The EMBO Journal vol.8 no.11 pp.3459-3466, 1989

Drosophila Sgs3 TATA: effects of point mutations on expression *in vivo* and protein binding *in vitro* with staged nuclear extracts

Angela Giangrande[1], Clément Mettling[2], Marianne Martin[3], Claude Ruiz and Geoff Richards

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France
[1]Present address: Department of Zoology, University of Washington, Seattle, USA
[2]Present address: Laboratory for Neurobiology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK
[3]Present address: INRA, Station de Pathologie Comparée, 30380 Saint Christol-lez-Alès, France

Communicated by P.Chambon

The *Drosophila* salivary gland secretion protein gene, *Sgs3*, has a consensus TATA sequence and gives rise to abundant stage and tissue-specific transcripts. Two TATA point mutations (TAAA and TAGA) reduce transcript levels ~50-fold when assayed in transgenic flies. This effect is reflected *in vitro*, in DNase I footprint and gel retardation assays where we observed TATA-probe-specific complexes that are not seen with TAAA, TAGA or non-specific probes. The binding patterns observed when using nuclear extracts from 0-2- and 0-20-h embryos (*Sgs3* inactive) differ from those seen with extracts from third instar salivary glands (*Sgs3* active). There are also differences in *in vitro* binding when using an *hsp70* TATA fragment, previously shown to substitute *in vivo* for the *Sgs3* TATA sequence, as probe. Together these observations suggest the possibility that more than one TATA box factor may be present in these extracts. We conclude that a wild-type TATA motif is crucial for the binding of a TATA box factor and all subsequent interactions with other factors bound to the proximal and distal regulatory sequences that are necessary for the normal expression of *Sgs3*.

Key words: Drosophila/gel retardation/*Sgs3* glue gene/TATA box/tissue-specific gene regulation

Introduction

The TATA box motif, located some 30 bp upstream of the transcription start point of many genes transcribed by polymerase B(II), was the first transcriptional regulatory element to be recognized in eukaryotes [see Breathnach and Chambon (1981) for review]. Analyses of TATA box mutations, *in vivo* or *in vitro*, have suggested that the TATA box is involved both in determining the RNA start site and levels of transcription [see Wasylyk (1988) for review].

Interactions between the TATA sequence and transcription factors have been studied in *in vitro* transcription systems, using partially purified fractions of cellular or nuclear extracts (e.g. Davison *et al.*, 1983; Parker and Topol, 1984a). Subsequent studies have shown that a factor (BTF1 or TFIID) interacts specifically with the TATA sequence to position RNA polymerase B(II) and its ancillary initiation factors which, at least in certain cases, have been shown to form a preinitiation complex (e.g. Davison *et al.*, 1983). The HeLa cell TATA box factor can be functionally substituted in this complex by a yeast factor, suggesting a strong evolutionary conservation of these proteins (Buratowski *et al.*, 1988; Cavallini *et al.*, 1988). However, for a given cell type, it remains uncertain as to whether there is a single TATA binding protein, a family of related proteins or a more complex situation.

In *Drosophila*, studies of the heat-shock protein genes (*hsps*) have revealed constitutive sequence-specific protein binding in the TATA box region (e.g. Parker and Topol, 1984b; Wu, 1984a, 1985; Thomas and Elgin, 1988) and have shown that interactions between specific transcription factors bound to upstream promoter sequences (the HSEs) following induction and those bound to the TATA box region are necessary for transcriptional activation.

Drosophila data concerning TATA box mutants are limited. In *D.melanogaster*, mobile elements inserted within the TATA box of *Sgs4* (McGinnis *et al.*, 1983) or larval cuticle protein III (Snyder *et al.*, 1982), severely reduced or eliminated expression respectively. In *D.mulleri*, an alcohol dehydrogenase (*Adh*) pseudogene lacking a TATA sequence is transcribed at ~10% of the level of *Adh2*, which has a classic TATA box sequence (Posakony *et al.*, 1985). In transformation experiments, Hirsh *et al.* (1986) showed that deletion of the supposed TATA sequence (TTTAAA) of the *D.melanogaster* dopa decarboxylase (*Ddc*) gene reduced expression to between 10 and 50% of the wild-type in the hypoderm without altering the RNA start site. Using *Drosophila* cell line extracts, Heberlein *et al.* (1985) showed that deletion of the TATA sequences of both the distal and proximal promoter of *Adh* essentially abolishes specific initiation *in vitro*.

The *in vivo* analysis of expression of the *D.melanogaster* salivary gland secretion protein gene *Sgs3*, by P element transformation, has revealed a complex regulation extending over some 2 kb of 5′ flanking sequence [see Martin *et al.* (1989a) for review]. Stage- and tissue-specific activity of *Sgs3* requires elements contained within a 98-bp fragment immediately 5′ to the start site (Martin *et al.*, 1989b) which includes a consensus TATA box sequence at −31 (TATAAA) (see Bucher and Trifonov, 1986). Here we have assayed the importance of the TATA motif for *Sgs3* expression *in vivo*, and have initiated an *in vitro* analysis of DNA-protein interactions of the TATA box region using nuclear extracts from embryos and third-instar salivary glands.

Results

Expression of Sgs3 TATA box mutants in vivo

In tranformed fly strains carrying the Formosa *Sgs3* construct C20g71 (Figure 1A), transcripts from the Formosa *Sgs3* allele (800 bases) are found at levels similar to those from the resident Oregon-like allele (1100 bases) and, as for the resident gene, this expression is restricted to the third-instar salivary gland (Bourouis and Richards, 1985a). The C20g71 construct carries 2.7 kb of 5′ flanking sequence, containing remote regulatory elements necessary for abundant transcript levels, and 0.3 kb of 3′ sequence. In the C20g71G and C20g71A constructs, the only change is the mutation of the second T of the TATA sequence to a G or A respectively.

The level of expression of *Sgs3* in constructs carrying >2 kb of 5′ flanking sequence varies some 2- to 3-fold for different insertions, presumably a result of position effects, e.g. g71 50−120% of resident allele (Bourouis and Richards, 1985a; see also Giangrande *et al.*, 1987). We therefore analysed the expression in late third-instar larvae of four insertions for both C20g71G and C20g71A (Figure 1B). Transcripts from the Formosa allele are seen as a weak band only upon extended autoradiographic exposure of the filters (Figure 1C). The maximum expression of the inserted gene is seen in line G2, and was estimated as being <2% of that of the resident gene (see Materials and methods). The *in vivo* expression of these TATA mutants is therefore severely reduced with respect to the wild-type.

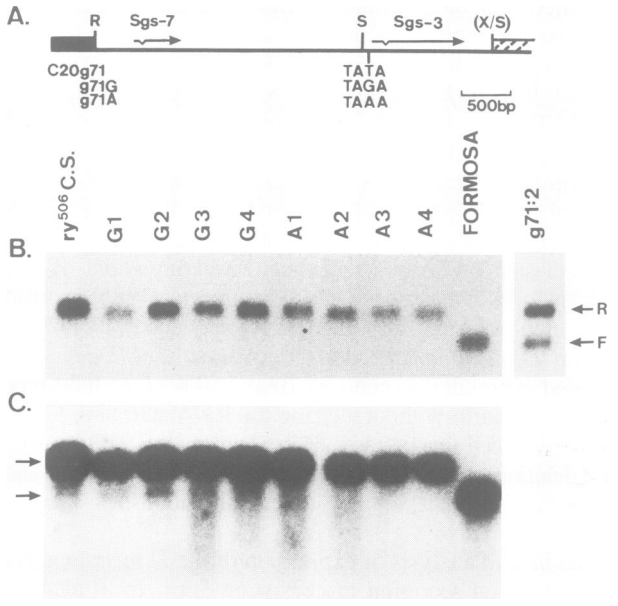

Fig. 1. The *in vivo* expression of the *Sgs3* gene after *in vitro* mutagenesis of the TATA sequence. (**A**) The 68C genomic sequence of the Formosa strain inserted between P (solid box) and ry^+ (hatched box) sequences in the Carnegie 20 (C20) transposon (Rubin and Spradling, 1983). C20g71 (Bourouis and Richards, 1985a) carries the wild-type sequence, C20g71G and g71A are point mutants in the second T of the *Sgs3* TATA sequence as shown. R = *Eco*RI, S = *Sal*I, X = *Xho*I. The *Sgs7* and *Sgs3* transcribed sequences are denoted by a short and long arrow respectively. (**B**) Northern analysis of 5 μg of total RNA from third-instar larvae of wild-type (ry^{506} *C.S.* and Formosa) and transformed strains (C20g71, g71:2; C20g71G, G1−G4; C20g71A, A1−A4). Transcripts from the resident (R) *Sgs3* gene or the Formosa (F) *Sgs3* gene are denoted by arrows. (**C**) A longer exposure of the autoradiogram of (B) showing low levels of expression of the transformed Formosa *Sgs3* gene (see in particular the G2 line).

DNA−protein interactions in the TATA-containing region in vitro

We asked whether we could detect DNA−protein interactions in the TATA-box-containing region *in vitro* and whether the *in vivo* effects of the TATA point mutations were reflected by alterations of specific DNA−protein interactions. We made nuclear extracts (NE) from third-instar salivary glands where *Sgs3* is expressed, and established an *in vitro* DNase I footprinting assay using a template carrying the wild-type TATA box (Figure 2). Like many other tissues and cell lines (e.g. Wu, 1984a), salivary glands contain endogenous nucleases. Bands due to these nucleases are shown by open circles in Figure 3.

Footprinting reactions on the upper strand using salivary gland NE show three regions where the pattern differs from that obtained without NE (naked DNA) as shown in Figure 3A (lanes 2 and 3, see brackets to the right of gel lanes). The first is centred on the TATA box and is seen as the partial protection against DNase I digestion at −29, −31 and −33. In addition a site due to endogenous nuclease appears at −34 (open circle). A second region is located downstream of the TATA box, between −15 and −21. Here, the site at −15 disappears and new sites appear when NE is added. The third region is downstream of the start site, between +8 and +16, where an increased sensitivity is evident. Addition of increasing amounts of a 30mer oligonucleotide carrying the *Sgs3* TATA sequence (Figure 2) in the footprint reaction specifically minimizes the footprint pattern in all three regions (Figure 3A, lanes 3−5). Experiments with the complementary strand reveal two protected regions with salivary gland NE (Figure 3B, lanes 2 and 3). The first protection, 3′ to the TATA box, is located between positions −9 and −16. The second is centred on the TATA box, between −27 (−28) and −33. Although the changes seen with NE are not striking, they are seen consistently with different NE preparations (not shown).

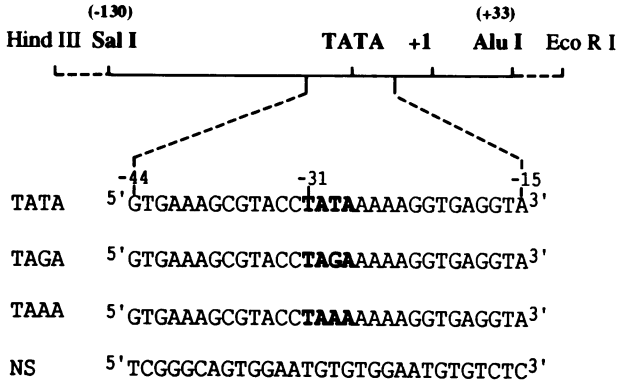

Fig. 2. Probes from the *Sgs3* proximal promoter used in this study. The *Sal*I (−130) to *Alu*I (+33) fragment in the pEMBL19 polylinker, labelled at the *Hin*dIII site or *Eco*RI site for the upper or lower strand respectively, footprint reactions (Figure 3). Below are shown the upper strand sequences of the wild-type and point mutant double-stranded oligonucleotides (30mers) used as probes in gel retardation assays and for competition reactions. The non-specific (NS) oligonucleotide, lacking a TATA motif, is also shown, as is the *hsp70* TATA region probe.

The specificity of protein interactions with the TATA sequence

We performed gel retardation assays with the oligonucleotide used in the footprint competition experiments. Incubation with NE results in the formation of a number of retarded DNA−protein complexes (Figure 4A, lanes 1 and 4). To investigate the specificity of these interactions we used two 30mers carrying the TAAA and TAGA mutations as well as a 30mer lacking a TATA-like sequence (NS = non specific, see Figure 2), either as probes or as competitors in gel retardation assays. We first determined whether the bands seen with the TATA probe are found with the other probes. Most bands are common to all four probes and are therefore not TATA specific (open triangles, Figure 4A), whereas one is detected only with the wild-type TATA probe (asterisk, Figure 4A, see lanes 4 and 7). Note that the relative abundance of the specific versus aspecific complexes varies between NEs and for this reason a TATA probe + NE sample is included in each gel as an internal standard (see Materials and methods).

Competition experiments, using the TATA 30mer as probe and increasing amounts (10-, 50-, 100- and 300-fold excess) of the TATA, TAGA, TAAA or NS oligonucleotides as competitors, show >50% competition of the TATA-specific band with 50-fold excess of unlabelled wild-type nucleotide, whereas for the mutants, a 100- to 300-fold excess is necessary to obtain a similar level of competition (Figure 4B). Significantly, the NS oligonucleotide does not compete, even at a 300-fold excess. The minor difference in the levels of competition between the TAAA and TAGA mutants may result from the use of a different NE in lanes 1−9 from that of lanes 10−20 (see above).

DNA−protein interactions in the TATA-box-containing region using embryonic nuclear extracts

To investigate whether the binding pattern found in the *in vitro* assays reflects DNA−protein interactions specific to the salivary gland where *Sgs3* is expressed, we prepared extracts from *Drosophila* embryos (0−20 h) where the gene is not expressed.

We performed footprinting experiments using the upper strand and compared the pattern with that obtained with the salivary gland NE (Figure 3A, lanes 7−9). Two out of three regions display a similar pattern, the most important difference being the absence, in the case of the embryonic NE, of the hypersensitive region downstream of the start

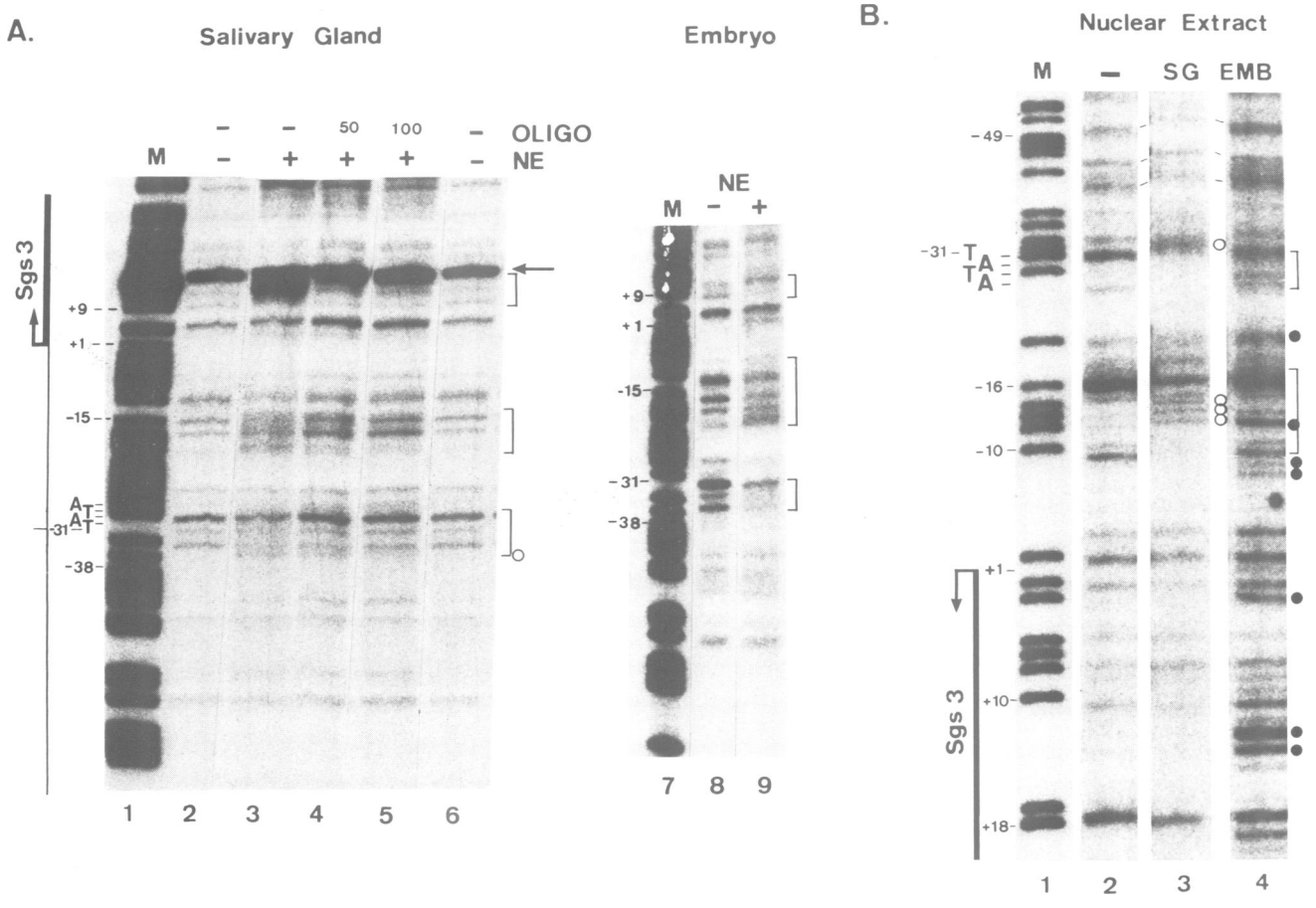

Fig. 3. DNase I *in vitro* footprinting patterns in the *Sgs3* TATA box region. (**A**) Digestion patterns on the upper strand (fragment *Sal*I−*Alu*I, see Figure 2) of naked DNA (**lanes 2** and **6**) or with 30 μg of salivary gland nuclear extract (NE) (**lanes 3−5**) and varying amounts of the *Sgs3* TATA oligonucleotide (0, 50- or 100-fold excess, OLIGO) as competitor, and those with or without 30 μg of embryonic nuclear extract (**lanes 9** and **8**) respectively. M denotes marker lanes (**lanes 1** and **7**) with A+G sequencing reactions (Maxam and Gilbert, 1977). The *Sgs3* gene and TATA box (−31) are shown schematically at the left. Nucleotides are numbered with respect to the start site. Regions showing a differential digestion pattern in the presence of NE are indicated by brackets to the right of the panels. Bands due to endogenous nuclease activity in the extracts are shown by open circles. The arrow denotes a degradation product of the probe. (**B**) Digestion patterns on the lower strand. Changes due to salivary gland (SG) or embryonic (EMB) nuclear extract are shown to the right. Other symbols, etc. as in (A), except for new DNase I hypersensitive sites seen with embryonic extract (**lane 4**) and indicated by solid circles.

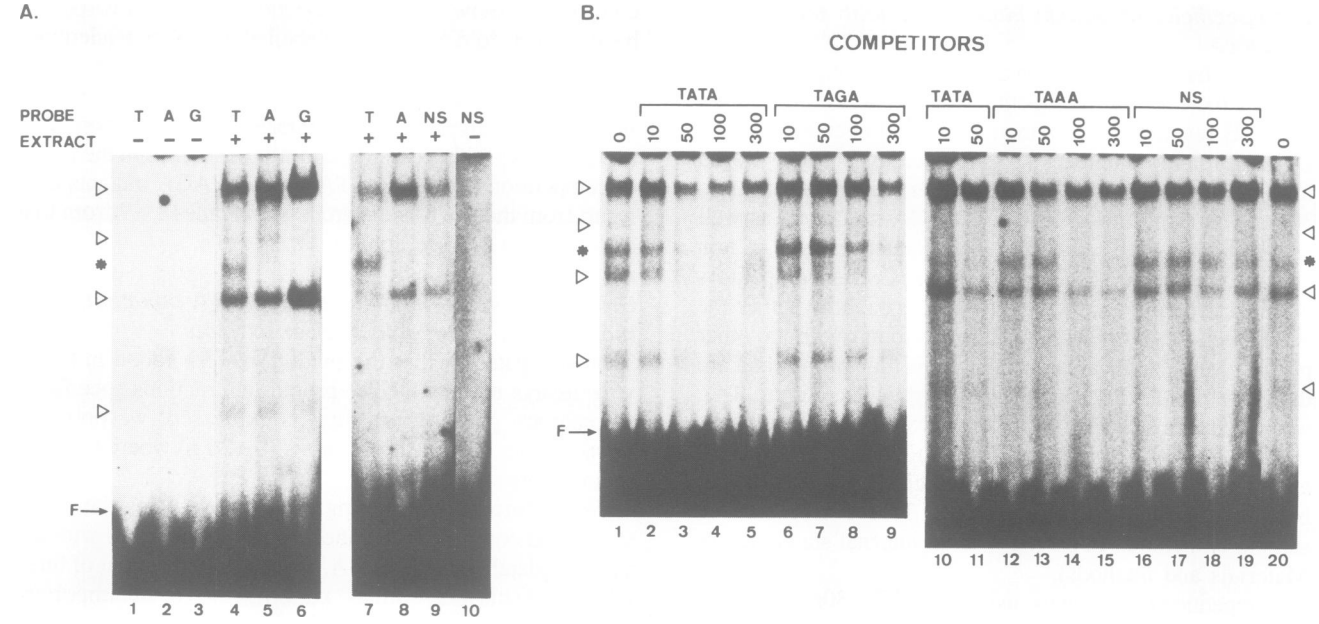

Fig. 4. DNA–protein complexes formed *in vitro* between *Sgs3* TATA region oligonucleotides and components of salivary gland nuclear protein extracts. (**A**) Radioactively labelled *Sgs3* TATA, TAAA, TAGA and NS oligonucleotides in the absence (**lanes 1–3** and **10**) and presence (**lanes 4–9**) of extract. **Lanes 4** and **7** (for TATA) and **5** and **8** (for TAAA) show complexes formed between oligonucleotides and two independent extracts (see Materials and methods and Results). The TATA-specific band is shown by an asterisk, non-specific bands seen to varying degrees wth all templates and extracts are shown by open triangles. F denotes the upper limit of free (uncomplexed) oligonucleotides. (**B**) Complex formation with the TATA oligonucleotide [**lanes 1** and **20**; experimental conditions as for lane 4 in (A)] is competed by the addition of a 10-, 50-, 100- or 300-fold excess of unlabelled TATA, TAGA, TAAA or NS oligonucleotides in the preincubation step. Other symbols as in (A).

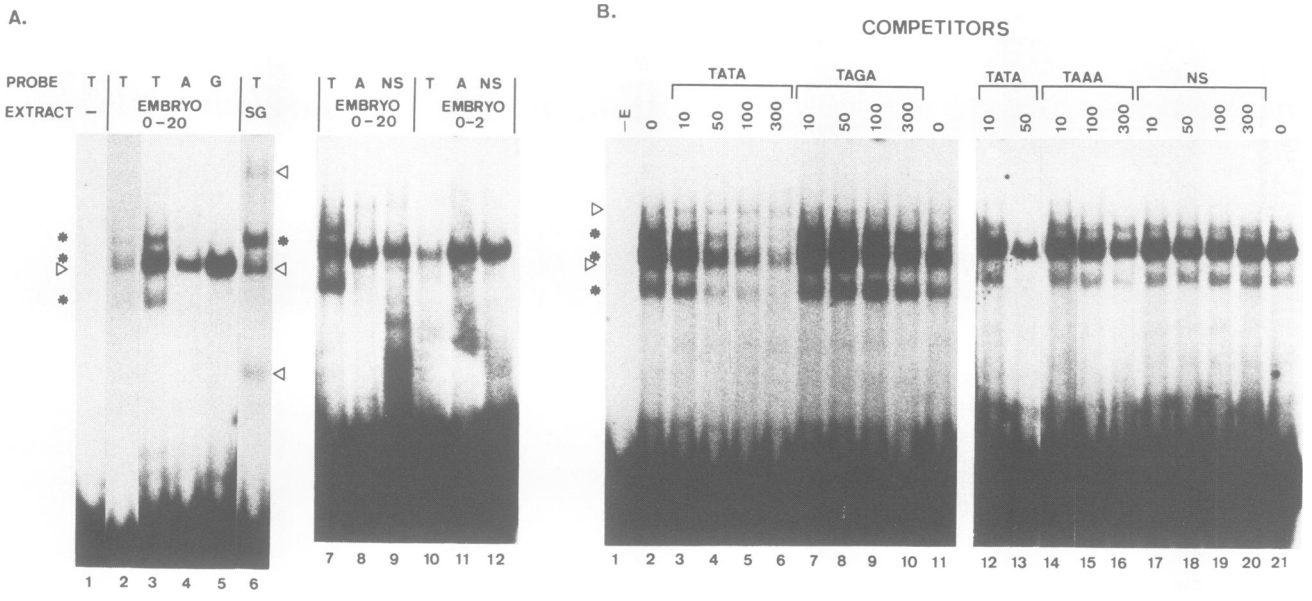

Fig. 5. Formation of DNA–protein complexes *in vitro* between *Sgs3* TATA region oligonucleotides and components of embryonic nuclear protein extracts. (**A**) Radioactively labelled *Sgs3* TATA (**lanes 2, 3** and **7**), TAAA (**lanes 4** and **8**), TAGA (**lane 5**) and NS (**lane 9**) oligonucleotides in the presence of 0–20-h embryonic extract. The *Sgs3* TATA oligonucleotide in the absence of extract (**lane 1**) or wth salivary gland extract (**lane 6**) is also shown. **Lane 2** is a shorter autoradiographic exposure of **lane 3**, where the central doublet is more evident. The salivary gland pattern of complexes is resumed to the right of **lane 6** as an aid to comparisons. In **lanes 10–12**, the experiment of **lanes 7–9** is repeated using a nuclear extract from 0–2-h embryos. (**B**) Complex formation of 0–20-h embryonic extract with the *Sgs3* TATA oligonucleotide [**lanes 2, 11** and **21**; experimental conditions as for lane 3 in (A)] is competed by the addition of a 10-, 50-, 100- or 300-fold excess of unlabelled TATA, TAGA, TAAA or NS oligonucleotides in the preincubation step. The TATA probe, without extract (–E) is shown in lane 1. Other symbols as in (A).

site. In contrast a footprint reaction on the complementary strand reveals a quite different pattern, as several hypersensitive sites are visible in the two regions protected when the salivary gland NE is used, and novel hypersensitive sites are visible downstream of the start site (Figure 3B, lane 4). This suggests the formation of different DNA–protein complexes in the TATA-box-containing region when using the two NEs.

In gel retardation assays we incubated the TATA, TAGA, TAAA or NS oligonucleotides with 0–20-h embryonic NE and compared the binding pattern with that obtained using gland NE (Figure 5A, lanes 1–9). The wild-type TATA

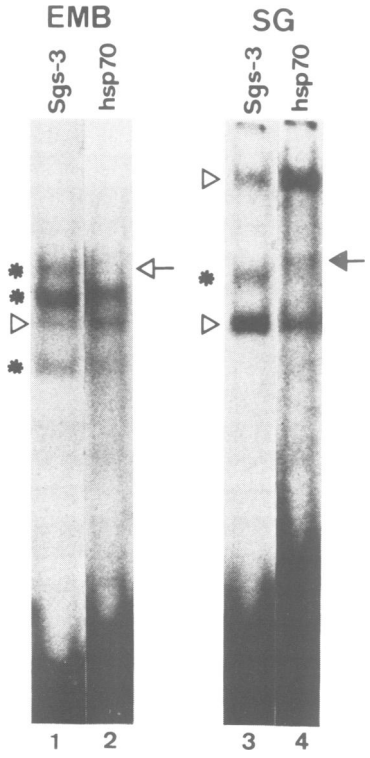

Fig. 6. Comparison of DNA−protein complexes formed *in vitro* between TATA-region oligonucleotides and components of embryonic and salivary gland nuclear protein extracts. Radioactively labelled *Sgs3* TATA (**lanes 1** and **3**) and *hsp70* TATA (**lanes 2** and **4**) oligonucleotides in the presence of 0−20-h embryonic extract (**lanes 1** and **2**) or salivary gland extract (**lanes 3** and **4**). Symbols for the *Sgs3* probe−protein complexes are as in Figures 4 and 5. Note the improved resolution of the doublet (see Materials and methods). For the *hsp70* probe, the 'missing' band with embryonic extract and the novel band with salivary gland extracts are marked by an open and solid arrow respectively.

30mer gave four bands, one of which, the lower band in the central doublet (open triangle, Figure 5A; see also Figure 6, lanes 1 and 2, where the resolution of this doublet is improved), is common to all four probes. Amongst the three TATA-specific bands (asterisks), the upper one comigrates with the salivary gland TATA-specific DNA−protein complex. If NE from 0−2-h embryos is used, only the non-TATA-specific band is clearly evident (Figure 5A, lanes 10−12), although faint TATA-specific bands appear in lane 10 upon extended exposure of the autoradiogram (data not shown). As transcription is low at this stage this suggests that the three complexes seen with 0−20-h embryonic NE do indeed reflect a stage-dependent increase in transcription factors.

In competition experiments using the 0−20-h NE, as for the salivary gland NE, the TATA 30mer competes >50% with a 50-fold excess, the two mutants, TAGA and TAAA require a 300-fold excess for a similar level of competition whereas the NS oligonucleotide does not compete (Figure 5B).

The Drosophila hsp70 TATA sequence

If the *Sgs3* fragment −50 to +10 (including the TATA motif) is replaced by the equivalent sequence of the *hsp70* gene, stage- and tissue-specific expression of the *Sgs3* promoter is maintained (Martin *et al.*, 1989b). It was therefore of interest to determine the DNA−protein complexes formed with the *hsp70* sequence *in vitro*. We repeated the gel retardation experiments with embryonic and salivary gland extracts (Figure 6) using a comparable *hsp70* oligonucleotide (Figure 2). With embryonic extracts we detected three bands, which migrate similarly to those seen with the *Sgs3* TATA oligonucleotide (Figure 6, lanes 1 and 2), except that the uppermost band (that co-migrates with that seen in salivary glands; Figure 5A, lanes 3 and 6) is dramatically reduced, if not absent. With salivary gland extracts, in addition to the non-TATA-specific bands, we detected a single band, that is distinct from the *Sgs3* TATA-specific band (Figure 6, lanes 3 and 4). These differences were observed with several independent NEs (not shown).

Discussion

The importance of the Sgs3 TATA motif in vivo

Two point mutations in the *Sgs3* TATA sequence reduce expression at least 50-fold (<2% of wild-type). This effect is one of the most important reported to date and should be compared with the 5-fold decrease seen with the same point mutations in the chick conalbumin promoter analysed by injection into *Xenopus* oocytes (Grosschedl *et al.*, 1981), or the 2- to 3-fold effect of point mutations in the TATA box of the mouse β-globin promoter analysed by transfection into HeLa cells (Myers *et al.*, 1986). More dramatic effects of mutations in the TATA box region have been reported from *in vivo* analyses [e.g. activity reduced to 3%, adenovirus E2aE1 (Zajchowski *et al.*, 1987)], but these result from 'linker scanning' mutations involving 10 bp substitutions and derive from assays in which multiple copies of the mutant templates are introduced into cells, in contrast to our assay in which a single copy is introduced into each cell by germline transformation. Our model is closest to the case of β-thalassemia studied by Orkin *et al.* (1983), in which modification of the ATA sequence resulted in a 10-fold decrease in RNA levels in the patient's erythroid cells. Our results confirm that the importance of the TATA sequence for levels of expression varies from gene to gene.

Protein−DNA interactions in the Sgs3 TATA region in vitro

Our footprinting experiments with salivary gland extracts revealed modifications in the DNase I digestion pattern extending from −33 to +16. Using *Drosophila* cell line extracts, Parker and Topol (1984a,b) demonstrated modifications extending from −50 to +30 for the histone H3 gene, −40 to +40 for the actin 5C gene and −37 to +29 for *hsp70*, each of the genes being a transcriptionally active configuration. Similar extended TATA region footprints have been described in other systems for transcriptionally active templates [e.g. the adenovirus ML promoter (Sawadogo and Roeder, 1985)].

As our TATA region footprints show only partial protection, a common finding when using crude nuclear extracts, we pursued the problem of altered DNA−protein interactions in the mutants by gel retardation assays. We concentrated on binding in the TATA box region, the oligonucleotide probes extending from −44 to −15. With salivary gland extracts we detect one DNA−protein complex that is specific to the wild-type probe. Its abundance is reduced in the presence of a 100- to 300-fold excess of the

TAAA or TAGA oligonucleotides suggesting that they have a reduced affinity for the protein. If this is the *Sgs3* TATA box factor, the *in vivo* consequence of these mutations is an important reduction in the binding of the transcription complex on the *Sgs3* TATA region. It is unlikely that TATA-specific complexes are formed with proteins similar to the African green monkey α-protein, that do not recognize specifically the TATA motif but rather (A+T)-rich regions of 5−6 bp, as the TAAA probe should provide a target of similar affinity for such a protein (Solomon *et al.*, 1986).

The results with embryonic extracts are more complicated, although those from the competition experiments are essentially similar to those with the salivary gland extract. Three bands are specific to the wild-type probe, one of which has a similar mobility to that seen with salivary gland extracts. This band may reflect the binding of the same protein seen in salivary gland extracts, although it is not known whether such a binding normally occurs on the *Sgs3* TATA box in embryonic nuclei. By analogy with the *hsps* (see Introduction), the presence of a TATA box factor may be constitutive for *Sgs3*, and activation in the salivary gland requires the addition of stage- and tissue-specific factors. The additional lower bands may result either from the binding of different TATA box factors (or families of closely related factors) to any TATA sequence presented *in vitro* or represent degradation products of a single factor modified by proteases present in the embryonic extracts.

A multiplicity of TATA factors?

In the absence of conflicting data, it has been assumed that the highly conserved TATA motif is recognized by an equally conserved cognate factor. A number of recent studies of promoters showing both constitutive and inducible activity (L.Wu *et al.*, 1987; Simon *et al.*, 1988; Chen and Struhl, 1988) have suggested that eukaryotic cells may contain more than one TATA box element and cognate factor, a distinction being made between consensus TATAAA elements and those carrying modifications (e.g. the TATTTAT of the early SV40 promoter). However, both Simon *et al.* (1988) and Nakajima *et al.* (1988) have cautioned that a single factor may interact with both TATA and TATA-related sequences, possible with different affinities, and give rise to different conformations that are differentially recognized by further transcription factors. Sequences flanking the TATA consensus may also be involved in the formation of such specific structures (Nakajima *et al.*, 1988).

We observe different DNA−protein interactions with *Drosophila* TATA boxes in our experiments with the *hsp70* TATA box probe. Although both *hsp70* and *Sgs3* carry the TATAAA consensus, the immediate flanking sequences, both 5′ and 3′, differ (Figure 2). While the *hsp70* probe clearly forms two complexes with embryonic extracts similar to those of *Sgs3*, the third, which comigrates with that which is specific to the *Sgs3* TATA probe with salivary gland extracts, is severely reduced or absent. The result with salivary gland extracts is complementary as the *hsp70* probe forms a novel complex, distinct from that of *Sgs3*. This suggests either that the *hsp70* probe is complexed with a factor in salivary gland extracts that differs from that associated with the *Sgs3* probe, or, as suggested above, the same factor is bound in a different conformation or is modified to a different extent prior to or following binding to different TATA sequences. These results, together with those of Martin *et al.* (1989b) raise the possibility that different TATA box factors may be incorporated into the complex necessary for *Sgs3* expression.

Interactions with Sgs3 upstream sequences

Recently we have shown that a deletion in the proximal upstream sequences (between −98 and −57) effectively eliminates *Sgs3* promoter activity (Martin *et al.*, 1989b), while a deletion in the distal regulatory region (centred at −600) severely reduces expression (<5%) (Ramain *et al.*, 1988). The present results, showing that the TATA motif is critical for expression, enable us to establish a hierarchy of interactions for *Sgs3* regulation. The simplest model is that factors bound to the proximal upstream sequences interact directly (or indirectly) with those of the TATA region. The activity of the cell- and stage-specific complex thus formed is enhanced by further interactions with factors bound to the distal regulatory region.

Such a model derives from studies in which interactions between upstream and TATA factors have been demonstrated by binding studies. In the case of the *Drosophila hsps*, the binding of the heat-shock transcription factor to the HSE is necessary for the activation of *hsps* (Wu, 1984b; Topol *et al.*, 1985; C.Wu *et al.*, 1987; Thomas and Elgin, 1988) and induces changes in the binding of a TATA factor. It is possible to change the components of these activation complexes. *In vivo* the *Drosophila hsp70* minimal promoter (lacking a HSE) interacts with heterologous upstream sequences to give developmentally specific expression [e.g. *ftz* (Hiromi and Gehring, 1987), *Adh* (Fischer and Maniatis, 1988), *Ddc* (Bray *et al.*, 1988) and *Sgs3* (Martin *et al.*, 1989b)]. Such a flexibility in promoter design has the advantage of allowing the assembly of promoters from relatively simple elements. Promoter fusions, e.g. the *Sgs3*−*hsp70* fusion, will be useful for a combined *in vivo*−*in vitro* study of the specificity of the DNA−protein interactions that establish the structure of such a promoter.

Materials and methods

Construction of transposons and isolation of transformed strains

The 1.3 kb *SalI*−*XhoI* Formosa genomic fragment containing *Sgs3* (see Figure 1) was subjected to oligonucleotide-directed mutagenesis using the oligonucleotides CGTACCTAGAAAAAGGTG or CGTACCTAAA-AAAAGGTG. The modified *SalI*−*XhoI* fragments were then inserted in C20g1ΔS (Bourouis and Richards, 1985b) to give C20g71G and C20g71A. Transposons (300 μg/ml) were co-injected with the integration-defective helper plasmid pπ25.7WC (300 μg/ml) into ry^{506} C.S. embryos and transformants selected as G1 ry^{+} flies. Chromosomal assignment and isolation of homozygous lines was by standard genetic crosses to appropriately marked balancer stocks. Southern analysis showed that each strain carried a single non-rearranged insertion of the transposon, except line A3 which carries two independent insertions (data not shown).

RNA analysis

RNA was extracted from third instar larvae and separated on formaldehyde−agarose gels as described previously (Richards *et al.*, 1983). After transfer to nitrocellulose the filters were hybridized with the nick-translated *Sgs3* cDNA clone pG3 and exposed for autoradiography. Activity was estimated by densitometric scanning of a series of autoradiographs as the film was rapidly saturated in the region of the resident allele (Figure 1C). See Bourouis and Richards (1985a) and Ramain *et al.* (1988) for discussion of the technique.

Embryonic nuclear extract

Nuclear extract was prepared from 1−2 g of fresh embryos following Wu (1984a,b), with minor modifications. Embryos were dechorionated in

Chlorox−0.02% Triton X-100 for 2 min at room temperature and then homogenized wth a B pestle on ice in 2 vol of freshly prepared solution A [0.3 M sucrose, 10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM EGTA pH 8.0, 0.5 mM DTT, 0.5 mM PMSF and protease inhibitors (pepstatin 5 g/ml, aprotinin 5 g/ml, leupeptin 30 g/ml, antipain 5 g/ml and chemiostatine 5 g/ml)]. The following steps were performed between 0 and 4°C. The homogenate was centrifuged for 1 min at 160 *g* in a swinging rotor to remove cuticle. The supernatant was centrifuged for 5 min at 2000 *g* to pellet crude nuclei which were resuspended in 2.5 vol solution A. An equal volume of solution A−1.7 M sucrose was added and the two phases were mixed to give a final concentration of 1 M sucrose. After centrifugation, 20 min at 7 K, the crude nuclear pellet was resuspended in 2−3 vol solution B (10 mM HEPES pH 7.9, 400 mM NaCl, 1.5 mM $MgCl_2$, 0.1 mM EGTA pH 8.0, 0.5 mM DTT, 5% glycerol, 0.5 mM PMSF and protease inhibitors as above). The NaCl concentration was adjusted to 0.4 M at the end. Nuclei were extracted by continuous stirring for 30 min and then centrifuged for 20 min at 14 K in a SS34 rotor. The supernatant was dialysed twice for 90 min against 100 ml of solution C (20 mM HEPES pH 7.9, 75 mM NaCl, 0.1 mM EDTA pH 8.0, 0.5 mM DTT, 20% glycerol, 0.5 mM PMSF) and cleared by centrifugation for 5 min at 14 K. Nuclear proteins were quantified (Bradford, 1976), frozen in liquid N_2 and stored at −80°C.

Salivary gland nuclear extract

Salivary glands (3−400 pairs), obtained by hand dissection for 1 h, were collected (in lots of 100) on ice in 100 μl of solution A. Homogenization used a resin-based pestle preformed to the shape of a 1.5 ml Eppendorf tube. The homogenates were pooled and centrifuged for 10 min at 2 K in a HB4 rotor to pellet the polytene nuclei. Lipids were removed with a tissue and the nuclei resuspended in 3 vol solution A. The sucrose concentration was adjusted to 1 M as above and all subsequent steps were as for the embryonic extract. Five to eight of these extracts were pooled to give one working extract. As salivary glands contain large amounts of endogenous nucleases, different protocols were tested (e.g. extraction and lysis buffer, time of dissection, sucrose concentration, etc.). The conditions chosen were the most satisfactory as judged by subsequent footprints, although endogenous nucleases were not completely inhibited. For both nuclear extracts all steps were monitored by optical microscopy to check the state of the nuclei.

DNase I footprinting assays

A pEMBL19 subclone carrying the *Sal*I (−130)−*Alu*I (+33) fragment was linearized at the *Hin*dIII site in the polylinker, treated with phosphatase, labelled at its 5′ extremities with [γ-^{32}P]ATP and phage T4 polynucleotide kinase and then recut with *Eco*RI at the other end of the polylinker (Figure 2). For the lower strand, labelling was at the *Eco*RI site and redigestion at the *Hin*dIII site. The labelled fragments were separated on polyacrylamide gels and electroeluted.

Nuclear extract (15−30 μg) was preincubated for 15 min at 0°C with 20−50 ng of denatured salmon sperm DNA and salts (1.6 mM $MgCl_2$, 25 mM KCl, final concentrations). Then, 5−15 000 c.p.m. (0.4 ng) of labelled fragment were added and incubated for 10 min at 20°C. DNase I was added and digestions were carried out for 90 min at 20°C. The amounts of DNase I were determined empirically. Digestion was stopped with 200 μl 0.3% SDS−150 mM NaCl. Proteins were extracted with phenol−chloroform and DNA was precipitated with 2.5 vol EtOH and 50 μl 3 M NaoAc contaning 10 μg *Escherichia coli* tRNA. For reactions on naked DNA, DNase I digestion was for 60 s at 20°C. Samples were heated for 2 min at 100°C before loading onto a 6 or 8% polyacrylamide sequencing gel.

Gel retardation assays

Assays were as described by Strauss and Varshavsky (1984) with minor modifications. Optimal binding was obtained in the absence of Mg^{2+}. A typical reaction contained 2−4 μg of nuclear extract, 1 μg of poly(dI·dC), 30 mM KCl in a final volume of 10 μl, completed with solution C. Preincubation was on ice for 15 min, before the addition of 9−16 fmol (1.5×10^5 c.p.m.) of the ^{32}P 5′ end-labelled DNA template. Incubation was at 20°C for 10 min and the DNA−protein complexes were then loaded onto a low ionic strength 5% polyacrylamide gel (30:1 cross-linking ratio) containing 6.7 mM Tris−HCl pH 7.5, 3.3 mM sodium acetate and 1 mM EDTA. Electrophoresis was 25 mA for 2 h with recirculation of buffer. In later experiments (e.g. Figure 6) band separation was improved by varying the rate of recirculation. The gel was soaked in 5% glycerol and dried for autoradiography. The relative intensity of bands varies with different NEs (see Results). In the figures each panel derives from a single gel. A space is left between panels showing results from related experiments that used different NEs. In both DNase I footprinting and gel retardation competition experiments, unlabelled double-stranded oligonucleotide was added during the preincubation step.

Acknowledgements

We thank Professor P.Chambon for his continuous support. We thank Frank Ruffenach and Adrien Staub for the synthesis of oligonucleotides. We are grateful to Bohdan Wasylyk for critically reading a draft of this manuscript, which was prepared with the help of B.Boulay. This work was supported by grants from the CNRS, the INSERM, the Fondation pour la Recherche Médicale, the Ministère de la Recherche et de l'Enseignement Supérieur and the Association pour la Recherche sur le Cancer. C.M. was funded by the Ministère de la Recherche et de la Technologie, M.M. by the Université Louis Pasteur, Strasbourg. A.G. is funded by the Consiglio Nazionale delle Ricerche Progetto Finalizzato 'Ingegneria Genetica' U.O.47 no. 8686/51, Istituto di Genetica, Università di Bari.

References

Bourouis,M. and Richards,G. (1985a) *Cell*, **40**, 349−357.
Bourouis,M. and Richards,G. (1985b) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 355−360.
Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248−254.
Bray,S.J., Johnson,W.A., Hirsh,J., Heberlein,U. and Tjian,R. (1988) *EMBO J.*, **7**, 177−188.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349−383.
Bucher,P. and Tifonov,E.N. (1986) *Nucleic Acids Res.*, **14**, 10009−10026.
Buratowski,S., Hahn,S., Sharp,P.A. and Guarente,L. (1988) *Nature*, **334**, 37−42.
Cavallini,B., Huet,J., Plassat,J.-L., Sentenac,A., Egly,J.-M. and Chambon,P. (1988) *Nature*, **334**, 77−80.
Chen,W. and Struhl,K. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2691−2695.
Davison,B.L., Egly,J.-M., Mulvihill,E.R. and Chambon,P. (1983) *Nature*, **301**, 680−686,
Fischer,J.A. and Maniatis,T. (1988) *Cell*, **53**, 451−461.
Giangrande,A., Mettling,C. and Richards,G. (1987) *EMBO J.*, **6**, 3079−3084.
Grosschedl,R., Wasylyk,B., Chambon,P. and Birnstiel,M.L. (1981) *Nature*, **294**, 178−180.
Heberlein,U., England,B. and Tjian,R. (1985) *Cell*, **41**, 965−977.
Hiromi,Y. and Gehring,W.J. (1987) *Cell*, **50**, 963−974.
Hirsh,J., Morgan,B.A. and Scholnick,S.B. (1986) *Mol. Cell. Biol.*, **6**, 4548−4557.
McGinnis,W., Shermoen,A.W. and Beckendorf,S.K. (1983) *Cell*, **34**, 75−84.
Martin,M., Mettling,C., Giangrande,A., Ruiz,C. and Richards,G. (1989a) *Dev. Genet.*, **10**, 189−197.
Martin,M., Giangrande,A., Ruiz,C. and Richards,G. (1989b) *EMBO J.*, **8**, 561−568.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560−564.
Myers,R.M., Tilly,K. and Maniatis,T. (1986) *Science*, **232**, 613−618.
Nakajima,N., Horikoshi,M. and Roeder,R.G. (1988) *Mol. Cell. Biol.*, **8**, 4028−4040.
Orkin,S.T., Sexton,J.P., Cheng,T., Goff,S.C., Giardina,P.J.V., Lee,J.I. and Kazazian,H.H. (1983) *Nucleic Acids Res.*, **11**, 4727−4734.
Parker,C.S. and Topol,J. (1984a) *Cell*, **36**, 357−369.
Parker,C.S. and Topol,J. (1984b) *Cell*, **37**, 273−283.
Posakony,J.W., Fischer,J.A. and Maniatis,T. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 515−520.
Ramain,P., Giangrande,A., Richards,G. and Bellard,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2718−2722.
Richards,G., Cassab,A., Bourouis,M., Jarry,B. and Dissous,C. (1983) *EMBO J.*, **2**, 2137−2142.
Rubin,G.M. and Spradling,A.C. (1983) *Nucleic Acids Res.*, **11**, 6341−6351.
Sawadogo,M. and Roeder,R.G. (1985) *Cell*, **43**, 165−175.
Simon,M.C., Fisch,T.B., Benecke,B.J., Nevins,J.R. and Heintz,N. (1988) *Cell*, **52**, 723−729.
Snyder,M.P., Kimbrell,D., Hunkapiller,M., Hill,R., Fristrom,J. and Davidson,N. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7430−7433.
Solomon,M.J., Strauss,F. and Varshavsky,A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1276−1280.
Strauss,F. and Varshavsky,A. (1984) *Cell*, **37**, 889−901.
Thomas,G.H. and Elgin,S.C. (1988) *EMBO J.*, **7**, 2191−2201.
Topol,J., Ruden,D.M. and Parker,C.S. (1985) *Cell*, **42**, 527−537.
Wasylyk,B. (1988) *CRC Crit. Rev. Biochem.*, **23**, 77−120.

Wu,C. (1984a) *Nature*, **309**, 229–234.
Wu,C. (1984b) *Nature*, **311**, 81–84.
Wu,C. (1985) *Nature*, **317**, 84–87.
Wu,C., Wilson,S., Walker,B., Dawid,I., Paisley,T., Zimarino,V. and Ueda,H. (1987) *Science*, **238**, 1247–1253.
Wu,L., Rosser,D.S.E., Schmidt,M.C. and Berk,A. (1987) *Nature*, **236**, 512–515.
Zajchowski,D.A., Boeuf,H. and Kédinger,C. (1987) *Gene*, **58**, 243–256.

Received on April 20, 1989; revised on July 13, 1989