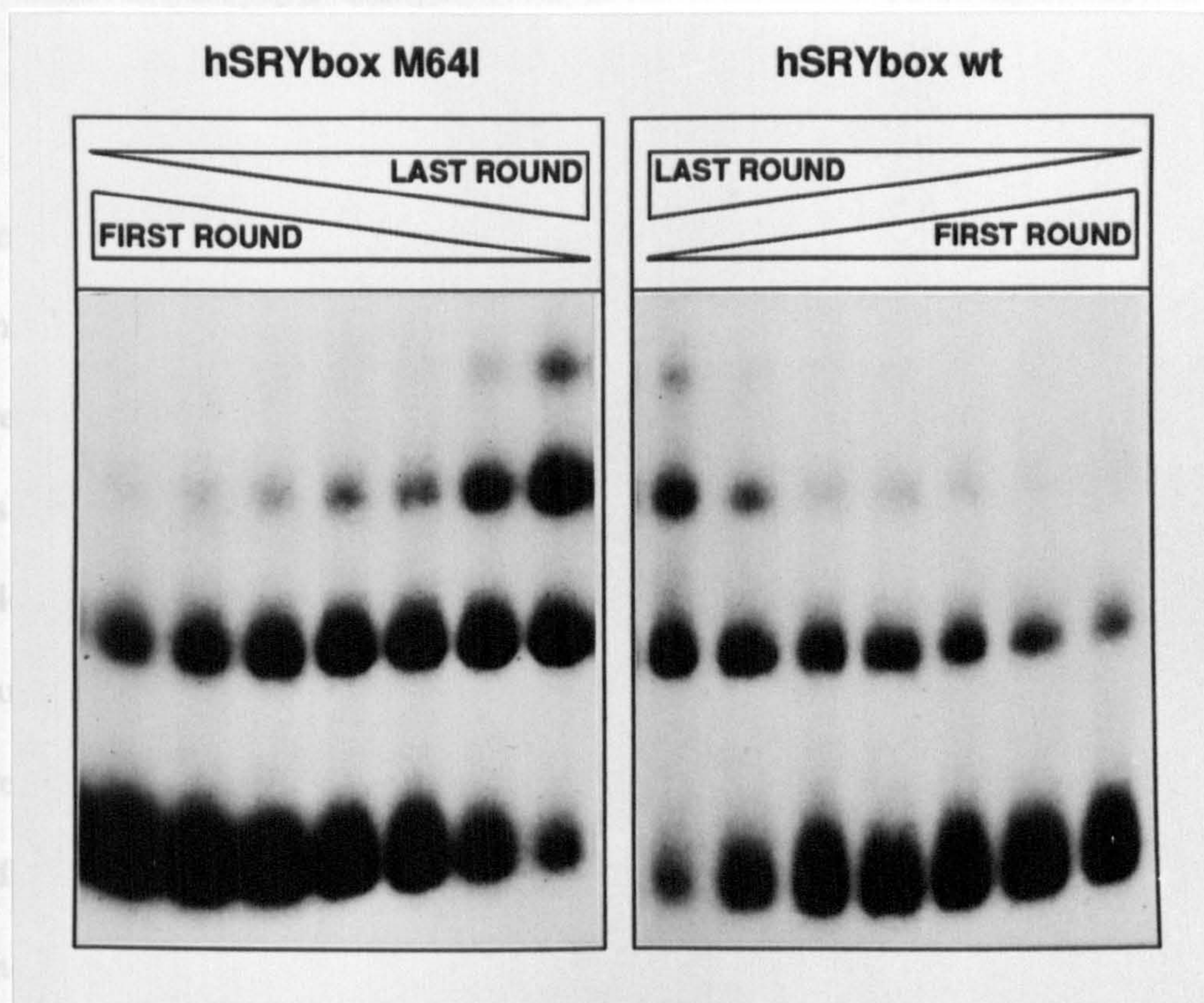


Fig. 12. Wt SRY box and mutant M64 recognise the same repertoire of sites. A population of 60 bp DNA molecules initially containing all possible 10 bp sequences at their centre was subjected to five rounds of selection/amplification by SRY box M64I (see materials and methods). At the end of the procedure, labelled templates amplified after the first round of selection by hSRY box M64I (first round) and after the fifth and last round were mixed in different ratios, mixed with wt SRY box or M64I SRY box, and electrophoresed. The same total amount of DNA (21 fmol) was used as probe in every lane (last round/first round: 0/21, 3/18, 7/14, 10.5/10.5, 14/7, 18/3, and 21/0). To 9  $\mu$ l of reaction mix containing the DNA probes, 20 ng of hSRY box M64I (panel above, right) or 5 ng of hSRY box wt (panel above, left) were added. After incubation on ice, samples were applied to vertical 6.5% polyacrylamide gels in 0.5X TBE as described in materials and methods. As expected, the population of DNA molecules amplified after the fifth round of selection/amplification with the mutant SRY box M64I are recognised much better by mutant M64I than the initial population. However, the same is true if the populations are challenged with the wt SRY box. Mixing the initial and final populations in different proportions is equivalent to performing a competition experiment between a specific binding site and aspecific DNA. As the figure shows, again no difference is apparent between the bandshifts produced by the wt SRY box and mutant M64I, an indication that their binding selectivity is the same.



proportions is equivalent to performing a competition experiment between a specific binding site and non specific DNA. The same amount of DNA was used as a probe in every lane and incubated with wt hSRY box or M64I hSRYbox and electrophoresed. As expected, the population of DNA molecules amplified after the fifth round of selection/amplification with the mutant M64I SRY box are recognised much better by mutant M64I than the initial population. However, the same is true if the populations are challenged with the wt SRY box (figure 10). As figure 12 shows, again no difference is apparent between the bandshifts produced by the wt SRY box and mutant M64I, an indication that their binding selectivity is the same. Similar results were also obtained using defined sequences rather than populations. We conclude that the mutant M64I and the wild type SRY boxes recognise the same sites, and that each specific “preferred” sequence is recognised with similar specificity over “nonpreferred” sequences.



## **2.5 DNA SEQUENCE AFFECTS DNA BENDING BY SRY**

The experiments reported in the previous sections establish that one specific DNA sequence is molded into different configurations by mutant M64I and by the wild type SRY. However, there is no guarantee that the CD3e site is a true SRY binding site; in fact, a physiological target for SRY has not been identified so far. Moreover, the DNA itself is possibly the major determinant in the final conformation of the nucleoprotein complex, and slight variations of the DNA sequence interacting with SRY might affect the geometry of the nucleoprotein complex quite substantially. We then measured by means of the circular permutation assay the angle of deflection produced by wild type and mutant SRY proteins on two sequences that diverge slightly from the CD3e sequence.

The sequence TAACAATG (MUT11) has been reported to bind SRY protein with higher affinity than the GAACAAAG sequence present in the enhancer of the CD3e gene, while the sequence GAACACAG (MUT0) was reported to be a poorer binding site (Giese et al., 1992). The two sequences were substituted for the CD3e sequence in plasmid pBendCD3e. This was obtained by annealing synthetic oligonucleotides CTAGAGAGCGCATTGTTATCAG and TCGACTGATAACAATGCGCTCT for MUT11, CTAGAGAGCGCTGTGTTCTCAG and TCGACTGAGAACACAGCGCTCT for MUT0. Double stranded molecules contain sticky ends for restriction endonucleases *Xba*I and *Sal*I.

These molecules were subcloned within the sites *Xba*I/*Sal*I of the vector pBend2; recombinant vectors were checked by sequencing. Cleavage with several restriction endonucleases yields a set of fragments of DNA of identical length and sequence, but a different distance from the DNA binding domain and the edges of the molecules.

The circularly permuted fragments obtained from the two new plasmids (called pBendmut11 and pBendmut0, respectively) were purified from agarose gels, labelled and tested electrophoretically with wild type and mutant hSRY boxes for binding affinity and bending features (Table 3 and figure 13), in the same way as described before.

The fragments bearing the sequence TAACAATG bound every hSRY box variant with about twofold better affinity than the GAACAAAG sequence. The circular permutation assay showed that the locus of flexure was the same as for pBendCD3e fragments, but the bending was slightly less.

The fragments bearing the sequence GAACACAG bound every SRY HMG-box variant about six fold less efficiently than the pBendCD3e fragments. Again, the protein-induced bending was centered around the same locus, showing that the GAACACAG sequence was preferred to any other sequence in the probe. However, the bending was dramatically reduced with the wild type SRY box, while mutant M64I appeared unable to induce any measurable bending of these fragments (figure 13).



| Probe | Sequence | wt                   |       | M64I                   |       |
|-------|----------|----------------------|-------|------------------------|-------|
|       |          | KD                   | Angle | KD                     | Angle |
| CD3e  | GAACAAAG | $2 \times 10^{-8}$ M | 76    | $6.5 \times 10^{-8}$ M | 56    |
| mut11 | TAACAATG | $10^{-8}$ M          | 73    | $3 \times 10^{-8}$ M   | 55    |
| mut0  | GAACACAG | $10^{-7}$ M          | 35    | $3.5 \times 10^{-7}$ M | <10   |

Table 3. Influence of the DNA sequence on the amplitude of the angle induced by SRY proteins. The dissociation constants for complexes between SRY boxes and DNA were calculated by titrating the SRY box polypeptides against 0.25 nM DNA and assaying the formation of complexes by electrophoretic mobility shift assays (see material and methods). "Angle" refers to the angle calculated by the algorithm from the circular permutation assays on the set of probes bearing the CD3e site. The reported values are the mean of at least two assays.



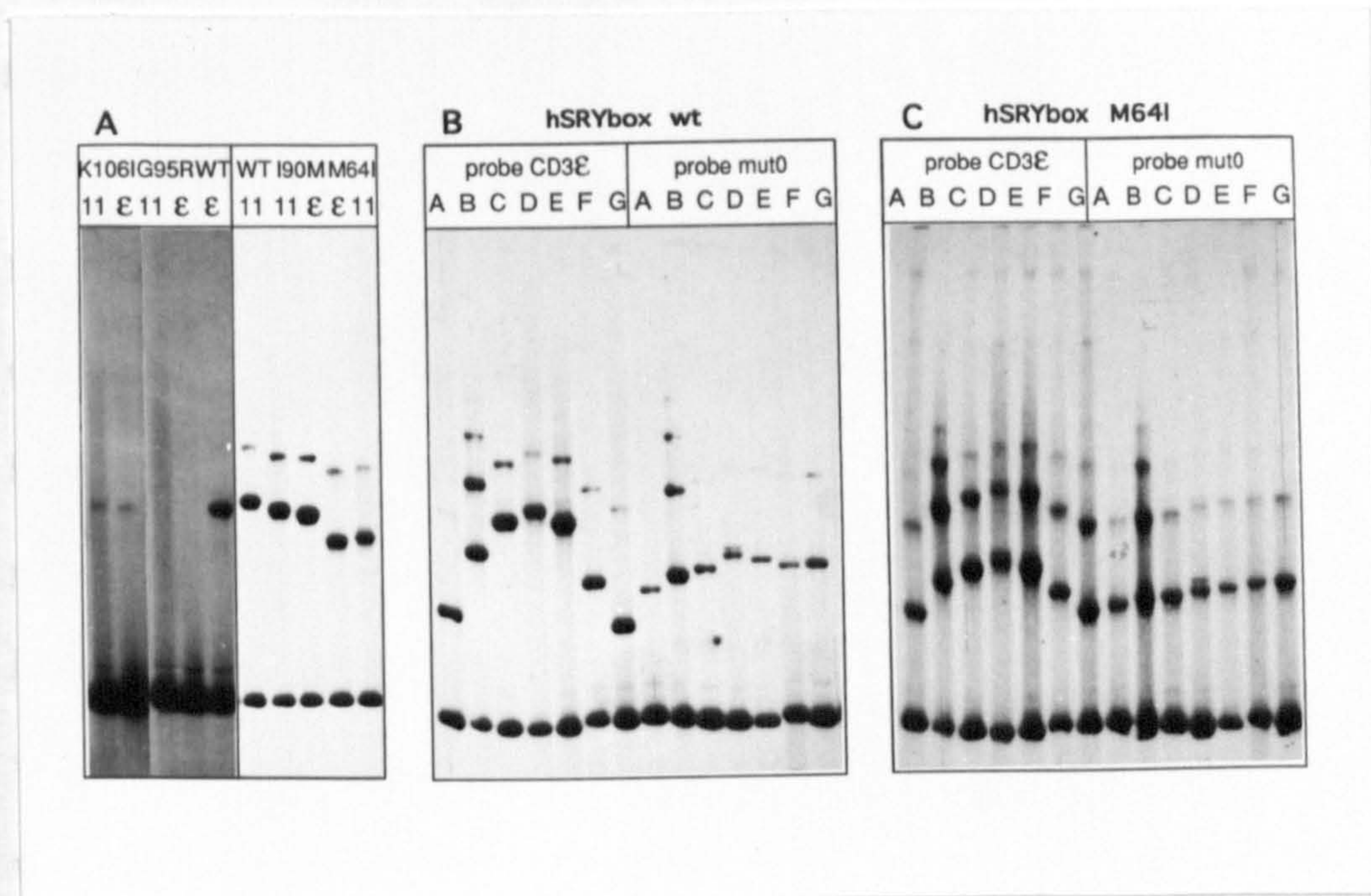


Fig. 13. DNA binding sites with different sequences are bent differently by SRY boxes. (A) Mobility shift assay of wild type and mutant SRY boxes on CD3e (e) and mut11 (11) binding sites. Probe D was used because it contains the DNA binding site (CD3e or mut11) in the middle of the fragment and the differences in electrophoretic mobilities are maximal. DNA binding mixtures (see materials and methods) were set up to contain 3 fmol of CD3e D probe or mut11 D probe in 9  $\mu$ l. To each mixture containing the CD3e D probe, 1  $\mu$ l of purified protein was added, containing 50 ng of hSRY box of G95R, 3 ng of hSRY box wild type, 10 ng of hSRY box I90M, 15 ng of hSRY box M64I, 25 ng of hSRY box K106I, respectively. About a quarter as much wild type hSRY box and half as much for each mutant SRY box were added to mixtures containing the mut11 D probe. Autoradiographic signals were also analysed with a PhosphorImager. No binding activity was found for hSRY box G95R. The same results were obtained also when using larger amounts of purified hSRY box G95R protein. (B,C) Circular permutation analysis of wild type hSRY box (panel B) and mutant M64I (panel C) on CD3e and mut0 binding sites. DNA probes A-G (3 fmol) were mixed in 9  $\mu$ l of standard DNA binding buffer (see materials and methods). To the various mixtures containing labelled DNA fragments 1  $\mu$ l of purified protein solution was added, containing 2 ng of hSRY box wild type (probe CD3e), 6 ng of hSRYbox wild type (probe mut0), 15 ng of hSRY box M64I (probe CD3e), 60 ng of hSRY box M64I (probe mut0). The samples were electrophoresed and autoradiographed as described in materials and methods. The retarded bands running behind the principal protein-DNA complex band are complexes containing more than one polypeptide molecule per DNA molecule. Unbound DNA probes containing mut0 site have the same electrophoretic mobility, indicating that the mut0 site, as well as the CD3e site, does not distort DNA on its own.



## **2.6 THE POSITIVE CONTROL: THE EVOLUTIONARY TREE OF PRIMATES.**

In the first part of this work, anomalous DNA protein complexes were shown to be correlated to anomalous biological function of SRY protein.

Indeed all single point mutations in the human SRY coding sequence that cause gonadal dysgenesis and sex reversal fall in the HMG box region (Berta et al., 1990; Hawkins et al., 1992 and 1994). In addition, the HMG box is the only part of SRY which is conserved during the evolution of mammals: the protein segments flanking the HMG box are variable in sequence and length and cannot be aligned except between closely related species (see introduction). All the DNA binding and bending activity resides in the HMG box region. If binding and bending are really important for the biological function of SRY proteins, these features will be expected to be positively selected and conserved during evolution. The purpose of the second part of this work is then to demonstrate that the architecture and stability of nucleoprotein complexes formed by the DNA-binding domains of SRY proteins is of primary importance and therefore conserved irrespective of the time of divergence during primate evolution and the number of residues modified during this evolution.

## **2.7 PRODUCTION OF SRY DNA BINDING DOMAINS OF DIFFERENT PRIMATES.**

We first chose to compare SRY HMG boxes (SRY boxes) of different primates on the basis of their sequence similarity. Figure 14 shows the sequences of predicted SRY boxes from 8 primates (Whitfield et al., 1993). Dots indicate agreement with the HMG box of Human SRY. All the residues modified during evolution map to a different position from those associated with sex reversal phenotypes in humans.

To study the behavior of the DNA protein complexes of SRYs boxes it was first necessary to express and purify the corresponding proteins. We obtained DNA fragments coding for SRY boxes by PCR of total genomic DNA with the same specific oligonucleotides used for sex reversal mutations. We amplified DNA coding for SRY proteins from genomic DNA of common chimpanzee (*Pan troglodytes*), pygmy chimpanzee (*Pan paniscus*), orang-utan (*Pongo pygmaeus*), gorilla (*Gorilla gorilla*), gibbon (*Hylobates lar*), baboon (*Papio spp.*) and marmoset (*Callithrix jacchus*). Briefly, to allow the synthesis of the boxes in *E. coli*, the first amino acids (aa) of the HMG box (N in marmoset, D for all the other boxes listed in figure 14 ) was changed to M to provide a translation start codon, and the last (N for chimpanzee, pygmy chimpanzee and marmoset, K for all the other boxes in figure 14) was changed to a stop codon. The same oligos, SRY box dir and SRY box rev, used for the amplification of human sex reversal SRYs were used for PCR according to conditions already reported in the previous section. Reaction products were cleaved with *NdeI* and *HindIII*, gel purified and cloned between *NdeI* and *HindIII* sites of plasmid pT7-7. The resulting plasmids were introduced in strain B121 (DE3), as described in the section concerning the construction of human SRY mutant proteins. The



|                                       |            |            |            |            |            |            |            |           |
|---------------------------------------|------------|------------|------------|------------|------------|------------|------------|-----------|
| HUMAN ( <i>Homo sapiens</i> )         | DRVKRPMNAF | IVWSRDQRRK | MALENPRMRN | SEISKQLGYQ | WKMLTEAEKW | PPFQEAQKLQ | AMHREKYPNY | KYRPRRKAK |
| CHIMPANZEE ( <i>Pan troglodytes</i> ) | .....      | F.....     | .....      | .....      | .....      | .....      | .....      | .....N    |
| PYGMY CHIMP. ( <i>Pan paniscus</i> )  | .....      | F.....     | .....      | .....      | ....A....  | .....      | .....      | .....N    |
| GORILLA ( <i>Gorilla gorilla</i> )    | .....      | .....      | .....      | .....      | .....      | .....      | .....      | .....     |
| ORANG ( <i>Pongo pygmaeus</i> )       | .....      | .....      | ....K...   | .....      | .....      | .....      | .....      | .....     |
| GIBBON ( <i>Hylobates lar</i> )       | .....      | .....      | ....K...   | .....R     | .....      | .....      | .....      | .....     |
| BABOON ( <i>Papio spp</i> )           | .....      | .....      | ....K...   | .....      | .....D..   | .....      | .....      | .....     |
| MARMOSET ( <i>Callitrix jacchus</i> ) | N.....     | .....      | ..V..D..   | ....R....  | ..L.....   | .....      | .....      | .....N    |
| SEX REVERSAL MUTATIONS                | .....I...  | .....      | .....      | ..M....R.. | .....I.    | .S.....    | .....      | .....     |

Fig. 14. Alignment of HMG-box sequences from SRY proteins of primates. Dots indicate agreement with the human aa sequence. Upper case letters indicate non-synonymous mutations with the resultant aa change shown. The gorilla and human SRY boxes are exactly identical in sequence. The last line reports some single-site substitutions in humans that cause sex reversal: the mutated aa is indicated.

corresponding peptides, which are called SRY boxes, were produced efficiently as described in the previous section, and were soluble. Recombinant proteins were purified in the same way as described in the first section of this work. Briefly, nucleic acids were removed by batch absorption to DEAE-cellulose at 0.45 M NaCl; the supernatant was filtered and applied to a Mono S HR5/5 FPLC column (Pharmacia) equilibrated with 20 mM HEPES pH 7.9, 0.2 mM EDTA. The column was eluted with a linear gradient from 0 M to 2 M NaCl in 20 mM HEPES pH 7.9, 0.2 mM EDTA. SRY boxes were eluted in fractions containing about 1 M NaCl and were more than 99% pure. Protein concentrations were estimated by reading at 280 nm and by scanning protein titrations on SDS-polyacrylamide gels.



## **2.8 ARCHITECTURE OF THE DNA COMPLEXES INDUCED BY PRIMATE SRY PROTEINS**

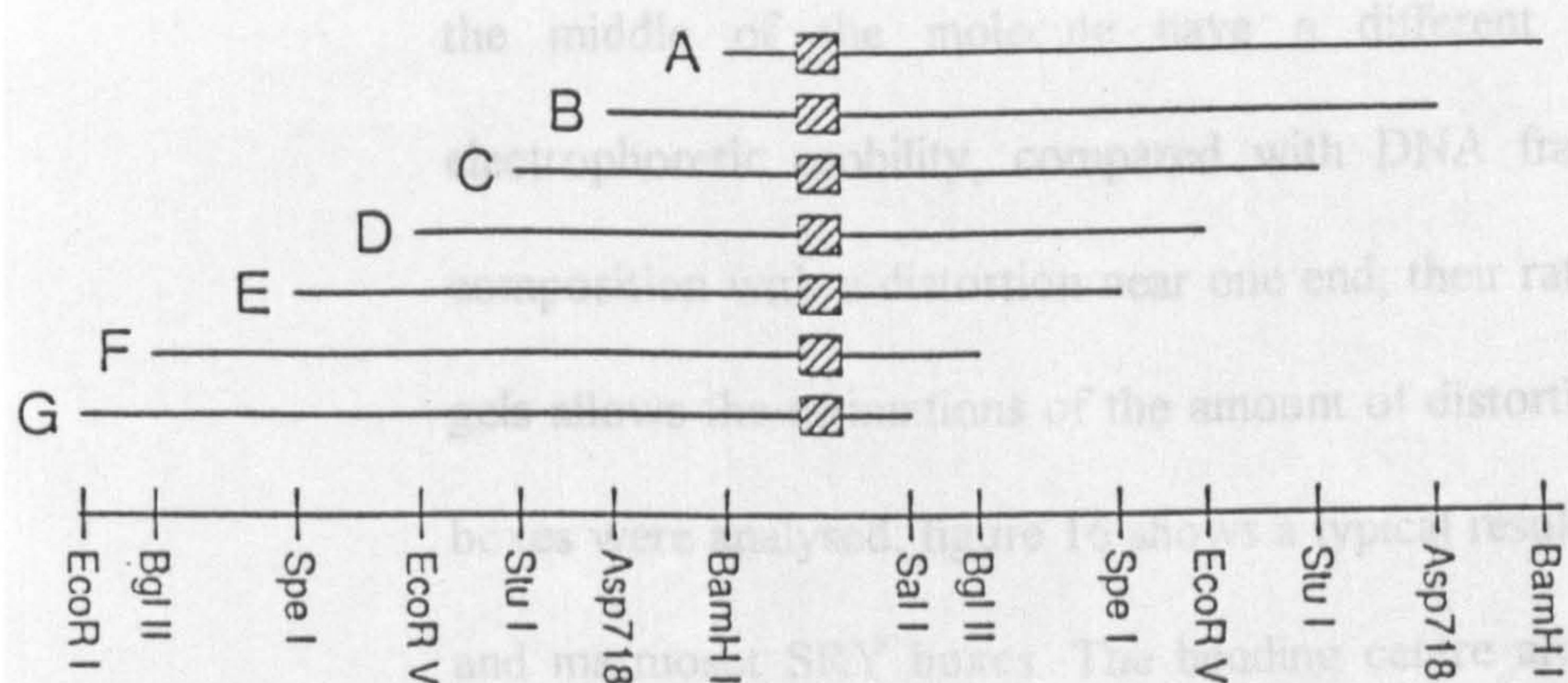
To estimate the dissociation constant of the complexes between SRY boxes and DNA, and the deformation induced in DNA, we performed bandshift and circular permutation assays.

The same short sequence, AACAAAG, cloned between directly repeated sequences in plasmid pBend2 shown in the previous section to be a good DNA binding site for wt human and mutant SRY proteins was used as a probe for bandshift and circular permutation assays. Experimental conditions are the same as reported with human SRY proteins.

Cleavage with several restriction endonucleases yielded the same set of fragments of identical length and circularly permuted sequence (figure 15). Fragments D and A were used to estimate the affinity of the various primates SRY boxes for the AACAAAG binding site. Individual SRY boxes were titrated on limiting amounts of DNA in a band shift assay: as described in the first section, under conditions of protein excess, the  $K_D$  corresponds to the concentration of SRY boxes where half of the input DNA is engaged in the specific complex. All SRY boxes had identical titration profiles (not shown). Figure 15 shows a bandshift assay performed using labelled DNA fragment D, which contains the AACAAAG binding site in the middle, with the same amount of pure SRY boxes from different primates. The amount of free and complexed DNA is very similar in all cases, and corresponds to a dissociation constant of  $10^{-8}$  M. The different electrophoretic mobility of DNA-protein complexes is due to the different bend induced in the DNA (see the first section).



A



B

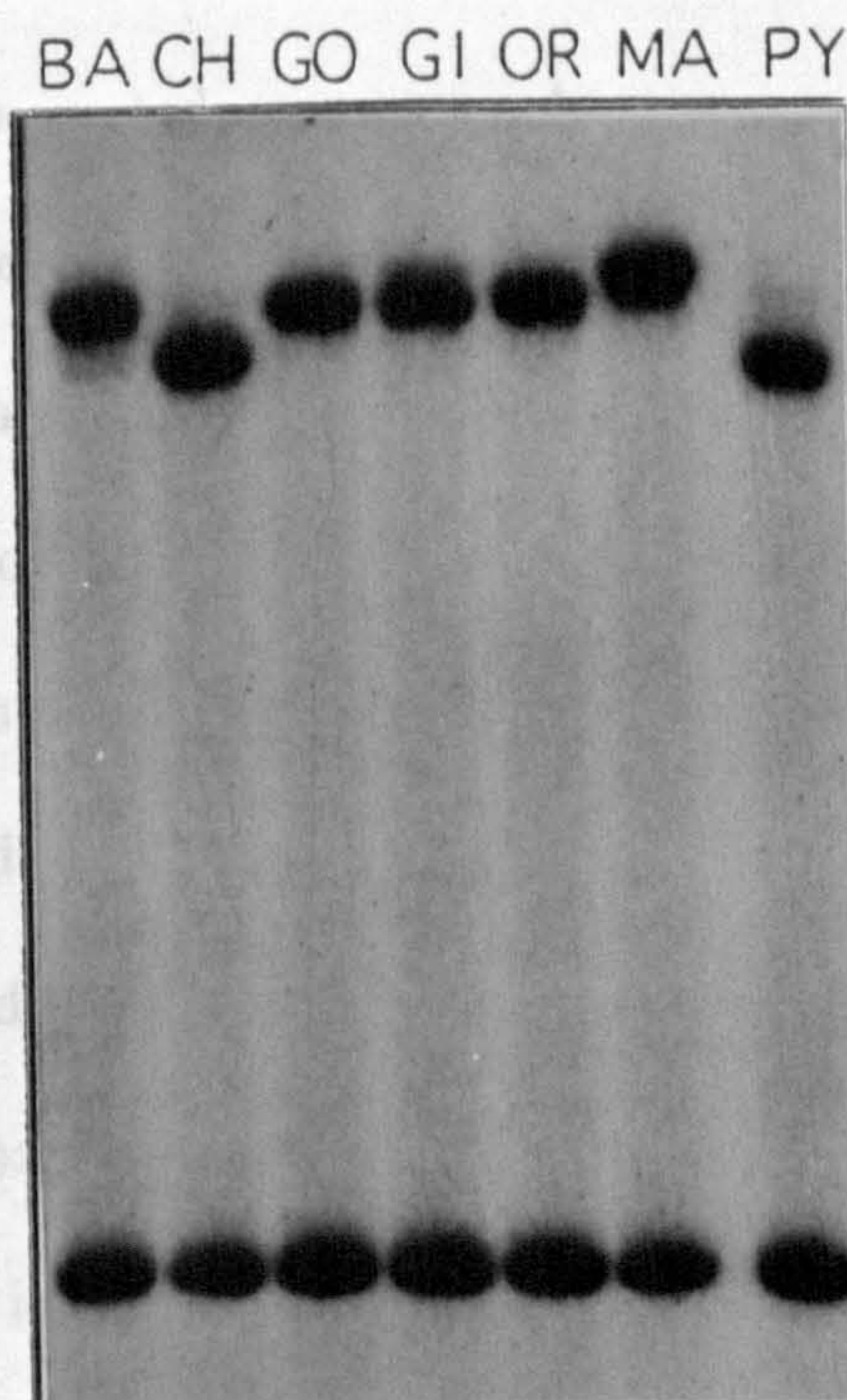


Fig. 15. DNA binding and bending properties of HMG-boxes of primate SRY proteins. (A) Construction of the probes for the electrophoretic mobility shift and circular permutation assays. Plasmid pBend2CD3e, containing the 5'-AACAAAG binding site (hatched box) flanked by tandemly duplicated DNA sequences, was cleaved at the restriction sites indicated in the map. The DNA fragments obtained in this way (designated A-G) contain circular permutations of the same sequence of 141 bp. (B) The different SRY-boxes have the same affinity for the 5'-AACAAAG binding site. SRY-box peptides (5 ng) of primates (BA, baboon; CH, chimpanzee; GO, gorilla; GI, gibbon; OR, orang; MA, marmoset; PY, pygmy chimpanzee) were mixed with labelled DNA fragment D (8 fmol) in 9  $\mu$ l of DNA binding buffer (8% Ficoll, 100 mM NaCl, 10 mM HEPES pH 7.9). After incubation for 10 min on ice, 5  $\mu$ l samples were subjected to electrophoresis at 11 V  $\text{cm}^{-1}$  at room temperature in vertical 10% PA gels containing 45 mM Tris-borate buffer, pH 8.3. The gels were dried and autoradiographed with Kodak XAR-5 films at -80C for 24-72 h. Under conditions of protein excess, these concentrations of SRY-boxes (10-8M) cause the shift of half of the input DNA, and correspond to the dissociation constant of the complex.



We then tested the electrophoretic mobility of the different SRY boxes with the complete set of labelled DNA fragments with circular permuted sequence. As described in more detail in the previous section, DNA fragments with a distortion in the middle of the molecule have a different shape, and hence a different electrophoretic mobility, compared with DNA fragments of identical length and composition with a distortion near one end; their rate of migration in polyacrylamide gels allows the estimations of the amount of distortion introduced in DNA. All SRY boxes were analysed; figure 16 shows a typical result obtained with the pygmy chimp. and marmoset SRY boxes. The bending centre always maps to the 5' AACAAAG binding site. Protein complexes with fragment A and G had identical mobility for all SRY boxes, while the mobility of the complex with fragment D was slightly variable. Moreover, the variation is small indeed, and the deduced DNA angles are very similar in the complex formed by all primate SRY boxes (Table 4). The small differences in the angles induced by SRY boxes of different primates does not correlate with the divergence times of the different species from the primate evolutionary tree. In fact the difference in the induced bend between human and chimpanzee is slightly more than between human and gibbon or between human and marmoset, contrary to the evolutionary divergence times: (chimp vs. human 7 millions of years (Myr), gibbon vs. human 19 Myr, marmoset vs. human 27 Myr. Nei, 1987). These data suggest that the differences are not significant, and that essentially the same geometry of the SRY-DNA complex is required for sex determination. Thus, the DNA binding and bending properties of the marmoset SRY box are not significantly different from those of the human SRY box, although it differs in 4 aas. Therefore, the alternative aa in the SRY boxes of the different primates must be considered neutral mutations. This has to be contrasted with the single point mutations associated with loss of function of SRY



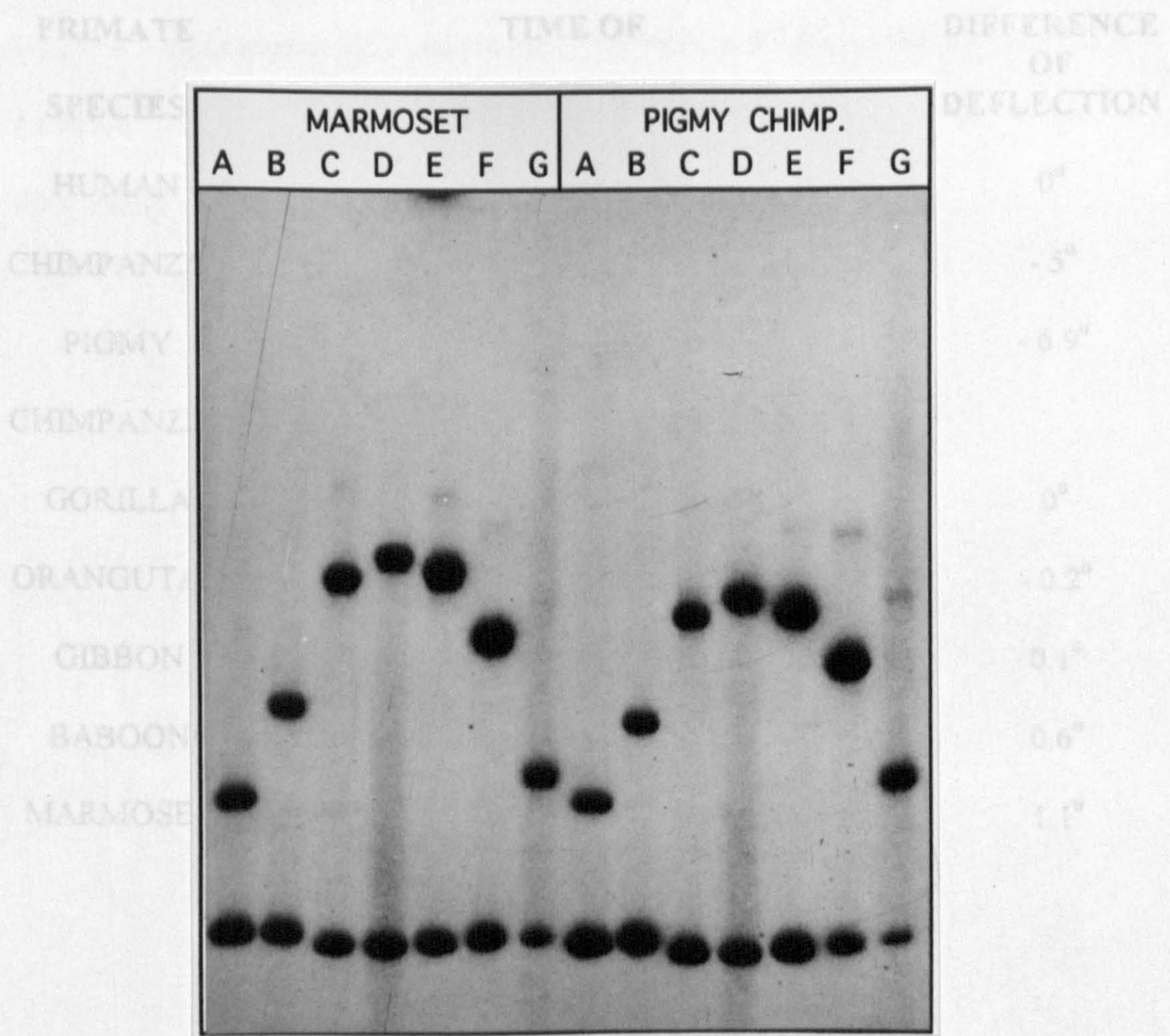


Table 4. No correlation between the time of divergence from the human line of descent and the amplitude of the deflection in DNA induced by SRY.

Fig. 16. Circular permutation analysis of the DNA flexure induced SRY-boxes of time of divergence from human (expressed in Myr) as reported by Nei (1987), marmoset and pygmy chimpanzee. Eight fmol of circular permuted DNA fragments (from A to G) were mixed with 5 ng of SRY boxes (first panel, marmoset; second, pygmy chimpanzee). After incubation on ice, the samples were assayed as described in material and methods. The difference in amplitude of the two curves corresponds to the small difference in DNA bending induced by the two proteins.



| PRIMATE SPECIES  | TIME OF DIVERGENCE | DIFFERENCE OF DEFLECTION |
|------------------|--------------------|--------------------------|
| HUMAN            |                    | 0°                       |
| CHIMPANZEE       | 6.5                | - 5°                     |
| PIGMY CHIMPANZEE | 6.5                | - 6.9°                   |
| GORILLA          | 8.3                | 0°                       |
| ORANGUTAN        | 13                 | - 0.2°                   |
| GIBBON           | 19                 | 0.1°                     |
| BABOON           | *                  | 0.6°                     |
| MARMOSET         | 27.5               | 1.1°                     |

Table 4. No correlation between the time of divergence from the human line of descent and the amplitude of the deflection in DNA induced by SRY.

Time of divergence from human (expressed in Myr) as reported by Nei (1987). However, the exact times are subject to some controversy and it is still undecided for baboon (\*).

Difference of deflection means the difference of angles induced in DNA by SRYs of monkeys compared to the human one (angle of monkey minus angle of human).

protein (sex reversal) in humans. In fact, as shown before, mutations M64I, I90M, G95R, K106I in human SRY reduce the DNA binding affinity to the AACAAAG site; mutation M64I also reduces significantly the angle induced in the DNA (see discussion).



# DISCUSSION

### **3.1 HOW DOES SRY WORK?**

The product of the sex-determining *SRY* gene is a DNA binding protein that controls the differentiation of pre-Sertoli cells during embryonic life. The only conserved part of the SRY protein in related animals is the DNA binding domain, which recognises a defined repertoire of sites. However, the sequence selectivity of SRY is rather low and indistinguishable from that of other proteins with similar DNA-binding domains but very different biological functions. Both we and others have shown that SRY induces a large alteration of the local geometry of DNA sites it binds to, and that indeed DNA molecules that already contain sharp bends or kinks are actually bound by SRY with high affinity and selectivity, irrespective of the sequences that they contain (Ferrari et al., 1992; Giese et al., 1992). Such unusual properties suggest that the DNA bending function might be central to the biological activity of SRY, as well as of other protein with similar HMG box DNA binding domains (Grosschedl et al., 1994; Bianchi, 1994). As illustrated in Figure 17, SRY would promote the interaction of flanking transcription factors by acting mechanically on the DNA molecule. Two predictions are central to the model: SRY does not promote transcription on its own, but depends on the presence of "left" and "right" transcription factors, which are only active if physically associated. In fact, known SRYs lack recognisable transcription activation domains, and their sequences are completely divergent outside the HMG box. In contrast, the existence of the predicted "left" and "right" factors rests on the analogy with the LEF-1/TCRa system (Giese et al., 1992; Ho and Leiden 1990; Giese and Grosschedl, 1993).

The left and right transcription factors would interact only in the presence of SRY, but not with any other protein. This is possible only if SRY organises the



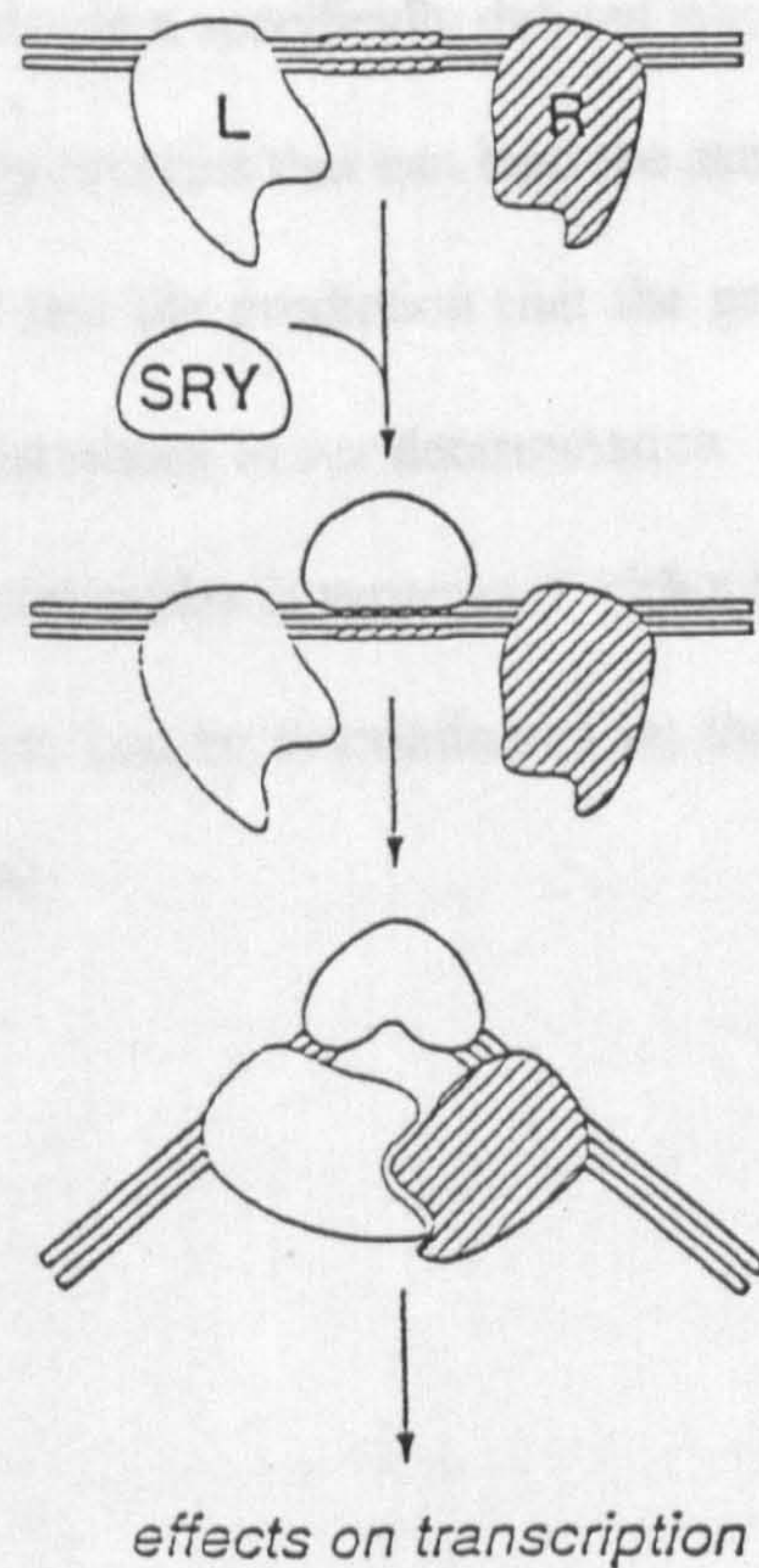


Fig. 17. A model for the action of SRY protein. The human testis-determining factor SRY is a DNA binding protein belonging to the HMG box class. SRY is most related to a number of gene-specific transcriptional regulators, which recognise specific sequences in DNA. However, the recognition takes place mainly through the minor groove and has a low selectivity. In addition, SRY proteins even from closely related mammals share similarity only within the HMG box DNA binding domain, and diverge considerably outside of it. This argues against the presence of additional domains involved in transcriptional transactivation. However, SRY bends the DNA considerably after binding to some sites, and this may promote protein-protein interactions between factors (L for "left" and R for "right") bound to either side of the SRY binding site. Thus, a mechanical action on DNA may induce the formation of a nucleoprotein complex endowed with high specificity (requiring at least three different sites bound by different proteins), high stability (the protein-protein and protein-DNA interactions would be cooperative and reinforce each other) and biological activity (through domains present on the L and R proteins). The experiments reported in this paper support the notion that the deformation induced by SRY must be geometrically precise to serve as a genetic switch.



nucleoprotein complex in a specifically defined way, which can neither be attained spontaneously nor by proteins that can bind the same sites as SRY.

The data we report test the prediction that the geometry of the SRY/DNA complex must be a prime determinant in sex determination. A mutation that alters the geometry of the nucleoprotein complex is associated with a loss of function of SRY. Moreover, different binding sites can be discriminated on the basis of the geometry they adopt when bound by SRY.



### **3.2 DISSECTION OF THE DNA BINDING AND DNA BENDING ACTIVITIES OF SRY PROTEIN**

The most obvious result of our analysis is that the DNA binding and DNA bending activities of SRY protein can be separated. The K106I amino acid substitution in SRY reduces the DNA binding affinity of the protein more than 100-fold, and yet still allows the bending of the DNA target to the same extent attained by the wild type protein. Conversely, the M64I amino acid substitution reduces the DNA binding affinity only 3-fold, but causes a reduction in DNA bending from about 75° to about 56°.

The non-correlation between the binding affinity of the SRY variants and the extent of deformation caused in DNA is at first sight surprising. How can a weakly binding protein like SRY K106I cause a large deformation in DNA? The simple answer is that mutant K106I binds weakly precisely because it still deforms DNA significantly. The binding affinity of a protein to its ligand is related to the difference in free energy at equilibrium between the bound and unbound states of the system that comprises the protein, the ligand and the aqueous environment. The free energy variation between the bound and unbound states of the SRY/DNA system can be decomposed (somewhat arbitrarily) in several contributions, as shown in the diagram in figure 18. One contribution corresponds to the establishment of chemical hydrogen bonds between SRY and the nucleic acid, irrespective of the sequence or the conformation of the DNA, in a non-specific mode. A second contribution corresponds to the formation of sequence-specific contacts between SRY and a favored binding site. The third contribution, the deformation of the DNA site, is negative: DNA bending takes

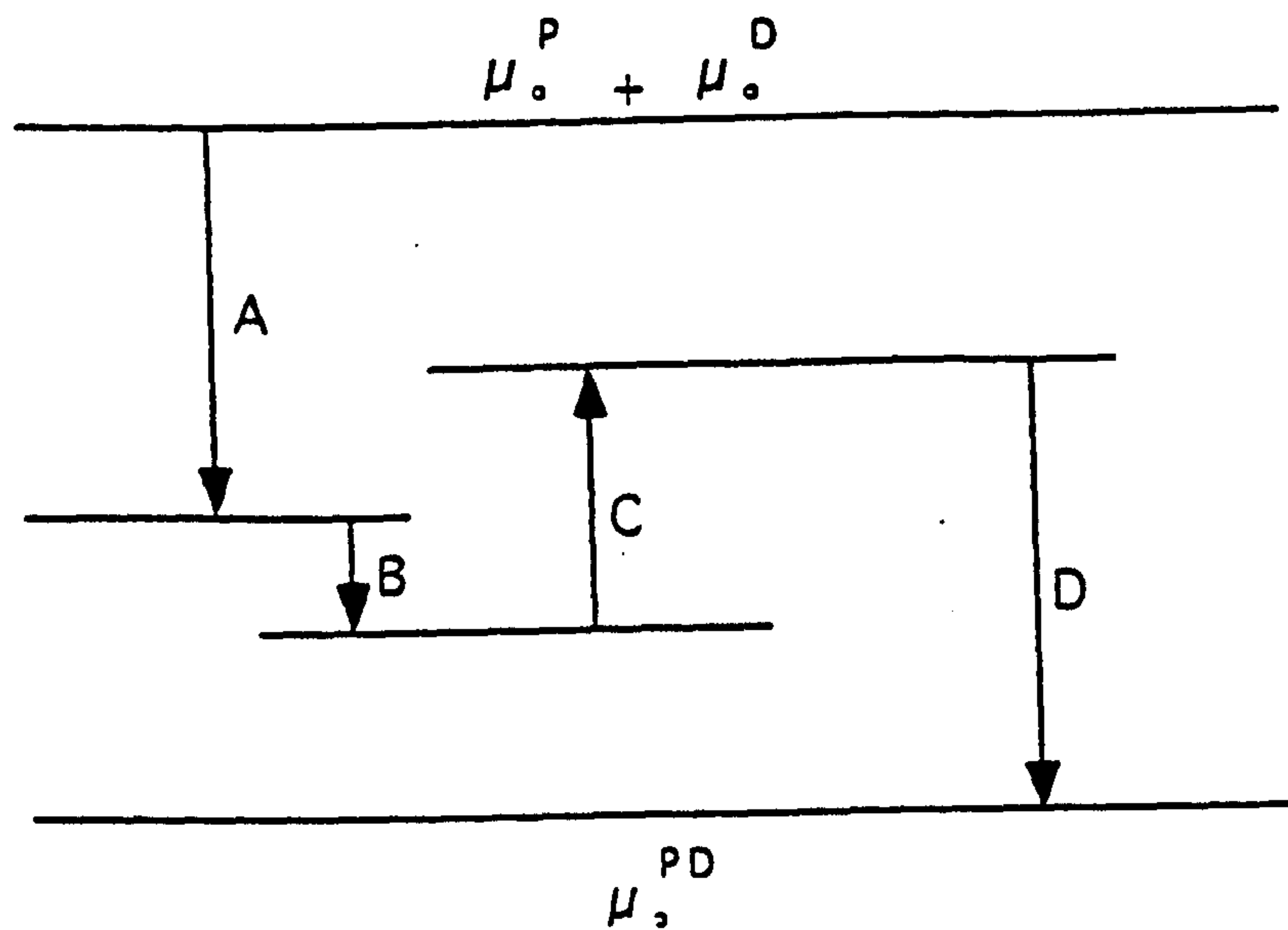


Fig. 18 Thermodynamics of the binding of SRY boxes to DNA. The chemical potential of SRY protein and of a linear DNA molecule containing a preferred target site are indicated as  $\mu_0^P$  and  $\mu_0^D$ , respectively; the chemical potential of the complex as  $\mu_0^{PD}$ . The difference in the energy levels  $\mu_0^P + \mu_0^D$  and  $\mu_0^{PD}$  represents the  $\Delta G^0$  variation in free energy associated with complex formation. The total free energy variation can be divided in several contributions: A, the free energy variation associated with non-specific protein-DNA interactions; B, that associated with sequence-specific interactions; C, the one associated with the deformation of the DNA; D, the one associated with conformation-specific interactions. The formation of the complex is thermodynamically favored so long as  $A+B+D$  is larger in absolute value than C.



up energy. A last contribution represents shape-specific interactions between the protein and the DNA, once this is deformed.

If the mutant K106I deforms the DNA in the same way as the wild type, the negative contribution will be the same. Then, if the mutation reduces any one of the other contributions, the sum of the free energy variations of all the contributions will be close to zero, and the protein will bind weakly to its ligand.

A second key result of our analysis is that the affinity of the various mutated SRY box domains for specific sites on linear DNA is variable, but the affinity for the four-way junction DNA is identical. This implies that the mechanistic details of duplex binding and junction binding are different, at least partially. We suggest that the SRY mutations we have analysed do not affect the shape-recognising properties of the HMG box, but one of the steps by which the linear DNA molecule is moulded into a specific shape. On the biological level, the equal recognition of four-way junctions by wild type and mutant SRY proteins formally rules out distorted structures as the physiological target of SRY, and suggests that sequence recognition is essential although probably not sufficient for the action of SRY in differentiation. The model in Figure 19 summarises our current thinking on SRY binding to DNA. We suggest that the DNA binding domain of SRY contains sites that allow sequence discrimination in linear DNA, sites that deflect the DNA, and sites that establish conformation-specific contacts with bent DNA. The correspondence between sites and specific amino acid residues need not be unequivocal: each site will be made up of several residues forming a surface, and a certain residue might contribute to more than one site at the same time. All three sites must be involved in the establishment of protein-DNA contacts in complexes containing the SRY box and the distorted double helix, and mutations affecting any of these sites have the potential to alter the affinity of the protein for a



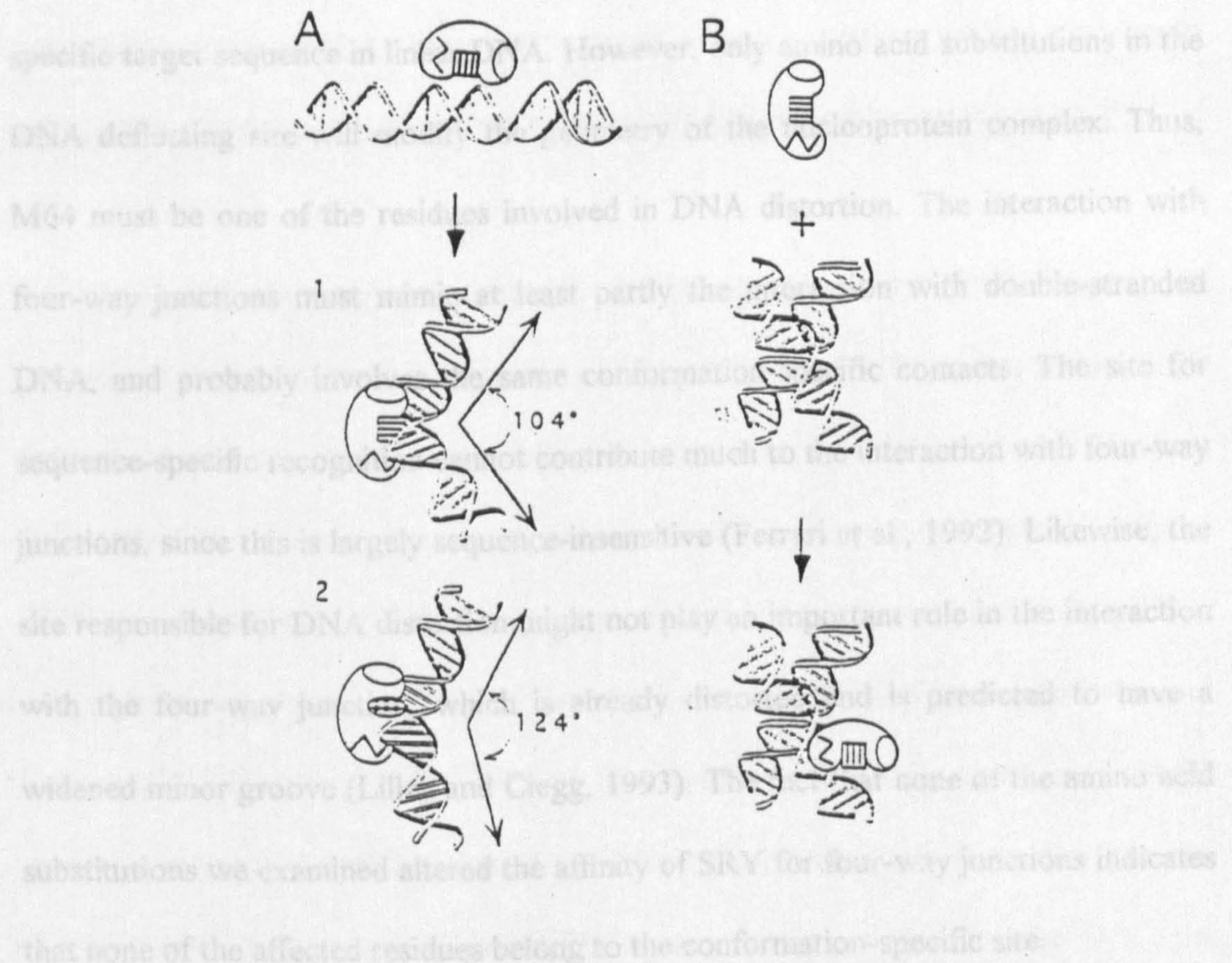


Fig.19. A schematic model for the interaction of SRY proteins with linear and four-way junction DNAs. The DNA-binding domain of wild type SRY is depicted here as a rounded object with three sites which are important for the interaction with DNA: the grey disc is involved in sequence-specific recognition, the zig-zag is associated with conformation-specific interactions, and the square with bars represents the site in the protein responsible for the deflection of linear DNA. In reality, each site may be composed of multiple residues; conversely, some residues may contribute to more than one of the sites that are shown here as independent. (A) Interaction with linear DNA. When wild type SRY interacts with a specific sequence in linear DNA (1), all three sites will establish contacts with the DNA in the complex, and the DNA deflection will be mainly determined by the burrowing of the site represented as the square with bars into the minor groove. The substitution of a residue in this site (chequered disc) will modify marginally the overall affinity of the protein for the DNA, but will have a profound effect on the overall geometry of the complex (2). (B) Interaction with four-way junctions. The affinity of SRY for four-way junctions does not depend on the sequence of the DNA. Moreover, the DNA already contains kinks. Therefore, we expect that the sequence-specific site and the DNA-deflecting site will not play important roles in this type of interaction, which will be mainly governed by the conformation-specific site. To stress this point, in the drawing the grey disc and the square with bars are drawn as not making contacts with the four-way junction, although of course this may not reflect the real situation.



specific target sequence in linear DNA. However, only amino acid substitutions in the DNA deflecting site will modify the geometry of the nucleoprotein complex. Thus, M64 must be one of the residues involved in DNA distortion. The interaction with four-way junctions must mimic at least partly the interaction with double-stranded DNA, and probably involves the same conformation-specific contacts. The site for sequence-specific recognition cannot contribute much to the interaction with four-way junctions, since this is largely sequence-insensitive (Ferrari et al., 1992). Likewise, the site responsible for DNA distortion might not play an important role in the interaction with the four-way junction, which is already distorted and is predicted to have a widened minor groove (Lilley and Clegg, 1993). The fact that none of the amino acid substitutions we examined altered the affinity of SRY for four-way junctions indicates that none of the affected residues belong to the conformation-specific site.

### **3.3 THE CONTRIBUTION OF THE DNA SEQUENCE TO THE ORGANISATION OF SRY-INDUCED NUCLEOPROTEIN COMPLEXES.**

As discussed in the previous paragraph, the biological activity of SRY appears to depend critically on the recognition of specific sequences. We have shown that target sites with altered sequences can still be recognised by SRY, but will be bent differently. This offers a solution to the puzzle of the limited sequence-specificity of SRY, as opposed to the required sequence-specificity of its action. A certain DNA sequence may turn out to be a key element in the regulation of gene expression not because SRY binds especially well to it, but because the SRY-induced angle is the correct one for the establishment of a productive complex involving other proteins (see figure 17). In fact, different HMG box proteins may recognise the same sequence, but they will produce different angles: while human SRY induces an angle of about  $80^{\circ}$ , LEF-1 produces an angle of about  $130^{\circ}$  (Giese et al., 1992).

At a mechanistic level, the intrinsic bendability of specific sequences may ultimately dictate whether SRY will bind to them or not, as suggested by King and Weiss (1993), and how much the DNA will be bent in the complex. Studies on the adaptation of double-helical DNA to the curved path around nucleosomes have indicated that different sequences have indeed different flexibility (reviewed by Travers, 1988). In any event, it is clear that the issue of bendability and the issue of angle-choice taken together, amount to the likely existence of yet another informational level in DNA readout: one that ultimately depends on the sequence, but in quite indirect and complex ways.



### **3.4 CORRELATION BETWEEN THE DNA BINDING PROPERTIES OF THE SRY BOX MUTANTS AND THE PHENOTYPE OF THE AFFECTED PATIENTS**

In this study, we have adopted the classical genetic approach of understanding a phenomenon by considering mutations that perturb it. To what extent, however, is there a causal relationship between the chemical-physical properties of the DNA binding domains of mutant SRY proteins and the lack of testes formation? A DNA-binding protein must bind to DNA if it is to work, and in this respect the biological consequences of mutations G95R and K106I correspond well to the reduced binding affinity of the corresponding protein, as had already been noted in a previous study (Harley et al., 1992). At the other extreme, mutations F109S and I90M can be transmitted from father to son and are indeed present in a number of physiologically differentiated males: they clearly behave as conditional mutations. From this point of view, it is quite logical that the mutated proteins differ little from wild type SRY with respect to their DNA binding and DNA bending properties. Their failure to function in some cases, but not all, must depend on the genetic background or the environment. It has already been shown that the genetic background is important in sex determination: in the mouse when the variant poschiavinus Y chromosome is present on a C57BL6 background, XY females result, but the same chromosome gives males in all other backgrounds.

Mutant M64I is at the same time more informative, and more difficult to interpret. This mutation is *de novo*, which suggests that it is always associated with sex-reversal. The 3-fold reduction in DNA binding affinity with respect to the wild type is unimpressive, and it is unlikely that a reduction of this extent would completely

subvert its function. On the other hand, a reduction in the deformation of the DNA translates into a significant dislocation of proteins bound a few helical turns from the SRY binding site. We cannot measure the exact difference in the architecture of the SRY-DNA complex responsible for sex determination for two reasons: circular permutation is a low-resolution assay and no bona fide target site for SRY has been identified. The first limitation is not very relevant, since in any event the differences in DNA distortion are large enough to be identified qualitatively. The second limitation is more critical, but it is clear that M64I always distorts the DNA differently from the wild type protein, whatever the sequence of the target site.

There is another limit in our understanding of the role of DNA distortion in SRY action: we do not know how tolerant to perturbations is the system. To obtain a rough idea of the degree of wobble allowed in SRY-induced nucleoprotein complexes, we compared the deflection angles induced by human SRY and its primate and mouse counterparts. Primate SRY proteins (gorilla, gibbon, baboon, chimpanzee, pygmy chimpanzee, orang-utan, and marmoset) all bend the AACAAAG site in the CD3e enhancer in a very similar way to human SRY (the maximum difference is 6 degrees), despite the accumulation of several amino acid substitutions in the DNA binding domain. Such evolutionary conservation suggests a very limited permissible wobble. On the other hand, the angle produced by mouse SRY appears to be quite different from that produced by human SRY (Giese et al., 1994), and this correlates with the inability of transgenic human *SRY* to cause male differentiation in mice (Koopman et al., 1991). The angle variation may be the key molecular defect of mutant M64I, although this deduction is not supported by direct experiments on transcriptional control. However, we have excluded the possibility that the M64I SRY mutant causes gonadal dysgenesis because its sequence specificity is different from



that of the wild type protein: both proteins select a comparable repertoire of binding sequences in selection/amplification experiments.

In conclusion, we have shown that the DNA binding and DNA bending activities of SRY protein can be dissociated by mutation. The DNA bending activity of SRY is also strongly affected by the underlying sequence of the binding site. We have also shown that the defect of one mutant SRY protein might be traced to its anomalous DNA bending activity. Our findings are in keeping with the hypothesis of a direct role of protein-induced DNA bending in certain types of transcriptional control and should stimulate increased efforts in the study of DNA structural plasticity and in the search of proteins that modulate such plasticity.

### **3.5 MOLECULAR BASIS OF HUMAN 46X, Y SEX REVERSAL REVEALED FROM THE THREE-DIMENSIONAL SOLUTION STRUCTURE OF THE HUMAN SRY-DNA COMPLEX**

While this work was in progress, the solution structure of the specific complex between the human SRY box and its hypothetical DNA target site was determined by multidimensional NMR spectroscopy. The SRY box has a twisted L shape that presents a concave surface (made up of three helices and the N- and C- terminal strands) to the DNA for sequence-specific recognition (Werner et al., 1995). Binding of SRY to its DNA target site occurs exclusively in the minor groove and induces a large conformational change in the DNA. The DNA in the complex has an overall 70°-80° bend and is helically unwound relative to classical A and B DNA (figure 18).

The structure of the complex reveals the effects of point mutations that cause sex reversal phenotype at the atomic level.

From the genetic stand point, naturally occurring point mutations in hSRY are of two types: inherited mutations with variable penetrance and de novo mutations with full penetrance (see introduction and table 1).

Clearly, the functional effects of the former must be considerably less severe than those of the latter. From a structural point of view, the point mutations also fall into two categories: those that affect the packing of residues within the protein core and those that involve residues that directly contact the DNA.

Two inherited point mutation, I90M (I 35 in Figure 21) and F109S (F 54 in Figure 21) result in packing defects that would be expected to destabilise the protein, but they are not so crucial for the correct packing as V60 (V5 in Figure 21) that serves to position K 73 and Y 74 for interaction with the DNA.



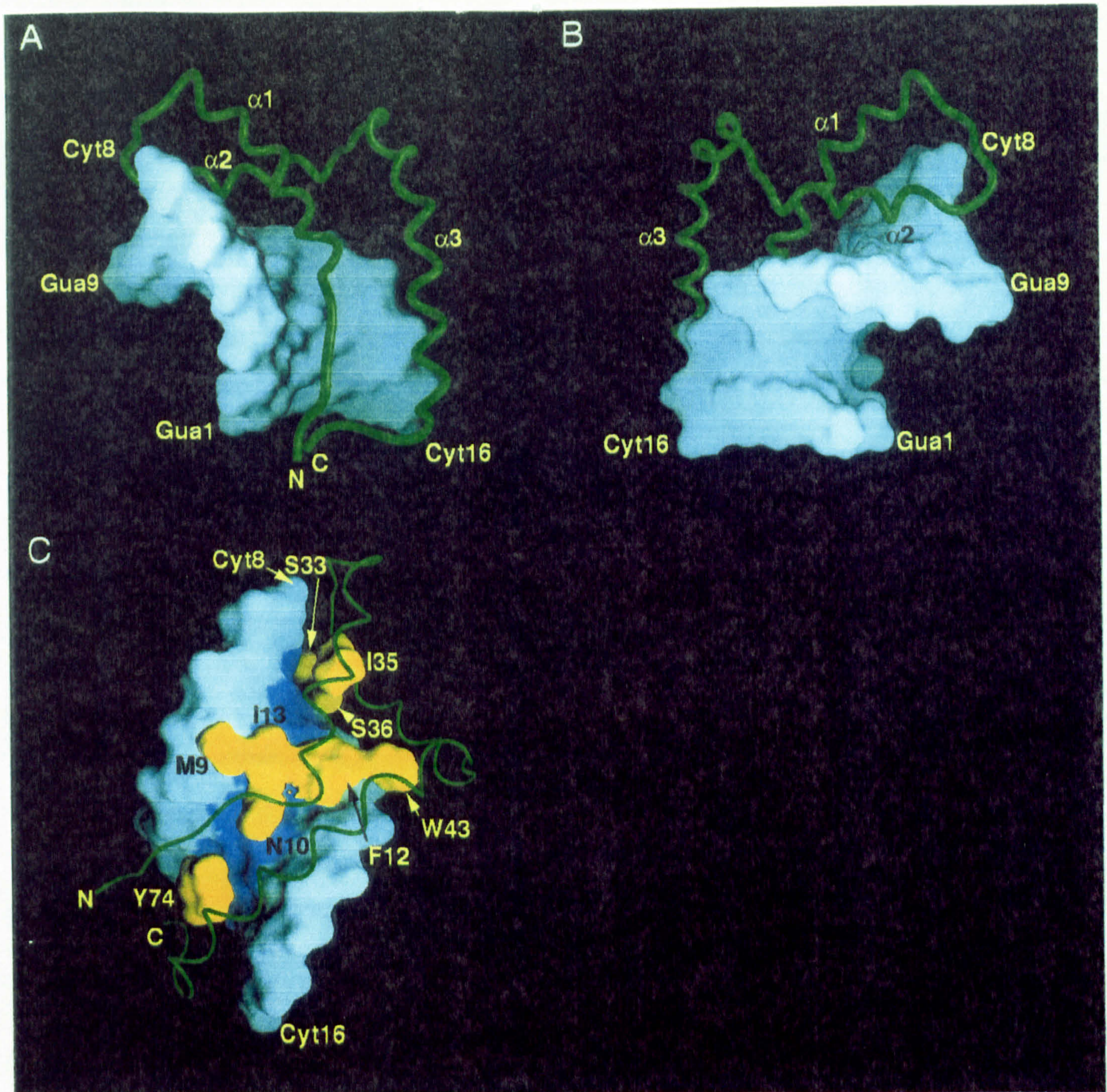


Fig. 20. The interaction of hSRYbox with DNA (from Werner et al.)

Three views (A-C) are displayed. The protein is shown as a schematic ribbon drawing in green, and the colour coding used for the DNA bases is red for A, lilac for T, dark blue for G and light blue for C. Side chains that contact the DNA bases are depicted in yellow in (C). (D) shows the same view as in (C) with the molecular surface of the protein shown in grey and the DNA atoms in yellow. The patches of blue on the protein surface indicate the location of the side chains of four of the seven residues that interact with the DNA bases.



F 109 (F 54 of Figure 21) is a buried residue within the hydrophobic core that packs against A 11, V 14, W 15 and F 55 as shown in the model reported in the Figure 21. Substitution by the polar S at this position would be anticipated to destabilise the protein, but this effect, however, may not be sufficient to perturb the DNA binding properties of the F109S mutant, but may accelerate its degradation within the cell.

The other packing defect point mutant, K106I (K 51 in Figure 21), plays a role in the interaction with the DNA binding core and the rest of the protein. The aliphatic portion of the side chain of K51 in Figure 21 is packed between L 46 and F 55. Introduction of a branched Ile side chain with two bulky methyl groups at this position may disrupt F 55 from the hydrophobic core and partially displace L 46, destabilising the packing of helices 2 and 3.

The remaining de novo point mutations involve residues that contact the DNA. The G95R (Arg 7 in Figure 22) mutant removes hydrophobic contacts with the sugar of C4 and a salt bridge to the phosphate of A5. Owing to shortening of the side chain, the M64I (Met 9 Ile of Figure 22) mutant will disrupt extensive van der Waals contacts with the deoxyribose of A5 and A6 and with the O3 atom of A6 that is located at the hinge point of the bend between base pair 5 and 6 (Figure 22).



Fig. 21. Ribbon drawing of hSRYbox.

The backbone ribbon is shown in green, and side chains with less than 20% and between 20% and 50% of their surface (relative to an extended Gly-X-Gly tripeptide segment) accessible to solvent are shown in red and blue, respectively.



MATERIALS AND METHODS

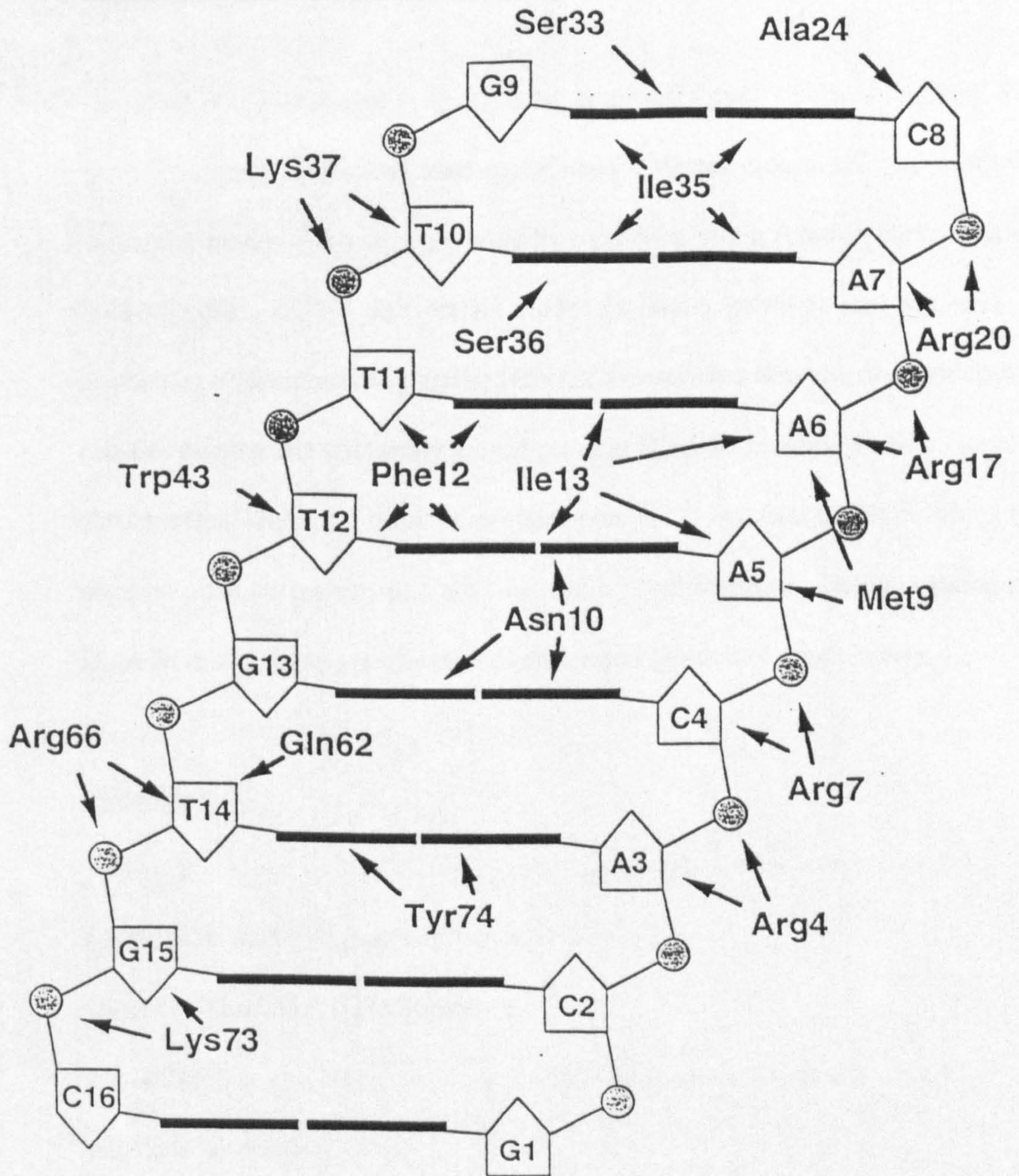


Fig. 22. Summary of the contacts between hSRYbox and DNA.

the DNA is represented as a cylindrical projection viewed in the minor groove with the bases depicted as thick black lines, the deoxyribose sugar rings as pentagons, and the phosphate as stippled circles.



# **MATERIALS AND METHODS**

# **MATERIALS AND METHODS**



## **4.1 ALKALINE LYSIS MINIPREP**

The alkaline lysis procedure is the most commonly used miniprep method. Plasmid DNA is prepared from small amounts of many different cultures (1 to 24) of plasmid-containing bacteria. Bacteria are lysed by treatment with a solution containing sodium dodecyl sulfate (SDS) and NaOH (SDS denatures bacterial proteins, and NaOH denatures chromosomal and plasmid DNA). The mixture is neutralized with potassium acetate, causing the covalently closed plasmid DNA to reanneal rapidly. Most of the chromosomal DNA and bacterial proteins precipitate--as does the SDS, which forms a complex with potassium--and are removed by centrifugation. The reannealed plasmid DNA from the supernatant is then concentrated by ethanol precipitation.

### **Materials**

LB medium containing appropriate antibiotic.

Glucose/Tris/EDTA (GTE) solution

TE buffer .

NaOH/SDS solution

Potassium acetate solution

95% and 70% ethanol

10 mg/ml DNase-free RNase.

1.5-ml disposable microcentrifuge tubes

### **Methods**

**1. Inoculate 5 ml sterile LB medium with a single bacterial colony. Grow to saturation (overnight).**

**2. Spin 1.5 ml of cells 20 sec in a microcentrifuge at maximum speed to pellet. Remove the supernatant with a Pasteur pipet.**

**The spins in steps 2 and 6 can be performed at 4C or at room temperature. Longer spins make it difficult to resuspend cells.**

**3. Resuspend pellet in 100 ul GTE solution and let sit 5 min at room temperature.**

**Be sure cells are completely resuspended.**

**4. Add 200 ul NaOH/SDS solution, mix by tapping tube with finger, and place on ice for 5 min.**

**5. Add 150 ul potassium acetate solution and vortex at maximum speed for 2 sec to mix. Place on ice for 5 min.**

**Be sure mixing is complete.**

**6. Spin 3 min as in step 2 to pellet cell debris and chromosomal DNA.**

**7. Transfer supernatant to a fresh tube, mix it with 0.8 ml of 95% ethanol, and let sit 2 min at room temperature to precipitate nucleic acids.**



8. Spin 1 min at room temperature to pellet plasmid DNA and RNA.

9. Remove supernatant, wash the pellet with 1 ml of 70% ethanol, and dry pellet under vacuum.

10. Resuspend the pellet in 30 ul TE buffer and store as in support protocol. Use 2.5 to 5 ul of the resuspended DNA for a restriction digestion.

## **4.2 PREPARATION OF CRUDE LYSATE BY ALKALINE LYSIS: MAXI PREP**

### **Materials**

LB medium or enriched medium (e.g., superbrot h or terrific broth) containing ampicillin or other appropriate selective agent .

Plasmid-bearing E. coli strain.

Glucose/Tris/EDTA solution.

25 mg/ml hen egg white lysozyme in glucose/Tris/EDTA solution (prepare fresh).

0.2 M NaOH/1% (w/v) SDS [prepare fresh from 10 M NaOH and 10% (w/v) SDS stocks]

3 M potassium acetate solution, pH ~5.5

Isopropanol

70% ethanol

Sorvall GSA, GS-3, or Beckman JA-10 rotor or equivalent

High-speed centrifuge tubes with  $\geq 20$ -ml capacity (e.g., Oak Ridge centrifuge tubes)

Sorvall SS-34 or Beckman JA-17 rotor or equivalent

### Methods

1. Inoculate 5 ml LB medium or enriched medium containing selective agent (most commonly ampicillin) with a single colony of *E. coli* containing the desired plasmid. Grow at 37C with vigorous shaking overnight.

2. Inoculate 500 ml LB medium or enriched medium containing selective agent in a 2-liter flask with  $\sim 1$  ml of overnight culture. Grow at 37C until culture is saturated (OD<sub>600</sub>  $\approx 4$ ).

To increase yields, maximize aeration using a flask with high surface area (whose volume exceeds the culture volume--i.e., is  $>2$  liters) and baffles and shake at  $>400$  rpm. Alternatively, treat cultures of cells growing logarithmically with chloramphenicol to amplify the plasmids. Growing the bacteria in medium that supports higher cell densities also increases the yield. These media include M9, terrific broth, and LB medium containing 0.1% (w/v) glucose. These media can increase plasmid yields 2- to 10-fold; different plasmids respond to the media differently. Most plasmids commonly used today, particularly derivatives of the pUC series, grow at a copy number high enough to routinely yield 1 to 5 mg plasmid DNA from a 500-ml culture grown in LB medium.



An important consideration when using enriched medium is the method to be used for final purification of plasmid DNA. Increased yield poses no problems when using CsCl/ethidium bromide or PEG purification. However, the capacity of some commercially available chromatography columns--e.g., the Qiagen-tip 2500 (Qiagen) and Wizard Maxiprep (Promega)--is easily exceeded. Therefore, the increased yield of plasmid DNA in the crude lysate will not result in increased recovery from the column. The pZ523 column (5 Prime->3 Prime) does not require that plasmid DNA bind to the column and can be used to purify larger amounts of DNA.

3. Collect cells by centrifuging 10 min at 6000 x g (~6000 rpm in Sorvall GSA/GS-3 or Beckman JA-10 rotors), 4C.

If necessary the pellets can be stored frozen indefinitely at -20 or -70C.

4. Resuspend pellet from 500-ml culture in 4 ml glucose/Tris/EDTA solution and transfer to high-speed centrifuge tube with  $\geq$ 20-ml capacity.

**Lyse the cells**

5. Add 1 ml of 25 mg/ml hen egg white lysozyme in glucose/Tris/EDTA solution. Resuspend the pellet completely in this solution and allow it to stand 10 min at room temperature.

Neither glucose nor lysozyme is absolutely necessary for the success of the procedure. Glucose serves as a buffer in step 6 when the pH of the solution is greatly increased

by addition of NaOH. Glucose provides buffering in the range of pH 12 and, by preventing the pH from rising too drastically in step 6, increases the efficiency of precipitation in step 7 (when the pH is lowered by addition of potassium acetate).

Lysozyme assists in the destruction of bacterial cell walls and subsequent release of plasmid DNA. Bacterial debris and soluble proteins are precipitated in step 7. One problem that can reduce recovery of plasmid DNA is inefficient separation of plasmid DNA from cellular debris. Lysozyme helps increase yield by reducing the amount of plasmid DNA trapped in partially degraded cell material and subsequently lost by precipitation at step 7.

The effort and expense required to include glucose and lysozyme in step 5 is negligible. The efficiency gained in streamlining the procedure by omitting them is also negligible. However, the potential for loss of plasmid DNA when these components are not included is measurable and worth avoiding. It should be noted that some commercially available chromatographic systems (e.g., Qiagen) rely on inefficient bacterial lysis to reduce contamination of plasmid DNA with chromosomal DNA. Although omitting lysozyme reduces the recovery of plasmid DNA, when using these products the manufacturer's recommendations should be followed.

When chromatographic methods are used for final purification of plasmid DNA, it is essential to degrade RNA that contaminates the lysate, and will copurify with plasmid DNA. Treating the lysate with RNase A is the most efficient and economical method for degrading RNA. This can be accomplished at any step in the preparation of crude



lysate, it is most convenient to do it at step 5, by adding RNase A to the glucose/Tris/EDTA solution to a final concentration of 50 ug/ml.

6. Add 10 ml freshly prepared 0.2 M NaOH/1% SDS and mix by stirring gently with a pipet until solution becomes homogeneous and clears. Let stand 10 min on ice.

The solution should become very viscous.

7. Add 7.5 ml of 3 M potassium acetate solution and again stir gently with a pipet until viscosity is reduced and a large precipitate forms. Let stand 10 min on ice.

8. Centrifuge 10 min at 20,000 x g (13,000 rpm in Sorvall SS-34; 12,500 rpm in Beckman JA-17), 4C.

A large, fairly compact pellet will form; this contains most of the chromosomal DNA, SDS-protein complexes, and other cellular debris. Plasmid DNA remains in the translucent supernatant.

Addition of ~0.5 ml chloroform before the centrifugation can help reduce floating material.

**Precipitate plasmid DNA**

9. Decant the supernatant into a clean centrifuge tube. Pour it through several layers of cheesecloth if any floating material is visible. Add 0.6 vol isopropanol, mix by inversion, and let stand 5 to 10 min at room temperature.

If the supernatant is cloudy or contains floating material, repeat centrifugation (step 8) before adding isopropanol.

10. Recover nucleic acids by centrifuging 10 min at 15,000 x g (11,500 rpm in SS-34 rotor; 10,500 rpm in JA-17 rotor), room temperature.

11. Wash the pellet with 2 ml of 70% ethanol; centrifuge briefly at 15,000 x g, room temperature, to collect pellet. Aspirate ethanol and dry pellet under vacuum.

The pellet can be stored indefinitely at 4C.

#### Materials

Potassium acetate solution (3 M), pH ~5.5

294 g potassium acetate (3 M final)

50 ml 90% formic acid (1.18 M final)

H<sub>2</sub>O to 1 liter

Store indefinitely at room temperature

#### Sucrose/Tris/EDTA solution

25% (w/v) sucrose

50 mM Tris-Cl, pH 8.0

100 mM EDTA, pH 8.0



Store indefinitely at 4C

### **4.3 TRANSFORMATION USING CALCIUM CHLORIDE**

Escherichia coli cells are grown to log phase. Cells are concentrated by centrifugation and resuspended in a solution containing calcium chloride. Exposure to calcium ions renders the cells able to take up DNA, or competent. Plasmid DNA is mixed with the cells and presumably adheres to them. The mixture of DNA and cells is then heat shocked, which allows the DNA to efficiently enter the cells. The cells are grown in nonselective medium to allow synthesis of plasmid-encoded antibiotic resistance proteins, then plated on antibiotic-containing medium to allow identification of plasmid-containing colonies.

#### **Materials**

Single colony of E. coli cells

LB medium

CaCl<sub>2</sub> solution, ice-cold

LB plates containing ampicillin

Plasmid DNA (pBend 2, pT7 hSRY box...)

Chilled 50-ml polypropylene tubes

Beckman JS-5.2 rotor or equivalent

42C water bath

All materials and reagents coming into contact with bacteria must be sterile.

#### **4.4 PREPARE COMPETENT CELLS**

##### **Method**

1. Inoculate a single colony of *E. coli* (DH5, B121 (DE3)) cells into 50 ml LB medium. Grow overnight at 37C with moderate shaking (250 rpm)

Alternatively, grow a 5-ml culture overnight in a test tube on a roller drum.

2. Inoculate 4 ml of the culture into 400 ml LB medium in a sterile 2-liter flask. Grow at 37C, shaking (250 rpm), to an OD<sub>590</sub> of 0.375.

This procedure requires that cells be growing rapidly (early- or mid-log phase). Accordingly, it is very important that the growing cells have sufficient air. A 1-liter baffle flask can be used instead of the 2-liter flask. Overgrowth of culture (beyond OD<sub>590</sub> of 0.4) decreases the efficiency of transformation.

3. Aliquot culture into eight 50-ml prechilled, sterile polypropylene tubes and leave the tubes on ice 5 to 10 min.

Cells should be kept cold for all subsequent steps.

Larger tubes or bottles can be used to centrifuge cells if volumes of all subsequent solutions are increased in direct proportion.

4. Centrifuge cells 7 min at 3000 rpm (1600 x g), 4C. Allow centrifuge to decelerate without brake.



We have not attempted to determine whether deceleration without braking is critical to this procedure. However, we do not use the brake for this step or for the subsequent centrifugation steps.

5. Pour off supernatant and resuspend each pellet in 10 ml ice-cold CaCl<sub>2</sub> solution.

Resuspension should be performed very gently and all cells kept on ice.

6. Centrifuge cells 5 min at 2500 rpm (1100 x g), 4C. Discard supernatant and resuspend each pellet in 10 ml cold CaCl<sub>2</sub> solution. Keep resuspended cells on ice for 30 min.

7. Centrifuge cells 5 min at 2500 rpm (1100 x g), 4C. Discard supernatant and resuspend each pellet in 2 ml of ice-cold CaCl<sub>2</sub> solution.

It is important to resuspend this final pellet well. The suspension may be left on ice for several days. For many strains (e.g., DH5) competency increases with increasing time on ice, and reaches a maximum at 12 to 24 hr. This is not true for B121 (DE3) cells, which should be frozen immediately.

8. Dispense cells into prechilled, sterile polypropylene tubes (250- $\mu$ l aliquots are convenient). Freeze immediately at -70C.

Assess competency of cells

9. Use 10 ng of pBR322 to transform 100 ul of competent cells according to the steps provided below. Plate appropriate aliquots (1, 10, and 25 ul) of the transformation culture on LB/ampicillin plates and incubate at 37C overnight.

10. The number of transformant colonies per aliquot volume (ul) x 10<sup>5</sup> is equal to the number of transformants per microgram of DNA.

Transformation efficiencies of 10<sup>7</sup>-10<sup>8</sup> and 10<sup>6</sup>-10<sup>7</sup> are obtained for E. coli MC1061 and DH5, respectively. Competency of strains decreases very slowly over months of storage time.

#### Transform competent cells

11. Aliquot 10 ng of DNA in a volume of 10 to 25 ul into a 15-ml sterile, round-bottom test tube and place on ice.

Plasmid DNA can be used directly from ligation reactions. When this is done, more DNA is usually used. However, if there is more than 1 ug of DNA in the ligation reaction, or if the ligation reaction is from low gelling/melting temperature agarose, it is wise to dilute the ligation mix

12. Rapidly thaw competent cells by warming between hands and dispense 100 ul immediately into test tubes containing DNA. Gently swirl tubes to mix, then place on ice for 10 min.



Competent cells should be used immediately after thawing. Remaining cells should be discarded rather than refrozen.

13. Heat shock cells by placing tubes into a 42C water bath for 2 min.

Alternatively, incubate at 37C for 5 min.

14. Add 1 ml LB medium to each tube. Place each tube on a roller drum at 250 rpm for 1 hr at 37C.

15. Plate aliquots of transformation culture on LB/ampicillin or other appropriate antibiotic-containing plates.

It is convenient to plate several different dilutions of each transformation culture. The remainder of the mixture can be stored at 4C for subsequent platings.

16. When plates are dry, incubate 12 to 16 hr at 37C.

### Materials

SOC medium

0.5% yeast extract

2% tryptone

10 mM NaCl

2.5 mM KCl

10 mM MgCl<sub>2</sub>

10 mM MgSO<sub>4</sub>

20 mM glucose

#### **4.5 AGAROSE GEL ELECTROPHORESIS**

Agarose gel electrophoresis is a simple and highly effective method for separating, identifying, and purifying 0.5- to 25-kb DNA fragments. The protocol can be divided into three stages: (1) a gel is prepared with an agarose concentration appropriate for the size of DNA fragments to be separated; (2) the DNA samples are loaded into the sample wells and the gel is run at a voltage and for a time period that will achieve optimal separation; and (3) the gel is stained or, if ethidium bromide has been incorporated into the gel and electrophoresis buffer, visualized directly upon illumination with UV light.

#### **4.6 RESOLUTION OF DNA FRAGMENTS ON STANDARD AGAROSE GELS**

##### **Materials**

Electrophoresis buffer (TAE or TBE )

Ethidium bromide solution

Electrophoresis-grade agarose

10x loading buffer

DNA molecular weight markers



Horizontal gel electrophoresis apparatus

Gel casting platform

Gel combs (slot formers)

DC power supply

### Method

#### Preparing the gel

1. Prepare an adequate volume of electrophoresis buffer (TAE or TBE ) to fill the electrophoresis tank and prepare the gel.

To facilitate visualization of DNA fragments during the run, ethidium bromide solution can be added to the electrophoresis buffer to a final concentration of 0.5 ug/ml. If buffer is prepared for the electrophoresis tank and the gel separately, be sure to bring both to an identical concentration of ethidium bromide.

2. Add the desired amount of electrophoresis-grade agarose to a volume of electrophoresis buffer sufficient for constructing the gel. Melt the agarose in a microwave oven or autoclave and swirl to ensure even mixing. Gels typically contain 0.8 to 1.5% agarose

Melted agarose should be cooled to 55C in a water bath before pouring onto the gel platform. This prevents warping of the gel apparatus. Gels are typically poured between 0.5 and 1 cm thick. The volume of the sample wells will be determined by both the thickness of the gel and the size of the gel comb.

**3. Seal the gel casting platform if it is open at the ends. Pour in the melted agarose and insert the gel comb, making sure that no bubbles are trapped underneath the combs and all bubbles on the surface of the agarose are removed before the gel sets.**

**Most gel platforms are sealed by taping the open ends with adhesive tape. As an added measure to prevent leakage, hot agarose can be applied with a Pasteur pipet to the joints and edges of the gel platform and allowed to harden.**

#### **Loading and running the gel**

**4. After the gel has hardened, remove the tape from the open ends of the gel platform and withdraw the gel comb, taking care not to tear the sample wells.**

**Most gel platforms are designed so that 0.5 to 1 mm of agarose remains between the bottom of the comb and the base of the gel platform. This is usually sufficient to ensure that the sample wells are completely sealed and to prevent tearing of the agarose upon removal of the comb. Low percentage gels and gels made from low gelling/melting temperature agarose should be cooled at 4C to gain extra rigidity and prevent tearing.**

**5. Place the gel casting platform containing the set gel in the electrophoresis tank. Add sufficient electrophoresis buffer to cover the gel to a depth of about 1 mm (or just until the tops of the wells are submerged). Make sure no air pockets are trapped within the wells.**



6. DNA samples should be prepared in a volume that will not overflow the gel wells by addition of the appropriate amount of 10x loading buffer. Samples are typically loaded into the wells with a pipettor or micropipet. Care should be taken to prevent mixing of the samples between wells.

**Be sure to include appropriate DNA molecular weight markers**

7. Be sure that the leads are attached so that the DNA will migrate into the gel toward the anode or positive lead. Set the voltage to the desired level, typically 1 to 10 V/cm of gel, to begin electrophoresis. The progress of the separation can be monitored by the migration of the dyes in the loading buffer.

8. Turn off the power supply when the Bromphenol Blue dye from the loading buffer has migrated a distance judged sufficient for separation of the DNA fragments. If ethidium bromide has been incorporated into the gel, the DNA can be visualized by placing on a UV light source and can be photographed directly.

Gels that have been run in the absence of ethidium bromide can be stained by covering the gel in a dilute solution of ethidium bromide (0.5 ug/ml in water) and gently agitating for 10 to 30 min. If necessary, gels can be destained by shaking in water for an additional 30 min. This serves to remove excess ethidium bromide which causes background fluorescence and makes visualization of small quantities of DNA difficult.

## **REAGENTS AND SOLUTIONS**

## **Agarose gel**

Gels typically contain ~1% agarose in 1x TAE or TBE . Electrophoresis-grade agarose powder is added to 1x gel buffer and melted by boiling for several minutes. Be sure all agarose particles are completely melted. To facilitate visualization of DNA fragments during the run, ethidium bromide can be added to 0.5 ug/ml in the gel.

### **Materials**

#### **Ethidium bromide solution**

1000x stock solution 0.5 ug/ml:

50 mg ethidium bromide

100 ml H<sub>2</sub>O

Working solution, 0.5 ug/ml:

Dilute stock 1:1000 for gels or stain solution

10x loading buffer

20% Ficoll 400

0.1 M Na<sub>2</sub>EDTA, pH 8

1.0% sodium dodecyl sulfate

0.25% Bromphenol blue

0.25% Xylene Cyanol (optional; runs ~50% as fast as Bromphenol blue and can interfere with visualization of bands of moderate molecular weight, but can be helpful for monitoring very long runs)



## **4.7 DIGESTING A SINGLE DNA SAMPLE WITH RESTRICTION ENDONUCLEASES**

Restriction endonuclease cleavage is accomplished simply by incubating the enzyme(s) with the DNA in appropriate reaction conditions. The amounts of enzyme and DNA, the buffer and ionic concentrations, and the temperature and duration of the reaction will vary depending upon the specific application.

### **Materials**

DNA sample in H<sub>2</sub>O or TE buffer.

10x restriction endonuclease buffers.

Restriction endonucleases.

10x loading buffer.

0.5 M EDTA, pH 8.0 .

### **Method**

1. Pipet the following into a clean microcentrifuge tube:

x ul DNA (0.1 to 4 ug DNA in H<sub>2</sub>O or TE buffer)

2 ul 10x restriction buffer.

18 - x ul H<sub>2</sub>O.

2. Add restriction endonuclease (1 to 5 U/ug DNA) and incubate the reaction mixture 1 hr at the recommended temperature (in general, 37C).

In principle, 1 U restriction endonuclease completely digests 1 ug of purified DNA in 60 min using the recommended assay conditions. However, crude DNA preparations, such as those made by rapid procedures, often require more enzyme and/or more time for complete digestion. The volume of restriction endonuclease added should be less than 1/10 the volume of the final reaction mixture, because glycerol in the enzyme storage buffer may interfere with the reaction.

3. Stop the reaction and prepare it for agarose or acrylamide gel electrophoresis by adding 5 ul (20% of reaction vol) 10x loading buffer.

The reaction can also be stopped by chelating Mg<sup>2+</sup> with 0.5 ul of 0.5 M EDTA (12.5 mM final concentration). If the digested DNA is to be used in subsequent enzymatic reactions (e.g., ligation or "filling-in" reactions), addition of EDTA should be avoided. Alternatively, many enzymes can be irreversibly inactivated by incubating 10 min at 65C. Some enzymes that are partially or completely resistant to heat inactivation at 65C may be inactivated by incubating 15 min at 75C. When the enzyme(s) is completely resistant to heat inactivation, DNA may be purified from the reaction mixture by extraction with phenol and precipitation in ethanol.

#### **4.8 DNA-DEPENDENT DNA POLYMERASES.**

##### **INTRODUCTION**



All DNA polymerases add deoxyribonucleotides to the 3'-hydroxyl terminus of a primed double-stranded DNA molecule

Synthesis is exclusively in a 5'->3' direction with respect to the synthesized strand. Each nucleotide that is incorporated during polymerization is complementary to the one opposite to it in the template (dA pairs with dT, dC with dG). The reaction requires the four deoxyribonucleoside triphosphates (dNTPs) and magnesium ions.

Many DNA polymerases have a 3'->5' exonuclease inherently associated with the polymerase activity. The 3'->5' exonuclease activity removes a single nucleotide at a time, releasing a nucleoside 5' monophosphate. In the absence of dNTPs, this activity will catalyze stepwise degradation from a free 3'-hydroxyl end of both single- and double-stranded DNA. In the presence of dNTPs, the exonuclease activity on double-stranded DNA is inhibited by the polymerase activity. During DNA synthesis, the exonuclease activity performs a proofreading function by removing misincorporated nucleotides.

In addition to the 3'->5' exonuclease activity, some DNA polymerases (e.g., *E. coli* DNA polymerase I) also have an associated 5'->3' exonuclease activity. This activity degrades double-stranded DNA from a free 5'-hydroxyl end .

The 5'->3' exonuclease activity removes from one to several nucleotides at a time, releasing predominantly nucleoside 5' phosphates, but also some larger oligonucleotides up to 10 nucleotides in length. The 5'->3' exonuclease activity of *E. coli* DNA polymerase I enables it to initiate synthesis from nicks in duplex DNA. The 5'->3' exonuclease degrades the DNA ahead of the synthesizing polymerase, resulting

in translocation of the nick. The 5'→3' exonuclease activity of E. coli DNA polymerase I, which is located at the N-terminus of the molecule, can be removed either by protease treatment or by deletion of the relevant part of the gene. The resulting DNA polymerase, which retains the 3' to 5' exonuclease, is referred to as the E. coli DNA polymerase I large fragment, or the Klenow fragment.

An important property of DNA polymerases is their processivity. Processivity is the ability of a polymerase molecule to incorporate nucleotides continuously on a given primer without dissociating from the primer template. Most DNA polymerases (e.g., E. coli DNA polymerase I, Klenow fragment, T4 DNA polymerase) have low processivity; they dissociate from a primer template after incorporating fewer than 10 nucleotides. In contrast, T7 DNA polymerase is highly processive; it can incorporate thousands of nucleotides from a given primer without dissociating from the primer template. This is a useful property when long stretches of DNA are being synthesized.



## **4.9 ENZYME KLENOW FRAGMENT OF ESCHERICHIA COLI DNA POLYMERASE I**

### **Introduction**

The Klenow fragment, molecular weight 76,000, consists of the C-terminal, 70% of E. coli DNA polymerase I. It retains the DNA polymerase and 3'->5' exonuclease activity of E. coli DNA polymerase I, but lacks the 5'->3' exonuclease activity.

## **4.10 LABELING CIRCULAR PERMUTE PROBS AT 3' ENDS OF DNA.**

### **Method**

1. In a 20- $\mu$ l reaction mixture, add 0.1 to 1  $\mu$ g of digested and purified DNA (fragments from circular permutations have a length of about 146 bps) with a restriction endonuclease that generates 5' overhanging ends

DNA fragments with blunt ends (e.g. probe D, digested with Eco RV) can be labeled inefficiently by replacement of the nucleotide at the 3'-hydroxyl terminus. For endonucleases that produce 3' overhanging ends, labeling of 3' termini must be carried out by replacement synthesis using T4 DNA polymerase.

2. Add 20  $\mu$ Ci of the desired [ $\alpha$ - $^{32}$ P]dNTP (3000 Ci/mmol) and 1  $\mu$ l of appropriate 5 mM 3dNTP mix. If higher specific activities are required, add 80  $\mu$ Ci of the radioactive dNTP at 5000 Ci/mmol.

Since the Klenow fragment incorporates nucleotides that are complementary to the single-stranded, 5' extensions, the choice of <sup>32</sup>P-labeled dNTP depends on the restriction endonuclease used to cleave the DNA. For example, labeling of BamHI ends (GATC, fragment A) can be accomplished with any of the radioactive precursors, whereas labeling of *EcoRI* ends (AATT, fragment G) requires either radioactive dATP or dTTP.

3. Add 1 U of the Klenow fragment and incubate 15 min at 30C.

It is unnecessary to change buffers, or to repurify the DNA prior to adding the Klenow fragment.

4. Stop the reaction with 1 ul of 0.5 M EDTA or by heating to 75C for 10 min. If desired, remove unincorporated dNTP precursors from labeled DNA .

#### Critical Parameters of Labeling 3' Ends

For DNA fragments that are 146 bp in length, the above procedure will generate DNA with a specific activity of about 200 cpm/ug. For situations where higher specific activities are desired, use radioactive dNTPs of higher specific activity ( 5000 Ci/mmol).

DNA can also be labeled selectively at one end by cleaving with two different restriction endonucleases and labeling with a <sup>32</sup>P-labeled dNTP that is complementary to only one of the two 5' extensions. For example, if DNA is cleaved with *EcoRI*



(AATT) and BamHI (GATC), the BamHI ends can be selectively labeled by using <sup>32</sup>P-labeled dGTP. Alternatively, DNA can be cleaved with one enzyme, radiolabeled, and then cleaved with the second enzyme. In this case, it is important to inactivate the Klenow fragment by heating to 75C for 15 min prior to addition of the second restriction enzyme.

#### **4.11 REPAIRING 3' OR 5' OVERHANGING ENDS TO GENERATE BLUNT ENDS.**

For many cloning experiments, it is necessary to convert the ends generated by restriction endonucleases into blunt ends.

##### **Method**

1. In a 20- $\mu$ l reaction, digest 0.1 to 4  $\mu$ g DNA with a restriction endonuclease.
2. Add 1  $\mu$ l of 0.5 mM each dNTP.
3. Add 1 to 5 U of the Klenow fragment and incubate at 30C for 15 min.

Repair of 5' extensions is carried out by polymerase activity, whereas repair of 3' extensions is carried out by 3' to 5' exonuclease activity. Due to the relative inactivity of exonuclease, this method is not desirable in cases where extensive repair of overhanging 3' ends is required. In such situations, T4 DNA polymerase (a much more expensive enzyme) or native T7 DNA polymerase are better choices.

4. Stop the reaction by heating to 75C for 10 min or by adding 1  $\mu$ l of 0.5 M EDTA.

For restriction fragments produced by cleavage with different endonucleases, it is possible to repair one end selectively. This is done by cleaving with enzyme 1, repairing the ends, inactivating the Klenow fragment by heat (75C for 10 min), and cleaving with enzyme 2.

## **4.12 ENZYME T4 DNA POLYMERASE**

### **Introduction**

T4 DNA polymerase, the product of gene 43 of bacteriophage T4, is either prepared from cells of E. coli that have been infected with the phage or has been overproduced from its cloned gene. T4 DNA polymerase is a single polypeptide of molecular weight 112,000. In addition to its DNA-dependent DNA polymerase activity, it has a very active single-stranded and double-stranded 3'->5' exonuclease. It lacks a 5'->3' exonuclease activity. T4 DNA polymerase by itself has low processivity; however, in the presence of several T4 accessory proteins, it becomes very processive. This polymerase has been produced from a clone that overexpresses its gene .

### **Materials**

For 50-ul reaction:

50 mM Tris-Cl, pH 8.0

5 mM MgCl<sub>2</sub>

5 mM DTT



2 ug DNA

100 uM 4dNTP mix

50 ug/ml BSA

10 U T4 DNA polymerase.

### Method

Incubate 20 min at 11C.

Stop reaction by adding 2 ul of 0.5 M EDTA or by heating to 75C for 10 min. The volume of reaction, concentration of 4 dNTPs, and the temperature of the reaction will vary, depending upon the individual application.

### Effect of Triphosphate Concentration

For reaction conditions that do not require radioactive labeling of the DNA, high concentrations (100 uM) of dNTPs are used to maximize the ratio of polymerase to exonuclease activity. In labeling experiments, the concentration of the labeled dNTP is reduced to 1 to 2 uM; this maximizes the specific activity of the DNA. Levels lower than 1 uM labeled dNTP should not be used because once the dNTPs are exhausted the exonuclease activity will degrade the DNA.

### Buffer Compatibility

For most applications, T4 DNA polymerase is used following digestion with restriction endonucleases. For many restriction enzymes, cleavage can be carried out

in T4 DNA polymerase buffer, and T4 DNA polymerase can be used directly. If buffers for the restriction enzyme and T4 DNA polymerase are incompatible, the DNA should be digested with the restriction enzyme and then repurified by phenol extraction, ethanol precipitation, and resuspension in TE buffer prior to treatment with T4 DNA polymerase.

#### **4.13 ENZYME ALKALINE PHOSPHATASES: BACTERIAL ALKALINE PHOSPHATASE AND CALF INTESTINE PHOSPHATASE**

##### **Introduction**

Bacterial alkaline phosphatase (BAP) from *E. coli* and calf intestine phosphatase (CIP) from veal are commonly used in nucleic acid research. Both enzymes catalyze the hydrolysis of 5'-phosphate residues from DNA, RNA, and ribo- and deoxyribonucleoside triphosphates. The dephosphorylated products possess 5'-hydroxyl termini which can subsequently be radioactively labeled using [ $\gamma$ - $^{32}\text{P}$ ]ATP and T4 polynucleotide kinase ( e.g., probe D, digested with *EcoRV*, which originates blunt ends).

Both phosphatases require  $\text{Zn}^{2+}$  for activity. The primary difference between them is the stability of the two enzymes. CIP is readily inactivated by heating to 70C for 10 min and/or extraction with phenol. On the other hand, BAP is much more resistant to these treatments. Thus, for most purposes, CIP is the enzyme of choice. Furthermore, CIP has a 10- to 20-fold higher specific activity than BAP.

We used vectors (pT7-7) dephosphorilated to prevent self ligation.



#### **4.14 CIP REACTION. CONDITION FOR DEPHOSPHORYLATION OF DNA.**

##### **Materials**

For 50- $\mu$ l reaction:

20 mM Tris-Cl, pH 8.0

1 mM MgCl<sub>2</sub>

1 mM ZnCl<sub>2</sub>

1 to 20 pmol DNA termini

0.1 U CIP

##### **Method**

Incubate at 37C for 30 min. Stop reaction by heating to 75C for 10 min or extracting with phenol, and then precipitate with ethanol. NOTE: CIP is heat labile; 10 min at 75C effectively inactivates CIP.

The volume of reaction and concentration of DNA will vary, depending upon the individual application. In general, phosphatase treatment can be done directly following cleavage by a restriction endonuclease, thus minimizing the number of manipulations.

#### **4.15 ENZYME T4 POLYNUCLEOTIDE KINASE**

The forward reaction of T4 polynucleotide kinase catalyzes the transfer of the terminal (gamma) phosphate of ATP to the 5'-hydroxyl termini of DNA and RNA. This reaction is very efficient and hence is the general method for labeling 5' ends or for phosphorylating oligonucleotides.

The exchange reaction of T4 polynucleotide kinase catalyzes the exchange of 5'-terminal phosphates. In this reaction, which requires an excess of ADP, the 5'-terminal phosphate is transferred to ADP and subsequently rephosphorylated by the transfer from the gamma phosphate of [ $\gamma$ - $^{32}\text{P}$ ]ATP. The exchange reaction is less efficient than the forward reaction; thus, it is rarely used.

Finally, polynucleotide kinase is a 3' phosphatase. Some commercial preparations of polynucleotide kinase are prepared from the phage T4 strain am N81 pseT1, which has a mutated pseT gene. This derivative lacks the 3' phosphatase activity.

#### **4.16 LABELING 5' ENDS BY THE FORWARD REACTION.**

For 30- $\mu\text{l}$  reaction:

##### **Materials**

50 mM Tris-Cl pH 7.5

10 mM MgCl<sub>2</sub>

5 mM DTT



1 to 50 pmol dephosphorylated DNA, 5' ends (fragment D, digested with *EcoRV*)

50 pmol (150 uCi) [ $\gamma$ -<sup>32</sup>P]ATP (specific activity >3000 Ci/mmol)

50 ug/ml BSA

20 U T4 polynucleotide kinase

### Method

Incubate the above amounts of enzyme, buffer DNA and radioactive ATP at 37C for 60 min.

Stop the reaction by adding 1 ul of 0.5 M EDTA or by heating to 75C for 10 min.

Extract with phenol/chloroform.

Separate the labeled DNA from the unincorporated labeled nucleotides by filtration on Sephadex G-100 or by centrifugation through a spin column containing Sephadex G-50 .

The volume of reaction and the concentration of DNA and [ $\gamma$ -<sup>32</sup>P]ATP will vary, depending upon the application.

To phosphorylate synthetic oligonucleotides by the forward reaction,

### Materials

for 30-ul reaction:

50 mM Tris-Cl, pH 7.5

10 mM MgCl<sub>2</sub>

5 mM DTT

1 to 10 ug oligonucleotide linker

1 mM ATP

50 ug/ml BSA

20 U T4 polynucleotide kinase

Incubate at 37C for 60 min. Stop reaction by adding 1 ul of 0.5 M EDTA. If desired, [ $\gamma$ -<sup>32</sup>P]ATP can be added to trace the reaction.

#### **4.17 LABELING 5' TERMINI BY THE EXCHANGE REACTION**

For 30- $\mu$ l reaction:

##### **Materials**

50 mM imidazole-Cl, pH 6.6

10 mM MgCl<sub>2</sub>

5 mM DTT

1 to 50 pmol phosphorylated DNA, 5' ends

5 mM ADP

60 pmol (180  $\mu$ Ci) [ $\gamma$ -<sup>32</sup>P]ATP (specific activity >3000 Ci/mmol)

50  $\mu$ g/ml BSA

20 U T4 polynucleotide kinase

##### **Method**

Incubate at 37C for 60 min. Stop reaction by adding 1  $\mu$ l of 0.5 M EDTA. Extract with buffered phenol. Separate the labeled DNA from the unincorporated labeled nucleotides by either Sephadex G-100 gel filtration or centrifugation through a spin column containing Sephadex G-50.

The volume of reaction and the concentration of DNA and [ $\gamma$ -<sup>32</sup>P]ATP will vary, depending upon the application.



## **4.18 ENZYME T4 DNA LIGASE**

### **Introduction**

T4 DNA ligase, the product of gene 30 of phage T4, was originally purified from phage-infected cells of *E. coli*. The phage T4 gene 30 has been cloned, and the enzyme is now prepared from overproducing strains. Using ATP as a cofactor, T4 DNA ligase catalyzes the repair of single-stranded nicks in duplex DNA and joins duplex DNA restriction fragments having either blunt or cohesive ends. It is the only ligase that efficiently joins blunt-end termini under normal reaction conditions.

### **Materials**

For 50- $\mu$ l reaction:

40 mM Tris-Cl, pH 7.5

10 mM MgCl<sub>2</sub>

10 mM DTT

1  $\mu$ g DNA

0.5 mM ATP

50  $\mu$ g/ml BSA

1 "Weiss" U T4 DNA ligase

### **Method**

Incubate at 12-30C for 1 to 16 hr. Stop reaction by adding 2  $\mu$ l of 0.5 M EDTA or by heating to 75C for 10 min. The volume of reaction, concentration of DNA, and the

temperature and time of the reaction will vary, depending upon the individual application. One Weiss unit is equivalent to 60 cohesive-end units.

Ligation of cohesive ends is usually carried out at 12-15°C to maintain a good balance between annealing of the ends and activity of the enzyme. Higher temperatures make it difficult for the ends to anneal, whereas lower temperatures diminish ligase activity. Blunt-end ligations are typically performed at room temperature since annealing is not a factor (the enzyme is not particularly stable above 30°C). Blunt-end ligations require about 10 to 100 times more enzyme than cohesive-end ligations to achieve an equal efficiency. T4 DNA ligase is not inhibited by tRNA, but it is strongly inhibited by NaCl concentrations >150 mM. Macromolecular exclusion molecules (e.g., PEG 8000) have been shown to greatly increase the rate of both cohesive-end and blunt-end joining by T4 DNA ligase. An inherent consequence of macromolecular crowding is that all ligations are intermolecular; thus, this technique is not suitable for the ligation and circularization of inserts and vectors that are required for most cloning experiments.

#### **4.19 SUBCLONING OF DNA FRAGMENTS**

In order to construct new DNA molecules, the starting DNAs are treated with appropriate restriction endonucleases and other enzymes if necessary (pT7-7 was digested with *Nde*I and *Eco*RI, p Bend 2 was digested with *Xba*I and *Sal*I). The individual components of the desired DNA molecule are purified by agarose or polyacrylamide gel electrophoresis, combined, and treated with DNA ligase. The products of the ligation mixture (along with control mixtures) are introduced into



competent *E. coli* cells, and transformants are identified by an appropriate genetic selection. DNA is prepared from the colonies or plaques and subjected to restriction endonuclease mapping in order to determine if the desired DNA molecule was created. All cloning experiments follow the steps outlined below.

### Materials

Calf intestine phosphatase (CIP) and buffer.

dNTP mix (0.5 mM each)

Klenow fragment of *E. coli* DNA polymerase I or T4 DNA polymerase

Oligonucleotide linkers.

10 mM ATP

0.2 mM dithiothreitol (DTT)

T4 DNA ligase (measured in cohesive-end units).

2x T4 DNA ligase buffer

### Methods

1. In a 20- $\mu$ l reaction mixture, cleave the individual DNA components with appropriate restriction endonuclease. Vector pT7-7 and fragment of PCR containing SRY boxes were digested with endonucleases *Nde*I and *Hind*III, pBend2 with *Xba*I and *Sal*I. After the reaction is complete, inactivate the enzymes by heating 15 min to 75C.

Reaction mixtures can be done in any volume; 20  $\mu$ l is convenient for gel electrophoresis. Many of the subsequent enzymatic manipulations can be carried out sequentially without further buffer changes.

2. To remove the 5' phosphates of vectors pT7-7 and p Bend 2 (to prevent self ligation), add 2 ul of 10x CIP buffer and 1 U CIP; incubate 30 to 60 min at 37C as described before. After the reaction is complete, inactivate CIP by heating 15 min to 75C. .

3. If one or both ends generated by a restriction endonuclease must be converted to blunt ends add 1 ul of a solution containing all 4 dNTPs (0.5 mM each) and an appropriate amount of the Klenow fragment of E. coli DNA polymerase I or T4 DNA polymerase; carry out the filling-in or trimming reaction. After the reaction is complete, inactivate the enzymes by heating 15 min to 75C.

4. Isolate the desired DNA segments by gel electrophoresis, or by other methods if appropriate.

Electrophoresis in agarose gels is the most common method. Purification is not essential for many cloning experiments, but it is usually very helpful.

5. Using longwave UV light for visualization of the DNA, cut out the desired band(s) and purify the DNA away from the gel material using the procedures described.

It is critical to use longwave UV light sources to prevent damage to the DNA. If low gelling/melting temperature agarose is used, ligation reactions can usually be performed directly in the gel slice (see alternative protocol).



**6. Set up the following ligation reaction:**

**7 ul component DNAs (0.1 to 5 ug: the molar ratio between vector and insert is 1:1)**

**10 ul 2x ligase buffer**

**1 ul 10 mM ATP**

**20 to 500 U (cohesive end) T4 DNA ligase**

**Incubate 1 to 24 hr at 15C.**

**Simple ligations with two fragments having 4-bp 3' or 5' overhanging ends require much less ligase than more complex ligations or blunt-end ligations. The quality of the DNA will also affect the amount of ligase needed.**

**For simplicity, appropriate DNAs should be added to each tube (adding water if necessary) so that the volume is 9 ul. Immediately prior to use, premix on ice the remaining ingredients (2x buffer, ATP, enzyme) in sufficient amounts for all the reactions. To start the ligation reactions, add 11-ul aliquots of the premix to each tube containing DNA.**

**8. Introduce 1 to 10 ul of the ligated products into competent E. coli cells and select for transformants by virtue of the genetic marker present on the vector.**

**9 From individual E. coli transformants, purify plasmid or phage DNAs by miniprep procedures and determine their structures by restriction mapping and sequencing.**

## **4.20 ALTERNATIVE PROTOCOL: LIGATION OF DNA FRAGMENTS IN GEL SLICES**

This alternate protocol saves considerable time in comparison to the basic protocol because it eliminates purification of the DNA fragments away from the gel matrix. It is suitable for most simple cloning experiments and is particularly valuable for carrying out a set of hybrid constructions involving a variety of different DNA fragments. However, as the cloning efficiency may be reduced, this method should be employed only when it is desired to make one (or relatively few) specific molecules.

### **Additional Materials**

Low gelling/melting temperature agarose (SeaPlaque, FMC Marine Colloids)

TAE buffer

### **Method**

1. Treat the starting DNAs with appropriate restriction endonucleases and other enzymes.
2. Subject the treated DNAs to electrophoresis in low gelling/melting temperature agarose using TAE buffer.

It is critical that the agarose be of high quality. SeaPlaque is a good choice. The agarose concentration should be kept as low as possible (0.7% is suitable for most applications).



3. Cut out the desired band of linear vectors and fragments in the smallest possible volume (20 to 50 ul) using a clean razor blade, and place the gel slice in a microcentrifuge tube.

4. Melt the gel slices containing DNA at 70C for at least 10 min.

This temperature is hot enough to melt the agarose without denaturing the DNA.

5. In separate tubes for each ligation reaction, combine the gel slices containing appropriate DNAs (and water if necessary) for a total volume of 9 ul. Place the tubes at 37C for a few minutes.

The gel slices should remain molten at 37C.

6. To each tube containing DNA, add 11 ul of an ice-cold mixture containing 2x buffer, ATP, and T4 DNA ligase. Mix immediately by flicking the tube and place on ice. Then incubate the reaction mixtures 1 to 48 hr at 15C.

DNA fragments can still be ligated even though the reaction mixture has resolidified into a gel.

7. After the ligation reaction is complete, remelt the gel slices 5 to 10 min at 73C and add 5 ul of the ligated products to 200 ul of competent E. coli cells.

The remelted gel must be diluted at least 30-fold so that it does not resolidify when the cells are placed on ice.

### Materials

2x T4 DNA ligase buffer

100 mM Tris-Cl, pH 7.5

20 mM MgCl<sub>2</sub>

20 mM DTT

## **4.21 CONSTRUCTION OF RECOMBINANT DNA FRAGMENT OF SRY BOXES BY POLYMERASE CHAIN REACTION.**

### **INTRODUCTION**

Any two segments of DNA can be ligated together into a new recombinant molecule using the polymerase chain reaction (PCR). The DNA can be joined in any configuration, with any desired junction-point reading frame or restriction site, by incorporating extra nonhomologous nucleotides within the PCR primers. Cloning by PCR is often more rapid and versatile than cloning with standard techniques that rely on the availability of naturally occurring restriction sites and require microgram quantities of DNA. It is not necessary to know the nucleotide sequence of the DNA being subcloned by this technique, other than the two short flanking regions (~20 bp) that serve as anchors for the two oligonucleotide primers used in the amplification



process. Moreover, PCR can be performed on low-abundance or even degraded DNA (or RNA) sources.

Important variables that can influence the outcome of PCR include the  $MgCl_2$  concentration and the cycling temperatures. Additives that promote polymerase stability and processivity or increase hybridization stringency, and strategies that reduce nonspecific primer-template interactions, especially prior to the critical first cycle, can greatly improve sensitivity, specificity, and yield.

#### **4.22 SUBCLONING DNA FRAGMENTS**

In this protocol, synthetic oligonucleotides incorporating new unique restriction sites are used to amplify a region of DNA to be subcloned into a vector containing compatible restriction sites. The amplified DNA fragment is purified, subjected to enzymatic digestion at the new restriction sites, and then ligated into the vector. Individual subclones are analyzed by restriction endonuclease digestion and sequenced.

##### **Materials**

**Sterile H<sub>2</sub>O**

**30 mM  $MgCl_2$**

**10x  $MgCl_2$ -free PCR amplification buffer**

**25 mM 4dNTP mix**

**50 uM oligonucleotide primer 1: 50 pmol/ul in sterile H<sub>2</sub>O (store at -20C)**

**50 uM oligonucleotide primer 2: 50 pmol/ul in sterile H<sub>2</sub>O (store at -20C)**

**Template DNA: 1 ug mammalian genomic DNA**

**5 U/ul Taq DNA polymerase**

**DMSO, cell culture grade**

**Glycerol**

**Taq DNA polymerase**

**Mineral oil**

**Automated thermal cycler**

**MgCl<sub>2</sub>-free PCR amplification buffer, 10x**

**500 mM KCl**

**100 mM Tris-Cl, pH 9.0 (at 25C)**

**0.1% Triton X-100**

**Store indefinitely at -20C**

**This buffer can be obtained from Promega; it is supplied with Taq DNA polymerase.**

**4dNTP mix**

**For 2 mM 4dNTP mix: Prepare 2 mM each dNTP in TE buffer, pH 7.5. Store up to 1 year at -20C in 1-ml aliquots.**

**For 25 mM 4dNTP mix: Combine equal volumes of 100 mM dNTPs (Promega). Store indefinitely at -20C in 1-ml aliquots.**



**TE-buffered phenol and chloroform**

**100% ethanol**

**TE buffer, pH 8.0**

**Klenow fragment of E. coli DNA polymerase I**

**Vector DNA**

**Calf intestinal phosphatase**

**Additional reagents and equipment for enzymatic amplification of DNA by PCR, agarose and polyacrylamide gel electrophoresis, DNA extraction and precipitation, purification of DNA by glass beads, electroelution from agarose gels, or from low-gelling/melting temperature agarose gels, restriction endonuclease digestion, ligation of DNA fragments, transformation of E. coli, plasmid DNA minipreps, and DNA sequence analysis.**

**Amplify the target DNA**

**Method**

**1. Prepare the template DNA. If using an impure DNA preparation, heat sample 10 min at 100C to inactivate nucleases.**

**Genomic DNA from patients and primates provided by Goodfellow's group were used as the source of target DNA.**

**2. Prepare oligonucleotide primers.**

Oligonucleotides SRYboxdir (CCACATATGCAGGATAGAGTGAAGCGA).and SRYboxrev (CGAAGCTTAACGACGAGGTCGATACTT) were synthesised.

Because the purity of the oligonucleotides does not seem to affect the PCR reaction, primer purification may not be necessary. Sites for restriction endonucleases *NdeI* and *HindIII* are underlined.

Oligonucleotides SRYboxdir and SRYboxrev were synthesised by the phosphotriester method.

3. Set up a standard amplification reaction and overlay with mineral oil. 1. Carry out PCR in an automated thermal cycler. PCR mixtures (50 µl) contained 50 pmol each of oligonucleotides SRYboxdir and SRYboxrev, 0.2 mM dNTPs, 400 ng purified total genomic DNA, 1 unit Taq DNA polymerase and 5 µl Taq polymerase 10 X buffer (Perkin-Elmer Cetus). Twenty five cycles of denaturation (30 s at 94C), annealing (60 s at 50C) and polymerisation (60 s at 72C) were performed. Extension of an additional 10 min at 72C in the last cycle made products as complete as possible.

Include negative controls of no template DNA and each oligonucleotide alone, as well as several different oligonucleotide:template ratios.

Recover the amplified fragment as described .

4. Analyze an aliquot (e.g., 4 to 8 ul) of each reaction mix by agarose gel electrophoresis to verify that the amplification has yielded the expected product.



5. Recover amplified DNA from PCR reaction mix. Remove mineral oil overlay from each sample, then extract sample once with buffered chloroform to remove residual mineral oil. Extract once with buffered phenol and then precipitate DNA with 100% ethanol.

6. Microcentrifuge DNA 10 min at high speed, 4C. Dissolve pellet in 20 ul TE buffer. Purify desired PCR product from unincorporated nucleotides, oligonucleotide primers, unwanted PCR products, and template DNA using glass beads, electroelution, or phenol extraction of low gelling/melting temperature agarose.

Unused oligonucleotide primers can inhibit the ability of the restriction enzymes to digest the amplified PCR product.

Prepare amplified fragment and vector for ligation.

7. Since primers contain unique restriction sites, digest half of the amplified DNA in 20 ul with the appropriate restriction endonucleases (*Hind* III and *Nde* I for SRY boxes). Use an excess of enzyme, and digest for several hours.

Reserve the undigested half for future use, if necessary.

8. Prepare the recipient vector pT7-7 for cloning by digesting 0.2 to 2 ug in 20 ul with *Nde*I-*Hind*III restriction enzymes; although not necessary, since the vector could not be completely digested, treat vector DNA with calf intestinal phosphatase to prevent recircularization during ligation.

9. Separate the linearized vector from uncut vector by agarose or low-gelling/melting temperature gel electrophoresis. Recover linearized vector from the gel by adsorption to glass beads, electroelution, or phenol extraction of low-gelling/melting temperature agarose.

10. Ligate the PCR fragments into the digested vector.

11. Transform an aliquot of each ligation into *E. coli*. Prepare plasmid miniprep DNA from a subset of transformants.

#### Analyze recombinant plasmids

12. Digest the plasmid DNA of the selected transformants with *NdeI* and *HindIII* restriction endonucleases. Analyze the digestions by agarose gel electrophoresis to confirm fragment incorporation.

13. Sequence the amplified fragment portion of the plasmid DNA to check for mutations.

This analysis is critical because the Taq DNA polymerase can introduce mutations into the amplified fragment.



## **4.23 DNA SEQUENCING STRATEGIES**

### **INTRODUCTION**

Any double-stranded template used in dideoxy sequencing should be denatured before being annealed to the primer. In general practice, alkali denaturation of plasmid DNA works better for sequencing than heat denaturation

## **4.24 ALKALI DENATURATION OF DOUBLE-STRANDED PLASMID DNA FOR DIDEOXY SEQUENCING**

Theoretically, it should be possible to denature double-stranded templates either by treating with alkali or by boiling and achieve equal results in sequencing. In practice, however, alkali denaturation of closed-circular double-stranded templates and  $\lambda$  templates usually gives superior results for dideoxy sequencing.

In the procedure described below, a recombinant plasmid is denatured using NaOH. After adjusting the pH to 7.0, the DNA is precipitated with ethanol, washed, and dried. The dried pellet is suitable to be added to annealing reactions prior to DNA sequencing reactions

### **Materials**

Recombinant plasmid DNA 2 M NaOH/2 mM EDTA

**3 M sodium acetate, pH 6.0**

**95% and 70% ethanol**

**0.5-ml microcentrifuge tubes**

**Method**

**1. Add ~0.5 pmol of recombinant plasmid DNA pT7-SRYbox-wt, M64I, I90M, G95R, K106I and F109S to a 0.5-ml microcentrifuge tube. If the volume is >20 ul, ethanol precipitate the DNA and redissolve in 20 ul water. If the volume is <20 ul, add water to bring the volume to 20 ul.**

**0.5 pmol of a 2.5-kbp plasmid is ~0.8 ug.**

**2. Add 2 ul of 2 M NaOH/2 mM EDTA and gently mix by drawing up and down with a pipet. Incubate 5 min at 25-37C.**

**3. Place sample on ice, add 7 ul water, and mix thoroughly by drawing solution up and down with the pipettor.**

**4. Add 7 ul of 3 M sodium acetate, pH 6.0 (to neutralize DNA solution). Mix thoroughly by drawing solution up and down with the pipettor. Check the pH of the solution by spotting 1 ul on pH paper. Add 3 M sodium acetate, pH 6.0 until the pH is =7.0.**

**5. Add 75 ul of 95% ethanol and place 10 min on dry ice.**

**6. Microcentrifuge 10 min, 4C, and carefully remove and discard the supernatant.**



Use caution when removing the ethanol because the DNA pellet may also be removed.

7. Add 400 ul 70% ethanol. Microcentrifuge 10 min at 4C, then carefully remove and discard the ethanol layer.

8. Dry pellet 10 min in a Speedvac evaporator. Store the pellet at 20C (up to several weeks) until used as template for dideoxy sequencing reactions

#### **4.25 LABELING/TERMINATION SEQUENCING REACTIONS USING SEQUENASE**

The labeling/termination sequencing protocol involves two steps. In the labeling step, primed DNA synthesis is initiated in the presence of limiting concentrations of all four dNTPs, including [ $\alpha$ -<sup>35</sup>S]dATP, and continues until one of the dNTP pools is depleted. At this point, the uniformly labeled DNA chains are distributed randomly in length from a few nucleotides to hundreds of nucleotides. In the second step, synthesis resumes in the presence of additional dNTPs and one ddNTP. Elongation of the DNA chains in this step is rapid and processive until termination occurs at specific bases after incorporation of the corresponding dideoxynucleotide. In this protocol, the average length of the radioactively labeled oligonucleotide products is modified by altering the concentration of dNTPs in the first step; however, it can also be regulated by altering the dNTP:ddNTP ratio in the termination reaction

This basic protocol uses Sequenase. The labeling/termination procedure can also be used with other polymerases; however, because each polymerase has different buffer and  $Mg^{++}$  concentration optima, and each discriminates to a different extent against ddNTPs, the concentrations of these components must be modified in each case.

### Materials

0.5 pmol single-stranded or denatured double-stranded DNA template

0.5 to 1 pmol/ul oligonucleotide primer in water (store at -20C)

10x Sequenase buffer

Sequenase termination mixes

Sequenase/pyrophosphatase mix

Sequenase diluent

Labeling mixes

10 mCi/ml [ $\alpha$ -<sup>35</sup>S]dATP (500 to 1200 Ci/mmol)

Stop/loading dye

0.5-ml microcentrifuge tubes

Heat-resistant microtiter plates.

Anneal primer and template

### Method

1b. Resuspend a dried pellet containing 0.5 pmol denatured double-stranded DNA in the following mixture:

1 pmol primer T7 for pT7 vector, M13 for pbend2 (pharmacia).



**1 ul 10x Sequenase buffer**

**H<sub>2</sub>O to 10 ul.**

**Mix gently by pipetting up and down (avoid creation of bubbles). Incubate 30 min at 37C, then keep at this annealing temperature until ready to proceed to step 2.**

**0.8 ug of 2500-bp double-stranded DNA molecule corresponds to 0.5 pmol of template molecules. Double-stranded DNA templates require twice as much primer as single-stranded DNA templates. For more stringent annealing conditions, incubate at 42-50C instead of 37C.**

**Set up termination reactions**

**2. While the primer is being annealed to the template, label four microcentrifuge tubes A, C, G, and T for each template to be sequenced.**

**For simultaneous sequencing of a large number of templates, the reactions can be carried out in heat-resistant, round-bottom, 96-well microtiter plates with heat-resistant lids**

**3. Add 2.5 ul each of A, C, G, and T Sequenase termination mixes to the bottom of the A, C, G, and T tubes, respectively.**

**Keep tubes closed because the small volumes evaporate rapidly when open to the air.**

**Carry out labeling reaction**

**4. Immediately before use, dilute the Sequenase/pyrophosphatase mix in Sequenase diluent to 1 to 2 U Sequenase/ul and keep on ice.**

**5. Add the following to the annealed primer and template:**

**2 ul labeling mix**

**0.5 to 1.5 ul 10 mCi/ml [ $\alpha$ -<sup>35</sup>S]dATP**

**2 ul diluted Sequenase/pyrophosphatase mix.**

**The total volume is 14.5 to 16 ul. Incubate 5 min at 25C (room temperature).**

**Choose the labeling mix appropriate for the lengths of sequencing products that are desired. Reaction times can be extended to 20 min without detriment, although the reaction is complete within 5 min, by which time the nucleotide pools are exhausted. This step can be performed at 37C; however, 25C limits the processivity of the Sequenase, an advantage during the labeling step, and maximizes its half-life during the sequencing reaction.**

**At least 3 pmol of [ $\alpha$ -<sup>35</sup>S]dATP are required for reactions containing the short labeling mixes whereas 15 pmol must be added to reactions containing the long labeling mixes. Amounts greater than these are not incorporated during the labeling step, because the other three dNTPs become limiting. The concentration of [ $\alpha$ -<sup>35</sup>S]dATP at 10 mCi/ml with a specific activity of 1000 Ci/mmol is 10 uM. Thus for**



such a preparation, 0.3 and 1.5 ul are sufficient for the labeling step with the short and long labeling mixes, respectively.

Carry out termination reaction

6. Add 3.5 ul of the labeling reaction mixture to the tube containing Sequenase termination mix A (from step 3). Mix the solution by gently pipetting up and down. Repeat this addition to the C, G, and T tubes, changing pipet tips each time. Incubate 5 to 10 min at 37C.

Reactions are complete within 2 to 3 min but incubations can be extended to 30 min without problems except when using dITP under some conditions

7. Add 4 ul stop/loading dye.

35S sequencing reactions may be stored for up to one week at -20C before electrophoresis.

8. Heat samples 2 min in a 95C water bath, then place on ice. Load 2 to 3 ul of each sample on a sequencing gel. Electrophorese the gel and read the sequence .

Excessive boiling of the completed reactions in formamide/dye solution may cause DNA chain breakage and smeared bands on the sequencing gel. If repeated loadings are planned, remove a 3 ul aliquot of each reaction to heat before each loading.

## **4.26 COLORIMETRIC METHODS FOR QUANTIFYING PROTEINS.**

### **4.27 BRADFORD METHOD**

The Bradford method depends on quantitating the binding of a dye, Coomassie brilliant blue, to an unknown protein and comparing this binding to that of different amounts of a standard protein, usually bovine serum albumin. It is designed to quantify 1 to 10 ug protein.

#### **Materials**

0.5 mg/ml bovine serum albumin (BSA)

0.15 M NaCl

Coomassie brilliant blue solution

#### **Method**

1. Into 8 microcentrifuge tubes aliquot duplicate amounts of 0.5 mg/ml BSA (5, 10, 15, and 20 ul) and with 0.15 M NaCl bring the volume in each to 100 ul. Into 2 microcentrifuge tubes, aliquot 100 ul of 0.15 M NaCl; these are blank tubes.

2. Add 1 ml Coomassie brilliant blue solution and vortex. Allow to stand 2 min at room temperature.

3. Determine the  $A_{595}$  using a 1-cm pathlength microcuvette (1 ml) and make a standard curve by plotting absorbance at 595 nm versus protein concentration.



Determine the absorbance for the unknown and use the standard curve to determine the concentration of protein in the unknown.

If the unknown protein concentration is too high, dilute the protein, assay a smaller aliquot of the unknown, or generate another standard curve in a higher concentration range (e.g., 10 to 100 ug).

## REAGENTS AND SOLUTIONS

0.5 mg/ml bovine serum albumin (BSA)

The concentration of BSA is determined using the  $A_{280} = 6.6$  for a 10 mg/ml solution of BSA measured in a 1-cm pathlength cuvette (e.g., a 0.5 mg/ml solution will have an  $A_{280} = 0.33$ ).

Coomassie brilliant blue solution

In a 1-liter volumetric flask, dissolve 100 mg Coomassie brilliant blue G-250 in 50 ml of 95% ethanol. Add 100 ml of 85% phosphoric acid. Bring to volume with water. Filter through Whatman No. 1 filter paper. Store at 4C.

Commercial kits are available from Pierce (#23200) and Bio-Rad (#500-0006).

## **4.28 GEL ELECTROPHORESIS**

### **4.29 DENATURING (SDS) DISCONTINUOUS GEL ELECTROPHORESIS: LAEMMLI GEL METHOD**

One-dimensional gel electrophoresis under denaturing conditions (i.e., in the presence of 0.1% SDS) separates proteins based on molecular size as they move through a polyacrylamide gel matrix toward the anode. The polyacrylamide gel is cast as a separating gel (sometimes called resolving or running gel) topped by a stacking gel and secured in an electrophoresis apparatus. After sample proteins are solubilized by boiling in the presence of SDS, an aliquot of the protein solution is applied to a gel lane, and the individual proteins are separated electrophoretically. 2-Mercaptoethanol (2-ME) or dithiothreitol (DTT) is added during solubilization to reduce disulfide bonds.

This protocol is designed for a vertical slab gel with a maximum size of 0.75 mm x 14 cm x 14 cm. For thicker gels, or minigels the volumes of stacking and separating gels and the operating current must be adjusted.

#### **Materials**

**Separating and stacking gel solutions**

**H<sub>2</sub>O-saturated isobutyl alcohol**

**1x Tris-Cl/SDS, pH 8.8 (dilute 4x Tris-Cl/SDS, pH 8.8)**

**Protein sample to be analyzed**



**2x and 1x SDS sample buffer.**

**Protein molecular-weight-standards mixture.**

**6x SDS sample buffer.**

**1x SDS electrophoresis buffer.**

**Electrophoresis apparatus: Protean II 16-cm cell (Bio-Rad) or SE 600/400 16-cm unit (Hoefer Pharmacia Biotech) with clamps, glass plates, casting stand, and buffer chambers**

**0.75-mm spacers**

**0.45- $\mu$ m filters (used in stock solution preparation)**

**25-ml Erlenmeyer side-arm flask**

**Vacuum pump with cold trap**

**0.75-mm Teflon comb with 1, 3, 5, 10, 15, or 20 teeth**

**25- or 100- $\mu$ l syringe with flat-tipped needle**

**Constant-current power supply**

**Pour the separating gel**

### **Method**

**1. Assemble the glass-plate sandwich of the electrophoresis apparatus according to manufacturer's instructions using two clean glass plates and two 0.75-mm spacers.**

**If needed, clean the glass plates in liquid Alconox or RBS-35 (Pierce). These aqueous-based solutions are compatible with silver and Coomassie blue staining procedures.**

2. Lock the sandwich to the casting stand.

3. Prepare the separating gel solution degassing using a rubber-stoppered 25-ml Erlenmeyer side-arm flask connected with vacuum tubing to a vacuum pump with a cold trap. After adding the specified amount of 10% ammonium persulfate and TEMED to the degassed solution, stir gently to mix.

The stacking gel is the same regardless of the separating gel used.

The desired percentage of acrylamide in the separating gel depends on the molecular size of the protein being separated. Generally, use 5% gels for SDS-denatured proteins of 60 to 200 kDa, 10% gels for SDS-denatured proteins of 16 to 70 kDa, and 15% gels for SDS-denatured proteins of 12 to 45 kDa

4. Using a Pasteur pipet, apply the separating gel solution to the sandwich along an edge of one of the spacers until the height of the solution between the glass plates is ~11 cm.

Use the solution immediately; otherwise it will polymerize in the flask.

Sample volumes <10 ul do not require a stacking gel. In this case, cast the resolving gel as you normally would, but extend the resolving gel into the comb to form the well. The proteins are then separated under the same conditions as used when a stacking gel is present. Although this protocol works well with single-concentration gels, a gradient gel is recommended for maximum resolution



5. Using another Pasteur pipet, slowly cover the top of the gel with a layer (~1 cm thick) of H<sub>2</sub>O-saturated isobutyl alcohol, by gently layering the isobutyl alcohol against the edge of one and then the other of the spacers.

Be careful not to disturb the gel surface. The overlay provides a barrier to oxygen, which inhibits polymerization, and allows a flat interface to form during gel formation.

The H<sub>2</sub>O-saturated isobutyl alcohol is prepared by shaking isobutyl alcohol and H<sub>2</sub>O in a separatory funnel. The aqueous (lower) phase is removed. This procedure is repeated several times. The final upper phase is H<sub>2</sub>O-saturated isobutyl alcohol.

6. Allow the gel to polymerize 30 to 60 min at room temperature.

A sharp optical discontinuity at the overlay/gel interface will be visible on polymerization. Failure to form a firm gel usually indicates a problem with the ammonium persulfate, TEMED (N, N, N', N'-tetramethylethylenediamine), or both. Ammonium persulfate solution should be made fresh before use. Ammonium persulfate should "crackle" when added to the water. If not, fresh ammonium persulfate should be purchased. Purchase TEMED in small bottles so, if necessary, a new previously unopened source can be tried.

**Pour the stacking gel**

7. Pour off the layer of H<sub>2</sub>O-saturated isobutyl alcohol and rinse with 1x Tris-HCl/SDS, pH 8.8.

Residual isobutyl alcohol can reduce resolution of the protein bands; therefore, it must be completely removed. The isobutyl alcohol overlay should not be left on the gel longer than 2 hr.

8. Prepare the stacking gel solution.

Use the solution immediately to keep it from polymerizing in the flask.

9. Using a Pasteur pipet, slowly allow the stacking gel solution to trickle into the center of the sandwich along an edge of one of the spacers until the height of the solution in the sandwich is ~1 cm from the top of the plates.

Be careful not to introduce air bubbles into the stacking gel.

10. Insert a 0.75-mm Teflon comb into the layer of stacking gel solution. If necessary, add additional stacking gel to fill the spaces in the comb completely.

Again, be careful not to trap air bubbles in the tooth edges of the comb; they will cause small circular depressions in the well after polymerization that will lead to distortion in the protein bands during separation.

11. Allow the stacking gel solution to polymerize 30 to 45 min at room temperature.



A sharp optical discontinuity will be visible around wells on polymerization.

**Prepare the sample and load the gel**

12. Dilute a portion of the protein sample to be analyzed 1:1 (v/v) with 2x SDS sample buffer and heat 3 to 5 min at 100C in a sealed screw-cap microcentrifuge tube. If the sample is a precipitated protein pellet, dissolve the protein in 50 to 100 ul of 1x SDS sample buffer and boil 3 to 5 min at 100C. Dissolve protein-molecular-weight standards mixture in 1x SDS sample buffer according to supplier's instructions as a control.

For dilute protein solutions, consider adding 5:1 protein solution/6x SDS sample buffer to increase the amount of protein loaded. Proteins can also be concentrated by precipitation in acetone, ethanol, or trichloroacetic acid (TCA), but losses will occur.

For a 0.8-cm-wide well, 25 to 50 ug total protein in <20 ul is recommended for a complex mixture when staining with Coomassie blue, and 1 to 10 ug total protein is needed for samples containing one or a few proteins. If silver staining is used, 10- to 100-fold less protein can be applied (0.01 to 5 ug in <20 ul depending on sample complexity).

To achieve the highest resolution possible, the following precautions are recommended. Prior to adding the sample buffer, keep samples at 0C. Add the SDS sample buffer (room temperature) directly to the 0C sample (still on ice) in a screw-

top microcentrifuge tube. Cap the tube to prevent evaporation, vortex, and transfer directly to a 100C water bath for 3 to 5 min. Let immunoprecipitates dissolve for 1 hr at 56C in 1x SDS sample buffer prior to boiling. DO NOT leave the sample in SDS sample buffer at room temperature without first heating to a 100C to inactivate proteases. Endogenous proteases are very active in SDS sample buffer and will cause severe degradation of the sample proteins after even a few minutes at room temperature. To test for possible proteases, mix the sample with SDS sample buffer without heating and leave at room temperature for 1 to 3 hr. A loss of high-molecular-weight bands and a general smearing of the banding pattern indicate a protease problem. Once heated, the samples can sit at room temperature for the time it takes to load samples.

13. Carefully remove the Teflon comb without tearing the edges of the polyacrylamide wells. After the comb is removed, rinse wells with 1x SDS electrophoresis buffer.

The rinse removes unpolymerized monomer; otherwise, the monomer will continue to polymerize after the comb is removed, creating uneven wells that will interfere with sample loading and subsequent separation.

14. Using a Pasteur pipet, fill the wells with 1x SDS electrophoresis buffer.

If well walls are not upright, they can be manipulated with a flat-tipped needle attached to a syringe.

15. Attach gel sandwich to upper buffer chamber using manufacturer's instructions.



16. Fill lower buffer chamber with the recommended amount of 1x SDS electrophoresis buffer.

17. Place sandwich attached to upper buffer chamber into lower buffer chamber.

18. Partially fill the upper buffer chamber with 1x SDS electrophoresis buffer so that the sample wells of the stacking gel are filled with buffer.

Monitor the upper buffer chamber for leaks and if necessary, reassemble the unit. A slow leak in the upper buffer chamber may cause arcing around the upper electrode and damage the upper buffer chamber.

19. Using a 25- or 100- $\mu$ l syringe with a flat-tipped needle, load the protein sample(s) into one or more wells by carefully applying the sample as a thin layer at the bottom of the wells. Load control wells with molecular weight standards. Add an equal volume of 1x SDS sample buffer to any empty wells to prevent spreading of adjoining lanes.

Preparing the samples at approximately the same concentration and loading an equal volume to each well will ensure that all lanes are the same width and that the proteins run evenly. If unequal volumes of sample buffer are added to wells, the lane with the larger volume will spread during electrophoresis and constrict the adjacent lanes, causing distortions.

The samples will layer on the bottom of the wells because the glycerol added to the sample buffer gives the solution a greater density than the electrophoresis buffer. The bromphenol blue in the sample buffer makes sample application easy to follow visually.

20. Fill the remainder of the upper buffer chamber with additional 1x SDS electrophoresis buffer so that the upper platinum electrode is completely covered. Do this slowly so that samples are not swept into adjacent wells.

Run the gel

21. Connect the power supply to the cell and run at 10 mA of constant current for a slab gel 0.75 mm thick, until the bromphenol blue tracking dye enters the separating gel. Then increase the current to 15 mA.

For a standard 16-cm gel sandwich, 4 mA per 0.75-mm-thick gel will run ~15 hr (i.e., overnight); 15 mA per 0.75-mm gel will take 4 to 5 hr. To run two gels or a 1.5-mm-thick gel, simply double the current. When running a 1.5-mm gel at 30 mA, the temperature must be controlled (10-20C) with a circulating constant-temperature water bath to prevent "smiling" (curvature in the migratory band). Temperatures <5C should not be used because SDS in the running buffer will precipitate. If the level of buffer in the upper chamber decreases, a leak has occurred.

22. After the bromphenol blue tracking dye has reached the bottom of the separating gel, disconnect the power supply.



## **Disassemble the gel**

**23. Discard electrode buffer and remove the upper buffer chamber with the attached gel sandwich.**

**24. Orient the gel so that the order of the sample wells is known, remove the sandwich from the upper buffer chamber, and lay the sandwich on a sheet of absorbent paper or paper towels.**

**25. Carefully slide one of the spacers halfway from the edge of the sandwich along its entire length. Use the exposed spacer as a lever to pry open the glass plate, exposing the gel.**

**26. Carefully remove the gel from the lower plate. Cut a small triangle off one corner of the gel so the lane orientation is not lost during staining and drying. Proceed with protein detection.**

**The gel can be stained with Coomassie blue or silver or proteins can be electroeluted.**

#### **4.30 ALTERNATIVE PROTOCOL 1: ELECTROPHORESIS IN TRIS-TRICINE BUFFER SYSTEMS**

Separation of peptides and proteins under 10 to 15 kDa is not convenient in the traditional Laemmli discontinuous gel system . This is due to the comigration of SDS and smaller proteins, obscuring the resolution. Two approaches to obtain the separation of small proteins and peptides in the range of 5 to 20 kDa are presented: this Tris-tricine method that follows and a system using increased buffer concentrations. The Tris-tricine system uses a modified buffer to separate the SDS and peptides, thus improving resolution. Several precast gels are available for use with the tricine formulations.

##### **Additional Materials**

Separating and stacking gel solutions

2x tricine sample buffer

Peptide molecular-weight-standards mixture

Cathode buffer

Anode buffer

Coomassie blue G-250 staining solution

10% (v/v) acetic acid

##### **Method**

1. Prepare and pour the separating and stacking gels.



2. Prepare the sample but make the following changes for tricine gels. Substitute 2x tricine sample buffer for the 2x SDS sample buffer. Dilute an aliquot of the protein or peptide sample to be analyzed 1:1 (v/v) with 2x tricine sample buffer. Treat the sample at 40C for 30 to 60 min prior to loading.

If proteolytic activity is a problem, heating samples to 100C for 3 to 5 min before loading the wells may be required. Use the peptide molecular-weight-standards mixture for peptide separations .

3. Load the gel and set up the electrophoresis apparatus with the following alterations. Remove comb and, using the tricine-containing cathode buffer, or water, rinse once and fill wells. Fill the lower buffer chamber with anode buffer, assemble the unit, and attach the upper buffer chamber. Fill the upper buffer chamber with cathode buffer and load the samples.

4. Connect the power supply to the cell and run 1 hr at 30 V (constant voltage) followed by 4 to 5 hr at 150 V (constant voltage). Use heat exchanger to keep the electrophoresis chamber at room temperature.

5. After the tracking dye has reached the bottom of the separating gel, disconnect the power supply.

Coomassie blue G-250 is used as a tracking dye instead of bromphenol blue because it moves ahead of the smallest peptides.

6. Disassemble the gel. Stain proteins in the gel for 1 to 2 hr in Coomassie blue G-250 staining solution. Follow by destaining with 10% acetic acid, changing the solution every 30 min until background is clear (3 to 5 changes). For higher sensitivity, use silver staining as a recommended alternative.

Prolonged staining and destaining will result in the loss of resolution of the smaller proteins (<10 kDa). Proteins diffuse within the gel and out of the gel, resulting in a loss of staining intensity and resolution.

### Materials

Anode buffer

121.1 g Tris base

500 ml H<sub>2</sub>O

Adjust to pH 8.9 with concentrated HCl

Dilute to 5 liters with H<sub>2</sub>O

Store at 4C up to 1 month

Final concentration is 0.2 M Tris-Cl, pH 8.9.

Cathode buffer

12.11 g Tris base

17.92 g tricine

1 g SDS

Dilute to 1 liter with H<sub>2</sub>O



**Do not adjust pH**

**Store at 4C up to 1 month**

**Final concentrations are 0.1 M Tris, 0.1 M tricine, and 0.1% (w/v) SDS.**

**Coomassie blue G-250 staining solution**

**200 ml acetic acid**

**1800 ml H<sub>2</sub>O**

**0.5 g Coomassie blue G-250**

**Mix 1 hr and filter (Whatman no. 1 paper)**

**Store at room temperature indefinitely**

**The final solution is 0.025% (w/v) Coomassie blue G-250 in 10% (v/v) acetic acid.**

**SDS electrophoresis buffer, 5x**

**15.1 g Tris base**

**72.0 g glycine**

**5.0 g SDS**

**H<sub>2</sub>O to 1000 ml**

**Dilute to 1x or 2x for working solution, as appropriate**

**Do not adjust the pH of the stock solution, as the solution is pH 8.3 when diluted.**

**Store at 4C until use (up to 1 month).**

Tricine sample buffer, 2x

2 ml 4x Tris-Cl/SDS, pH 6.8 (Table 10.2.1)

2.4 ml (3.0 g) glycerol

0.8 g SDS

0.31 g DTT

2 mg Coomassie blue G-250

Add H<sub>2</sub>O to 10 ml and mix

Final concentrations are 0.1 M Tris, 24% (w/v) glycerol, 8% (w/v) SDS, 0.2 M DTT, and 0.02% (w/v) Coomassie blue G-250.

#### **4.31 ALTERNATIVE PROTOCOL: RAPID COOMASSIE BLUE STAINING**

Protein bands stained using this protocol can be detected within 5 to 10 min after adding rapid Coomassie staining solution. Because the Coomassie blue concentration is lower than that used in the basic protocol, the gel background never stains very darkly and the bands can be seen even while the gel remains in the staining solution. Another difference from the basic protocol is that isopropanol is substituted for methanol in the fixing and destaining solutions. This method is slightly less sensitive than the basic protocol.

##### **Additional Materials**



**Isopropanol fixing solution**

**Rapid Coomassie staining solution**

**10% acetic acid**

**Method**

1. Place the polyacrylamide gel in a plastic or glass container. Cover the gel with isopropanol fixing solution and shake gently at room temperature. For a 0.7-mm-thick gel, shake 10 to 15 min; for a 1.5-mm thick gel, shake 30 to 60 min.

2. Pour out fixing solution. Cover the gel with rapid Coomassie blue staining solution and shake gently until desired intensity is reached, 2 hrs to overnight at room temperature.

Bands will become visible even in the staining solution within 5 to 30 min, depending on gel thickness. The gel background will never stain very darkly.

3. Pour out staining solution. Cover the gel with 10% acetic acid to destain, shaking gently 2 hr at room temperature until a clear background is obtained.

4. If necessary, pour out 10% acetic acid and add more. Continue destaining until clear background is obtained. Store gel in 7% acetic acid or water, or in plastic wrap at 4C.

It is usually unnecessary to add additional destaining solution.

5. If desired, photograph or dry the gel as in step 7 of the basic protocol.

See steps 6 and 7 above concerning photographing or drying the gel.

#### **Coomassie blue staining solution**

50% methanol (v/v)

0.05% (w/v) Coomassie brilliant blue R-250 (Bio-Rad or Pierce)

10% (v/v) acetic acid

40% H<sub>2</sub>O

Dissolve the Coomassie brilliant blue R in methanol before adding acetic acid and water. Solution can be stored for 6 months. If precipitate is observed following prolonged storage, filter to obtain a homogeneous solution.

#### **Destaining solution**

5% methanol

7% acetic acid

88% H<sub>2</sub>O

Can be stored ~1 month at room temperature

### **4.32 SILVER STAINING**

Detection of protein bands in a gel by silver staining depends on binding of silver to various chemical groups (e.g., sulfhydryl and carboxyl moieties) in proteins. The



detection limit is 2 to 5 ng/protein band. In this procedure, proteins separated in a polyacrylamide gel are successively fixed in methanol/acetic acid and glutaraldehyde. After exposure to silver nitrate, the gel is treated with developer to control the level of staining. When the desired staining intensity is reached, the gel is fixed, photographed, and dried.

### Materials

#### Fixing and destaining solutions

10% (v/v) glutaraldehyde (freshly prepared from 50% stock; Kodak #1200534)

Silver nitrate solution

Developing solution

Kodak Rapid Fix Solution A (#8323917)

### Method

1. Place the polyacrylamide gel in a plastic container and add 5 gel volumes of fixing solution. Agitate slowly 30 min on an orbital shaker.
2. Pour out fixing solution. Fix the gel with 5 gel volumes of destaining solution for 60 min, agitating slowly.

No destaining is taking place in this step; fixation continues using the same solution as was used for destaining in the Coomassie blue staining protocol.

3. Pour out destaining solution. Add 5 gel volumes of 10% glutaraldehyde and agitate slowly 30 min in a fume hood.

4. Pour out the glutaraldehyde. Wash the gel 4 times with water, 30 min for each wash and preferably overnight for the last wash. Agitate slowly with each wash.

Washing is done by adding the water, pouring it off, and adding more water.

5. Pour out the water. Stain the gel with ~5 gel volumes of silver nitrate solution (to cover the gel) for 15 min with vigorous shaking.

6. Transfer the gel to another plastic box and wash 5 times with deionized water, exactly 1 min for each wash. Agitate slowly with each wash.

7. Dilute 25 ml developing solution with 500 ml water. Transfer gel to another plastic box, add enough diluted developer to cover the gel during agitation, and shake vigorously until the bands appear as intense as desired. If the developer turns brown, change to fresh developer.

Development should be stopped immediately when gel background starts to appear.

8. Transfer to Kodak Rapid Fix Solution A for 5 min.

If necessary, swab gel surface with soaked cotton to remove residual silver deposits.



9. Pour off Rapid Fix Solution and wash the gel exhaustively in water (4 to 5 times).

10. Photograph the gel.

It is useful to use a blue-green filter such as a Wratten #58 filter with Kodak T-Max 100. Gels should be photographed as soon as possible because there may be slight changes in color intensity and increases in nonspecific background. The silver-stained proteins remain clearly visible for at least 18 hr.

11. Dry the gel to maintain a permanent record as in step 7 of the first basic protocol or store in sealable plastic bag (will last 6 to 12 months).

#### **4.33 PURIFICATION OF SRY BOXES.**

**Day 1**

- Inoculate BL21(DE3) containing pT7 hSRY box in 10 ml of LB + 100ug/ml ampicillin + 0.4% glucose. Grow o.n. at 37 degrees.

**Day 2**

- Inoculate 1.5 ml of o.n. culture in one 1 liter flask containing 150 ml of LB + 100 ug/ml ampicillin. Grow at 37 degrees with strong agitation.

- Prepare mono S FPLC column (Pharmacia) : wash 2X with program 2/3 and buffers A and B. Pressure limit 2.0 Mpa.

- When A600=0.6 -0.9 add 1 mM IPTG. Grow at 37° C for 90-100 min (see results formore details).
- Centrifuge cells in GSA rotor 6K rpm for 10 min. at 4° C. Resuspend pellet in 7.5 ml buffer L2 and transfer in 50 ml Nalgene tubes.
- Sonicate the suspension until it clarifies partially. Keep the tube in ice and apply a maximum of 8 cycles of 20 seconds (power 100 W); lower amounts of cells require fewer cycles.
- Add 1.5 ml of NaCl 5 M and mix thouroughly.
- Add 7.5 ml of a 50% v/v suspension of DEAE-sephadex in buffer L2 (ice temperature). Mix well and keep on ice 5 min.
- Spin at 10K rpm 10 min. in SS-34 rotor at 4° C. Discard pellet and transfer the supernatant in a new Nalgene tube.
- Add to the supernatant solid ammonium sulphate to final 70% solubility. Let stand the tube on ice 20 min. Centrifuge at 10 Krpm 10 min in SS-34 rotor at 4° C.
- Discard the pellet; tranfer the supernatant in to a 50ml Falcon tube.
- Load the supernatant in a superloop of 50 ml. Run column with program 2/2. Select the fraction size of 1 ml.
- At the end of the column run take tubes and transfer to fridge.

### Materials

buffer L2

50 mM Tris-HCl pH 8.0

20 mM EDTA

0.5mM PMSF

buffer A



20 mM HEPES pH 7.9

0.5 mM DTT

0.2 mM EDTA

buffer B

20 mM HEPES pH 7.9

0.5 mM DTT

0.2 mM EDTA

2M NaCl

program 2/3

|                |      |
|----------------|------|
| 0.0 conc % B   | 0    |
| 0.0 ml/min     | 0.25 |
| 0.0 valve pos  | 2.1  |
| 0.0 valve pos. | 1.1  |
| 0.0 ml/mark    | 1.0  |
| 0.0 cm/ml      | 0.25 |
| 0.0 port set   | 6.0  |
| 10 conc %B     | 100  |
| 15 conc B      | 100  |

program 2/2

|                |      |
|----------------|------|
| 0.0 conc % B   | 0    |
| 0.0 ml/min     | 0.25 |
| 0.0 valve pos  | 2.1  |
| 0.0 valve pos. | 1.1  |
| 0.0 ml/mark    | 1.0  |
| 0.0 cm/ml      | 0.50 |
| 2.0 valve pos  | 1.2  |
| 2.0 ml min.    | 0.2  |
| 2+x valve pos  | 1.1  |
| 4+x conc%B     | 100  |
| 4+x port set   | 6.1  |
| 14+x conc %B   | 100  |
| 17+x conc %B   | 100  |

#### **4.34 MOBILITY SHIFT DNA-BINDING ASSAY USING GEL ELECTROPHORESIS.**

##### **INTRODUCTION**

The DNA-binding assay using mobility shift polyacrylamide gel electrophoresis (PAGE) is a simple, rapid, and extremely sensitive method for the detection of sequence-specific DNA-binding proteins in crude extracts. This assay also permits the quantitative determination of the affinity, abundance, association rate constants, dissociation rate constants, and binding specificity of DNA-binding proteins. Proteins that bind specifically to an end-labeled DNA fragment retard the mobility of the fragment during electrophoresis, resulting in discrete bands corresponding to the individual protein-DNA complexes.

The protocol can be divided into four stages: (1) an end-labeled DNA probe containing a particular protein binding site is prepared; (2) a low-percentage, low-ionic-strength polyacrylamide gel is prepared; (3) a protein mixture is bound to the DNA probe; and (4) the binding reactions are electrophoresed through the gel, which is then dried and autoradiographed.

##### **MOBILITY SHIFT ASSAY USING LOW-IONIC-STRENGTH PAGE**

###### **Materials**



DNA with desired binding site (cruciform DNAs, cD3e, mut11 or mut 0 probes).

TE buffer

4x binding buffer

TBE running buffer

30% ammonium persulfate

TEMED (N,N,N',N'-tetramethylethylenediamine)

DNA-binding protein

Whatman 3MM or equivalent filter paper

### Method

Prepare the DNA probe. Digest, purify and label as described before.

. Prepare enough 0.5X TBE electrophoresis buffer to fill the tank.

. Assemble washed 20-cm-long glass plates and 0.8-mm spacers for casting the gel.

Use siliconized glass plates to aid removal of the gel after electrophoresis. Wash away all traces of detergent since this will disrupt protein-DNA interactions.

. Add 100  $\mu$ l of 30% ammonium persulfate and 34  $\mu$ l TEMED to 40 ml low-ionic-strength gel mix. Swirl gently to mix.

A large volume of premixed low-ionic-strength gel mix can be prepared and stored for several months at 4C. Allow the gel mix to reach room temperature before adding ammonium persulfate and TEMED. It is not necessary to degas the gel mix.

. Pour the gel mix between the plates and insert a comb. For optimal results, use a comb with teeth that are 4 mm wide.

The polyacrylamide gel contains an 30:1 ratio of acrylamide to bisacrylamide.

. Allow the gel to completely polymerize for 20 min. Remove the comb and bottom spacer and attach the plates to the electrophoresis tank after filling the lower reservoir with low-ionic-strength electrophoresis buffer. Fill the upper reservoir of the tank with low-ionic-strength electrophoresis buffer. With a bent-needle syringe, remove air bubbles trapped beneath the gel and flush out the wells

. Pre-run the gel at 100 V for at least 90 min.

At 100 V the gel should initially draw a current of ~22 mA, and should decrease to ~18 mA after 90 min.

**Prepare the binding buffer 4x:**

20% Ficoll, 400 mM NaCl, 40mM Hepes pH 7.9, bromphenol blue, xylene cyanol.

. In a microtiter combine the following

10,000 cpm DNA probe

~4/15 ng of protein from a pure fraction.

2.5 ul of buffer 4x



The final reaction volume is 10 ul.

Binding reactions must contain ~5% Ficoll to minimize mixing of the sample with the gel electrophoresis buffer and to ensure that the samples sink to the bottom of the wells. While some mixing occurs during loading, this is usually not a problem since the electrophoresis buffer is of lower ionic strength than the reaction buffer. This increases the stability of most nucleic acid-protein interactions and results in even longer-lived complexes.

. Incubate the binding reaction mix 5 min in ice.

#### Run the gel

. Load a small volume of 10x loading buffer into one of the wells. Allow the dyes to run into the gel and flush the wells before loading the samples.

The dyes can be used to monitor the progress of electrophoresis.

. Load each binding reaction into the appropriate well with a 10-ul glass capillary pipet and a Clay-Adams screw-top loader or with a pipettor.

. Electrophorese at ~30 to 35 mA until the bromphenol blue approaches the bottom of the gel (~2.5 to 3 hr).

If electrophoresis is performed at room temperature, the glass plates should be allowed to become only slightly warm. Decrease the voltage if the plates become any hotter. For probes <70 bp, do not run the bromphenol blue to the bottom of the gel. To run the gel faster, put the apparatus in a cold room. Higher voltages may then be used without heating the glass plates. In addition, colder temperatures cause a contraction of the gel, increasing its sieving properties. As a result, protein-DNA complexes may appear as sharper bands.

. Remove the glass plates from the gel box and carefully remove the side spacers.

. Using a spatula, slowly pry the glass plates apart, allowing air to enter between the gel and the glass plate. The gel should remain attached to only one of the plates.

Prying the plates apart too quickly may tear the gel or cause it to stick to both plates. If this occurs or if the gel has become distorted, squirt a stream of water underneath it. This will reduce the stickiness of the gel. Be careful not to let the gel slide off the plate.

. Lay the glass plate (with the gel attached) on the bench with the gel facing up. Place 3 sheets of Whatman filter paper cut to size on top of the gel.

. Support both sides with your hands and carefully flip the sandwich over so that the Whatman paper is on the bottom and the glass plate is on the top.



. Carefully lift up one end of the glass plate. Peel the Whatman paper with the gel attached to it from the plate.

. Cover the gel with plastic wrap and dry under vacuum.

. Autoradiograph the dried filter. Visualize the protein-DNA complex after overnight exposure without an intensifying screen or after 5-hr exposure with an intensifying screen.

Low-ionic-strength electrophoresis buffer

0.5 X T.B.E.

Low-ionic-strength electrophoresis gel

8% Acrylamide/Bisacrylamide (30/1) in 0.5 X TBE.

Filter through 0.4-um filter

Store at 4C for several months

#### **4.35 MAKING THE LABELED CROSS**

Four-way junction *c* is composed of four strands of 30, 35, 40 and 46 nucleotides. As controls for structure-specific binding, two linear duplex DNAs were used, called *a* and *b*. Details on the construction of these molecules are given in Bianchi et al. (1989).

Four-way junction *z* is composed of the four 30-mer oligos

1) AGCGCTCTCACACGGGCCTCCGCCAGCTG,

2) CAGCTGGGCGGAGGGCGGACGTTAACCCC,

3) GGGGTTAACGTCCGCGGTAATCTGGTAGA,

4) TCTACCAGATTACCCCGTGTGAGAGCGT. Control duplex az was constructed by annealing oligo 1 with its complement oligo 5 (CAGCTGGGCGGAGGCCCGTGTGAGAGCGCT); the sequences contained in the resulting duplex were shown to be poor binding sites for SRY (Harley et al., 1992). Control duplex bz was constructed by annealing oligo 2 with its complement; the resulting duplex is identical to the allmut probe, which was shown previously to be a poor binding sites for SRY (Harley et al

#### Mix

20 pmoles of oligo 1

2 ul of kinase bf 10 X

12.5 ul of <sup>32</sup>P-ATP(5000 uCi/mMmol)

1.5 ul of T4 polynucleotide Kinase

water to 20 ul.

incubate 1 hr at 37° C.

#### Add

30 pmoles of oligo arm 2, 3, 4.

water to 30 ul.

Put the mix in a dry block at 100° C. To allow annealing, switch the mix off the dry block, cover it with a lid of polystyrene foam and let it cool until 30° C. Remove and throw away condensed vapour from the inner cap.



Make a 6.5 % polyacrylamide gel in 0.5 X TBE, 1 mm thick. Prerun at RT, 250 V 2 hours.

Apply cruciform mix to 2 wells of prerun gel. Run at RT, 250 V for 1.5-2 hours.

Cruciform structure will be at the level of xylene cyanol or slightly higher.

Wrap gel in Saran wrap. Autoradiograph it for 1 min. with Fuji film. Mark the film with a felt tip pen; markings should extend outside the film into the Saran wrap around the gel.

Cut out the band corresponding to cruciform structure associated to high level of radioactivity and put gel slice in Eppendorf. Autoradiograph again as a check.

Pierce a hole with a needle in the bottom of the eppendorf tube containing the gel slice. Put into a second eppendorf tube and spin a few seconds in centrifuge at 4° C.

Add to the ground PA gel in the lower tube 600 ul of TE. On a rotary wheel shake the tube at 4° C for 4 hours. After recover supernatant and pool. Filter the supernatant through glass wool in a pierced eppendorf. Filter at 4° C. Assuming the recovery is 50%, the final cruciform concentration will be 10 nM or 10 fmoles/ul.

#### **4.36 CIRCULAR PERMUTATION ASSAY.**

Plasmids pBend2 containing DNA sequences CD3e, MUT11 and MUT0 were prepared by insertion of annealed synthetic oligonucleotides (CTAGAGAGCGCTTTGTTCTCAG and TCGACTGAGAACAAAGCGCTCT for pBend2CD3e, CTAGAGAGCGCATTGTTATCAG and TCGACTGATAACAATGCGCTCT for pBend2MUT11, CTAGAGAGCGCTGTGTTCTCAG and TCGACTGAGAACACAGCGCTCT for pB2MUT0) between the *Xba*I and *Sal*I restriction sites in plasmid pBend2.

Preparation of probes, electrophoresis and data analysis were carried out as previously reported

#### **4.37 CALCULATION OF DNA BEND PARAMETERS**

For circular permutation analysis, the mobilities of protein-DNA complexes were normalized to the mobility of free DNA ( $R_{\text{bound}}/R_{\text{free}}$ , vertical axis of the graph in the figures of results). The distances between the 5' end of the probe and the apparent center of flexure were normalized to the total length of the probe (flexure displacement, horizontal axis of the graph in the figures of results). The points in the graph were interpolated with a second-order equation (a parabola) by means of least squares algorithm (application Cricket Graph on a Macintosh computer).

To analyze the electrophoretic mobilities of the protein-DNA complexes, we adopted the Lumpkin-Zimm reptation model, in which the DNA chains migrate in wormlike fashion among the gel fibers. The DNA chains are confined to a tube, composed of a sequence of segments which connect the consecutive points of contact between the DNA and the gel fibers. The mobility of the chain,  $R$ , is proportional to the center-of-mass velocity,  $v_{\text{cm}}$ , of the chain in the direction of the electric field, such that:

$$R = \langle v_{\text{cm}} \rangle / E = \langle h_x^2 / L^2 \rangle Q / z \quad (\text{equation 1, Levene and Zimm, 1989})$$

where the field of strength  $E$  is along the  $x$  axis,  $Q$  is the total charge of the DNA,  $z$  is the friction constant for motion along the tube,  $h_x$  is the component in the  $x$  direction of the tube's end-to-end vector,  $L$  is the contour length of the tube, and the angle brackets denote an average over an ensemble of conformations. In this model,  $\langle h_x^2 / L^2 \rangle$  can be smaller than unity for two reasons: the introduction of a fixed, oriented bend in the DNA molecule, or the presence of an ensemble of non-fixed,



non-oriented bends, such as those allowed by a loose hinge. Levene and Zimm (1989) have computed with Monte Carlo simulations  $R$  and  $\langle hx^2/L^2 \rangle$  for straight and bent chains, and have compared them to the experimental electrophoretic mobilities of bent DNA molecules. They found that the results are closely approximated by the model if one introduces an additional, independently adjustable elastic force constant,  $B_{eff}$ , which accounts for the relative deformability of gel and DNA. In our analysis, we have drastically simplified the mathematical complexity of the real situation and of the Levene-Zimm model by adopting the following assumptions: (1) for circularly permuted DNA chains of fixed length,  $Q$  and  $z$  are constant; (2) the flexure introduced in the DNA chain by the binding of the protein is an angle,  $q$ , with a well-identified vertex and a fixed amplitude (as opposed to a continuous bend and a dynamically averaged ensemble  $\langle q \rangle$  of amplitudes) and in addition (3) due to the limited length of our probes and the ionic strength of the gel system, DNA chains are essentially rigid, so that  $\langle hx^2/L^2 \rangle$  can be approximated by  $hx^2/L^2$ ; (4)  $B_{eff}$  does not vary substantially with the displacement of the angle along the circularly permuted DNA molecules. Condition 2 is probably the most critical, since situations have been found in which the protein-induced flexure resembles more closely a loose hinge than a rigid angle (Gartenberg and Crothers, 1988; Kerrpola and Curran, 1991a). Conditions 2 and 4 can be partially checked by running the complexes in gels of different polyacrylamide concentration: the calculated angle amplitudes should not vary significantly. Under the conditions specified, equation 1 reduces to:  $R_{bound}/R_{free} = k_{bound} (hx_{bound}^2/L^2) / k_{free} (hx_{free}^2/L^2) = K hx_{bound}^2/L^2$  (equation 2)

where  $k$  (and hence  $K$ ) are constants, and  $hx_{free} = L$  for a straight rod. How does  $hx_{bound}$  depend on  $D$ , the distance of the vertex of the angle  $q$  from the 5' end of the

DNA molecule ? In a triangle, the length of the three sides a, b and c and the angle g subtended by a and b are related by the formula

$$c^2 = a^2 + b^2 - 2ab \cos g \quad (\text{equation 3}).$$

In our model:

$$h_{\text{bound}}^2 = D^2 + (L-D)^2 - 2D(L-D) \cos q \quad (\text{equation 4}).$$

Substituting equation 4 in equation 2 yields

$$\begin{aligned} R_{\text{bound}}/R_{\text{free}} &= K [D^2 + (L-D)^2 - 2D(L-D) \cos q] / L^2 = \\ &= 2K(1+\cos q)(D/L)^2 - 2K(1+\cos q)(D/L) + K \end{aligned} \quad (\text{equation 5}).$$

Thus in our model  $R_{\text{bound}}/R_{\text{free}}$  is a quadratic function of  $D/L$ , and the experimental values for  $R_{\text{bound}}/R_{\text{free}}$  can be interpolated by a parabola, whose minimum identifies the locus of flexure. In addition, the amplitude of  $q$  can be readily derived from the parameters for the second-order and first-order terms of the equation, both equal to  $2K(1+\cos q)$ ;  $K$  is the zero-order parameter of the same equation. In fact, the comparison of the two estimates of the angle  $q$  derived from the first- and second-order parameters is a good test of the model.

A similar geometrical treatment of the problem was used by Thompson and Landy (1988). They derived the formula  $\mu_M/\mu_E = \cos a/2$ , which relates the angle of deviation from linearity of the DNA,  $a$ , to the relative mobilities of complexes with a flexure exactly in the middle,  $\mu_M$ , or at the end of the molecule,  $\mu_E$ . Empirically, the formula  $\mu_M/\mu_E = \cos a/2$  was found to give a good fit to observed values for  $a$  angles between 0 and 140 degrees ( $q$  angles between 180 and 40). We find that our solution gives results numerically similar to Thompson and Landy's and is more accurate and robust, since it considers many data points rather than two ( $\mu_M$  and  $\mu_E$ ).



#### **4.38 DETERMINATION OF COMPLEX DISSOCIATION CONSTANTS.**

Wild type and mutant SRY boxes were titrated into binding mixtures containing a fixed amount (0.25 nM) of labelled cruciform z or of labelled fragment D from pBend2CD3e, pBend2CD3eMUT0 or pBend2CD3eMUT11. Samples were electrophoresed at 4C in 10% polyacrylamide gels as described. The radioactivity present in the bands was measured by exposing the wet gel to PhosphorImager screens (Molecular Dynamics). Under conditions of protein excess, the dissociation constant is equivalent to the concentration of polypeptide where half of the input DNA is taken up in the complex (see Results).

#### **4.39 DETERMINATION OF PROTEIN-DNA SEQUENCE SPECIFICITY BY PCR-ASSISTED BINDING-SITE SELECTION.**

Binding-site selection is used to determine the target specificity of a sequence-specific DNA-binding protein. The technique has a number of applications, ranging from identifying DNA target sequences for proteins with unknown DNA-binding specificities to providing additional information on the protein-DNA interactions of previously characterized DNA binding domains.

As indicated in the protocol, a pool of random-sequence oligonucleotides is used as the source of potential binding sites . The oligonucleotide pool is made double stranded, labelled with  $^{32}\text{P}$  and incubated in a binding reaction containing the DNA-binding protein of interest. Protein-DNA complexes are isolated by electrophoresis in a vertical 7% polyacrylamide gel in 0.5X TBE and electrophoresed at 11 V cm<sup>-1</sup> at

room temperature. The gel is autoradiographed with Kodak XAR-5 film at 4C for a few hours. Bound oligonucleotides are recovered, amplified by the polymerase chain reaction (PCR), and used as input DNA for a further round of binding, recovery, and amplification. After four rounds of selection, progress of the procedure is monitored again by mobility shift analysis of the selected oligonucleotide pools. After five rounds, individual binding site isolated from the appropriate complex on a mobility shift gel are cloned into plasmids and examined by sequencing.

### Materials

Random-sequence oligonucleotide R60-mer template,

5'CTGGTCGGGTGAATTCGTGTCGTGG(A/G/C/T)<sub>10</sub>CCGACCCAGCGAATTC  
AGAGCATG;

upper primer F, CTGGTCGGGTGAATTCGTGTCGTGG;

lower primer R, ACATGCTCTGAATTCGCTGGGTCGG.

10x Taq DNA polymerase buffer

0.5 mM 3dNTP mix (minus dCTP)

40 uM and 0.5 mM dCTP

10 mCi/ml [ $\alpha$ -<sup>32</sup>P]dCTP (800 Ci/mmol)

5 U/ul Taq DNA polymerase

Elution buffer (see recipe)

Glycogen carrier (e.g., Boehringer Mannheim)

TE buffer, pH 7.5 to 8.0

Binding buffer (see recipe), with and without 50 ug/ml BSA

Carrier DNA: e.g., 100 ng/ul

Purified DNA-binding proteins



**Scintillation vials and counter**

**17 x 100-mm polystyrene centrifuge tube with snap-cap**

**Tumbler or rotating wheel**

**Whatman 3MM paper**

**Additional reagents and equipment :**

**phenol/chloroform extraction and ethanol precipitation , and mobility shift DNA-binding assay**

**Prepare labeled double-stranded random-sequence oligonucleotide**

### **Method**

**1. Use Purified oligonucleotides R60, primer F, and primer R. Dilute R60 to 50 ng/ul and primers F and R to 80 ng/ul in water.**

**2. Set up the following reaction in a 0.5-ml microcentrifuge tube (20 ul total):**

**2 ul 50 ng/ul oligonucleotide R76**

**2 ul 10x Taq DNA polymerase buffer**

**2 ul 0.5 mM 3dNTP mix (minus dCTP)**

**2 ul 40  $\mu$ M dCTP**

**2 ul 80 ng/ul primer F**

**2 ul 10 Ci/ml (3000 Ci/mmol) [ $\alpha$ -<sup>32</sup>P]dCTP**

**7 ul H<sub>2</sub>O**

1 ul 5 U/ul Taq DNA polymerase.

3. Carry out one PCR cycle using the following cycling parameters:

1 min 94C (denaturation)

3 min 55C (annealing)

9 min 72C (extension)

Chase extension by adding 2 ul of 0.5 mM dCTP, then heating 9 min at 72C.

This reaction will generate double-stranded R60 oligonucleotide (dsR60) labeled to a specific activity of 3200 Ci/mmol, which is appropriate for use in a gel mobility shift assay .

Probes with higher specific activity can be generated by increasing the ratio of labeled to unlabeled dCTP in the reaction. R60 may also be rendered double-stranded by annealing primer F and extending with Klenow fragment of E. coli DNA polymerase I

4. Purify dsR60 on an 8% nondenaturing polyacrylamide gel visualizing it by autoradiography

A clearly visible dsR60 band can be obtained after a 60- to 90-sec exposure.

5. Excise gel slice containing labeled dsR60 using a clean scalpel and place in a 1.5-ml microcentrifuge tube containing 250 ul elution buffer. Incubate overnight at 37C.



6. Remove elution buffer to a fresh 1.5-ml microcentrifuge tube, add 1 ug of glycogen carrier, and ethanol precipitate.

7. Resuspend pellet in 10 ul TE buffer and measure 1 ul into a scintillation vial. Measure Cerenkov counts in a scintillation counter to determine cpm.

Because the molecular weight and specific activity of the probe oligonucleotide are known, the amount of dsR60 can be quantitated by assuming that 106 cpm is approximately equivalent to 1 uCi. The double-stranding reaction in step 2 and the PCR amplification reaction in step 22 contain [ $\alpha$ -<sup>32</sup>P]dCTP and unlabeled dCTP at concentrations such that four labeled C nucleotides are incorporated into every dsR60 oligonucleotide. The specific activity of dsR76 is therefore four times the specific activity of the [ $\alpha$ -<sup>32</sup>P]dCTP itself ( $4 \times 3000 \text{ Ci/mmol} = 12000 \text{ Ci/mmol}$ ).

8. Prepare a 0.2 ng/ul dilution of dsR60 in TE buffer for use in binding reaction

9. Incubate probes with proteins as describe for band shift assay.

10. Apply the binding mix to a 7% acrylamide gel, and allow it to run for two hours at 30 mAs.

11. Purify dsR60-protein complex visualizing it by autoradiography

A clearly visible dsR60 band can be obtained after two hours of exposure.

12. Excise gel slice containing labeled dsR60 using a clean scalpel and place in a 1.5-ml microcentrifuge tube containing 250 ul elution buffer. Incubate overnight at 37C.

13. Remove elution buffer to a fresh 1.5-ml microcentrifuge tube, perform a phenol extraction followed by a chloroform extraction, add 1 ug of glycogen carrier, and ethanol precipitate.

Determination of the proportion of the input probe associated to the protein compared to the unbound DNA allows quantitation of selected DNA. In addition, the proportion of input DNA recovered after each round of selection often gives a good indication of the progress of the site selection.

#### Amplify selected DNA

14. Prepare the following reaction mixture (20 ul total):

2 ul 10x Taq DNA polymerase buffer

3.2 ul 0.5 mM 3dNTP mix (minus C)

2 ul 40 uM dCTP

2 ul 80 ng/ul primer F

2 ul 80 ng/ul primer R

1 u 10 mCi/ml (800 Ci/mmol) [ $\alpha$ -32P]dCTP

7.3 ul H<sub>2</sub>O

0.5 ul 5 U/ul Taq DNA polymerase.



Add this mixture to 1 pg of selected DNA in a 0.5-ml microcentrifuge tube.

15. Carry out 30 of the following radioactive PCR amplification cycles:

1 min 94C (denaturation)

1 min 55C (annealing)

30 sec. 72C (extension)

This step is optimized for amplification of oligonucleotide R60 with primers F and R. It is important to carefully calibrate the amplification reaction with respect to the amount of input DNA and number of PCR cycles performed particularly when using different random oligonucleotide/primer combinations.

16. Dilute PCR reaction to 150 ul with TE buffer and phenol extract

17. Add 1 ug of glycogen carrier and 1/4 vol of 5 M ammonium acetate (to 1 M).

Ethanol precipitate

18. Purify PCR product by electrophoresis on a nondenaturing polyacrylamide gel as described in steps 4 to 6.

Because maximum recovery is not absolutely essential, the elution step can be cut to 2 hr at 45C.

19. Following ethanol precipitation, resuspend pellet in 10 ul TE buffer and measure 1 ul for Cerenkov counts. Quantitate selected, amplified dsR76, then resuspend at 0.1 ng/ul.

This procedure typically yields 10 to 20 ng labeled amplified selected dsR60 oligonucleotide.

20. Use the selected and amplified oligonucleotide pool in another round of binding-site selection. After five selection cycles, monitor the success of the binding-site selection procedure with a band shift assay.

If successful, four rounds of selection should yield oligonucleotide pools that are capable of forming visible and abundant protein-DNA complexes on a mobility shift gel

Controls are critical, and should include binding reactions in which a oligonucleotide containing the ipotetical binding site is incubated with the protein under investigation (positive control).

Binding reactions with probes derived from consecutive rounds of selection should be loaded adjacent to the positive control on the mobility shift gel. This allows the gradual appearance of specific complexes to be easily visualized.



#### **4.40 ISOLATION AND ANALYSIS OF BOUND OLIGONUCLEOTIDES FROM MOBILITY SHIFT GELS**

Once the success of the site selection procedure has been confirmed by mobility shift analysis, bound oligonucleotides are isolated from appropriate mobility shift complexes by direct amplification from gel slices. Bound oligonucleotides are then digested with restriction enzyme *EcoRI* and subcloned for sequence analysis of individual selected sites.

Molecules are ligated into the vector and transformed into *E. coli*. Sequence analysis of individual selected sites was performed.

**Elution buffer**

**0.5 M ammonium acetate**

**1 mM EDTA**

**0.1% SDS**

**Store several months at room temperature**

**Recovery buffer**

**50 mM Tris-Cl, pH 8**

**100 mM sodium acetate**

**5 mM EDTA**

**0.5% SDS**

**Store several months at room temperature**

# REFERENCES



1. Bardoni,B., Zanaria,E., Guioli,S., Florida,G., Worley,K.C., Tonini,G., Ferrante,E., Chiumello,G., McCabe,E.R.B., Fraccaro,M., et al. (1994). A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal. *Nat Genet*, 7, 497-501.
2. Behringer, R.R., Finegold, M.J., and Cate, R. L. (1994). Mullerian- inhibiting substance function during mammalian sexual development. *Cell* 79, 415-425.
3. Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990). Genetic evidence equating SRY and the testis-determining factor. *Nature*, 348, 448-450.
4. Bianchi,M.E. (1994). The HMG box domain. In Lilley, D.M.J. (ed.), *Proteins that manipulate DNA structure*. Oxford University Press, Oxford, in press.
5. Bianchi,M.E., Beltrame,M. and Falciola,L. (1992a) In Eckstein,F. and Lilley,D.M.J. (eds.), *Nucleic Acids and Molecular Biology*. Springer Verlag, Berlin, Vol. 6, pp. 112-128..
6. Bianchi,M.E., Falciola,L., Ferrari,S. and Lilley,D.M.J. (1992b). The DNA binding site of HMG1 protein is composed of two similar segments (HMGboxes), both of which have counterparts in other eukaryotic regulatory proteins.*EMBO J.*, 11, 1055-1063.
7. Burgoyne, P.S., Buehr, M., Koopman, P., Rossant, J., and McLaren, A. (1988). Cell-autonomous action of the testis determining gene: Sertoli cells are exclusively XY in XX-XY chimeric mouse testis. *Development* 102, 443-450.
8. Bustin, M., Lehn, D.A., Landsman, D. (1990). Structural features of the HMG chromosomal proteins and their genes. *Biochim. Biophys. Acta*, 3, 231-243.

9. Capel, B., Swain, A. Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P., and Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene SRY in adult mouse testis. *Cell* 73, 1019-1030.
10. Da Silva, S.M., Hacker, A., Harley, V., Goodfellow, P., Swain, A. and Lovell-Badge, R. (1996). Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nat Genet* 14, 62-68.
11. Falciola, L., Murchie, A.I.H., Lilley, D.M.J. and Bianchi, M.E. (1994). Mutational analysis of the DNA binding domain A of chromosomal protein HMG1. *Nucleic Acids Res.*, 22, 285-292.
12. Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R. and Bianchi, M.E. (1992). SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.*, 11, 4497-4506.
13. Foster, J.W., Brennan, F.E., Hampikian, G. K., Goodfellow, P. N., Sinclair, A.H., Lovell-Badge, R., Selwood, L., Renfree, M. B., Cooper, D. W., and Marshall Graves, J.A. (1992). Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. *Nature* 359, 531-533.
14. Foster, J.W., Dominiguez-Steglich, M.A., Guioli, S., Kowk G., Weller P.A., Stevanovic M., Weissenbach J., Mansour S., Young I.D., Goodfellow P.N. et al. (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in a SRY-related gene. *Nature*, 372, 525-530.
15. Giese, K., Cox, J. and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*, 69, 185-195.



16. Giese, K., Amsterdam, A. and Grosschedl, R. (1991). DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes. Dev.*, 12B, 2567-2578.
17. Giese, K. and Grosschedl, R. (1993). LEF-1 contains an activation domain that stimulates transcription only in a specific context of factor-binding sites. *EMBO J.*, 12, 4667-4676.
18. Giese, K., Pagel, J. and Grosschedl, R. (1994). Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors. *Proc. Natl. Acad. Sci. USA*, 91, 3368-3372.
19. Ginsburg, M., Snow, M.H.L., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* 110, 521-528.
20. Grosschedl, R., Giese, K. and Pagel, J. (1994). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet.*, 10, 94-100.
21. Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Mönsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryologically expressed genes. *Nature*, 346, 245-250.
22. Gubbay, J., Vivian, N., Economou, A., Jackson, D., Goodfellow, P., and Lovell-Badge, R. (1992). Inverted repeat structure of the SRY locus in mouse. *Proc. Natl. Acad. Sci. USA* 89, 7953-7957.
23. Harley, V.R., Jackson, D.I., Hextall, P.J., Hawkins, J.R., Berkovitz, G.D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P.N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. *Science*, 255, 453-456.
24. Harley, V.R., Lovell-Badge, R. and Goodfellow, P.N. (1994). Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res.*, 22, 1500-1501.

25. Hawkins, J.R., Taylor, A., Berta, P., Levilliers, J., Van der Auvera, B. and Goodfellow, P.N. (1992a). Mutational analysis of SRY: nonsense and missense mutations in XY sex reversal. *Hum. Genet.*, 88, 471-474.
26. Hawkins, J.R., Taylor, A., Goodfellow, P.N., Migeon, C. J., Smith, K.D. and Berkovitz, G.D. (1992b). Evidence for increased prevalence of SRY mutations in XY females with complete rather than partial gonadal dysgenesis. *Am. J. Hum. Genet.*, 51, 979-984.
27. Hawkins, J.R. (1994). Sex determination. (1994). *Hum Mol Genet*, 3, 1463-1467.
28. Ho, I.C. and Leiden, J.M. (1990). Regulation of the human T-cell receptor alpha gene enhancer: multiple ubiquitous and T-cell-specific nuclear proteins interact with four hypomethylated enhancer elements. *Mol. Cell. Biol.*, 10, 4720-4727.
29. Honda, S.I., Morohashi, K.I., Nomura, M., Takeda, H., Kitajima, M. and Omura, T. (1993). Ad4BP-regulating steroidogenic P-450 gene is a member of steroid hormone receptor superfamily. *J. Biol. Chem.* 268, 7494-7502.
30. Jager, R.L., Harley, V.R., Pfeiffer, R.A., Goodfellow, P.N. and Scherer, G. (1992). A familial mutation in the testis-determining gene SRY shared by both sexes. *Hum. Genet.*, 90, 350-355.
31. Jantzen, H.M., Admon, A., Bell, S.P., and Tjian, R. (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature*, 344, 830-836.
32. Jost, A. (1953). Studies on sex differentiation in mammals. *Rec. Prog. Horm. Res.* 8, 379-418.
33. Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., and Koopman, P. (1996) A male-specific role for SOX9 in vertebrate sex determination. *Development*, 9, 2813-2822.



34. Kerppola, T.K. and Curran, T. (1991). DNA bending by Fos and Jun: the flexible hinge model. *Science*, 254, 1210-1214.
35. King, C.-Y. and Weiss, M.A. (1993). The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. *Proc. Natl. Acad. Sci. USA*, 90, 11990-11994.
36. Koopman, P., Munsterberg, A., Capel, B., Vivian, N., and Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* 348, 450-452.
37. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for SRY. *Nature*, 351, 117-121.
38. Kuhn, A., Voit, R., Stefanovsky, V., Evers, R., Bianchi, M.E. and Grummt, I. (1994). Functional differences between the two splice variants of the nucleolar transcription factor UBF: the second HMG box determines specificity of DNA binding and transcriptional activity. *EMBO J.*, 13, 416-424.
39. Ikeda, Y., Shen, W.H., Ingraham, H.A. and Parker, K.L. (1994). Developmental expression of mouse steroidogenic factor 1, an essential role of the steroid hydroxylase. *Mol Endocrinol* 8, 654-662.
40. Lala, D.S., Rice, D.A. and Parker, K.L. (1992). Steroidogenic factor 1, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu factor 1. *Mol. Endocrinol.* 6, 1249-1258.
41. Levene, S.D. and Zimm, B.H. (1989). Understanding the anomalous electrophoresis of bent DNA molecules: a reptation model. *Science*, 245, 396-399.
42. Lilley, D.M.J. (1992). DNA-protein interactions. HMG has DNA wrapped up. *Nature*, 357, 282-283.

43. Lilley, D.M.J. and Clegg, R.M. (1993). The structure of branched DNA species. *Annu. Rev. Biophys. Biomol. Struct.*, 22, 299-328.
44. Liu-Johnson, H.-N., Gartenberg, M.R. and Crothers, D.M. (1986). The DNA binding domain and bending angle of *E. coli* CAP protein. *Cell*, 47, 995-1005.
45. Lovell-Badge, R., and Robertson, E. (1990). XY female mice resulting from a heritable mutation in the primary testis-determining gene, TDY. *Development* 109, 635-646.
46. Luo, X., Ikeda, Y. and Parker, K.L. (1994). A cell specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* 77, 481-490.
47. McLaren, A. (1991). Development of the mammalian gonad: the fate of the supporting cell lineage. *Bioessays* 13, 151-156.
48. McStay, B., Hu C.H., Pikaard C.S. and Reeder R.H. (1991). xUBF and Rib 1 are both required for formation of a stable polymerase I promoter complex in *X. laevis*. *EMBO J*, 10, 2297-2303.
49. Nasrin, N., Buggs, C., Fu Kong, X., Carnazza, J., Goebel, M. and Alexander-Bridges, M. (1991) DNA binding properties of the product of a testis-determining gene and a related protein. *Nature*, 354, 317-320.
50. Onate, S.A., Prendergast, P., Wagner, J.P., Nissen, M., Reeves, R., Pettijohn, D.E., Edwards, D.P. (1994). The DNA-bending protein HMG-1 enhances progesterone receptor binding to its target DNA sequences. *Mol Cell Biol*, 14, 3376-3391.
51. Palmer, M. S., Sinclair, A.H., Berta, P., Hellis, N. A., Goodfellow, P.N., Abbas, N.E., and Fellous, M. (1989). Genetic evidence that ZFY is not the testis determining factor. *Nature* 342, 937-939.



52. Palmer, S.J., and Burgoine, P.S. (1991). In situ analysis of fetal, prepuberal and adult XX----XY chimaeric mouse testes: Sertoli cells are predominantly, but not exclusively, XY.. *Development* 112, 265-268.
53. Palmer, S.J., and Burgoine, P.S. (1991). XY follicle cells in the ovaries of XO/XY and XO/XY/XYY mosaic mice. *Development* 1991 111, 1017-1019
54. Parisi M.A. and Clayton D.A. (1991). Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science*, 252, 965-969.
55. Patek, C., Kerr, J.B., Gosden, R. G., Jones, K.W., Hardy, K., Muggleton-harris, A.L., Handyside, A.H., Whittingham, D. G., and Hooper, M. L. (1991). Sex, chimeras, fertility, and sex determination in the mouse. *Development* 113, 311-325.
56. Pikaard, C.S., Smith, S.D., Reeder, R.H., Rothblum, L. (1990). rUBF, an RNA polymerase I transcription factor from rats, produces DNase I footprints identical to those produced by xUBF, its homolog from frogs. *Mol Cell Biol* 10, 3810-3812.
57. Pikaard, C.S., Pape, L.K., Henderson, S.L., Ryan, K., Paalman, M., H., Lopata, M., A., Reeder, R., H. and Sollner-Webb. (1990). Enhancer for RNA polymerase I in mouse ribosomal DNA. *Mol Cell Biol* 10, 4816-4825.
58. Rice, D.A., Mouw, A.R., Bogerd, A.M. and Parker, K.L. (1991). A shared promoter element regulates the expression of three steroidogenic enzymes. *Mol. Endocrinol.* 5, 1552-1561.
59. Shen, W., Moore, C., Ikeda, Y., Parker, K.L., and Ingrham, H. A. (1994). Nuclear receptor steroidogenic factor 1 regulates the Mullerian-inhibiting substance gene: a link to the sex determination cascade. *Cell* 77, 651-661.
60. Sheridan, P.L., Sheline, C.T., Cannon, K., Voz, M.L., Pazin, M.J., Kadonaga, J.T. and Jones K.A. (1995). Activation of the HIV-1 enhancer by the LEF-1 HMG protein on nucleosome-assembled DNA in vitro. *Genes Dev*, 9, 2090-2104.

61. Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R. and Goodfellow, P.N. (1990) A gene from the human sex determining region encodes a protein with homology to a conserved DNA binding motif. *Nature*, 346, 240-244.
62. Staben, C. and Yanofsky, C. (1990). *Neurospora crassa* a mating type region. *Proc. Natl. Acad. Sci*, 87, 4917-4921.
63. Studier, F.W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol*, 21, 37-44.
64. Sudbeck, P and Scherer, G. (1997). Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. *J.B.C*, 272, 27848-27852.
65. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y. and Yamamoto, M. (1991). *Schizosaccharomyces pombe* *stell1+* encodes a transcription factor with an HMG motif that is a critical regulator of sexual development. *Genes Dev.* 5, 1990-1999.
66. Swain, A., Zanaria, E., Hacker, A., Lovell-Badge R. and Camerino G. (1996). Mouse *Dax1* expression is consistent with a role in sex determination as well as in adrenal and hypothalamus function. *Nat Genet* 12, 404-409.
67. Thompson, J.F. and Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res.*, 16, 9687-9705.
68. Travers, A.A. (1988) In Eckstein, F. and Lilley, D.M.J. (eds.), *Nucleic Acids and Molecular Biology*. Springer Verlag, Berlin, Vol.2, pp.136-148.
69. Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. *Genes Dev.* 5, 880-894.



70. Tucker, P.K. and Lundrigan, B.L. (1993). Rapid evolution of the sex determining locus in Old World mice and rats. *Nature*, 364, 715-717.
71. van de Wetering, M. and Clevers, H. (1992) Sequence specific interaction of the HMGbox protein TCF-1 and SRY occur within a minor groove of a W. C. double helix. *EMBO J.*, 11, 3039-3044.
72. van de Wetering, M., Oosterwegel M. and Clevers, H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMGbox. *EMBO J.*, 10, 123-132.
73. Wagner, T., Wirth, J., Meyer, J., Zabel, B., Heid, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Huster, E., Wolf, U., Tommerup, N., Schempp, W., and Scherer, G. (1994). Autosomal sex reversal campomelic dysplasia are caused by mutations in and around the SRY-related gene Sox9. *Cell* 79, 1111-1120.
74. Waterman, M.L., Fischer, W., H. and Jones, K.A. (1991). A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev*, 5, 656-669.
75. Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995). Molecular basis of human 46,XY sex reversal revealed from the three dimensional solution structure of the human SRY DNA complex. *Cell*, 81, 705-714.
76. Whitfield, L.S., Lovell-Badge, R. and Goodfellow, P.N. (1993). Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature*, 364, 713-715.
77. Wright, E., Hargrave, M.R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A., and Koopman, P. (1995). the SRY related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nature Genet.* 9, 15-20.
78. Wu, H.-M. and Crothers, D.M. (1984) *Nature*, 308, 509-513.

79. Zanaria, E., Muscatelli, F., Bardoni, B., Strom, T., Guioli, S., Guo, W., Lalli, E., Moser, C., Walker, A., McCabe, E., Meitinger, T., Monace, A., Sassone-Corsi, P., and Camerino, G. (1994). An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature* 372, 635-641.
80. Zwilling, S., Koning, H. and Wirth, T. (1995). High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. *EMBO J* 14, 1198-1208.
81. Zwingman, T., Erickson, R.P., Boyer, T., and Ao, A. (1993). Transcription of sex-determining region genes SRY and Zfy in the mouse preimplantation embryo. *Proc. Natl. Acad. Sci. USA* 90, 814-817.