The EMBO Journal vol.10 no.9 pp.2533–2541, 1991

Genomic targets of the *serendipity* β and δ zinc finger proteins and their respective DNA recognition sites

François Payre and Alain Vincent

Centre de Recherche de Biochimie et de Génétique Cellulaires, 118 route de Narbonne, 31062 Toulouse Cédex, France

Communicated by A.Sentenac

The closely related *Drosophila serendipity* (*sry*) *beta* (β) and *delta* (δ) Cys2–His2 zinc finger proteins show partly overlapping *in vitro* DNA binding specificities and distinct patterns of binding sites on polytene chromosomes. Using a newly developed procedure, we identified genomic DNA targets for these two proteins. Both the *sry* β and δ proteins protect an 18–22 base region from DNase I digestion within each analysed genomic binding site, that includes a 13 bp consensus sequence. The consensus recognition sites *sry* β 5′-YCAGAGATGCGCA-3′ and *sry* δ 5′-YTAGAGATGGRAA-3′ thus differ by nucleotides at four out of 13 positions. They are determinant for specific binding of the *sry* β and δ proteins, respectively, produced in *Escherichia coli* or present in *Drosophila* embryos. We further show that *sry* β is the major (if not exclusive) *Drosophila* nuclear protein that specifically binds the 5′-CAGAGTGCGC-3′ sequence. The identified *sry* β genomic targets are all contained within single-copy DNA in euchromatic regions of the genome. Two out of the five characterized in detail map at cytological positions coincident with binding sites of the native *sry* β protein on polytene chromosomes.

Key words: consensus binding sites/Cys2/His2 zinc fingers/DNA binding proteins/genomic targets/*serendipity* genes

Introduction

One of the major recent contributions of molecular biology concerns the description, at the molecular level, of mechanisms of gene regulation. Studies on control of transcription have shown the fundamental role of binding of regulatory proteins upstream of, or close to, the transcriptional start site (reviews by Ptashne, 1988; Mitchell and Tjian, 1989). Families of transcription factors can be defined according to the type of conserved DNA binding domain they display in common (Mitchell and Tjian, 1989). One such domain is the Cys2–His2 zinc finger made of 28–30 amino acids with two cysteine and two histidine residues at fixed positions which stabilize the 'finger' by tetrahedrally coordinating a zinc ion (Klug and Rhodes, 1987; Berg, 1988). Characteristic of this DNA binding domain is its tandem repetition occurring a variable number of times per protein. Connected to variable finger domains, other conserved modules located at the N-terminal of Cys2–His2 finger proteins have been identified (Knöchel *et al.*, 1989). This multiple modular design thus offers an extensive number of combinatorial possibilities for the specific recognition of DNA on one hand, and interaction with other components of the transcription machinery on the other. It lends itself to the rapid acquisition of new regulatory functions, to parallel the increasing complexity of the genome and its expression in multicellular organisms during evolution. Indeed, Cys2–His2 finger proteins, found in a variety of higher and lower eukaryotes, appear to be involved in the control of diverse cellular and developmental processes (see Evans and Hollenberg, 1988; Pankratz and Jäckle, 1990 for reviews).

To assess the relative contributions of structural and expression parameters to functional speciation of zinc finger proteins during evolution, we use as a model system the *Drosophila sry* β and δ proteins. These proteins are so far the only identified pair of closely related Cys2–His2 finger proteins within a single species amenable to both genetic and molecular analyses. The *sry* β and δ genes result from a duplication event (Vincent *et al.*, 1985) with the most extensively conserved region between the two proteins corresponding to the predicted DNA binding domain which includes six contiguous fingers (Vincent, 1986; Payre *et al.*, 1990). The presence of acidic domains in both proteins supports the prediction that they are transcription factors, since a significant net negative charge is characteristic of the transcription activating region of various gene control proteins (Ptashne, 1988). The *sry* β and δ proteins are maternally inherited and present in embryonic nuclei at the onset of zygotic transcription, suggesting that they may be involved in this process (Payre *et al.*, 1989, 1990). Each of the characterized *sry* δ mutant alleles (Lindsley and Zimm, 1990) which display embryonic lethal phenotypes, is rescued by transformation with a DNA fragment containing the *sry* δ gene but not an extra copy of the *sry* β gene. Yet, the three *sry* δ mutant alleles sequenced to date display a single amino acid exchange at residues conserved between the *sry* β and δ proteins (Crozatier,M., Kongsuan,K., Ferrer,P., Merriam,J.R., Lengyel,J.A. and Vincent,A., submitted). Recent data have shown that changes in zygotic expression patterns and DNA binding specificity *in vitro* and *in vivo* have accompanied the functional divergence of the *sry* β and δ genes (Payre *et al.*, 1990).

Gaining further insight into the respective functions of these two proteins now requires the identification and characterization of the downstream genes they regulate. During this work, we developed a new procedure for cloning genomic targets of DNA binding proteins that should be especially useful for characterizing downstream genes of *sry* β and δ, and applicable to any other protein with specific DNA binding properties for which antibodies are available. We report here the DNA recognition sequences determined for the *sry* β and δ proteins. We have concentrated our efforts on the *sry* β protein because its sites of interaction with chromatin have all been mapped on polytene chromosomes (Payre *et al.*, 1990). We show the chromosomal position of several potential genomic targets of the *sry* β gene product.

© Oxford University Press

Results

Identification of genomic syr β binding sites

Purification of sequence-specific DNA binding proteins on oligonucleotides immobilized on Sepharose resin, originally described by Kadonaga and Tjian (1986), demonstrated that affinity chromatography based on DNA–protein interactions is very efficient. We reasoned that this methodology could be adapted in reverse for the purification of genomic DNA fragments specifically interacting with a given DNA binding protein.

We chose a purification scheme involving two different steps (thus avoiding any bias inherent in a repetitive procedure). First, *Drosophila melanogaster* genomic DNA cut with *Mbo*I was incubated with purified *sry β* protein made in *Escherichia coli* and protein–DNA complexes were immunoselected (see Materials and methods). DNA bound by the *sry β* protein was not dissociated up to 0.2 M KCl salt conditions. DNA eluted at the final elution step (0.4 M KCl, 0.01% of the total input, Figure 1A) was cloned into pZ2Not, a purposely designed vector (see Materials and methods). The second step was an immunoprecipitation of *sry β*–DNA complexes, in increasing amounts of competitor DNA, done on pools of 24 randomly picked recombinant clones. Approximately one out of four clones was found positive for interaction with *sry β* in this immunoprecipitation assay (Figure 1B). As an internal control for the specificity of this assay, a LacR–β-galactosidase fusion protein (i.e. the product of the *LacI–Z* fusion gene) was added to the DNA/*sry β* protein mix. Immunoprecipitation without antibody, with *sry β* antibody, or with anti-β-gal antibody, resulted in no immunoprecipitated DNA, precipitation of DNA fragments specific for *sry β* or immunoprecipitation of only the vector sequence (which contains a *lac* operator, binding site for the LacR protein), respectively. The immunoprecipitated DNA inserts were used directly for identification of recombinant clones. Each isolated clone was further validated as containing a *sry β* binding site using a gel retardation assay performed on the *Not*I digested plasmid (data not shown).

At this stage, we kept 10 independent clones with genomic inserts of small size for characterization by nucleotide sequencing and DNase footprinting. Nucleotide sequence comparison revealed that one insert, pBZ4.1, had been cloned three times and another, pBZ3.1, twice. Southern blot analysis done with each of the seven different fragments indicated that they all correspond to single-copy DNA in the *D.melanogaster* genome (data not shown). A search of nucleotide sequence databases (EMBL and Genbank)

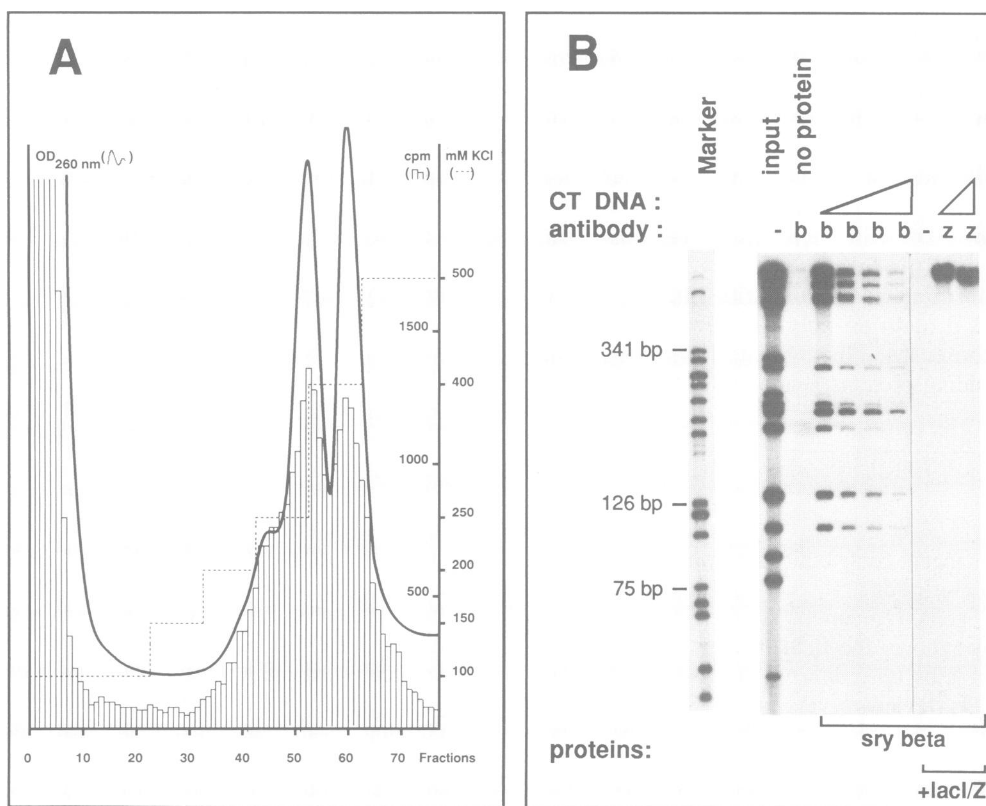

Fig. 1. *sry β* immunoselection of *Drosophila* genomic DNA. (**A**) The elution profile of *Drosophila* genomic DNA bound to the *sry β* protein was monitored by UV spectrophotometry (260 nm) on an FPLC system (Pharmacia). After extensive washing at 100 mM KCl, the protein bound DNA was stepwise eluted with increasing salt concentration. The corresponding fractions contained: 5 ng of DNA in the 150 mM KCl wash, 12 ng of 200 mM KCl, 54 ng of 250 mM KCl, 35 ng of 400 mM KCl, and no detectable DNA at 500 mM KCl and above. Full line, absorbance at 260 nm; dotted line, salt concentration of washing solution; histogram, radioactivity counted in each fraction. (**B**) Specific binding of *sry β* protein on DNA from one pool of 24 recombinant clones derived from immuno-selected genomic fragments. DNA–protein complexes were separated by immunoprecipitation. Input and immunoprecipitated DNA was analysed on denaturing polyacrylamide gel. Lanes: b, immunoprecipitation with anti-*sry β* antibody; –, no antibody; z, anti-β-galactosidase antibody. The amount of competitor DNA (CT DNA) was, from left to right, 100 ng, 400 ng, 1.5 μg and 6 μg; 400 ng and 6 μg. All lanes, except control, contained 100 ng of *sry β* protein, supplemented or not with 10 ng of LacI–Z fusion protein. The pZ2NOT plasmid vector contains a binding site for LacI–Z protein (*lac* operator).

indicated that these sequences have not been previously reported.

Definition of the sry β recognition sequence

The sequence elements reponsible for *sry* β binding on the isolated clones were determined by DNase I footprinting (Galas and Schmitz, 1978) in the presence of purified *sry* β protein (bacterially expressed), or BSA as control (see Materials and methods). Figure 2A shows the results of DNase footprinting on both strands obtained with a representative clone, pBZ2.2. The footprinted DNAs exhibited a single protected region ranging from 18 to 22 nucleotides in size as measured by comparison with a Maxam and Gilbert sequence track. The protected regions on the five footprinted clones (Figure 2B) led to the definition of a consensus *sry* β binding site of 13 nucleotides: 5′-YCAGAGATGCGCA-3′, (with Y meaning C or T) corresponding to the only motif shared by the five DNA sequences. All the binding sites had a motif consisting of at least 10−13 nucleotides identical to the consensus, with variations at only some positions (Figure 2C).

To demonstrate that the 5′-YCAGAGATGCGCA-3′ sequence was sufficient for specific binding of the *sry* β protein we performed gel retardation experiments with a 16 bp synthetic oligonucleotide (B22c) encompassing this sequence (see Materials and methods). Results (Figure 3A) show that the *sry* β protein binds efficiently to the B22c oligo. The specificity of this interaction was shown by competition experiments. Neither calf thymus DNA, nor the 28 bp oligonucleotide IRL which does not contain the consensus sequence, lessened formation of the *sry* β protein−oligonucleotide complex. In contrast, a 100-fold molar excess of cold B22c oligo displaced the protein−radiolabelled oligo complex. Finally, addition of affinity-purified monospecific anti-*sry* β antibody specifically resulted in a much more retarded complex, presumably composed of oligomer bound by *sry* β protein bound in turn to specific IgG. This demonstrates that retardation is due to the *sry* β protein.

5′-TCAGAGATGCGCA-3′ binds the sry β protein in Drosophila nuclear extracts

As the *sry* β protein shows its maximal level of accumulation during development in early embryos (Payre *et al.*, 1990), we chose embryonic nuclear extracts to look for a *Drosophila* binding activity specific to the *sry* β consensus sequence. In gel retardation assay, the B22c oligo is shifted to a retarded complex (Figure 3B). We verified that the retardation is specific and involves the *sry* β protein by using different types of competitors (see above) and by the shift in mobility of the complex formed with purified anti-*sry* β antibody,

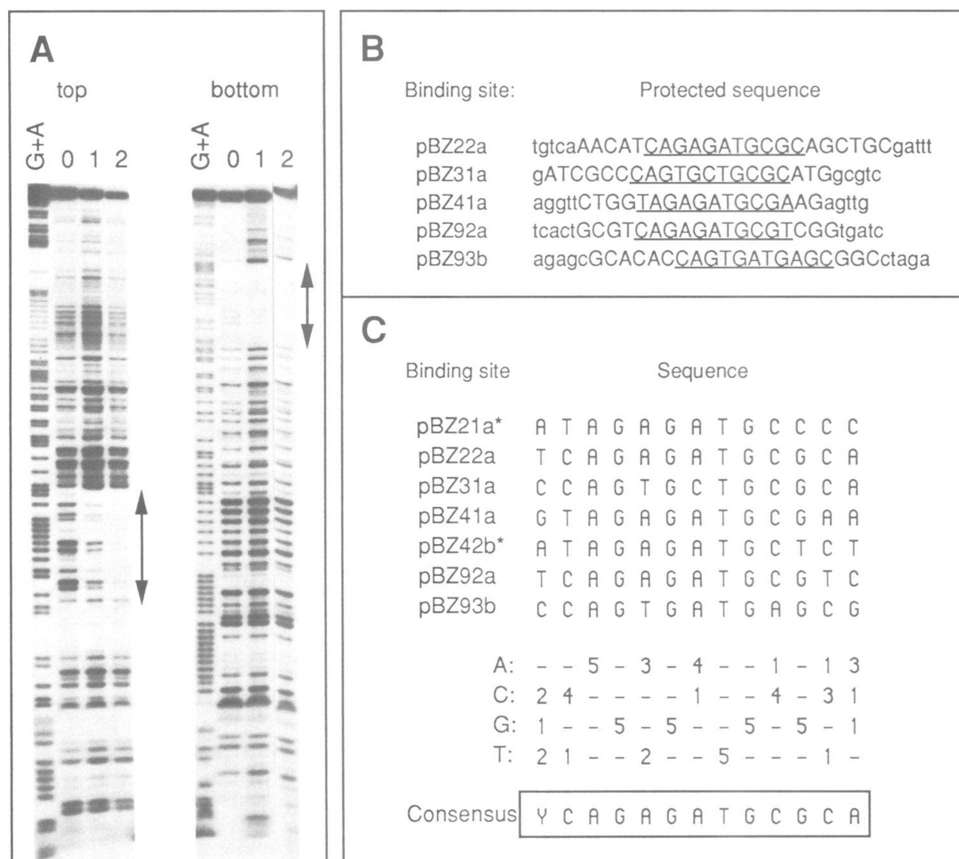

Fig. 2. Identification of *sry* β-specific binding sites on immunoselected DNA clones. (**A**) DNase I footprint of purified *sry* β protein bound to clone pBZ2.2. Amounts of protein were 400 ng (1), 1 μg (2). Lanes 0: control reaction containing 5 μg of BSA; lanes G+A: product of purine-specific sequencing reactions. Vertical bars mark regions protected from DNase I digestion. (**B**) Nucleotide sequences of the *sry* β protected regions on five different immunoselected genomic DNA frgments. The putative *sry* β recognition motifs are underlined. Protected sequences are read on top (a) or bottom (b) strands, lower case letters correspond to nucleotides flanking the protected region. (**C**) Sequence alignment of the putative *sry* β recognition motifs. The occurrence of a given nucleotide at each position is reported. The consensus sequence is defined by nucleotides present at least three out of five times at the position. Y, pyrimidine residue; R, purine residue. *Clones pBZ2.1 and pBZ4.2 have not been footprinted.

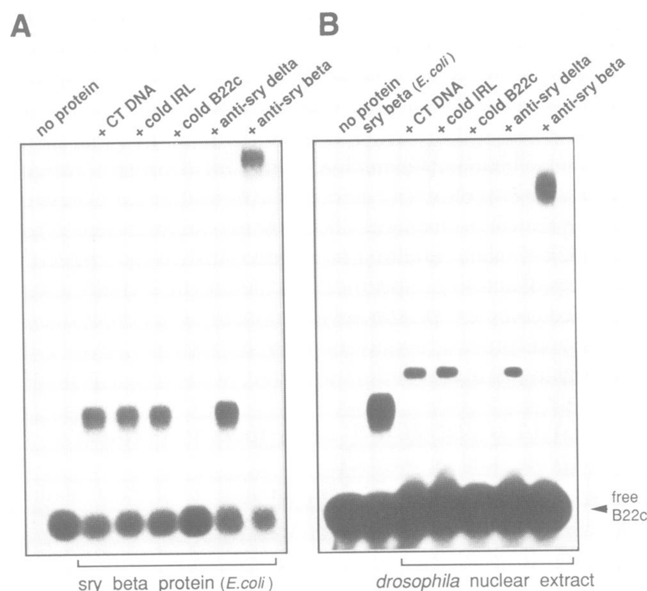

Fig. 3. Sequence-specific binding of the *sry* β protein. Gel mobility-shift analysis of protein binding to a 20 bp synthetic oligonucleotide (B22c) containing the *sry* β consensus binding site. The duplex B22c oligo was incubated with (**A**) purified *sry* β protein made in *E.coli* or (**B**) a *Drosophila* nuclear extract, and subjected to agarose gel electrophoresis. Specificity of the observed protein–DNA complexes was shown by addition to the incubation mix of various types of cold competitor DNA as indicated on the top. From left to right: control without protein; no competitor DNA; 10 000-fold mass excess of sonicated calf thymus (CT) DNA; 300 ng of cold IRL oligo; 30 ng of cold B22c oligo. The presence of the *sry* β protein in the retarded DNA–protein complexes was demonstrated by addition of anti-*sry* δ or anti-*sry* β antibody.

but no other antibody such as anti-β-galactosidase or anti-*sry* δ was added to assay.

The sry β DNA recognition sequence

Having verified that our immunopurification allowed an efficient selection of *sry* β DNA targets, we used the same procedure to clone genomic fragments containing binding sites for the *sry* δ protein (designated herein as pDZn.n). Complete characterization of these fragments, still in progress, will be reported elsewhere. Initial characterization of a small subset of them, however, by DNA sequencing and DNase I protection assays in the presence of purified *sry* δ protein identified *sry* δ binding sites (Figure 4A). The protected region on four footprinted fragments and nucleotide sequencing of additional immunoprecipitated short fragments (Figure 4B and data not shown) led to the definition of a consensus *sry* δ binding site of 13 nucleotides: 5′-YTAGA-GATGGRAA-3′, (with R standing for A or G) (Figure 4C). Specific binding of *sry* δ to this consensus site was assayed by gel retardation experiments using the 21 bp oligonucleotide D524 encompassing the consensus sequence (see Materials and methods), and the purified *sry* δ protein made in *Escherichia coli*. Results (Figure 5A) show that the *sry* δ protein binds efficiently to the D524 oligo. This binding is not displaced by adding either the oligonucleotide IRL or the *sry* β-specific oligo B22C. In contrast, a 100-fold molar excess of cold D524 oligonucleotide abolished completely formation of the radiolabelled oligonucleotide–protein complex. Addition of affinity-purified monospecific anti-*sry* δ antibody resulted in a much more retarded complex, presumably composed of a ternary oligonucleotide–*sry* δ protein–antibody complex, demonstrating that retardation is due to the *sry* δ protein. Identical results were obtained when a nuclear embryonic extract was used in place of purified *sry* δ protein. Furthermore, Western blot analysis of the material retrieved in the retarded complex demonstrated the presence of the *sry* δ protein (Figure 5B) but not the *sry* β protein (data not shown).

The sry β and sry δ DNA binding sites differ at four out of 13 nucleotide positions

Sequence comparison of the *sry* β and *sry* δ consensus binding sites shows that they differ by nucleotides at three, or possibly four, positions and share a central motif of seven nucleotides (Figures 2C and 4C). In order to determine whether or not the *sry* β and *sry* δ actually bind to either one or both of these sequences, we performed gel retardation assays using the bacterially made proteins and the two oligos B22c and D524 (Figure 6). Using standard conditions of stringency and competitor DNA, no cross specificity is detectable. In conditions where the *sry* β protein shifts $>80\%$ of the B22 oligo, no binding is observed on the D524 oligo. Conversely, the *sry* δ protein binds efficiently the D524 oligo but does not bind the B22c oligo. These results demonstrate that, while still closely related structurally, *sry* β and *sry* δ have diverged in their DNA recognition properties to the point where they are able to discriminate in an all or none manner between their respective DNA targets.

The sry β protein and 5′-TCAGAGATGCGCA-3′ nuclear binding activity co-purify

A lower apparent mobility of the B22c retarded complex was observed with the nuclear extract than with the *sry* β protein made in *E.coli* (compare lanes 2 and 3, Figure 3B). It correlates with the difference in mobility observed for the protein synthesized in *E.coli* from a full length *sry* β cDNA and the *Drosophila sry* β protein itself (Figure 7 bottom). This difference suggests that the *sry* β protein is subject, in *Drosophila* ovaries and/or embryos, to post-translational modifications altering its electrophoretic migration in denaturing gel conditions. In order to verify this assumption as well as to determine whether several proteins in *Drosophila* nuclei are capable of specific binding to the 5′-TCAGAGATGCGCA-3′ sequence, the band shift assay with the B22c oligo was performed using 70 fractions from MonoQ chromatography of a nuclear extract purified by heparin–Sepharose chromatography (H4 extract). The B22c binding activity was found in a single peak of elution corresponding to fractions 59–62 (Figure 7 top). Western blotting experiments revealed that the *sry* β protein is eluted precisely in these fractions (Figure 7 bottom). The co-purification of the *sry* β protein and the B22c binding activity indicates that *sry* β is the major (if not exclusive) protein in *Drosophila* nuclei that specifically binds the sequence 5′-TCAGAGATGCGCA-3′.

Compared cytological location of the genomic targets and binding sites of sry β on polytene chromosomes

The *sry* β protein has been shown to bind to ~ 54 cytological sites on polytene chromosomes of salivary glands of third instar larvae (Payre *et al.*, 1990). We compared the chromosomal locations of the five footprinted DNA target fragments with those of these sites. We first isolated

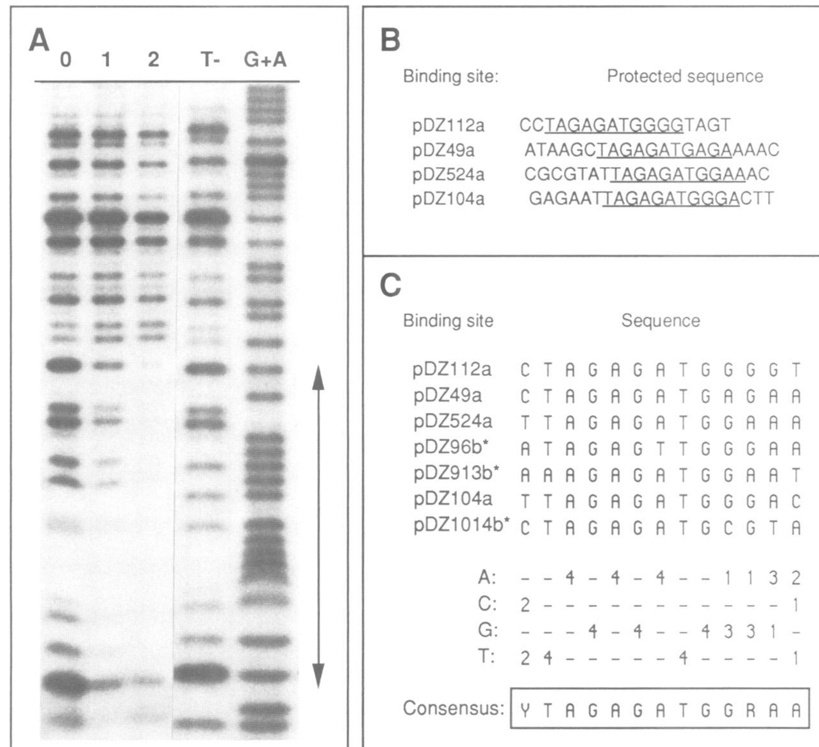

Fig. 4. Identification of *sry* δ-specific binding sites on immunoselected DNA clones. (**A**) DNase I footprint of purified *sry* δ protein bound to clone pDZ5.24. Amounts of protein were 400 ng (1), 1 μg (2). Lanes O and T−: control reaction containing respectively 5 μg of BSA or 5 μg of protein extract from bacteria containing the vector plasmid pAR3040; lanes G+A: product of purine-specific sequencing reactions. Vertical bars mark regions protected from DNase I digestion. (**B**) Nucleotide sequences of the *sry* δ protected regions on four different immunoselected genomic DNA fragments. The putative *sry* δ recognition motifs are underlined. Protected sequences are read on top (a) or bottom (b) strands. (**C**) Sequence alignment of the putative *sry* δ recognition motifs. The occurrence of a given nucleotide at each position is reported. The consensus sequence is defined by nucleotides present at least three out of four times at the position, R, purine residue; Y, pyrimidine residue. *Clones pDZ9.13, pDZ9.6 and pDZ10.14 have not been footprinted.

recombinant bacteriophage containing each of these targets, from a *Drosophila* genomic library (Maniatis *et al.*, 1978). The cytological localization of each phage was determined by *in situ* hybridization to polytene chromosomes. All phage map in euchromatic regions of polytene chromosomes. Phages FB5 and FB37, containing respectively the insert of pBZ3.1 and pBZ9.2, map at 65D and 13E, two regions which contain a *sry* β protein binding site (Figure 7, and Payre *et al.*, 1990). Phage FB8 corresponding to the insert from pBZ4.1 maps at 56B4-C8, a site of fixation of a fusion protein containing the *sry* β finger region (Noselli *et al.*, manuscript in preparation). The other two phages, corresponding to DNA inserts from pBZ22 and pBZ93, map at positions 64B and 1B, respectively (data not shown). The observation that two *sry* β DNA targets display cytological positions coincident with binding sites of the protein on polytene chromosomes suggest that at least a fraction of the DNA targets of *sry* β protein isolated could correspond to functional sites at this, or other developmental stages.

Discussion

We have chosen the *Drosophila sry* β and δ genes as a model for evolutionary studies on the structure and function of Cys2−His2 finger proteins within a given species. Unravelling their biological activity and the biochemical basis of their functional divergence requires identification of their downstream genes. However, if it has become relatively straightforward to identify which regulatory protein(s) interact with a specific DNA fragment (e.g., see Kadonaga and Tjian, 1986; Singh *et al.*, 1988), there is still a lack of methods allowing one to go in the opposite direction, i.e. from a DNA binding protein with unknown recognition specificity to its genomic targets. We report here the identification and cloning of genomic targets for the *sry* β and *sry* δ proteins, using a newly developed procedure which involves two affinity-purification steps. Our method differs from others recently described (Biedenkapp *et al.*, 1988; Kinzler and Vogelstein, 1990; Sompayrac and Danna, 1990) in that it does not include repetitive steps (thus avoiding amplification of any bias inherent to each step) nor DNA cloning before selection of specific DNA−protein complexes (eliminating the problem of sequence representation in the starting DNA source).

Several lines of evidence indicate that our method may efficiently purify single copy genomic targets for DNA binding proteins. Although we started from whole genomic DNA corresponding to a theoretical number of $\sim 6 \times 10^5$ different *Mbo*I fragments, we observed some redundancy among the clones obtained after a single selection step. For example, the fragment contained in clone pBZ4.1 was present in seven out of 500 clones derived from DNA specifically bound by the *sry* β protein and eluted at 0.4 M KCl (data not shown). Therefore, one can estimate that a 10^4-fold enrichment of some targets may be accomplished in this first step. After the second immunoselection step, all selected single-copy (Figure 8 and data not shown) genomic fragments contained a strong *sry* β binding site (Figure 2).

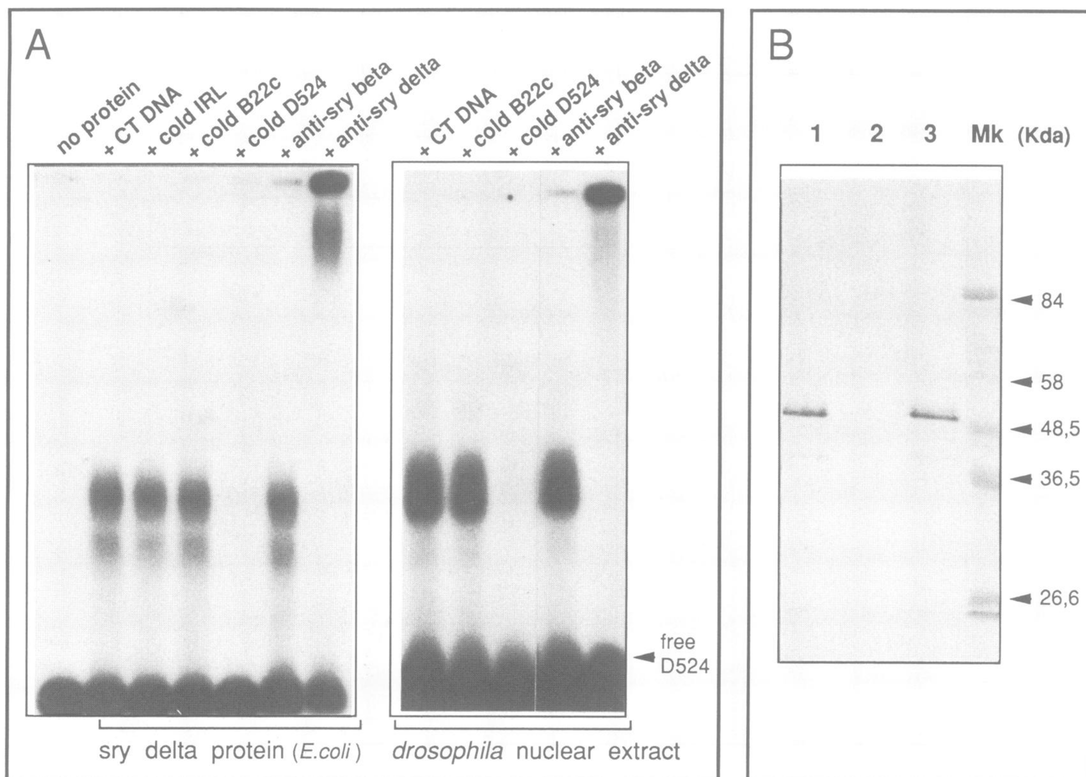

Fig. 5. Sequence-specific binding of the *sry* δ protein. (**A**) Gel mobility-shift analysis of protein binding to a 21 bp synthetic oligonucleotide (D524) containing the *sry* δ consensus binding site. The duplex D524 oligo was incubated with (left panel) purified *sry* δ protein made in *E.coli* or (right panel) a *Drosophila* nuclear extract, and subjected to agarose gel electrophoresis. Specificity of the observed protein–DNA complexes was shown by addition to the incubation mix of various types of cold competitor DNA as indicated on the top. From left to right: control without protein; no competitor DNA; 10 000-fold mass excess of sonicated calf thymus (CT) DNA; 300 ng of cold IRL oligo; 300 ng of cold B22c oligo; 30 ng cold D524 oligo. The presence of the *sry* δ protein in the retarded DNA–protein complexes was demonstrated by addition of anti-*sry* β or anti-*sry* δ antibody. (**B**) Detection of the *sry* δ protein in retarded complexes by Western blot analysis using anti-*sry* δ antibody after separation of the proteins by 10% polyacrylamide gel electrophoresis. lanes 1 and 3: DNA–protein complexes formed in the presence of *sry* δ protein made in *E.coli* and *Drosophila* nuclear extract, respectively; lane 2: nuclear proteins migrating at the same position as the *sry* δ protein–DNA complex on an agarose gel (lane without DNA).

The five *sry* β binding sites characterized in more detail all showed DNase I protected regions extending from 18 to 22 nucleotides on both strands and centred around the 13 bp sequence 5′-YCAGAGATGCGCA-3′ (Figure 2). This consensus sequence was found also in the other selected fragments that were sequenced. That the 5′-YCAGAGATGCGCA-3′ sequence is sufficient for specifically binding the *sry* β protein was demonstrated by gel retardation assays, using a synthetic oligonucleotide and either the protein made in bacteria or crude nuclear extracts from *Drosophila* embryos (Figure 3). By the same criteria, the *sry* δ protein was shown to bind specifically to the sequence 5′-YTAGAGATGGRAA-3′ (Figures 4 and 5). The *sry* β and δ DNA binding consensus sequences are strongly related as seven nucleotides at the centre of these sequences are identical for both proteins. Nevertheless, the *sry* β and δ proteins are able to discriminate fully between their respective binding sites in band shift assays. This binding specificity is observed with proteins made in *E.coli* as well as crude embryonic nuclear extracts (Figures 5 and 6). Furthermore, chromatographic purification of nuclear extracts resulting in the elution of the *sry* β and δ proteins in separate fractions allowed confirmation of this binding specificity (Figure 6 and data not shown).

As our method was designed to select the strongest DNA binding sites, however, it remains entirely possible that genomic sequences intermediate between the *sry* β and δ consensus recognition sequences are capable of binding both proteins. Along this line, it is interesting to recall that some λ phage DNA restriction fragments are bound by only one

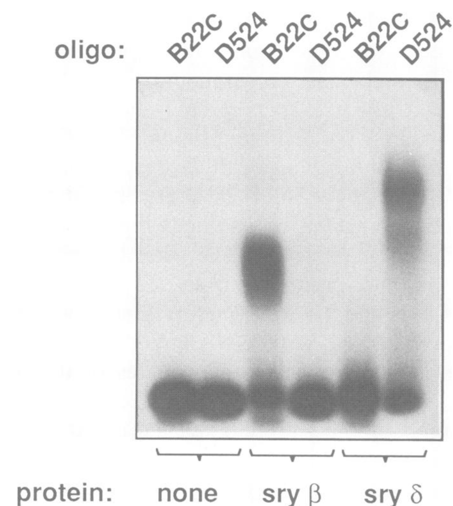

Fig. 6. Compared DNA binding specificities of the *sry* β and *sry* δ proteins. Gel mobility-shift assays were done in conditions given in Figures 2 and 5, using purified *sry* proteins made in *E.coli*.

of the two proteins while some others are bound by both (Payre *et al.*, 1990). All (weakly) bound fragments contain a sequence that matches at, minimally, 8/13 positions in these consensus sequences. In fragments bound by both proteins, the matching nucleotides are preferentially located within the 'core' recognition sequence shared by *sry* β and δ (unpublished data).

The size of the *sry* β and δ recognition sites, and of the regions these proteins protect from DNase I digestion confirms that the initial assumption of a binding site of 5.5 bp per finger is not a general rule for Cys2−His2 zinc finger proteins (Rhodes and Klug, 1986; see discussion in Payre and Vincent, 1988). Modelling of interactions between DNA and fingers linked by the widespread *Krüppel*-type TGEKP H−C link (Schuh *et al.*, 1986; Gibson *et al.*, 1988; Nardelli *et al.*, 1991) indicated 2−3 bp of DNA bound per finger (see Berg, 1990). It led to the proposal that the nature of the H−C link could be a major determinant of this spacing. As both *sry* β and δ (like TFIIIA) present H−C links which differ from the general sequence TGEKP, structural analysis of their binding to DNA could provide new insights on modalities of interaction between DNA and Cys2−His2 finger proteins.

The *sry* β and δ recognition sites both differ from other reported binding sites for transcription factors, including the GAGA factors recently described (Biggin and Tjian, 1988; Gilmour *et al.*, 1989). They were not found among the *Drosophila* nucleotide sequences already present in the databases (corresponding mostly to cDNA sequences). The absence of the *sry* β and δ recognition sites from what is predominantly protein coding sequence, together with proximity to other putative *cis*-regulatory elements (data not shown), is consistent with a location in gene upstream regions. It suggests that *sry* β and δ do bind to gene control regions not sequences so far.

Gel retardation assays using nuclear extract purified by two chromatographic steps showed that the *sry* β protein not only specifically recognizes the 5′-YCAGAGATGCGCA-3′ sequence, but is the major if not the only *Drosophila* nuclear protein which efficiently binds to it. The endogenous *sry* β protein is detected at ~54 specific sites on third instar larval polytene chromosomes (Payre *et al.*, 1990). Cytological mapping of the five cloned genomic targets of *sry* β characterized in detail showed that there is a correlation between the positions of three of these targets and *in vivo* binding sites of (or a derivative of, see below) the protein (Figure 7). The absence of correlation for the other two fits the assumption that only a subset of recognition sites for a given DNA binding protein is occupied in a specific tissue at a given developmental stage (e.g. depending upon chromatin structure, Berg and von Hippel, 1988). Indeed, while two fragments map at positions of binding sites of the *sry* β protein (65D and 13E, respectively), the third one maps to the position (56BC) of a major binding site for a β-gal fusion protein containing the *sry* β zinc fingers which is not a detectable binding site for the native *sry* β protein (Payre *et al.*, 1990; Noselli,S., Payre,F. and Vincent,A., in preparation). Experiments towards determining whether or not the *in vivo* binding sites of the proteins and the *sry* β recognition sites in the cloned fragments are the same, and characterizing the nearby transcription units identified by Northern blot analysis (A.Vincent, unpublished results) have been undertaken.

The method developed in this study should provide a rather direct way to identify and characterize the downstream genes of *sry* β and δ and potentially many other transcription factors, including those in organisms where a comprehensive genetic analysis is not possible. Knowledge of the *sry* β and *sry* δ DNA recognition sequences should help in unravelling the respective roles of *sry* β and *sry* δ in development.

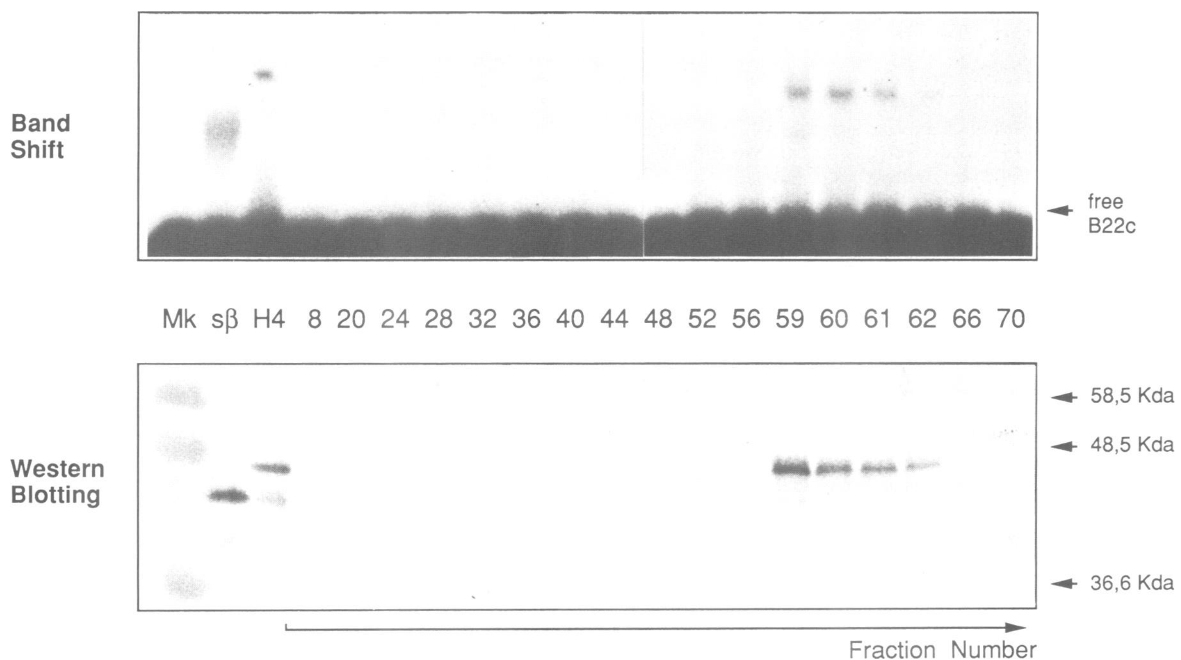

Fig. 7. Correlation between 5′-YCAGAGATGCGCA-3′-specific binding activity and the presence of the *sry* β protein in a fractionated nuclear extract. **Top**: gel retardation assay using the B22c oligo. **Bottom**: detection of the *sry* β protein by Western blot analysis using anti-*sry* β antibody, after separation of the proteins by 10% polyacrylamide−SDS gel electrophoresis. Mk, size markers; sβ, 5 ng of *sry* β protein made in *E.coli*; H4, heparin−Sepharose bound nuclear extract; 8−70, numbers of fractions from MonoQ chromatography.

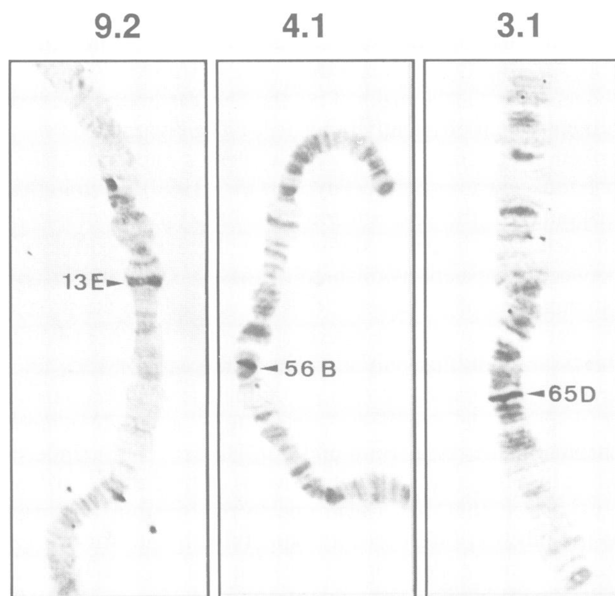

Fig. 8. Cytological localization of three *sry* β genomic targets (inserts from plasmids pBZ3.1, pBZ4.1 and pBZ9.2), two of which potentially correspond to binding sites of the *sry* β protein on polytene chromosomes of salivary gland cells of third instar larvae (Payre *et al.*, 1990). Chromosome arms are oriented distal up.

Materials and methods

Plasmid constructs and DNA handling

A *Mlu*I–*Hinc*II DNA fragment containing a full size *sry* β cDNA (Payre *et al.*, 1990) was cloned in the plasmid pBKS+ (Stratagene) cut by *Sma*I, isolated from this vector as a *Bam*HI fragment, and cloned in the pAR3038 vector (Rosenberg *et al.*, 1987) to give rise to the pSBA2 protein producing plasmid.

The introduction of two *Not*I sites flanking unique *Bam*HI and *Cla*I restriction sites in the polylinker of pGEM7Z+ (Promega) has been achieved as follows. The plasmid vector was digested by *Sma*I and *Sac*I, and used for linker tailing with two oligonucleotides (5′-CGCGGCCGCATCGA-TTGGATCCTGCGGCCGC-3′ and 5′-GCGGCCGCAGGATCCAAT-CGATGCGGCCGCGAGCT-3′). After ligation, denaturation and rehybridization, this vector, designated pZ2Not, was introduced into *E.coli*. The structure of the new polylinker: T7 promoter–*Apa*I, *Aat*II, *Sph*I, *Xba*I, *Xho*I, *Eco*RI, *Kpn*I, ***Not*I**, *Bam*HI, *Cla*I, ***Not*I**, *Sac*I, *Bst*XI, *Nsi*I restriction sites–Sp6 promoter, was verified by nucleotide sequencing.

A Canton S *Drosophila* genomic library constructed in the bacteriophage λ Charon 4 (Maniatis *et al.*, 1978), was screened using the oligo random priming procedure. Purification of recombinant phages, extraction of DNA and other basic DNA techniques were done according to standard protocols (Sambrook *et al.*, 1989).

Proteins and antibody

Production in *E.coli* and purification of the *sry* β and δ proteins (purity >80%, as estimated by SDS–PAGE followed by silver staining) and preparation of anti-*sry* β and anti-*sry* δ antibodies were as previously described (Payre *et al.*, 1989, 1990). 2 mg of *Drosophila* nuclear extract prepared from 2–12 h Oregon R embryos, according to Biggin and Tjian (1988), was loaded on a 500 μl bed volume heparin–Sepharose (Pharmacia) column, equilibrated in TpA buffer [20 mM Tris–HCl pH 7.6, 100 mM KCl, 10% (v/v) glycerol, 1 mM dithiotreitol, 50 μM $ZnSO_4$, 0.2 mM phenyl mercuryl sulphonate fluoride, 5 μg/ml leupeptin, 5 μg/ml aprotinin]. After washing with TpA, heparin bound proteins were eluted with TpB (same as TpA but 1 M KCl), and subsequently diluted until the conductivity of the sample was equivalent to 100 mM KCl (H4 fraction) before loading onto a HR5/5 MonoQ column (Pharmacia). Bound proteins were eluted into 70 fractions (250 μl each) using a 0.1–1 M KCl linear gradient. 5 μl of each fraction (corresponding to 1–5 μg, as determined by Bradford assays) were used for either gel retardation assay or Western blotting (as previously described in Payre *et al.*, 1990).

Isolation of genomic DNA sequences which bind to sry β and δ

Genomic DNA from *D.melanogaster* strain Oregon R was extracted and purified by centrifugation to equilibrium in CsCl gradients following standard protocols. In the first screening step, *sry* protein–DNA interaction was done in 2 ml of HK100 buffer (20 mM HEPES (*N*-2-hydroxyethyl-piperazine-*N*′-2-ethanesulphonic acid), pH 7.5; 100 mM KCl; 50 μM $ZnSO_4$; 1 mM dithiothreitol; 0.05% NP-40; 20% glycerol) containing: 100 μg/ml of tRNA, and 1.5 mg of cold and 2 μg of nick translated (4×10^7 c.p.m.) genomic DNA digested by *Mbo*I. 100 μg of purified bacterial *sry* protein was added, and this mix was incubated for 1 h at 4°C. 25 μg of affinity-purified antibody pre-adsorbed on 200 μl (bed volume) of protein A–Sepharose (Pharmacia), was then added. After 30 min incubation, the Sepharose beads were packed in an HR5/5 column (Pharmacia). Following extensive washing with HK100 buffer, the *sry* protein bound DNA was step eluted by increasing the ionic strength of the HK buffer (to 500 mM KCl). The *sry* β bound DNA eluted at 400 mM KCl and *sry* δ bound DNA eluted at 250 mM KCl were cloned in the pZ2NOT plasmid (cut by *Bam*HI and dephosphorylated). Around 400 recombinant plasmids were obtained for 1/50 of each DNA.

In the second screening step, pools of 24 recombinant plasmids were digested by *Not*I, and end-labelled using Klenow enzyme and [α^{32}P]dGTP. The labelled DNA (50 ng, ~20 000 c.p.m.) was then subjected to standard immunoprecipitation protocols (McKay, 1981; Payre *et al.*, 1990). The immunoprecipitated DNA was either analysed on denaturing 6% polyacrylamide–7 M urea gels, or directly used as a probe for colony hybridization on the 24 corresponding clones, using rapid hybridization medium from Amersham.

DNase I footprinting

DNA inserts were labelled on one strand or the other, utilizing the pZ2Not unique restriction sites *Cla*I and *Xho*I. Plasmid DNA cut with one of these enzymes was labelled by filling-in using the Klenow fragment and [α^{32}P]dTTP. The labelled insert was cut out by digestion with the second enzyme and purified from a 6% polyacrylamide gel. 20 000 c.p.m. of probe DNA (~2 ng) was mixed with 2 μg of calf thymus sonicated DNA or poly(dI-dC), 2 μg of tRNA in 25 μl of HK100 buffer. Incubations with varying quantities of *sry* β or δ protein were for 30 min at 4°C, followed by 1 min at 25°C, just prior to addition of 25 μl of digestion solution (20 mM $MgCl_2$ and 5 mM $CaCl_2$) and 1 min incubation at 25°C. DNase I (1/100 unit, Pharmacia) was added and incubated for 1 min at 25°C. DNase digestion was stopped by addition of 50 μl of stop buffer (200 mM NaCl, 20 mM EDTA, 20 mM Tris–HCl pH 8.0, 0.1% SDS, containing 0.1 mg/ml calf thymus DNA), followed by phenol and phenol–chloroform extractions, and ethanol precipitation. Pellets were dissolved in formamide loading buffer (80% deionized formamide, 10 mM EDTA, 0.02% bromophenol blue, 0.02% xylene cyanol) and 10 000 c.p.m. of each sample were analysed on a 6% polyacrylamide–8 M urea sequencing gel. Marker lane contained a chemical cleavage reaction for A + G (Maxam and Gilbert, 1980), done on the same DNA fragment immobilized on MG paper (Amersham).

Gel retardation assay

Oligonucleotides:

```
B22c: 5'-GATCATCAGAGATGCGCAGC-3'
            TAGACTCTACGCGTCGCTAG-5'
IRL:  5'-GATCGGTGATGCTGCCAACTTACTGATTTATG-3'
            CCACTACGACGGTTGAATGACTAAATACCTAG-5'
D524: 5'-GATCGCGCGTATTAGAGATGGAAAC-3'
            CGCGCATAATCTCTACCTTTGCTAG-5'
```

were (or not) kinase-labelled using [γ^{32}P]ATP (5000 Ci/mmol) and hybridized with the complementary oligonucleotide. Duplex oligonucleotides were then purified from 10% polyacrylamide gel.

Labelled oligonucleotide (0.3 ng or 0.1 pmol) was incubated for 30 min at 4°C in 17 μl of HK100 buffer containing (unless indicated otherwise), 3 μg of sonicated calf thymus DNA, 2 μg of tRNA and 20 ng of *sry* β and δ protein, or *Drosophila* nuclear extract. After addition of 3 μl of loading buffer [15 μg/μl of BSA in HK100, 50% (w/v) Ficoll, 0.02% bromophenol blue, 0.02% xylene cyanol], the complexes were analysed by electrophoresis on 1% agarose gel in 0.5× TBE buffer. After migration (250 V for 1.5 h at 4°C), the gel was dried out in Whatman DE81 paper and auto-radiographed.

In situ hybridization to polytene chromosomes

Recombinant lambda phage were cold labelled with bio-14-dATP (BRL) using the nick translation kit from Amersham. Dissection of salivary glands from mature third instar larvae and spreading of polytene chromosomes

were done according to Engels *et al.* (1986), hybridization with biotinylated probes according to Langer-Safer *et al.* (1982). Hybridization signals were revealed using DETEK l-hrp Signal Generating System (ENZO) with DAB (Polysciences) as the substrate. Following this reaction, chromosomes were counterstained with Giemsa 1/100 for 30 s, dried and mounted in Eukit. Microphotographs were taken using Nikon equipment and EKTAR 25 (Kodak) or FP4 Plus (Ilford) film.

Acknowledgements

We are grateful to Claudia Kücherer for gifts of embryonic nuclear extracts, and David Cribbs, François Karch and André Sentenac for critical reading of the manuscript. We wish to thank Edmond Barbey for photography and Joëlle Maurel for editorial assistance. This work was supported by CNRS and the Association pour la Recherche sur le Cancer, grant No. 6457 to A.V.. F.P. was supported by fellowships from Ministère de la Recherche et de l'Enseignement Supérieur and Ligue Nationale Contre le Cancer.

References

Beng,J.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 99–102.
Berg,J.M. (1990) *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 405–421.
Berg,O.G. and Von Hippel,P.H. (1988) *Trends Biochem. Sci.*, **13**, 207–211.
Biedenkapp,H., Borgmeyer,U., Sippel,A.E. and Klempnauer,K.H. (1988) *Nature*, **335**, 835–837.
Biggin,M.D. and Tjian,R. (1988) *Cell*, **53**, 699–711.
Engels,W.R., Preston,C.R., Thompson,P. and Eggleston,W.B. (1986) *Focus*, **8**, 6–8.
Evans,R.M. and Hollenberg,S.M. (1988) *Cell*, **52**, 1–3.
Galas,D.J. and Schmitz,A. (1978) *Nucleic Acids Res.*, **5**, 3157–3170.
Gibson,T.J., Postma,J.P.M., Brown,R.S. and Argos,P. (1988) *Protein Eng.*, **2**, 209–218.
Gilmour,D.S., Thomas,G.H. and Elgin,S.C.R. (1989) *Science*, **245**, 1487–1490.
Kadonaga,J.T. and Tjian,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5889–5893.
Kinzler,K.W. and Vogelstein,B. (1990) *Mol. Cell. Biol.*, **10**, 634–642.
Klug,A. and Rhodes,D. (1987) *Trends Biochem. Sci.*, **12**, 464–469.
Knöchel,W., Pöting,A., Köster,M., El-Baradi,T., Nietfeld,W., Bouwmeester,T. and Pieler,T. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6097–6100.
Langer-Safer,P.R., Levine,M. and Ward,D.C. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4381–4385.
Lindsley,D.L. and Zimm,G. (1990) *Drosophila Information Service*, **68**, 232.
McKay,R.D.G. (1981) *J. Mol. Biol.*, **145**, 471–488.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connel,D., Quon,G., Sim,K. and Estrafiadis,A. (1978) *Cell*, **15**, 687–701.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Nardelli,J., Gibson,T.J., Vesque,C. and Charnay,P. (1991) *Nature*, **349**, 175–178.
Pankratz,M.J. and Jäckle,H. (1990) *Trends Genet.*, **6**, 287–292.
Payre,F. and Vincent,A. (1988) *FEBS Lett.*, **234**, 245–250.
Payre,F., Yanicostas,C. and Vincent,A. (1989) *Dev. Biol.*, **135**, 469–489.
Payre,F., Noselli,S., Lefrere,V. and Vincent,A. (1990) *Development*, **110**, 141–149.
Ptashne,M. (1988) *Nature*, **335**, 683–689.
Rhodes,D. and Klug,A. (1986) *Cell*, **46**, 123–132.
Rosenberg,A.H., Lade,B.N., Chui,D.S., Lin,S.W., Dunn,J.J. and Studier,F.W. (1987) *Gene*, **56**, 125–135.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual.* 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schuh,R., Aicher,W., Gaul,U., Coté,S., Preiss,A., Maier,D., Seifert,E., Nauber,U., Schröder,C., Kemler,R. and Jäckle,H. (1986) *Cell*, **47**, 1025–1032.
Singh,H., LeBowitz,J.H., Balwin,A.J. and Sharp,P.A. (1988) *Cell*, **52**, 415–423.
Sompayrac,L. and Danna,K.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3274–3278.
Vincent,A. (1986) *Nucleic Acids Res.*, **14**, 4385–4391.
Vincent,A., Colot,H.V. and Rosbash,M. (1985) *J. Mol. Biol.*, **186**, 149–166.

Received on December 28, 1990; revised on May 29, 1991