The EMBO Journal Vol.2 No.9 pp.1431 – 1440, 1983

Coordinately expressed chorion genes of *Bombyx mori*: is developmental specificity determined by secondary structure recognition?

Kostas Iatrou* and Sonia G. Tsitilou[1]

Department of Medical Biochemistry, Faculty of Medicine, The University of Calgary, Calgary, Alberta T2N 1N4, Canada and [1]Department of Biology, Athens University, Panepistimiopolis, Athens 621, Greece

Communicated by T.Jovin
Received on 11 February 1983; revised on 30 May 1983

Short inverted repeat sequences have been observed in the DNA flanking the 5′ and 3′ termini of a pair of coordinately expressed chorion structural genes of the silkmoth *Bombyx mori*. When superhelical cloned DNA containing the two chorion genes is digested with S1 nuclease, several sites are specifically cleaved including those around the centers of the putative cruciform structures resulting from the short inverted repeat sequences. The possible implication of conformational changes in the DNA surrounding the chorion structural genes in the process of determination of the developmental specificity and the coordinate transcriptional regulation of these genes is discussed.

Key words: Bombyx mori/chorion genes/developmental specificity

Introduction

The chorion (eggshell) of the silkmoth follicle is a complex structure composed of >150 proteins which are the products of distinct structural genes whose regulation is under strict developmental control (Kafatos *et al.*, 1977). Chorion proteins are evolutionarily related and chorion genes are members of a small number of multigene families, A, B, C and high-cysteine (Hc) (Regier *et al.*, 1978a, 1978b; Rodakis, 1978; Nadel and Kafatos, 1980; Iatrou *et al.*, 1982; Rodakis and Kafatos, 1982; Rodakis *et al.*, 1982).

Genetic studies on the chorion system of the commercial silkmoth *Bombyx mori* have shown that the majority, and probably all, of the chorion genes are clustered within a few map units on chromosome 2 of the moth (Goldsmith and Basehoar, 1978; Goldsmith and Clermont-Rattner, 1979). The clustering of chorion genes occurs according to the timing of transcriptional expression, genes of late developmental specificity, for example, being clustered together and separately from genes of middle or early developmental specificity (Iatrou *et al.*, 1980). These conclusions have been confirmed and extended by chromosomal walking along a *Bombyx* chromosome locus bearing a large number of chorion structural genes (Eickbush *et al.*, 1981; Eickbush and Kafatos, 1982). The global organization of the chorion genes of *Bombyx* was shown to consist of a large number of gene pairs that probably originated from repeated duplications and extensive diversification of an ancestral gene. As a rule, in the part of the chorion locus containing genes of middle developmental specificity, a member of the A multigene family is paired with a member of the B family, while in the late region both genes in each pair appear to belong to the Hc multigene family (Iatrou *et al.*, 1980; Eickbush and Kafatos, 1982). In general, both members of each pair appear to be coordinately regulated. Thus, timing of expression rather than evolutionary descent seems to be the basis of genomic chorion gene organization.

The general features of chorion genomic organization in *B. mori* appear to be very similar to the characteristics of the global organization described previously for the chorion genes of the American silkmoth *Antheraea polyphemus* (Jones and Kafatos, 1980). Although extensive analysis has been performed on several pairs of coordinately regulated genes of *Antheraea* (Jones and Kafatos, 1980), no such studies have yet been reported for the chorion genes of *Bombyx*.

Here we report part of the detailed characterization of one pair of coordinately expressed *Bombyx* chorion genes. We have paid particular attention to the features of the DNA sequences flanking the origins of the chorion structural genes, because such sequences may play an important role in controlling the transcriptional activation and in determining the developmental specificity of these genes. We have detected short regions of dyad symmetry around the origins of the two structural genes. These regions have the potential of inducing the appearance of mirror stem/loop structures at the gene origins. Both the structural genes and their putative control sequences contribute to the postulated secondary structures. Finally, we present evidence that such cruciform structures can exist in superhelical cloned DNA and that they are specifically recognized and cleaved by S1 nuclease.

Results

Isolation and characterization of pairs of coordinately expressed chorion genes

A chromosomal DNA library from wild-type *B. mori* genomic DNA was screened for clones containing specific chorion genes through hybridizations to a number of well characterized chorion cDNA clones. Three chorion cDNA clones of late developmental specificity, m2132, m2282 and m2574 (Iatrou *et al.*, 1982), were used as probes to identify genomic clones containing the corresponding or closely related genes. The cDNA clones we used correspond to sequences sharing very limited, if any, sequence homologies (Iatrou *et al.*, 1982), and all of them correspond to mRNAs whose synthesis is abolished by the Gr^B mutation (Iatrou *et al.*, 1980). One genomic clone, clone 150, which appeared to hybridize to all three probes, was further characterized. Clone 150 contains four chorion genes organized in two pairs of divergently orientated transcriptional units (Figure 1). The location of the chorion structural genes was determined by Southern hybridizations and R-loop analysis, and the transcriptional orientations by hybridizations of follicular RNA to single-end-labeled DNA fragments containing the chorion structural genes (Tsitilou, Kafatos and Iatrou, in preparation).

The *Eco*RI fragments of clone 150 were subcloned at the single *Eco*RI site of pBR322 and the subclones were characterized by hybridization, electron microscopy and nucleotide sequencing (Tsitilou *et al.*, in preparation). Three subclones were used in this study: subclone 8 containing the

*To whom reprint requests should be sent.

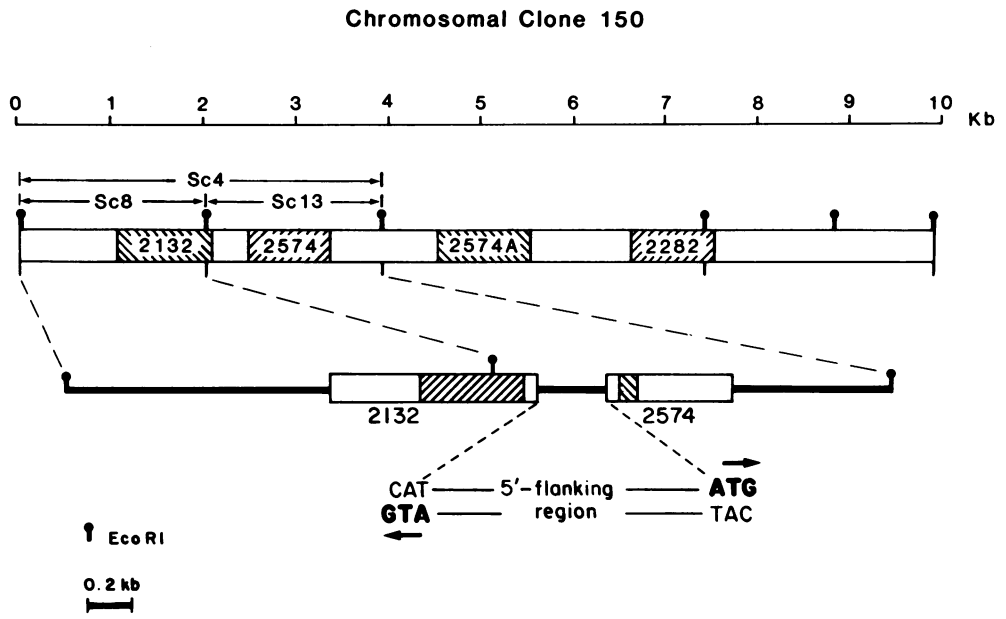

Fig. 1. Chorion gene organization in chromosomal clone 150. The arrangement of two chorion gene pairs and the transcriptional orientation of the genes in one of the two pairs (2132/2574) is shown. Hatched segments on genes 2132 and 2574 indicate the corresponding intervening sequences. The common 5′-flanking sequence has a length of 268 bp (cap site of gene 2132 to cap site of gene 2574). Subclones used in this study are shown on clone 150 with sc4, sc8 and sc13. Heavy arrows indicate transcriptional orientation.

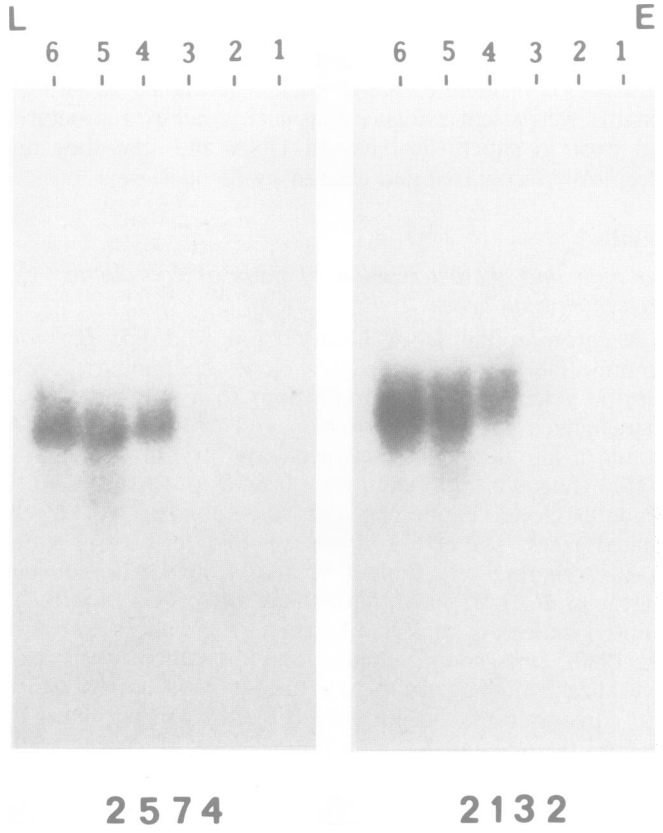

Fig. 2. The developmental specificity of genes 2132 and 2574. Northern hybridizations of stage-specific follicular RNA resolved on methylmercury agarose gels to portions of clone 150 containing the coding sequences of genes 2132 and 2574. Numbers 1–6 indicate choriogenic stages from early (E) to late (L).

left 1.9-kb fragment of clone 150 (gene 2132), subclone 13 comprising the adjacent 1.9-kb fragment (part of gene 2132 and gene 2574) and subclone 4 containing both of the above fragments in the order that they occur in clone 150 (Figure 1).

The developmental specificity of genes 2132 and 2574 was examined by Northern hybridizations of the chorion genes to stage-specific follicular RNA preparations (Figure 2). Messenger RNA sequences complementary to both genes appear in the follicular cells for the first time at stage 4 of choriogenesis, increase at stage 5 and peak at stage 6 (for stage definitions see Iatrou *et al.*, 1982). It appears, therefore, that genes 2132 and 2574 are coordinately expressed.

The entire nucleotide sequence of subclone 13 was determined (see Appendix). Subclone 8 was also sequenced from the side neighbouring subclone 13 (containing gene 2132, data not shown). The results of the gene localizations and transcriptional orientations are shown schematically in the map of Figure 1. Each structural gene contains a relatively short intervening sequence near the 5′ terminus. The size of the intervening sequences is not the same for both genes (80 bp and 482 bp for genes 2574 and 2132, respectively), but their location is the same, with the intron/exon splicing junctions being located between amino acids 17 and 18 of the respective signal peptides, in a manner reminiscent of the intron sequences of the chorion genes of *A. polyphemus* (Jones and Kafatos, 1980). Finally, both chorion genes were found to belong to the Hc multigene family, but each one to a distinct subfamily (Rodakis and Kafatos, 1982; Iatrou and Tsitilou, in preparation).

Analysis of the 5′-flanking sequences of the gene pair 2132/2574

Of particular interest to us is the common developmental regulation of this pair of genes (Figure 2). It is possible that control elements involved in the determination of the developmental specificity and the common transcriptional regulation of these chorion genes reside in the DNA sequences flanking their 5′ termini. To see whether common sequence elements flank both genes in the pair, the two strands of the 268 bp of DNA flanking the starts of the two

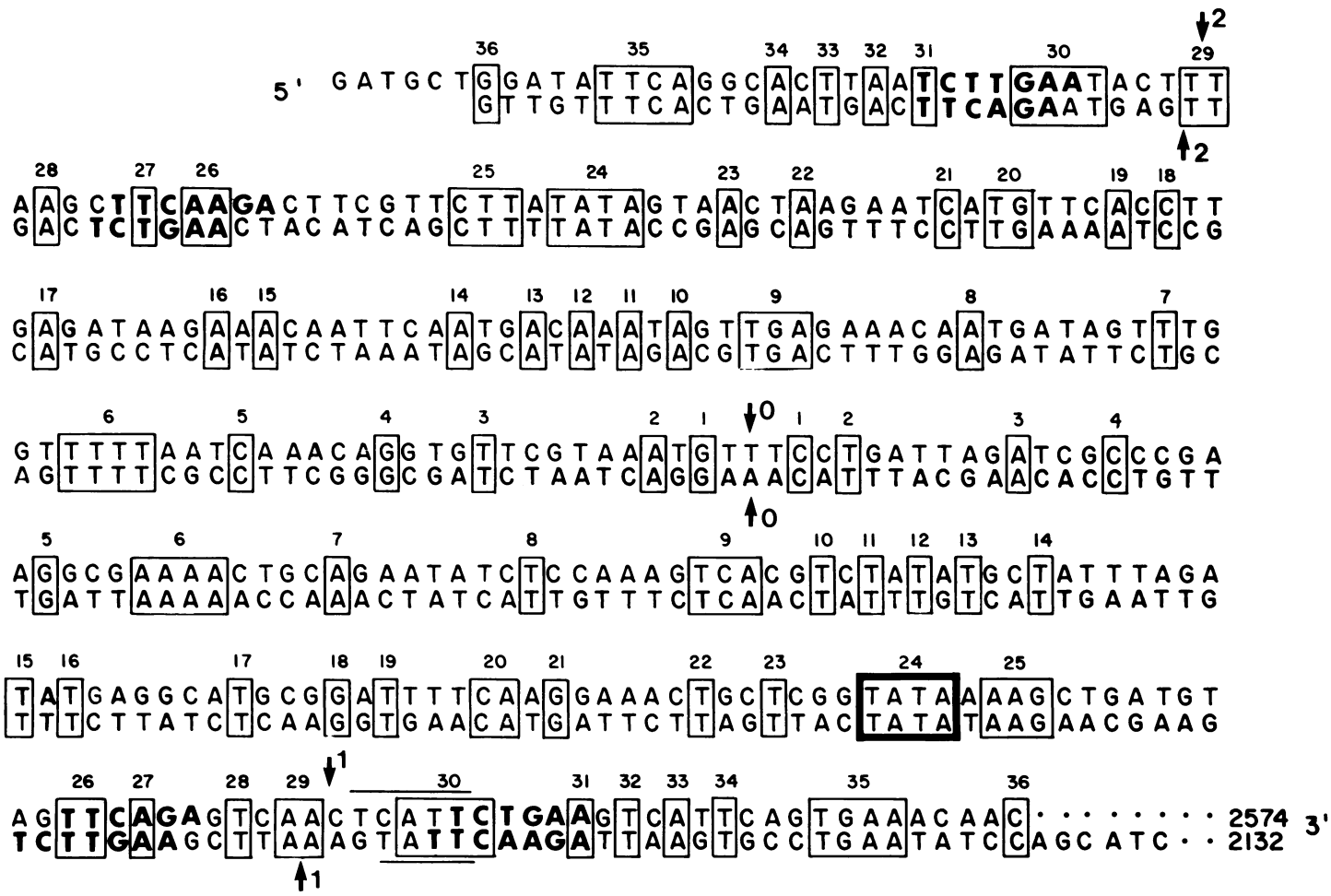

Fig. 3. The nucleotide sequences of the 5′-flanking sequences of genes 2132 and 2574. Both sequences are written in the 5′- to 3′-orientation and also comprise the entire 5′-non-coding regions of the two genes before the ATG codons for the start of mRNA translation. Center 0 (arrows) represents the middle of the sequences and is the center of dyad symmetry for the nucleotides that are the same for both genes (boxes). Centers 1 (arrows) represent the axes of dyad symmetry for the nucleotides of the short inverted repeat sequences (bold letters). Centers 2 (arrows) are the centers of dyad symmetries whose presence is dictated by the presence of centers 1 in the complementary strands. The TATA boxes of the two genes used as sequence alignment points are boxed with heavy lines. Bars indicate the cap sites of the two genes.

genes were compared with each other in the 5′- to 3′-orientation. The 5′-flanking and the 5′-non-coding sequences for both genes at the same orientation are shown in Figure 3. The two sequences have been aligned with their TATA boxes as the point of register. Center 0 represents the middle of the common 5′-flanking sequence. The control sequences (downstream from center 0) share only a minimum degree of overall homology (33.4%, boxes in Figure 3), quite insignificant in view of the base composition of the DNA sequences (59.26% and 69.05% A+T content for gene 2574 and gene 2132 flanking sequence, respectively). However, the presence of a short palindrome-like subsequence downstream from the TATA box was observed in both genes (bold type letters in Figure 3). The nucleotides involved in the symmetries differ from gene to gene. For each of the two genes the dyad symmetry involves a portion of the putative control sequence downstream from the TATA box and part of the 5′-non-coding sequence of the structural gene itself. Finally, both symmetries are located at equivalent positions with respect to the TATA boxes and the cap sites of the genes [determined through S1 protection experiments (Berk and Sharp, 1978), data not shown].

The presence of dyad symmetries at the beginning of the two genes dictates the presence of complementary symmetries at the opposite ends of the common 5′-flanking sequence (Figure 3). Center 0 is, therefore, the center of dyad symmetry for the nucleotides that are the same in the two control regions. Centers 2 are the centers of dyad symmetries dictated by the presence of centers 1 in the putative promoter sequences of genes 2132 and 2574.

The symmetries illustrated in Figure 3 predict the possibility of cruciform structures with similar features for both genes (Figure 4). The two hairpin structures at the beginning of the structural genes may be of potential significance in the regulation of their transcriptional activity (see Discussion).

Specific sites in the vicinity of the chorion structural genes are recognized by S1 nuclease

To see whether the postulated cruciform structures at the origins of genes 2132 and 2574 can be induced by supercoiling, covalently closed circular (ccc) plasmid DNA from subclone 13 was digested with S1 nuclease. The resulting digests were subsequently restricted with *Eco*RI nuclease, resolved on agarose gels, and Southern transferred to a solid support for hybridizations. Fragments representing various regions of the chorion DNA insert or the entire pBR322 portion of subclone 13 were used as radioactive probes for the hybridizations. In Figure 5 typical results from such an ex-

Fig. 4. Secondary structures predicted from the presence of short inverted repeat sequences. The cruciform structures predicted from the presence of inverted repeat sequences at the promoter/5′-non-coding region junctions of the genes (illustrated in Figure 3) are shown. The TATA boxes and the structural portions of both genes (RNA strands, 3′ downstream from the cap sites) are in bold letters. The sequences of the structural genes are written up to the first ATG codons. Note that the TATA boxes and the cap sites of the two genes are at equivalent positions outside the hydrogen bonded regions of the stems and that the 5′-non-coding portions of the structural genes participate in the predicted secondary structures. Arrows indicate cleavage sites for *Hind*III and *Hinc*II used for the excision of the presumed regulatory sequence in the experiment illustrated in Figure 7.

periment are presented. The degree to which subclone 13 DNA was digested by S1 nuclease is shown by the ethidium bromide staining pattern of the DNA obtained at the end of the digestions (Figure 5A). The superhelical DNA was digested to an extent resulting in, ~10%, 50% or 90% conversion to the linear form, the remainder being converted to relaxed (nicked) circle. The S1-treated DNAs were further restricted with *Eco*RI endonuclease (Figure 5B). In Figure 5C the results of Southern hybridizations of the digests shown in Figure 5B to six different probes are shown. It is clear that S1 nuclease cleaves both the pBR322 portion and the chorion insert of subclone 13 in a non-random manner which results in the appareance of discrete hybridizing fragments.

We calculated the lengths of all the major hybridizing fragments and identified the positions of the S1 nuclease cleavage sites along the chorion insert (Figure 6B). We analyzed, subsequently, the sequences surrounding the cleavage sites for the presence of short inverted repeat sequences and found that the cleavage sites represent loop areas in putative cruciform structures (Figure 6A). The lengths of the stems and loops in those structures varied from 5−7 bp and 1−25 nucleotides, respectively.

The first of the predicted cruciform structures in the control region of gene 2132 (Figure 4) appears to be recognized efficiently by S1 nuclease yielding an S1/*Eco*RI hybridizing fragment of 220 bp (bottom dot of first two hybridization sets in upper panel of Figure 5C; Figure 6B), while the second one, at the origin of gene 2574, appears also to be cleaved by S1 nuclease (0.5 kb S1/*Eco*RI hybridizing fragment, bracketed dot in upper panel of Figure 5C; Figure 6B) although with a much reduced efficiency. This reduced efficiency may be due to the presence of another short inverted repeat sequence 50 bp upstream from the one at the cap site of gene 2574. This latter inverted repeat sequence resulting in the formation of another putative cruciform stucture is cleaved very efficiently by S1 nuclease yielding a 410-bp hybridizing fragment (second dot from bottom in first two hybridization sets in upper panel of Figure 5C; Figure 6B). Finally, a number of other sites appear to be cleaved by the enzyme yielding additional discrete hybridizing fragments (see Discussion). Subclone 13 DNA linearized by *Eco*RI and subsequently digested with S1 nuclease does not yield any hybridizing fragments besides the 4.36-kb pBR322 and 1.93-kb insert bands (Figure 5D). Therefore, the postulated structures can be induced in superhelical DNA only, and the resulting S1 nuclease fragments are not due to other nuclease activities present in the enzyme preparation used in these experiments.

Developmental specificity of the 5′-flanking sequences

An additional experiment was undertaken to correlate the putative promoter region of genes 2132 and 2574 with the timing of their transcriptional activation. The DNAs of 21 *B. mori* clones spanning 270 kb of the chorion locus (Eickbush and Kafatos, 1982) were probed with three distinct probes by a modification of the Benton and Davies (1977) procedure (Figure 7). Panel A of Figure 7 shows the clones hybridizing to a chorion sequence belonging to the B chorion multigene family and having middle developmental specificity. Panel B indicates the clones hybridizing to a member of the Hc proteins having late developmental specificity. In panel C the results of the hybridization of the genomic clones to the excised promoter region are shown. The sequence excision points (arrows in Figure 4) are five and three nucleotides upstream from the cap sites of genes 2132 and 2574, respectively. Thus, the probe is completely devoid of any structural gene sequences. It is clear that the 5′-flanking sequence of the 2132/2574 gene pair is homologous to sequences of clones of late developmental specificity only. No hybridization is observed with any of the clones that are expressed at the middle stages of choriogenesis. Finally, in panel D, the results of the hybridization of the genomic clones to the same promoter region under more stringent conditions are presented. Only two clones, both containing the 2132/2574 gene pair from which the probe originated, hybridize very efficiently. A third clone also hybridizes to the probe, but less intensely. Therefore, the control elements associated with the gene pairs of late developmental specificity are similar to one another but not identical to that of the 2132/2574 pair.

Discussion

The chromosomal clone used in this study spans 10 kb of the Gr^B deletion of the *B. mori* chorion locus and contains four Hc chorion genes organized in two pairs (Tsitilou *et al.*, in preparation). All four genes are expressed during the last half of choriogenesis (Iatrou *et al.*, 1982) and the members of each pair were found to be transcribed in opposite and

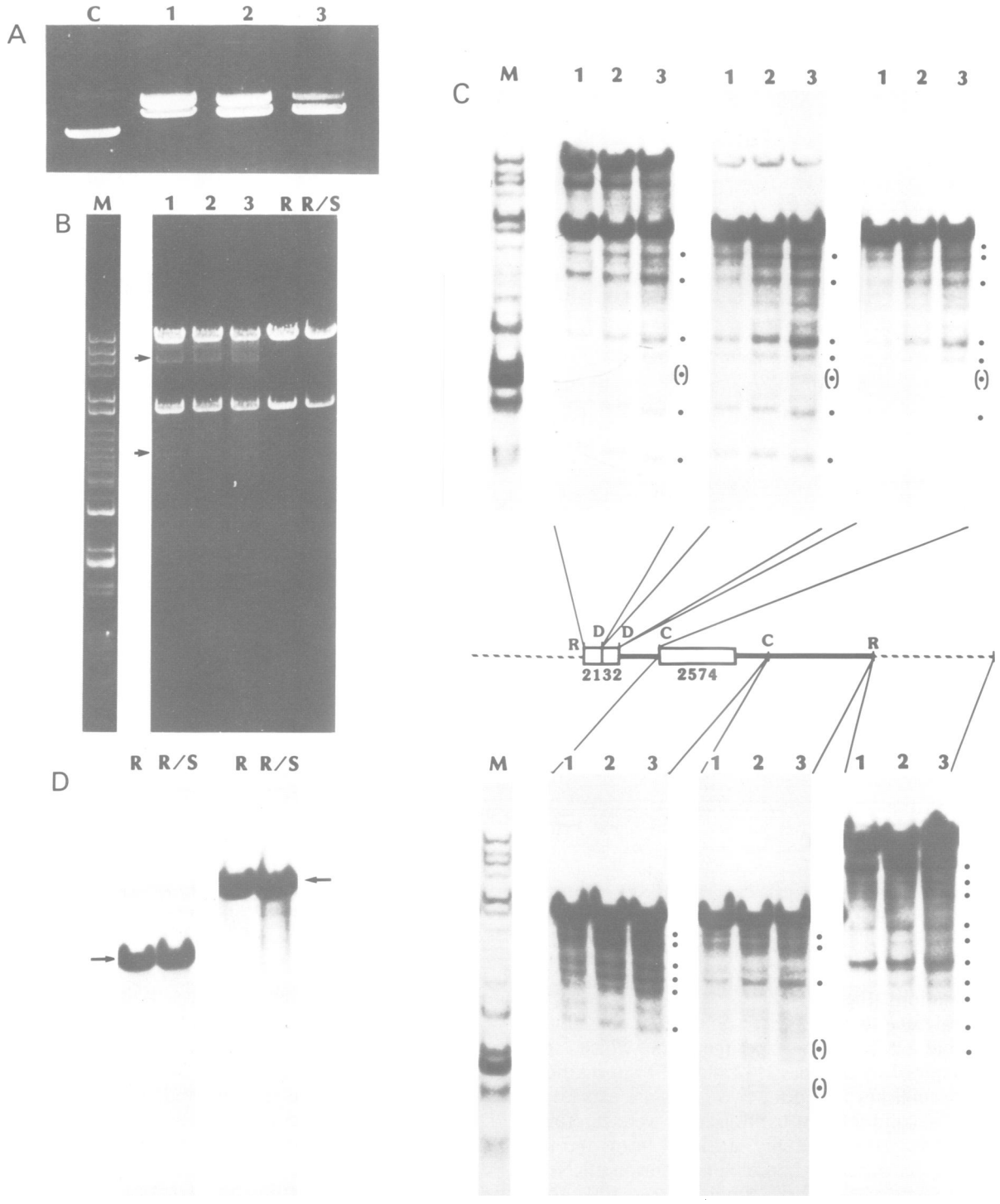

Fig. 5. S1 nuclease cleavage of chorion and pBR322 DNA. S1 nuclease digestions were carried out at 24°C in the presence of 150 mM NaCl as described in the experimental procedures. **Panel A**: ethidium bromide staining pattern of subclone 13 DNA digested to different degrees with S1 nuclease and analyzed on a 1% agarose gel; **lane C**: undigested ccc control; **lanes 1, 2** and **3**: S1 digestions converting the ccc DNA to ~10%, 50% and 90% linear form, respectively. **Panel B**: the DNA from the S1 digestions illustrated in **panel A** was subsequently digested with *Eco*RI and the digests resolved on a 1.5% agarose gel and stained with ethidium bromide; **lane M:** *Hind*III/*Hinc*II digest of phage lambda DNA mixed with a *Hae*III digest of pMB9 DNA and used as mol. wt. markers; **lanes 1, 2** and **3:** *Eco*RI digests of subclone 13 DNA digested with S1 nuclease to 10%, 50% and 90% linear conversion; **lane R**: subclone 13 DNA treated with *Eco*RI nuclease only (no S1 digestion); **lane R/S**: subclone 13 DNA cleaved first with *Eco*RI and then with S1 nuclease; arrows indicate the positions of the fragments resulting from the S1 digestion of the pBR322 cruciform structure reported by Panayotatos and Wells (1981). **Panel C**: Southern hybridizations of the S1/*Eco*RI-treated subclone 13 DNA shown in **panel B** to nick-translated portions of the chorion DNA insert and to the entire pBR322 portion of subclone 13 as shown in the map in the middle of the panel; **lanes 1, 2, 3** and **M** are as in **panel B**; boxes in the map indicate genes 2132 (5′ portion) and 2574; thick lines, the *B. mori* DNA flanking the structural genes; broken lines, the pBR322 moiety of subclone 13; dots and dots in parentheses indicate the positions of the hybridizing fragments discussed in the text (Results and Discussion). Note that with progressing S1 nuclease digestion the length of the hybridizing fragments decreases, presumably due to the trimming of the nucleotides in the loops. R, D and C on the map indicate cleavage points of *Eco*RI, *Hind*III and *Hinc*II used for generating the radioactive probes. **Panel D**: Southern hybridizations of control digests of subclone 13 DNA [*Eco*RI alone (R) or *Eco*RI followed by S1 nuclease (R/S)] to the entire chorion insert **(left)** and the pBR322 moiety **(right)** of subclone 13. S1 nuclease does not recognize specific sites on linear subclone 13 DNA. Arrows indicate the positions of the chorion insert and pBR322 portions of subclone 13 hybridizing to the probes.

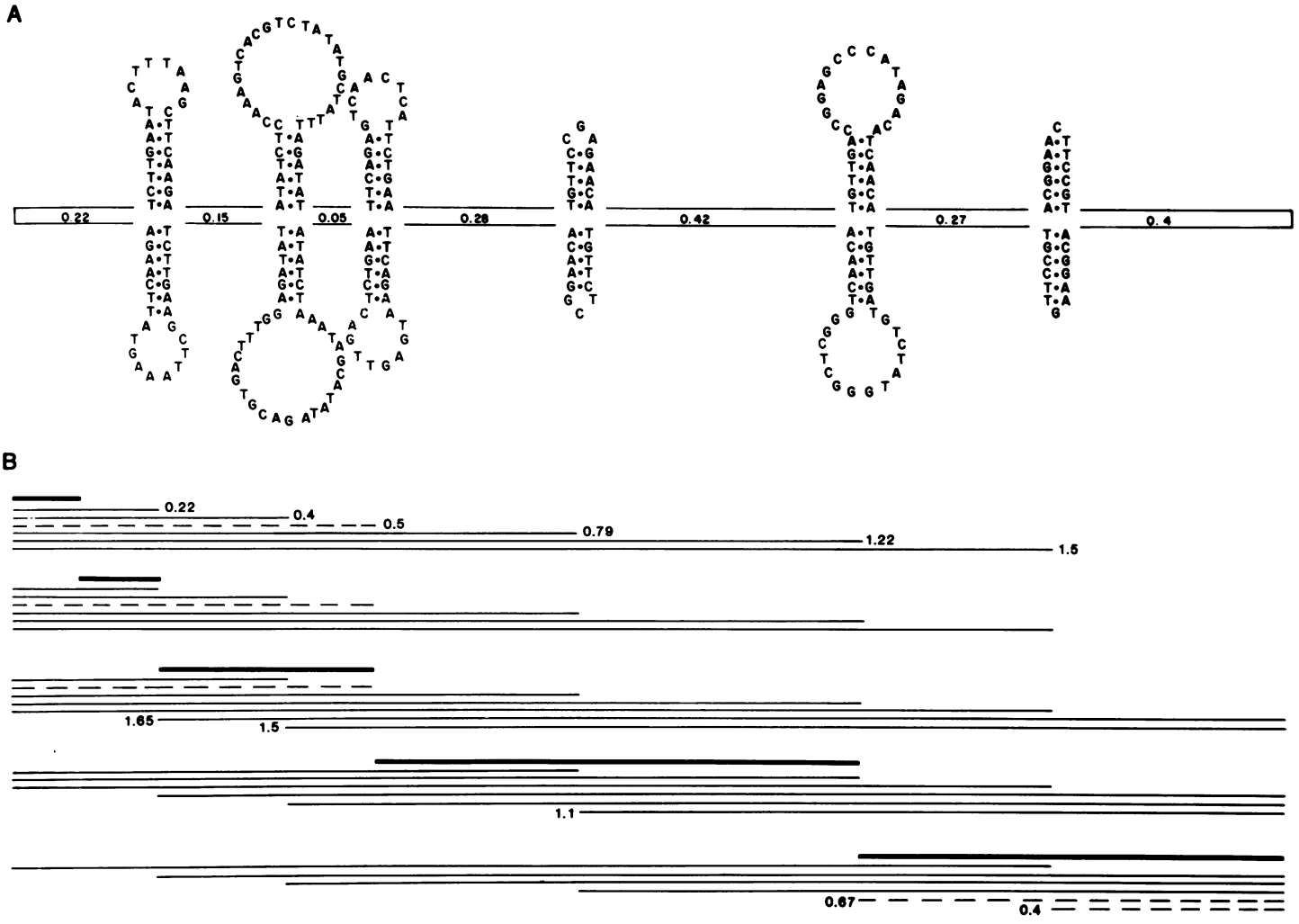

Fig. 6. S1 nuclease cleaves cloned supercoiled chorion DNA at sites flanked by short inverted repeat sequences. Schematic representation of the chorion insert of subclone 13 DNA indicating the sequences around the points recognized and cleaved by S1 nuclease (**panel A**). **Panel B** illustrates the length of the predominant (solid lines) and minor (broken lines) fragments hybridizing to the radioactive probes (thick lines) in the experiment shown in Figure 5C. Numbers in **panels A** and **B** indicate kb of DNA length.

divergent orientations (Figure 1).

Our structural analysis was focused on the pair containing genes 2132 and 2574. The experiment in which the developmental specificity of genes 2132 and 2574 was examined (Figure 2), establishes that these two genes are expressed coordinately. The conditions of hybridization were carefully chosen, so that complications arising from cross-hybridizations of our probes to homologous chorion mRNAs were minimal. Our hybridization conditions – 50% formamide, 3 x SET, 70°C – are sufficiently stringent ($\sim 3-5$°C below T_m) to ensure that only true hybrids are allowed to form and that cross-hybridizations to closely related species are practically eliminated (Iatrou *et al.*, 1982; Kafatos *et al.*, 1979, 1981).

It is possible that the common developmental regulation of genes 2132 and 2574 as well as the regulation of genes in other pairs of the chorion locus is determined by interactions between the shared 5′-gene flanking sequences of such pairs and specific factors appearing in the follicular cells at defined choriogenic stages. The examination of the presumptive control sequence of genes 2132 and 2574 led to the determination of the centers of dyad symmetry illustrated in Figure 3. Under appropriate conditions, these centers of symmetry could allow the 5′-non-coding sequences of the two structural genes and their 5′-flanking sequences to assume mirror structures. The symmetries are such that the location of the TATA boxes and the cap sites are equivalent in both genes (Figure 4). If a regulatory molecule is involved in the recognition of transcriptional structures/signals of one gene, the same molecule should be able to recognize mirror image structures/signals on the other.

The predicted hairpin/loop structure at the origin of gene 2132 is cleaved quite efficiently by S1 nuclease (220-bp hybridizing fragment, bottom dot of upper half hybridizations in Figure 5C). The second predicted structure at the origin of gene 2574 (500-bp fragment, bracketed dot of the upper half hybridizations in Figure 5C) is cleaved less efficiently than the first one. We believe that the reduced cleavage of the 2574 gene stem/loop structure may be due to the presence of an S1 cleavage site centered ~80 bp upstream from the cap site of gene 2574. This latter site appears to be cleaved very efficiently by S1 nuclease. An observation supportive of our reasoning for the reduced susceptibility of the predicted hairpin/loop structure at the origin of gene 2574 to S1 nuclease attack comes from the results obtained recently by Lilley (1981). He showed that the introduction of a ColE1

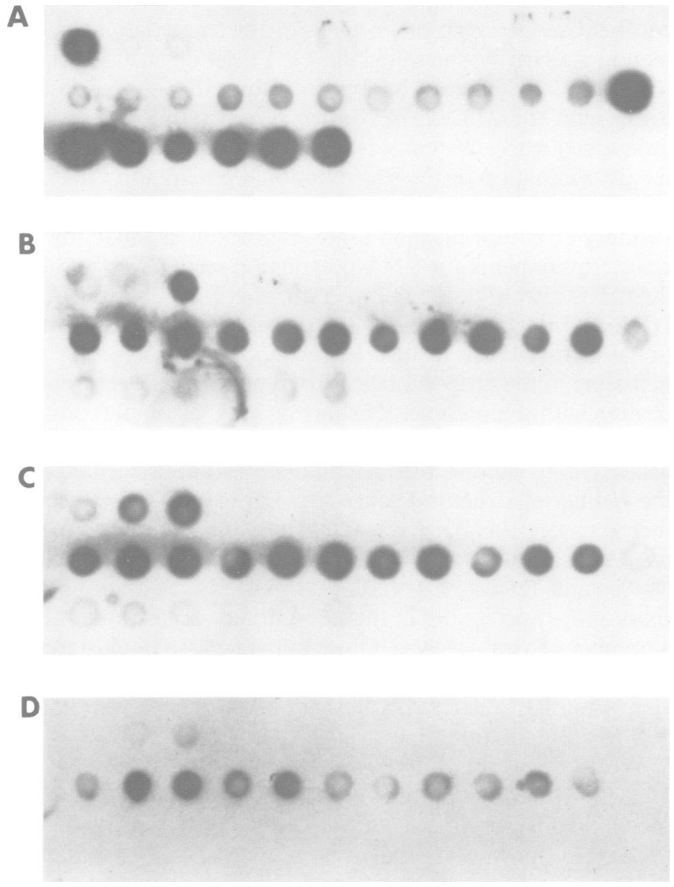

Fig. 7. The 5′-flanking region of the 2132/2574 gene pair is related to genes of the same developmental specificity. Hybridizations of genomic clones spanning 270 kb of the chorion locus (Eickbush and Kafatos, 1982) to three different radioactive probes. Phage lysates from 21 overlapping chromosomal clones were aliquoted in the wells of a microtiter plate. A bacterial lawn was subsequently infected with the lysates transferred from the microtiter wells with the aid of a replicator. Following an overnight incubation, plaques were lifted onto nitrocellulose filters according to Benton and Davies (1977). The order of the 21 overlapping chorion genomic clones is the same in all panels; from left to right, upper row: clones 1097, 285 and 645; middle row: clones 610, 150, 444, 654, 609, 4, 109, 781, 961, 791, 374 and 410; lower row: clones 345, 1285, 1035, 1142, 281 and 80. **Panel A**: hybridization to a cDNA clone belonging to the B multigene family (Iatrou *et al.*, 1982) – all genomic clones containing chorion genes of middle developmental specificity hybridize. **Panel B**: hybridization to a cDNA clone representing a high cysteine sequence expressed late in choriogenesis – only genomic clones of late developmental specificity hybridize. **Panel C**: hybridization to the 5′-flanking region of the late gene pair 2132/2574 (sequence comprised between arrows in Figure 4) at 70°C – only clones containing genes of late developmental specificity hybridize. **Panel D**: hybridization as in **Panel C** but at 78°C. The control elements associated with the late genes are similar but not identical to that of the 2132/2574 pair.

inverted repeat into pBR322 DNA results in a dramatic decrease in the ability of the major pBR322 inverted repeat sequence at position 3065 to assume a stem/loop configuration and to be cleaved by S1 nuclease.

Lilley (1980) and Panayotatos and Wells (1981) reported that S1 nuclease can specifically cleave pBR322 superhelical DNA at one major and two minor positions centered around 3065, 3219 and 3123 bp, respectively, from the single *Eco*RI site, when the DNA is digested by S1 nuclease to ~50% linear form and 50% relaxed circle. All three cleavage sites were found to be flanked by inverted repeat sequences that may form cruciform structures with 9–11 bp in the stem and 3–6 nucleotides in the loop. In our studies, a variety of experimental conditions were tried before establishing appropriate digestion parameters (temperature, salt and enzyme concentration) which resulted in the appearance of discrete *Eco*RI/S1 nuclease fragments. Under our optimum digestion conditions, we were able to detect a number of additional S1 nuclease recognition sites on pBR322 DNA (Figure 5C). However, since the entire pBR322 portion of subclone 13 was used as a hybridization probe rather than specific plasmid fragments, we cannot decide the exact location of the S1 cleavage sites along the pBR322 moiety. Judging from the intensities of hybridization in the resulting fragments we conclude that the additional S1 nuclease recognition sites on pBR322 DNA are cleaved 10–20 times less frequently than the major pBR322 site first described by Lilley (1980) and Panayotatos and Wells (1981).

In addition to the predicted hybridizing fragments at 220 bp and 500 bp from the left end of the chorion insert due to the hairpin structures at the promoter/coding region junctions of the genes, a number of additional hybridizing fragments resulting from cleavages by S1 nuclease were also observed (Figure 5C). Based on the size of the hybridizing fragments, the distances of the cleavage sites from the *Eco*RI sites of the insert were determined. It should be noted that the length of the radioactive fragments was derived from their mobility in lane 1 of each hybridization set (Figure 5), and that length determinations are accurate only within certain limits (probably ±20 bp for the smaller fragments and ±50–100 bp for the larger ones). In all cases small inverted repeat sequences with cruciform potential were found at the determined cleavage points (Figure 6). The new sites include one at the common 5′-flanking region of the 2132/2574 gene pair, ~50 bp upstream from the predicted cruciform structure of gene 2574, one within the coding region of gene 2574 and two more at the 3′-flanking region of the same gene (Figure 6). Another site within gene 2574 was also observed. This site results in a hybridizing fragment of 640–650 bp (Figure 5, hybridizations in upper part of panel C, dot just above the bracketed one) and is, apparently, located at the end of the intervening sequence of gene 2574. The sequence around the cleavage point is TTAAAAATATTTAATTTTTT. This site was not recorded in Figure 6 because it is likely that cleavage may be due to breathing of the DNA in that region (100% A+T over a length of 20 bp; see Appendix, nucleotides 629–648).

The left-end portion of the *B. mori* insert DNA used as the probe in the hybridization presented in Figure 5C (top left set of reactions) was apparently contaminated by pBR322 DNA sequences, since the bands corresponding to linear pBR322 and to the pBR322 derived 3.0-kb fragment above the insert band, hybridized quite efficiently. This contamination, however, in no way interferes with the interpretation of our results, since pBR322 by itself does not hybridize to any fragments of sizes smaller than 700 bp (Figure 5C, bottom right hybridization set).

The presence of secondary structures recognized by S1 nuclease does not appear to depend on the length of the DNA in the insert. We performed additional experiments in which subclone 13 (1.93-kb insert) and subclone 4 (3.85-kb insert, Figure 1) were digested with S1 and *Eco*RI and subsequently Southern hybridized to the same labeled portions of the chorion DNA insert of subclone 13 used for the experiments

shown in Figure 5. We found that the single-stranded regions recognized by S1 nuclease in subclone 13 are also recognized in subclone 4 which contains an additional 1.9 kb of chorion DNA adjacent to that present in subclone 13 (data not shown). Therefore, the induction in supercoiled DNA of particular secondary structures and the subsequent recognition by S1 nuclease of specific conformational changes in the DNA appears to be independent of the length of the DNA flanking the regions with the altered conformation. This is in agreement with the results obtained by Lilley (1981).

We have not as yet used sequences from the 3'-flanking region of gene 2132 as probes, and we do not know whether sequences at positions equivalent to the ones within or at the 3'-flanking region of gene 2574 are also cleaved by S1 nuclease in gene 2132.

Based on the symmetries observed at the origins of the genes 2132 and 2574, we postulate that their coordinate expression may be controlled, by the common structural features around the origins of the structural genes shared by both members of the pair. The kinetics of the S1 digestion experiments (Figure 5C) indicates that these structures are not stable. Stabilization of the putative structures by stage-specific factors acting as transcriptional activators or repressors may be responsible for the coordinate and stage-specific transcriptional activation of the genes. Thus, one level of transcriptional control may reside entirely in the 5'-flanking region of the structural genes and its interaction with nuclear macromolecules like RNA polymerase (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1980; Gluzman *et al.*, 1980), while the second level, the fine tuning of initiation of transcription, may involve the interaction of a portion of the structural gene and its adjacent 5'-flanking region with other nuclear factors. Whether the stabilized structures may act as enhancers of transcriptional activity or as attenuators is an open question. Homologous *in vivo* transcriptional experiments should determine which, if either, of the two alternatives is correct.

In S1 nuclease digestions of chicken chromatin, Larsen and Weintraub (1982) obtained results which indicate that S1 nuclease can be used as a probe for the detection of genomic elements associated with the transient expression of eukaryotic genes. Moreover, they showed that sites similar to those seen in erythrocyte chromatin are also cleaved when cloned globin DNA in superhelical form is digested with S1 nuclease. Their results support our notion that the postulated cruciform structures at the promoter/5'-non-coding gene sequence junctions of the 2132/2574 gene pair existing in supercoiled DNA may also be involved in the *in vivo* transcriptional regulation of those genes. The specificity with which supercoiled chorion DNA is cleaved by S1 nuclease (Figure 5) strongly suggests that the postulated cruciform structures are actually present in superhelical cloned DNA and possibly in chromosomal DNA as well.

The association of the putative regulatory element for genes 2132 and 2574 with the part of the chorion locus of late developmental specificity (Figure 7) is of particular interest. We do not as yet know whether all gene pairs in the late region are associated with such an element. The 13 genomic clones cross-hybridizing to the late promoter comprise 15 pairs of genes of late developmental specificity (Eickbush and Kafatos, 1982), and additional experiments involving Southern hybridizations of restricted DNA from the particular genomic clones are required to ascertain the exact patterns of hybridization. Nonetheless, the results from this experiment suggest that other genes of the same (late) developmental specificity spanning the Gr^B deletion are associated with control elements sharing pronounced sequence homologies with the 2132/2574 control region rather than with a unique sequence. We consider that this supports our idea that the determination of the transcriptional potentiation may not be dictated by direct primary structure recognition alone. Primary structure may be more important in satisfying the basic requirements for RNA polymerase recognition and binding, but also in determining conformational changes in the DNA outside and within the structural genes. The conformational changes may in turn be recognized by molecules participating in the process of cell differentiation. Several reports dealing with the presence of short or long terminal repeat sequences and their effect on the activation of normally silent genes tend to support the idea that conformational changes in the vicinity of structural genes may have profound effects on their transcriptional activity (Hayward *et al.*, 1981; Joyner *et al.*, 1982; Payne *et al.*, 1982; Levinson *et al.*, 1982). Finally, Müller and Fitch (1982) recently reported that hairpins with six or more nucleotides in the stems do not occur randomly in a number of viral DNA genomes but appear to be clustered in areas presumed to have regulatory significance. Their conclusion was that such structures are likely to be relevant to biological regulation.

We are now in the process of analyzing equivalent sequences from other pairs of genes of late and middle developmental expression to see whether the symmetry properties of the putative promoter sequences presented here occur in other genes of the same developmental specificity, and whether a different developmental specificity may be associated with 5'-flanking gene sequences whose properties are entirely different from those of the late gene sequences.

Our hypothesis must, ultimately, be tested through a meaningful functional assay. The importance of the structures and interactions predicted by our model will be tested by *in vitro* and *in vivo* transcription experiments in which authentic, truncated and chemically mutagenized pairs of genes will be used as templates for chorion RNA synthesis.

Materials and methods

Materials

The construction of chorion cDNA clones and of the partial *Eco*RI chorion genomic library has been described (Iatrou *et al.*, 1980; Eickbush and Kafatos, 1982). Plasmid pBR322 and *Escherichia coli* HB 101 were used for the subcloning of DNA fragments from the chorion genomic clones. Restriction enzymes were from BRL, Boehringer-Mannheim and New England Biolabs. *E. coli* DNA polymerase I, T4 polynucleotide kinase and T4 ligase from BRL; S1 nuclease from Sigma and BRL; [α-^{32}P]dNTPs from New England Nuclear; [γ-^{32}P]ATP (crude) from ICN; agarose from Sigma and Bio Rad; acrylamide from Bio Rad; methyl mercuric hydroxide from Alfa Products; hydrazine and dimethyl sulfate from Aldrich; piperidine from Fisher. Calf intestinal phosphatase was purchased from Worthington and purified according to Efstratiadis *et al.* (1977).

Subcloning

10 μg of clone 150 and 8 μg of pBR322 DNA were digested with endonuclease *Eco*RI. The *Eco*RI-treated pBR322 was dephosphorylated with an excess of phosphatase and both DNAs were deproteinized by phenol and Sevag extractions and ethanol precipitated. DNA pellets were washed with 70% ethanol and dried. Each DNA preparation was taken up in water and mixed in a ligase reaction mixture containing 30 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10 mM dithiothreitol, 50 μg/ml bovine serum albumin, 1 mM ATP and 2.5 units of T4 ligase in a final volume of 200 μl. The reaction was incubated at 13°C for 3 h. The reaction mixture was subsequently diluted to 1 ml with ligation buffer, an extra 2.5 units of T4 ligase were added and in-

cubation continued for 10 h more at 13°C. At the end of the second incubation period the reaction mixture was dialyzed extensively against 10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$, 10 mM $MgCl_2$ and aliquots of the dialysate used directly for *E. coli* transfection as described previously (Iatrou *et al.*, 1980). Selection of transformants was carried out on tetracycline-containing plates. Subclone classification was done by hybridizations of subclones picked on microtiter plates to radioactive *Eco*RI fragments of the chorion insert of clone 150 DNA according to Grunstein and Hogness (1975), as described previously (Iatrou *et al.*, 1980).

S1 nuclease digestions

S1 nuclease digestions were carried out in a standard buffer containing 40 mM sodium acetate pH 4.6, 50 mM NaCl and 1 mM $ZnCl_2$. Wherever noted, extra NaCl was added to the reaction mixture to bring the final concentration to 150 mM. The concentration of DNA was maintained at 50 μg/ml and that of the S1 nuclease at 750 units/ml (manufacturer's specification – 1 μg of single-stranded DNA in standard buffer rendered acid soluble per min by 1 unit of enzyme at 37°C) throughout. The enzyme preparations were tested for double-stranded DNase activity and no such activity could be detected. Incubations at 37°C or 24°C were for time periods (usually 5 – 90 min) appropriate to convert the desired proportion of ccc DNA into the linear form. Reactions were stopped with the addition of EDTA to 50 mM. If further restriction of the DNA was to follow, the S1-treated DNA was deproteinized, ethanol-precipitated and washed, dried and dissolved in the components of the secondary digestion.

Hybridizations

Southern hybridizations (Southern, 1975) were performed as described previously (Iatrou *et al.*, 1980, 1982) except that Dextran sulfate was also present in the hybridization mixture at a final concentration of 10% (w/v). Unless otherwise noted, standard hybridization temperature was 72°C and salt concentration 0.3 M NaCl (2 x SET). Grunstein-Hogness (1975) hybridizations were under the same conditions, but with hybridizing temperatures at 70°C and 78°C. Northern hybridizations of RNA preparations (Iatrou *et al.*, 1980, 1982) analyzed on methyl mercury hydroxide gels (Bailey and Davidson, 1976) and transferred onto DBM paper (Alwine *et al.*, 1977) were performed in the presence of 50% formamide, 3 x SET, 10% Dextran sulfate at 70°C.

Other methods

^{32}P end-labelling with T4 polynucleotide kinase was according to Lillehaug *et al.* (1976) and nick-translation according to Maniatis *et al.* (1975). Nucleotide sequencing was performed using the chemical method of Maxam and Gilbert (1977) and sequencing reactions were analyzed on 85 cm long 6% polyacrylamide gels.

Acknowledgements

We would like to thank Dr.Glen Andrews for drawing our attention to the paper by Panayotatos and Wells; Drs.G.A.Schultz, R.B.Church, J.C.States, H.van de Sande, J.Jankowski and F.C.Kafatos for useful discussions and suggestions; Dr.T.H.Eickbush for free access to all chorion genomic clones; Anne Vipond and Ken Ito for excellent technical assistance; Susan Carlson and Gretchen Haakenstad for skillful secretarial assistance. This work was supported by grants from the Medical Research Council of Canada and the Alberta Heritage Foundation for Medical Research.

References

Alwine,J.C., Kemp,D.J. and Stark,G.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5350-5354.
Bailey,J.M. and Davidson,N. (1976) *Anal. Biochem.,* **70**, 75-85.
Benoist,C. and Chambon,P. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 3865-3869.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.),* **196**, 180-182.
Berk,A.J. and Sharp,P.A. (1978) *Proc. Natl. Acad. Sci. USA,* **75**, 1274-1278.
Efstratiadis,A., Vournakis,J.N., Donnis-Keller,H., Chaconas,G., Dougall, D.K. and Kafatos,F.C. (1977) *Nucleic Acids Res.,* **4**, 4165-4174.
Eickbush,T.H., Jones,C.W. and Kafatos,F.C. (1981) *ICN-UCLA Symp. Mol. Cell. Biol.,* **23**, 135-153.
Eickbush,T.H. and Kafatos,F.C. (1982) *Cell,* **29**, 633-643.
Gluzman,Y., Sambrook,J.F. and Frisque,R.J. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 3898-3902.
Goldsmith,M.R. and Basehoar,G. (1978) *Genetics,* **90**, 291-310.
Goldsmith,M.R. and Clermont-Rattner,E. (1979) *Genetics,* **92**, 1173-1185.
Grosschedl,R. and Birnstiel,M.L. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 7102-7106.
Grunstein,M. and Hogness,D.S. (1975) *Proc. Natl. Acad. Sci. USA,* **72**, 3961-3965.
Hayward,W.S., Neel,B.G. and Astrin,S.M. (1981) *Nature,* **290**, 475-480.
Heckman,J.E. and RajBhandary (1979) *Cell,* **17**, 583-595.
Iatrou,K., Tsitilou,S.G., Goldsmith,M.R. and Kafatos,F.C. (1980) *Cell,* **20**, 659-669.
Iatrou,K., Tsitilou,S.G. and Kafatos,F.C. (1982) *J. Mol. Biol.,* **157**, 417-434.
Jones,W.C. and Kafatos,F.C. (1980) *Cell,* **22**, 855-867.
Joyner,A., Yamamoto,Y. and Bernstein,A. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 1573-1577.
Kafatos,F.C., Regier,J.C., Mazur,G.D., Nadel,M.R., Blau,H.M., Petri, W.H., Wyman,A.R., Gelinas,R.E., Moore,P.B., Paul,M., Efstratiadis, A., Vournakis,J.N., Goldsmith,M.R., Hunsley,J.R., Baker,B., Nardi,J. and Koehler,M. (1977) in Beerman,W. (ed.), *Results and Problems in Cell Differentiation,* Springer-Verlag, Berlin, pp. 45-145.
Kafatos,F.C., Jones,C.W. and Efstratiadis,A. (1979) *Nucleic Acids Res.,* **7**, 1541-1552.
Kafatos,F.C., Thireos,G., Jones,W.C., Tsitilou,S.G. and Iatrou,K. (1981) in Chirikjian,J.G. and Papas,T.S. (eds.), *Gene Amplification and Analysis,* vol. **2**, Elsevier/North Holland, pp. 537-550.
Larsen,A. and Weintraub,H. (1982) *Cell,* **29**, 609-622.
Levinson,B., Khoury,G., Van de Woude,G. and Gruss,P. (1982) *Nature,* **295**, 568-572.
Lillehaug,J.R., Kleppe,R.K. and Kleppe,K. (1976) *Biochemistry (Wash.),* **15**, 1858-1865.
Lilley,D.M.J. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 6468-6472.
Lilley,D.M.J. (1981) *Nucleic Acids Res.,* **9**, 1271-1289.
Maniatis,T., Jeffrey,A. and van de Sande,H. (1975) *Biochemistry (Wash.),* **14**, 3787-3794.
Maxam,A. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 560-564.
Müller,U.R. and Fitch,W.M. (1982) *Nature,* **298**, 582-585.
Nadel,M.R. and Kafatos,F.C. (1980) *Dev. Biol.,* **75**, 26-40.
Panayotatos,N. and Wells,R.D. (1981) *Nature,* **289**, 466-470.
Payne,G.S., Bishop,J.M. and Varmus,H.E. (1982) *Nature,* **295**, 209-214.
Regier,J.C., Kafatos,F.C., Goodfliesh,R. and Hood,L. (1978a) *Proc. Natl. Acad. Sci. USA,* **75**, 390-394.
Regier,J.C., Kafatos,F.C., Kramer,K.J., Heinrikson,R.L. and Keim,P.S. (1978b) *J. Biol. Chem.,* **253**, 1305-1314.
Rodakis,G.C. (1978), Ph.D. Thesis, University of Athens, Greece.
Rodakis,G.C., Moschonas,N.K. and Kafatos,F.C. (1982) *Mol. Cell. Biol.,* **2**, 554-563.
Rodakis,G.C. and Kafatos,F.C. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 3551-3555.
Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.
Tsai,S.Y., Tsai,M.-J. and O'Malley,B.W. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 879-883.
Tsujimoto,Y., Hirose,S., Tsuda,M. and Suzuki,Y. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 4838-4842.
Wasylyk,B., Kedinger,C., Corden,J., Brison,O. and Chambon,P. (1980) *Nature,* **285**, 367-373.

Appendix

The complete nucleotide (NT) sequence of subclone 13 is given opposite. The localization of genes 2132 and 2574 was as follows:

NT1 – 149: 5′ portion of gene 2132 intervening sequence.
NT150 – 200: leader sequence of gene 2132.
NT201 – 233: 5′-untranslated sequence of gene 2132.
NT234 – 501: 5′-flanking sequence of genes 2132 and 2574.
NT502 – 529: 5′-untranslated sequence of gene 2574.
NT530 – 580: leader sequence of gene 2574.
NT581 – 661: intervening sequence of gene 2574.
NT662 – 1006: coding sequence of gene 2574.
NT1007 – 1933: 3′-untranslated and 3′-flanking sequences of gene 2574.

```
          10        20        30        40        50        60
GAATTCGACCTCATGTCTCAAGATTTTTTCTACGTTTTCGTTTTTATTACGATGATAATA

          70        80        90       100       110       120
GAGAACAATGAAATGGTGCAAAAGAGGTCTACCGCTATTAAGTTAAAAATATTGCTTACA

         130       140       150       160       170       180
TTCGGAAGCTTTTAAAAAATATTACTTACTTGGATCAGGCAGCCCTGAACGCAGAGAAGG

         190       200       210       220       230       240
AGAAGAGCGAAGGTAAACATGATGCTGGATATTCAGGCACTTAATCTTGAATACTTTAAG

         250       260       270       280       290       300
CTTCAAGACTTCGTTCTTATATAGTAACTAAGAATCATGTTCACCTTGAGATAAGAAACA

         310       320       330       340       350       360
ATTCAATGACAAATAGTTGAGAAACAATGATAGTTTGGTTTTTAATCAAACAGGTGTTCG

         370       380       390       400       410       420
TAAATGTTTCCTGATTAGATCGCCCGAAGGCGAAAACTGCAGAATATCTCCAAAGTCACG

         430       440       450       460       470       480
TCTATATGCTATTTAGATATGAGGCATGCGGATTTTCAAGGAAACTGCTCGGTATAAAAG

         490       500       510       520       530       540
CTGATGTAGTTCAGAGTCAACTCATTCTGAAGTCATTCAGTGAAACAACATGGCCGCTAA

         550       560       570       580       590       600
ACTCATTGTCTTCGTCTGCGCCATCGCCCTCGTGGCTCAGGTAAGTCTTTTAATTTTAAA

         610       620       630       640       650       660
TTTAATTAATTGAACCTAAAATGATTACTTAAAAATATTTAATTTTTTCATTTGTTCATA

         670       680       690       700       710       720
GTCCGTTTTGGGCACTGGTTGTGGCTGCTGCTGTAGAGGTTGTGGCTGCGGTTGTGGAGG

         730       740       750       760       770       780
TTGTGGCTGTGGATGTTGTGAAAATTTCCGTGTATGTTCCAATTCGGCTGCTCCTACTGG

         790       800       810       820       830       840
TCTAAGTATATGTTCCGAGAACAGGTACAAGGGTGATGTCTGTGTCTGCGGTGAAGTGCC

         850       860       870       880       890       900
TTTCCTTGGTACCGCTGATGTCTGTGGCAACATGTGCAGCTCGGGTTGTGGTTGTATCGA

         910       920       930       940       950       960
CTACGGATGCGGAAACGGTTGCGTCGGTATCACCCGAAGCTGCGGCGGGTGCGGATGCGG

         970       980       990      1000      1010      1020
CTGTGGAGGCTGCGGATGCGGCTGTGGAGGCTGCGGTTGCTGCTAAGGATGCAGTTGCGG

        1030      1040      1050      1060      1070      1080
ACGTTCATGTTGTGGCGTCTATTAAAAAGCTCATTGAGTTATGTATTATTGAAAACATCA

        1090      1100      1110      1120      1130      1140
ATAAAACCCACATTCAGTAATTATCCTCATTGTCTTTTCAGTACCACAATTGTATTTTAT

        1150      1160      1170      1180      1190      1200
TGTAATGTGTAAAATGTTCCTGAATACGTGTTAATTGCTTATTCTTTTAATTGCTTAATT

        1210      1220      1230      1240      1250      1260
GGTGGACGAGGCCCACCTGATGTTATGTGTTGACCGGAGCCCATAGACATCAACAACGTA

        1270      1280      1290      1300      1310      1320
AATGCCGCCCTACCCGTCTTGAGATATAAGTTCCTAATCTCAGTTTTATAGTACAACTAG

        1330      1340      1350      1360      1370      1380
GTGTACCCGTCCGCTTCGCTGGGCATTTAAAATTAACATTATTATATCTCACCCCCACAA

        1390      1400      1410      1420      1430      1440
AGATTCTCATCATTAACGCCCCCGGAACTGGTGTAGGGAATCCAACACTCATATAAATAT

        1450      1460      1470      1480      1490      1500
TAGCCTATCCATTAAGTACATGTATTTTCTACATGAATACCAAGTTTCAAGTCAATCGGA

        1510      1520      1530      1540      1550      1560
TGCATGGTTCAGTAGTTATAACGGAACTTCCGTAAAAACCACTGTAGATTTATATATTAG

        1570      1580      1590      1600      1610      1620
TATAGATGGCTGCCCCACCCTTTAAATAGCCATGTGTTATTTGTCAAGGCTAAAATGGAC

        1630      1640      1650      1660      1670      1680
GGGGGTTGATACTTAACCATGAGATATTCCAGTATGTTTTATGGTTGGATAAAACTTGGC

        1690      1700      1710      1720      1730      1740
CGACAAATATTTTGAAGAATCACTGTCACATCAGAATCTGATATCAAATATAACTAATTA

        1750      1760      1770      1780      1790      1800
GGCGATAGGAGCTTACAATCTTTTTCAGAAAAACTAATAGTTTAAAAAACTGAATGGTTA

        1810      1820      1830      1840      1850      1860
AATGTCAGGATAGATTATTAGATTTAATTTATTTGTTAAATACAGATAGTAGCTTGGTCG

        1870      1880      1890      1900      1910      1920
TATGCCCCTGGCATAATTGATGTATTCCAATTTAATGCACAAGGACAGGCATTATTCCAT

        1930      1940      1950      1960      1970      1980
AACAATTGAATTC
```