Development 122, 4045-4056 (1996)
Printed in Great Britain © The Company of Biologists Limited 1996
DEV5095

Quantitative functional interrelations within the *cis*-regulatory system of the *S. purpuratus Endo16* gene

Chiou-Hwa Yuh, James G. Moore and Eric H. Davidson

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

SUMMARY

Embryonic expression of the *Endo16* gene of *Strongylocentrotus purpuratus* is controlled by interactions with at least 13 different DNA-binding factors. These interactions occur within a *cis*-regulatory domain that extends about 2300 bp upstream from the transcription start site. A recent functional characterization of this domain reveals six different subregions, or *cis*-regulatory modules, each of which displays a specific regulatory subfunction when linked with the basal promoter and in some cases various other modules (C.-H. Yuh and E. Davidson (1996) *Development* 122, 1069-1082). In the present work, we analyzed quantitative time-course measurements of the CAT enzyme output of embryos bearing expression constructs controlled by various *Endo16* regulatory modules, either singly or in combination. Three of these modules function positively in that, in isolation, each is capable of promoting expression in vegetal plate and adjacent cell lineages, though with different temporal profiles of activity. Models for the mode of interaction of the three positive modules with one another were tested by assuming mathematical relations that would generate, from the measured single module time courses, the experimentally observed profiles of activity obtained when the relevant modules are physically linked in the same construct. The generated and observed time functions were compared, and the differences were minimized by least squares adjustment of a scale parameter. When the modules were tested in context of the endogenous promoter region, one of the positive modules (A) was found to increase the output of the others (B and G), by a constant factor. In contrast, a solution in which the time-course data of modules A and B are multiplied by one another was required for the interrelations of the positive modules when a minimal SV40 promoter was used. One interpretation is that, in this construct, each module independently stimulates the basal transcription complex. We used a similar approach to analyze the repressive activity of the three *Endo16 cis*-regulatory modules that act negatively in controlling spatial expression. The evidence obtained confirms that the repressive modules act only by affecting the output of module A (C.-H. Yuh and E. Davidson (1996) *Development* 122, 1069-1082). A new hierarchical model of the *cis*-regulatory system was formulated in which module A plays a central integrating role, and which also implies specific functions for certain DNA-binding sites within the basal promoter fragment of the gene. Additional kinetic experiments were then carried out, and key aspects of the model were confirmed.

Key words: gene regulation, sea urchin embryo, *Endo16* gene, *cis*-regulatory system

INTRODUCTION

The *cis*-regulatory systems that control spatial and temporal gene expression are typically composed of subelements that function positively or negatively in different embryonic territories, or are utilized in different temporal phases of development (Kirchhamer et al., 1996). Here, we approach the mechanism by which the upstream regulatory subelements of the *Endo16* gene of *S. purpuratus* interact, producing an integrated transcriptional output. The embryonic pattern of expression of this gene is generated by the outputs of six different subelements of the overall *cis*-regulatory system, which we regard as modular components of the control system. Multiple interactions with diverse transcription factors occur within each of these modules. Each module displays a particular regulatory function when linked to an expression construct, either alone or in combination with other modules, and tested by gene transfer (Yuh and Davidson, 1996). Modular organization is a property shared by many embryonic *cis*-regulatory systems that have been studied in detail, in *Drosophila*, mouse and sea urchins (reviewed by Davidson, 1994; Kirchhamer et al., 1996).

The *Endo16* gene encodes a polyfunctional glycoprotein (Soltysik-Espanola et al., 1994) which in the late embryo is a secreted product of the midgut. The *Endo16* gene is transcriptionally activated in late cleavage (Godin et al., 1996) and, throughout the blastula stages, it is expressed in descendants of the veg_2 lineage, which constitute the vegetal plate. *Endo16* transcripts are initially observed throughout the archenteron, to which the vegetal plate gives rise by invagination. After gastrulation is complete, however, expression is silenced in the foregut and hindgut, but is stepped up in the

midgut (Nocente-McGrath et al., 1989; Ransick and Davidson, 1993). Control of all of these aspects of the expression pattern is primarily transcriptional, since the CAT mRNA products of *Endo16•CAT* expression constructs display the same temporal and spatial expression profiles as does endogenous *Endo16* mRNA (Ransick and Davidson, 1993; Yuh et al., 1994; Yuh and Davidson, 1996). The *cis*-regulatory system of the *Endo16* gene required for the complete embryonic pattern of expression is included in a 2300 bp DNA sequence extending upstream from the transcription start site, within which we have identified more than 30 target sites for at least 13 different highly specific DNA-binding factors (Yuh et al., 1994). The disposition of the six different modular subelements (G-A) resolved in our recent functional analysis (Yuh and Davidson, 1996), and of the protein-binding sites (Yuh et al., 1994), are summarized diagrammatically in Fig. 1.

Modules G, B and A function positively. When linked to the endogenous basal promoter or the SV40 promoter, each is independently capable of causing expression in the vegetal plate and later the archenteron, though in isolation all three modules also promote ectopic expression in the territories adjacent to the vegetal plate. Module G functions only weakly by itself, but, when present in an expression construct utilizing the SV40 promoter, it synergistically boosts the otherwise low activities of modules A and/or B to a significant extent (Yuh and Davidson, 1996). Modules A and B also function synergistically in combination with one another, producing an increased level of expression throughout development, with either the endogenous or the SV40 promoter (Yuh and Davidson, 1996). Module A is largely responsible for early expression in the vegetal plate. The late rise in expression of *Endo16* is due largely to module B, which by itself is capable of promoting midgut expression in the postgastrula stage embryo. Modules F, E and DC function negatively. The function of modules F and E is to preclude ectopic expression of *Endo16* in the ectoderm that lies above the upper boundary of the vegetal plate. Module DC similarly prevents expression in the cells deriving from the skeletogenic progenitors, which are initially located across the lower boundary of the vegetal plate. All three of these modules are sensitive to treatment of cleavage-stage embryos with LiCl, which expands the vegetal plate and the domain of *Endo16* expression at the expense of the overlying ectoderm (Ransick and Davidson, 1993). LiCl converts all three negative modules to positive function. Yuh and Davidson (1996) showed that both the negative (i.e., spatial control) functions of modules F, E and DC, and their LiCl sensitivity require the presence of module A. The function of the negative modules thus appears to be to eliminate the output of modules A, B and G in cells across the upper and lower boundaries of the vegetal plate.

In the following, we use a quantitative approach to address two specific issues by testing models of regulatory system function against temporal expression data. The first of these issues is the nature of the interactions amongst the positively acting modules, and between them and the basal transcription apparatus. These interactions are evidenced by the synergistic effects observed by Yuh and Davidson (1996) when two or more of the positive modules are physically linked in an expression construct. The second issue addressed here concerns the mode of action of the negatively acting modules. The manipulations that we describe provide strong additional

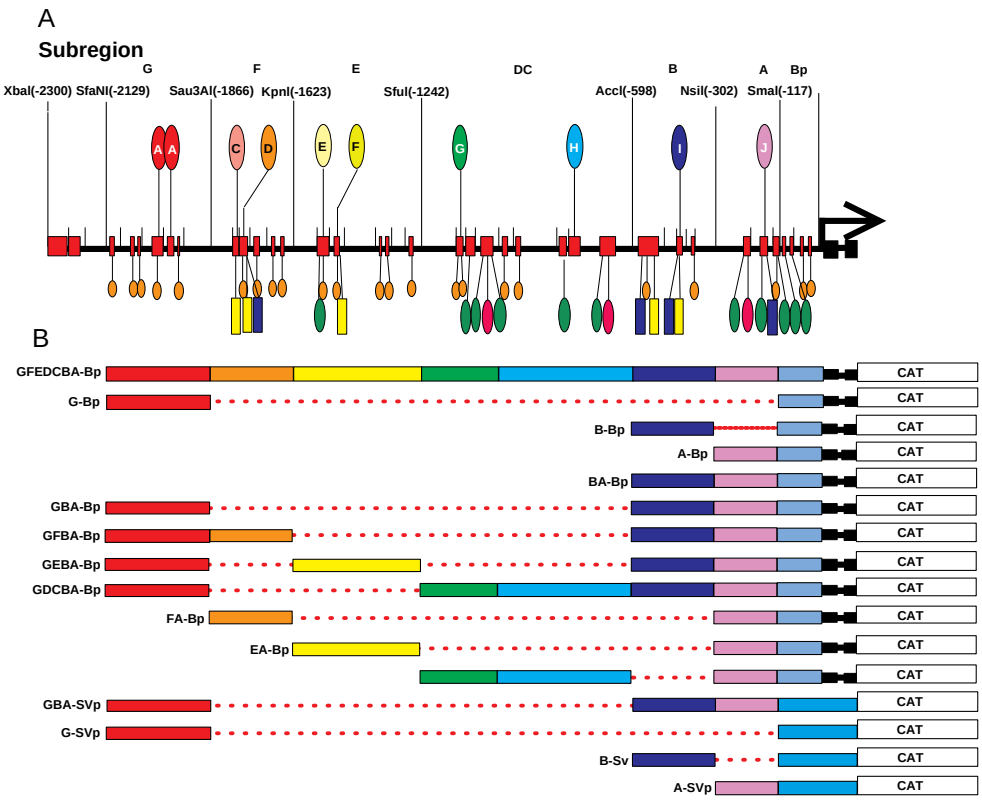

Fig. 1. Summary of *Endo16 cis*-regulatory system and fusion constructs utilized in this work. (A) Diagram of DNA-binding proteins and target sites in the upstream regulatory domain that is necessary and sufficient for accurate expression of *Endo16* fusion genes, from Yuh et al. (1994). Each color represents a different protein, as determined by oligonucleotide affinity chromatography, target site probe blots, and oligonucleotide gel shift competition analyses. Proteins indicated above the line representing the DNA bind uniquely within the subregions designated by the capital letters at the top of the diagram, while proteins indicated below the line bind in multiple regions. Many of the DNA-binding proteins remain uncharacterized; they are shown here to indicate the diversity and distribution of their specific target sites. Modular elements G-A are indicated, from Yuh and Davidson (1996); Bp, basal promoter element. (B) Fusion constructs utilized in this work (see Yuh and Davidson, 1996, for details).

evidence that these negative functions operate by interference with the positive function of module A. By focusing on the quantitative interrelations among the modules, we have been led to a hierarchical model of the *Endo16 cis*-regulatory system, which in turn provoked further experiments that confirmed key features of this model.

MATERIALS AND METHODS

Expression constructs

Most of the expression constructs used in this work are described by Yuh and Davidson (1996). Eggs were prepared and injected with the constructs, and CAT assays performed, exactly as described there. These constructs are shown diagrammatically in Fig. 1. Additional constructs including the 'J_m' and '$(CG)_2$' mutations were generated for the present studies. J_m refers to a mutated site for the 'J' factor of module A (see Fig. 1), and CG to a factor that binds at a number of locations in the regulatory domain, indicated in Fig. 1 by green ovals below the line representing the DNA.

Briefly, constructs *GBA(J_m)-BpCAT* and *BA(J_m)-BpCAT* were assembled from a cloned insert that linked G, B and A to Bp, into which the mutated J target site had been introduced by PCR. The outside primers used for this purpose were the Bluescript vector forward and reverse primers. The complementary inside primers included the normal *Endo16* sequence immediately upstream of the J target site (Yuh et al., 1994), with an appended sequence that replaces this site with an *Xba*I site:

J site: 5′--GTAGGATTAAG--3′
3′--CATCCTAATTC--5′

J_m: 5′--GTATCTAGAAG--3′
3′--CATAGATCTTC--5′

*Xba*I site

The *Xba*I site served later to check the success of the construction. *GBA(J_m)-BpCAT* was assembled from *GBA(J_mXba)* + *(XbaJ$_m$)A-Bp* + *Bluescript CAT*, where *J_mXba* and *XbaJ$_m$* represent the two parts of module A to be joined by ligation at the *Xba*I site. Construct *BA(J_m)-BpCAT* was assembled similarly, from *BA(J_mXba)* and *(XbaJ$_m$)A*.

Constructs *A$(CG)_2$-SVpCAT* and *B$(CG)_2$-SVpCAT* were derived from *GBA$(CG)_2$-SVpCAT*. This was generated by ligating into *GBA-SVpCAT* (Yuh and Davidson, 1996) a double-stranded oligonucleotide which represents the two CG factor target sites that appear between positions −64 and −109 of the *Endo16* basal promoter region (Yuh et al., 1994). The sequence of this oligonucleotide is as follows (CG target sites boxed).

```
   −109                                                    −69
5′     CTGTTTGAGTTTCGTCTCCTGATTGTGCTATCAAAGACAAA          3′
3′ GTACGACAAACTCAAAGCAGAGGACTAACACGATAGTTTCTGTTTCTAG 5′
   SphI                                          BglII
```

Data reduction and mathematical procedures

All the smooth curves shown were generated by means of a derivative matching algorithm (spline interpolation available in Mathcad Plus 6.0b). The calculations used to generate Figs 3-6 and 8, and Tables 1 and 2, were carried out using data shown in Fig. 2, as indicated in text.

For each model to be tested a single free parameter, λ, was used as a scale factor to match calculated functions to observed data. The value of this parameter was minimized by a homogeneous least squares procedure. The procedure was to determine the closest possible match between an observed time course ('target time course') and a time course calculated by applying a mathematical operation to other observed time courses (from the same data set; in the following these are the 'calculation time course(s)'). Thus, where d_0 represents the target time-course data set, and d_{0_i} an individual time point in this data set; and d_c represents the calculation data set(s),

$$d_0 = \lambda \bullet f(d_c) \; . \tag{1}$$

The minimum least square value of λ is given by

$$\lambda = \frac{\Sigma\,(d_{0_i}) \bullet f(d_{c_i})}{\Sigma\,[f(d_{c_i})]^2} \; . \tag{2}$$

The root mean square error, ε is then calculated as

$$\varepsilon = \sqrt{\frac{\Sigma\,(d_{0_i} - \lambda \bullet f(d_{c_i}))^2}{N}} \; . \tag{3}$$

The calculation was carried out at the times occupied by data points, and only the data, not the interpolated values were used for the calculation. Values of λ and ε were reported as shown in Tables 1 and 2. Spline interpolations were imposed on the generated data following the calculations, and these curves are shown in Figs 2-7 together with an envelope representing ±ε, portrayed as fraction of maximum value. We stress that the cubic spline curves shown are merely to improve the ease with which the individual time courses can be followed by eye. These smooth curves are not meant to indicate the actual time courses, either measured or calculated, in the intervals between the measured or calculated points.

Simplification based on low stability of CAT protein and mRNA

A representation of CAT enzyme production in embryos bearing CAT expression vectors would require solution of the following relations:

$$R_{(t)} = \int_0^t [S_{(t')} - k_{DR} \bullet R_{(t-t')}]dt' \tag{4}$$

$$C_{(t)} = \int_0^t [k_T \bullet R_{(t')} - k_{DC} \bullet C_{(t-t')}]dt' \; . \tag{5}$$

Here R denotes molecules of CAT mRNA per embryo; S is transcription rate for the expression vectors per embryo, i.e., molecules•min^{-1} synthesized (more precisely, molecules of mRNA flowing into the cytoplasm•min^{-1}, if processing is efficient as is usually the case in sea urchin embryos [Cabrera et al., 1984]. This is the same as transcription rate); C is CAT enzyme molecules per embryo; k_T is translation rate, i.e., CAT protein molecules•min^{-1}•R^{-1}; and k_{DC} and k_{DR} are the first order decay rate constructs for CAT enzyme and CAT mRNA, respectively. However, Flytzanis et al. (1987) showed that the half life of CAT enzyme is only ~40 minutes in sea urchin embryos, and that of CAT mRNA is less than or equal to this. This is a very short interval on the scale of the experiments shown in Figs 2-5, measurements for which extend over a period exceeding 50 hours. Therefore, averaged over several hours, C of equation 2 will always be proportional to S over the interval measured:

$$C = \alpha S \tag{6}$$

where α is a proportionality constant. The value of α approximates $k_T \bullet k_{DC}^{-1} \bullet k_{DR}^{-1}$. We can take k_T=2 molecules of protein•min^{-1}•mRNA^{-1} (Davidson, 1986); k_{DC}=ln2/40 min^{-1}; $k_{DR} \leq k_{DC}$ (Flytzanis et al., 1987). Thus for peak values of ~4×10^6 molecules of CAT enzyme/embryo late in development, when there are ~100 expressing midgut cells (Yuh and Davidson, 1996) S per nucleus would be ≥6 molecules of CAT mRNA•min^{-1}, a reasonable rate for a moderately active expression construct present in multiple copies, but limited by

the amount of transcription factors present, as observed earlier for other expression constructs (Livant et al., 1988; Franks et al., 1990).

The short half-lives of CAT mRNA and protein reduces the representation of CAT enzyme expression shown in equations 1 and 2 to the simple proportionality of equation 3. Thus we are able to carry out the operations indicated by the models to be tested directly on the data points ($C_{(t)}$ values) and interpret them as immediate indicators of the rates of expression construct transcription.

RESULTS

Time-course data and experimental approach

Measurements of Yuh and Davidson (1996) showed that modules G, B and A of the *Endo16* *cis*-regulatory system function positively and synergistically. Time-course data on CAT expression (Fig. 3 of Yuh and Davidson, 1996) indicated that, when associated with the *Endo16* basal promoter (Bp) in the expression construct *A-BpCAT*, module A by itself generates a profile of expression that peaks at about 40 hours and then declines, while module B alone generates relatively low activity until after 60 hours, when the level of CAT enzyme produced by the *B-BpCAT* construct rises sharply (see Fig. 1 for constructs discussed here and in the following). The construct *G-BpCAT* produces an almost flat, low level profile of CAT expression. All three of these modules individually suffice to produce vegetal plate and gut expression, though with differing temporal and quantitative profiles of activity. However, when physically combined in the construct *GBA-BpCAT*, the level of CAT enzyme expression is higher than that produced by any of the individual constructs. The first series of experiments that we discuss in this paper were designed to illuminate the nature of the synergistic interrelations amongst these three positively acting upstream regulatory modules.

Time-course data used in this study are reproduced in Fig. 2. A single batch of fertilized eggs deriving from a different female was used for each set of measurements. CAT enzyme activity was determined in pools of 100 normally developing embryos at four to seven time points between 20 and 72 hours postfertilization, depending on the experiment. The smooth curves shown are interpolations between these points (see Materials and Methods). Not all embryos deriving from injected eggs develop normally. Thus, for example, injection of ~5000-6000 eggs per experiment was required to obtain the 3500 morphologically normal embryos needed for experiments such as shown in Fig. 2A, which included seven constructs each assayed at five time points.

The approach that we followed in the initial set of experiments was to test various simple models that might provide interpretations for the synergism observed when the single modules are physically linked in expression constructs. Each model was tested by calculating the output that would be generated by the linked construct from the individually measured time courses of its constituents according to that model, and the generated values were then compared to the data obtained experimentally for that linked construct. Because of the short half life of CAT protein and mRNA (Flytzanis et al., 1987), the CAT enzyme levels at the time points measured are directly proportional to the transcription rate around that time (this argument is shown in equations 4-6 of Materials and Methods). For each model, the values at each time point of the respective activity profiles were multiplied or added, as required by the model, and the result was multiplied by a scale factor, λ (see equations 1-3 in Materials and Methods). The best value for λ was obtained by minimizing the difference between the calculated and observed data, using a least squares procedure. The value of λ, and of the root mean square error ε, were reported, and these data are summarized in Tables 1 and 2.

Synergistic interrelations amongst modules G, B and A in constructs utilizing the endogenous *Endo16* basal promoter

Time-course measurements for constructs composed of modules G, B and A tested singly and in various combinations, and linked to the endogenous *Endo16* basal promoter (Bp), are shown in Fig. 2A-C. The embryos used for the data set in Fig. 2A were somewhat more active than those used for the data set in Fig. 2B, but the curves representing the activity of each construct are of similar form. Comparison with figure 3 of Yuh and Davidson (1996) show that the forms of these time courses are the same as measured earlier as well. The smooth curves shown were interpolated through the points by a derivative matching procedure (see Materials and Methods). The data of Fig. 2A and B were averaged and interpolated to generate the single set of curves shown in Fig. 2C. This averaged data set was then used for the following operations.

The first and simplest case considered was the relation between modules A and B, when they are linked in construct *BA-BpCAT* (dark blue, green and gray curves in Fig. 2A-C). We considered three different models, as indicated in the top section of Table 1: (*i*) that module A increases the output of module B by a constant synergism factor, λ. This model is symbolized $\overline{BA}=B\bullet\lambda$, where $\overline{BA}$ indicates the time course generated by *BA-BpCAT*; (*ii*) that the activity of *BA-BpCAT* is

Table 1. Synergistic models for modules G, B and A linked to endogenous basal promoter

Model†	ε (% max)*	λ‡	λ/function#
$\overline{BA}=B\bullet\lambda$	0.227 (2%)	4.2	4.2
$\overline{BA}=(B+A)\bullet\lambda$	9.07 (24%)	1.6	1.6
$\overline{BA}=A\bullet\lambda$	6.49 (17%)	0.69	0.83
$\overline{GBA}=\overline{GB}\bullet\lambda$	1.99 (8%)	3.1	3.1
$\overline{GBA}=\overline{BA}\bullet\lambda$	4.35 (17%)	0.78	0.78
$\overline{GBA}=\overline{BA}\bullet G\bullet\lambda$	3.58 (14%)	0.39	0.62
$\overline{GBA}=A\bullet B\bullet G\bullet\lambda$	4.65 (18%)	0.26	0.64
$\overline{GBA}=\overline{GB}\bullet A\bullet\lambda$	3.97 (15%)	0.50	0.70
$\overline{GBA}=(G+B+A)\bullet\lambda$	4.40 (17%)	1.23	1.23
$\overline{GBA}=B\bullet A\bullet\lambda$	3.09 (12%)	0.59	0.77
$\overline{GBA}=\overline{GBA}\ (J_m)\bullet\lambda$	7.0 (9%)	1.42	1.42

†Overline indicates physical linkage of modules in expression construct. The left side of the propositions, which are indicated as equations, are the target time-course data imported from Fig. 2C. The right side of propositions indicates mathematical operations performed on data points from the time-course measurements indicated, also from Fig. 2C, to obtain calculated time points. These were then compared to data (see Materials and Methods). For the synergism scale factor λ see text.

*ε, root mean square error (see Materials and Methods). (% max) denotes [(ε/max)×100] where max is the highest data point on the respective time courses.

‡Minimum least squares values of λ.

#$n\sqrt{\lambda}$ where n is number of time-course data sets in calculated function; e.g., for $B\bullet\lambda$ n=1; for $B\bullet A\bullet\lambda$ n=2.

the sum of the activities of *B-BpCAT* and *A-BpCAT*, multiplied by the synergism scale factor λ, symbolized $\overline{BA}$=(B+A)•λ; *(iii)* that the activity of *BA-BpCAT* is the product of the activities of *B-BpCAT*, *A-BpCAT* and the synergism factor λ, symbolized $\overline{BA}$=B•A•λ. The result is in this case clear and obvious: $\overline{BA}$=B•λ produces an excellent fit to the observed data, as shown in Fig. 3B, while the other two models produce very poor fits, with relatively large root mean square (RMS) errors (Table 1). The value of the scale factor for $\overline{BA}$=B•λ is about 4. Thus we may conclude that when linked together in *BA-BpCAT* module A simply amplifies about four-fold the quantitative output of module B over the whole time course of the experiment. The converse model, that B amplifies the output of A by a constant factor, produces a worse fit than any of the others listed in the first portion of Table 1.

Fig. 4 shows the best solution that we found for the output of the combined fusion *GBA-BpCAT*. This is given by the model $\overline{GBA}$=$\overline{GB}$ •λ; i.e., that the activity of this construct is approximated by the activity of the construct *GB-BpCAT* amplified throughout by a constant scale factor of about 3 (Table 1). In other words, exactly as in the construct *BA-BpCAT*, module A functions in a particularly simple way, amplifying the output several fold by a constant factor. Since the time course of *G-BpCAT* is almost a constant itself (Fig. 2C), it is not possible to resolve the form of the contribution of module G *per se*, i.e., to separate its contribution out from the term λ. The main synergism obtained in the *GBA-BpCAT* construct is between modules A and B. In fact, Fig. 2C shows that the activity of *BA-BpCAT* is actually higher later in development than that of *GBA-BpCAT*, though overall these two time functions are quite similar. Table 1 shows that the models $\overline{GBA}$=$\overline{BA}$ •λ and $\overline{GBA}$=$\overline{BA}$ •G•λ, in which λ<1, give significantly worse fits than does $\overline{GBA}$=$\overline{GB}$ •λ, and the forms of these model curves are not coherent with the target data; nor is any model in which the output of *GBA-BpCAT* is conceived as a product of the time courses of module A and other modules useful, e.g., $\overline{GBA}$=A•B•G•λ; or $\overline{GBA}$=$\overline{GB}$•A•λ. $\overline{BA}$ •G•λ comes closest, when module G acts as a depressant by a factor of ~0.6 (Table 1). This cannot be excluded even though the RMS error is twice that for $\overline{GBA}$=$\overline{GB}$•λ, but is not an attractive interpretation, since we know that *G-BpCAT* in fact functions as a positive expression construct that promotes reporter gene transcription in gut and that module G acts as a synergistic element in various SV40 promoter (SVp) constructs tested by Yuh and Davidson (1996), rather than as a depressant. Nor can the activity of *GBA-Bp•CAT* be described as the sum of the activities of the component modules; Table 1 shows that $\overline{GBA}$=(G+B+A)•λ also gives a poor fit.

We conclude that module A probably acts the same in the context of *GBA-BpCAT* as in the context of *BA-BpCAT*. It functions to increase the output of its active partners by a constant factor of three to four. *A-BpCAT* is transcribed in vegetal plate and archenteron (as well as ectopically in the ectoderm and skeletogenic mesenchyme; Yuh and Davidson, 1996), beginning in blastula stage. This activity is highest at 40 hours. It then declines and becomes difficult to detect late in development. These kinetics are obviously distinct from the constant three- to four-fold synergistic amplification attributable to module A in the present experiments. Therefore, the synergistic function seen in combinations of modules A and B might depend on a different interaction within module A than that which causes vegetal plate expression and which peaks at 40 hours. We know (unpublished experiments) that the spatial and temporal expression function mediated by module A when it is isolated in the *A-BpCAT* construct, depend on interactions at the specific target site for the factor 'J' of module A (see Fig. 1, and Yuh et al., 1994). Thus the synergistic amplification function of module A may depend on one or more of the four other interactions that occur within this module. In Fig. 1 (top), these are symbolized as colored symbols (green, purple, pink) beneath the line representing the DNA of module A. Target sites for each of these interactions appear in several different regions of the *Endo16 cis*-regulatory domain, and for two of them, *viz* the SpGCF1 sites and those for the factor symbolized by the green ovals, in the Bp region of the gene as well.

Synergistic interactions amongst modules G, B and A in constructs utilizing the SV40 early region basal promoter

As discussed in detail by Yuh and Davidson (1996), the SV40 early region basal promoter (SVp) generates qualitatively similar but quantitatively much more feeble temporal patterns of activity when linked singly to G, B or A modules than is observed when these same modules are linked singly to the endogenous basal promoter. This can be seen again in Fig. 2D (compare Fig. 2C). However, when these modules are linked together a very large synergistic effect is observed, and thus *GBA-SVpCAT* is about as active as is *GBA-BpCAT*. Similar observations are reported by Yuh and Davidson (1996), who also found a very large synergistic effect when *GA-SVpCAT, GB-SVpCAT* and *BA-SVpCAT* were compared to *G-SVpCAT, B-SVpCAT* or *A-SVpCAT*. These combined SVp constructs all function at about the same level as do the corresponding Bp constructs. Two of the six Sp1 sites included in SVp happen to be strong sites for the sea urchin SpGCF1 factor, while Bp contains, in addition to two SpGCF1 sites, two sites for another protein that also binds within the module A sequence, as well as elsewhere in the *Endo16 cis*-regulatory domain (symbolized by the green ovals in Fig. 1). Yuh and Davidson (1996) supposed that, in some way, synergistic interactions amongst modules G, B and A in the combined SVp constructs substitute for interactions mediated by the additional proteins binding within the endogenous Bp region. In any case, the intermodule synergism observed in SVp constructs is, in quantitative terms, about 10× that seen with Bp constructs, comparing the output of single module constructs to that of the GBA constructs.

The results shown in Table 2 indicate that the form of the models for *GBA-SVpCAT* that provide reasonable approximations, using the time courses for *G-SVpCAT, B-SVpCAT* and *A-SVpCAT*, are essentially different from those shown in Table 1 and Figs 3 and 4 for the Bp constructs. The best solution, shown in Fig. 5A, is given by the model $\overline{GBA}_s$=A_s•B_s•λ, and almost as good is $\overline{GBA}$s=A_s•B_s•G_s•λ (Table 2). This is reasonable, since the *G-SVpCAT* time course is low and almost flat; the only effect of including the *G-SVpCAT* time course is to decrease the value of λ (per function) from about 5.2-fold to 3.7-fold amplification. The main point is that, in these models, the time courses of the individual A and B module constructs are multiplied by one another. The levels of output of each module that are observed when they are tested singly, presumably reflect directly the occupancy of their transcription

factor target sites through time; the results of Table 2 show that, for each module, these levels are proportional through time to its synergistic function when all three are physically combined in an SVp construct. The model that is effectual with the *GBA-BpCAT* target, i.e., $\overline{GBA}_S = B_S \bullet \lambda$, produces a terrible fit when applied to the SVp time courses, as do the other models of the same form (since the time courses for G and G_S are essentially flat, GB•λ is of the same form as B•λ). This is shown in Fig. 5B. Nor do models dependent on the *G-SVpCAT* time functions work and addition rather than multiplication of

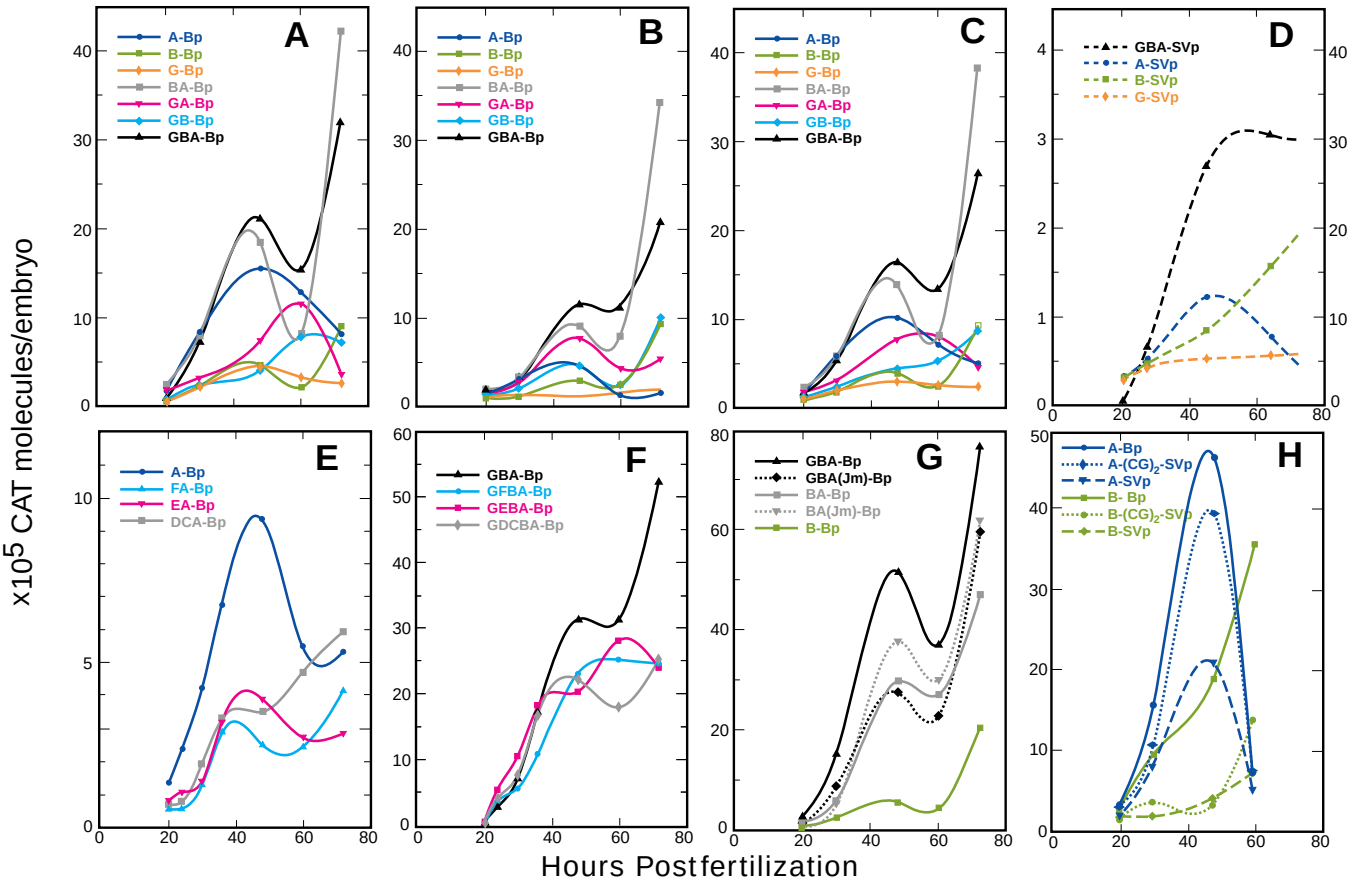

Fig. 2. Time-course measurements of CAT enzyme production by embryos bearing *Endo16•CAT* expression constructs. Each panel represents a set of measurements carried out on a single batch of eggs injected with the indicated constructs at the same sitting, and raised simultaneously under identical conditions. 100 embryos displaying normal morphology for their stage were collected at the indicated times and CAT enzyme activity was measured in pooled lysates of these embryos to generate the data points shown. Procedures were as described by Yuh and Davidson (1996). (A,B) Two independent data sets were obtained with constructs *GBA-BpCAT*, *BA-BpCAT*, *GA-BpCAT*, *GB-BpCAT*, *B-BpCAT*, *A-BpCAT* and *G-BpCAT* (see Fig. 1). Measurements were made at 20, 30, 48, 60 and 72 hours. Data points are indicated by symbols and the solid lines are cubic spline interpolations (see Materials and Methods). These interpolated curves are shown only to improve the ease with the individual time courses can be distinguished and are not used for any calculations in the paper. The interpolations may not accurately represent the time courses in the intervals between measurements. (C) Combined data set obtained by point-for-point averaging of data sets in A and B. The average data points were then used to support the construction of a new set of interpolated curves. (D) Data set obtained with constructs employing the SV40 early region promoter rather than the endogenous promoter. Constructs tested were *GBA-SVpCAT*, *A-SVpCAT*, *B-SVpCAT* and *G-SVpCAT* (Fig. 1). Data were collected at 20, 27, 45 and 64 hours. Curves *A-SVpCAT*, *B-SVpCAT* and *G-SVpCAT* refer to the left-hand ordinate, which is 10× expanded, while *GBA-SVpCAT* refers to the right-hand ordinate, which is the same as in the other panels. (E) Data set obtained with constructs of the form *XA-BpCAT*, where *X* denotes one of the negatively acting *Endo16* spatial control modules of Yuh and Davidson (1996). Constructs tested were *A-BpCAT*, *DCA-BpCAT*, *EA-BpCAT* and *FA-BpCAT* (Fig. 1). Data were collected at 20, 24, 30, 36, 48, 60 and 72 hours. (F) Data set obtained with constructs of the form *GXBA-BpCAT*, where *X* again represents the negative modules, viz: *GBA-BpCAT*, *GDCBA-BpCAT*, *GFBA-BpCAT* and *GEBA-BpCAT*. Data were collected at 20, 24, 30, 36, 48, 60 and 72 hours from the same batch of embryos as used for E, to which these data can be directly compared. (G) Average of two data sets obtained with J_m constructs, where J_m represents mutated target sites for factor J of module A (Yuh et al., 1994; see Materials and Methods). The coherence of the individual data sets to one another (not shown) was similar to that illustrated in Fig. 2A,B; i.e., the form of the time course measured for each construct was the same in both data sets, while the overall magnitude was higher in one data set than is the other. Constructs tested were *GBA-BpCAT*, *GBA(J_m)-BpCAT*, *BA-BpCAT*, *BA(J_m)-BpCAT* and *B-BpCAT*; measurements were made at 20, 30, 48, 60 and 72 hours. (H) Data set obtained with $(CG)_2$ constructs, where $(CG)_2$ represents a fragment of DNA from the *Endo16* Bp that includes two target sites for the CG factor (Yuh et al., 1994; see Materials and Methods). $(CG)_2$ was inserted immediately upstream of the SV40 promoter, at the junction between it and upstream regulatory modules A or B. A second individual data set (not shown) was entirely coherent in form for every construct, and produced exactly similar results. Constructs tested are *A($CG)_2$-SVpCAT*, *A-BpCAT*, *A-SVpCAT*, *B($CG)_2$-SVpCAT*, *B-BpCAT* and *B-SVpCAT*; measurements were made at 20, 30, 48 and 60 hours (first data set) and 20, 30, 48 and 72 hours (second data set not shown).

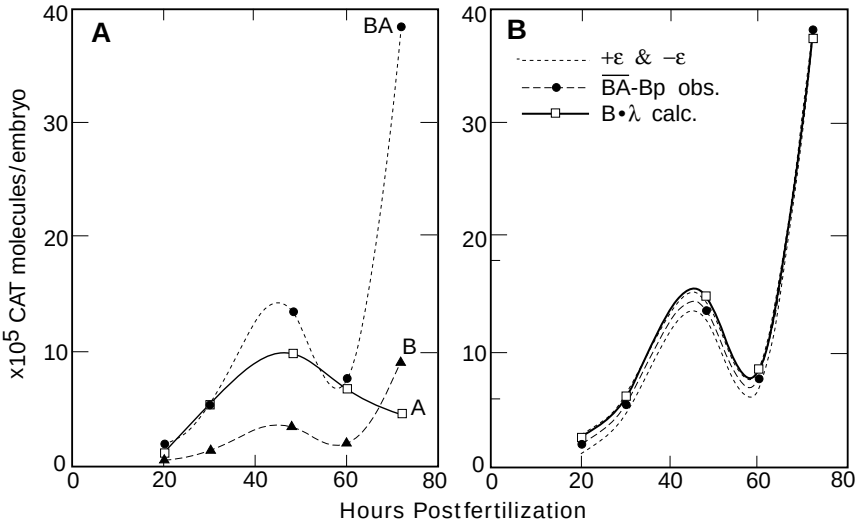

Fig. 3. Test of model $\overline{BA}$=B•λ for time course of *BA-BpCAT*. (A) Data for *BA-BpCAT*, *B-BpCAT* and *A-BpCAT* excerpted from Fig. 2C. (B) The curve generated by B•λ, where λ is set at its best least squares value, is shown as a solid line, i.e., a spline interpolation from the calculated data points. These are indicated as open boxes. The target curve, the observed time course for *BA-BpCAT*, is reproduced from A, using the same symbols. The dashed lines indicate the envelope generated by the calculation of ±ε as % max (see legend to Table 1). The value of λ for this solution is 4.2 (Table 1).

the time courses of the individual SVp constructs also fails (Table 2). We consider in Discussion the significance of the conclusion that use of SVp requires *multiplication* of the individual A and B module time courses.

Interaction of negatively acting *Endo16 cis*-regulatory elements with module A

As summarized above, Yuh and Davidson (1996) found that there are three negatively acting *cis*-regulatory modules, each functioning to exclude expression within the territories that abut the boundaries of the vegetal plate: their functions are required because all three of the positive modules G, B and A are active not only in the vegetal plate, but also in the adjacent regions, as shown by the occurrence of ectopic as well as vegetal plate expression in embryos bearing *A-BpCAT*, *B-BpCAT* or *G-BpCAT*, and in embryos bearing *GBA-BpCAT* (Yuh and Davidson, 1996). The three negative modules are alike in several respects: (*i*) none has appreciable transcription-stimulating activity on its own; (*ii*) each *decreases* expression about two-fold when associated with the combined GBA activator and more than two-fold when linked to module A alone; (*iii)* none produces any repressive effects when linked to modules B or G alone; (*iv*) all are converted into positively acting elements that increase expression by LiCl treatment of the embryos; (*v*) module A is required to confer LiCl sensitivity on all three negative modules just as it is required to confer repressive activity in untreated embryos. Yuh and Davidson (1996) concluded that, in the skeletogenic and adjacent ectoderm territories, the negative modules act through module A.

To further explore the mode of function of the three negative modules, we sought to extract from the data shown in Fig. 2E the time courses of their repressive activities. The repressive activity is expressed in different cells than is the positive activity of constructs that include both negative and positive modules. Therefore, we subtracted from the time-course data for *A-BpCAT* shown in Fig. 2E the time-course data for *DCA-BpCAT*, *FA-BpCAT* and *EA-BpCAT*. Note that the time course measured for *A-BpCAT* in this series is very similar in form to that shown for *A-BpCAT* in Fig. 2A-D (and in Yuh and Davidson, 1996). The results of these subtractions are shown in Fig. 6. Here we see that the three different curves representing the repressive activities of modules DC, F and E are remarkably similar to one another, and to the time course of positive module A function itself, which is superimposed in Fig. 6. This result is consistent with the conclusion that negative module function is directly dependent on the activity of module A, i.e., that these elements function by means of interactions with module A. Fig. 2F shows that when in the

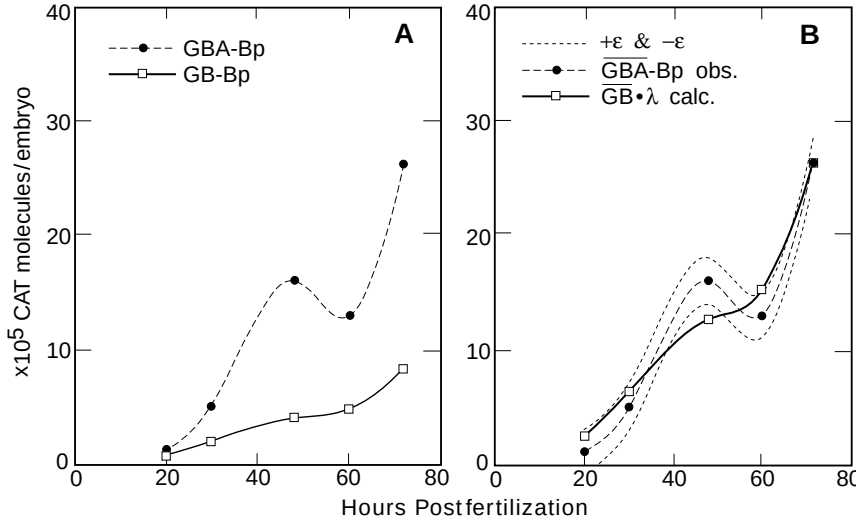

Fig. 4. Test of model $\overline{GBA}$=$\overline{GB}$ •λ, for time course of *GBA-BpCAT*. (A) Data for *GBA-BpCAT* and *GB-BpCAT* imported from Fig. 2C. (B) Comparison of observed and calculated functions. The curve generated by $\overline{GB}$ •λ, where λ is set at the minimal least squares value (see Table 2), is shown as a solid line. This is superimposed on the target curve observed for *GBA-BpCAT*, from A. The dashed lines represent the ±ε envelope, as in Fig. 3.

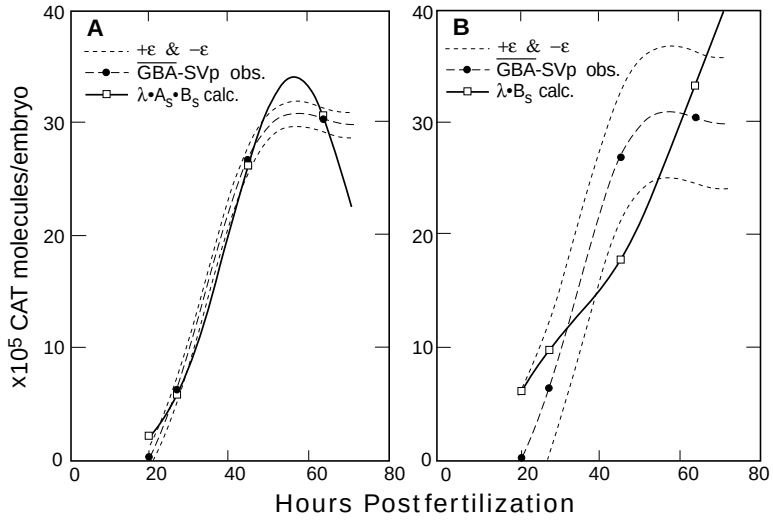

Fig. 5. Models for *GBA-SVpCAT*. (A) Test of model $\overline{GBA}_s=A_s•B_s•\lambda$. Presentation and symbolism are as in Figs 3 and 4. Data are shown in Table 2. Data for *GBA-SVpCAT, A-SVpCAT, B-SVpCAT* and *G-SVpCAT* are shown in Fig. 2D. The function generated by $A_s•B_s•\lambda$ and its envelope ±ε are indicated by the solid line and open boxes, superimposed on the target curve observed for *GBA-SVpCAT*. (B) Test of model $\overline{GBA}_s=B_s•\lambda$. Though solutions of this form are useful for *BA-BpCAT* (Fig. 3) and *GBA-BpCAT* (Fig. 4), they are not applicable to *GBA-SVpCAT*.

context of all three positive modules each of the negatively acting modules depresses the overall level of activity during the 40-60 hour period when module A alone is most active. Where X represents F, E or DC, the time courses of *GXBA-BpCAT* are generally similar to those of *XA-BpCAT* (Fig. 2E,F). Thus we believe that the relations that hold in the context of the complete positive regulatory system are similar to those which can be seen clearly in Fig. 6 for the interaction of the negative modules with module A alone. However, a much greater density of data would be required to test this point directly.

Experimental tests of inferences derived from successful models of Tables 1 and 2

Several mutations of expression constructs utilized for the measurements shown in Fig. 2 were built, in order to test two key functional inferences. The first of these was the inference that one of the other factors binding in module A, rather than the J transcription factor, is responsible for the constant synergistic effect through developmental time of module A on module B. To test this, we mutated the J target site (J_m) and constructed *GBA(J_m)-BpCAT* and *BA(J_m)-BpCAT* (see Materials and Methods). The mutation that was inserted totally destroys binding of natural J factor that had been partially purified by affinity chromatography (data not shown). The activity of the J_m constructs was compared with that of *GBA-BpCAT, BA-BpCAT* and *B-BpCAT* in two independent time-course experiments, the average of which is shown in Fig. 2G. The main result of this experiment is evident by inspection. *BA(J_m)-BpCAT* and *BA-BpCAT* are expressed almost identically and display an almost identical amplification with respect to *B-BpCAT*. This demonstrates that module A elements other than the factor J target site are indeed responsible for the amplification function of module A or module B. *GBA-BpCAT* is expressed a little more actively than is *GBA(J_m)-BpCAT*. Thus, as Table 1 shows, a calculation of the model $\overline{GBA}=\overline{GBA(J_m)}•\lambda$ gives an excellent fit with λ only about 1.4. This modest linear difference might indicate a mild synergistic function on the part of the J factor, which is observed only when module G is present; i.e., it is possible that J factor amplifies the positive output of module G, but we doubt the significance of this result since as shown in Fig. 2A-C *GBA-BpCAT* is only marginally more active than is *BA-BpCAT*. What is certain is that the much stronger synergism observed between modules B and A does not depend on factor J.

Another inference that we challenged concerns the difference between the SV40 and the *Endo16* promoter elements used in these studies. The experiments of Table 2 showed that the positive modules G, B and A function independently and multiplicatively when combined with SVp, but that their synergism is linear and not multiplicative when the same elements are combined with Bp. Furthermore, as shown by Yuh and Davidson (1996) and in Fig. 2D of this paper, a major difference between SVp and Bp is the relatively very low activity of SVp when driven by module A or module B alone. We inferred above that these differences could be due to the presence of two target sites within the *Endo16* Bp fragment for the yet unidentified factor provisionally called 'CG' (Yuh et al., 1994), which is indicated by green ovals in Fig. 1.

To test the idea that the difference between the endogenous *Endo16* Bp and *SVp* is due at least in part to the CG target site sequences, we sought to convert the SV40 promoter into a promoter that would behave like the *Endo16* Bp by inserting CG target sites just upstream of SVp in several of our expression constructs. Thus, as described in Materials and Methods, we created *A($CG)_2$-SVpCAT* and *B($CG)_2$-SVpCAT*, and compared their activities to *A-BpCAT, B-BpCAT, A-*

Table 2. Synergistic models for modules G, B and A linked to SV40 early region basal promoter[1]

Model	ε (% max)	λ	λ/function
$\overline{GBA}_s=B_s•A_s•\lambda$	1.08 (4%)	27	5.2
$\overline{GBA}_s=G_s•B_s•A_s•\lambda$	1.38 (5%)	51	3.7
$\overline{GBA}_s=G_s•A_s•\lambda$	5.95 (20%)	52	7.2
$\overline{GBA}_s=G_s•B_s•\lambda$	5.4 (18%)	41	6.4
$\overline{GBA}_s=(A_s+B_s+G_s)•\lambda$	5.6 (19%)	9.6	3.2
$\overline{GBA}_s=B_s•\lambda$	5.8 (19%)	21.4	21.4
$\overline{GBA}_s=A_s•\lambda$	7.7 (25%)	24.9	24.9
$\overline{GBA}_s=G_s•\lambda$	8.97 (30%)	41.6	41.6

[1]Denotation as in Table 1, except that subscript s denotes use of SV40 early region promoter rather than endogenous Bp.

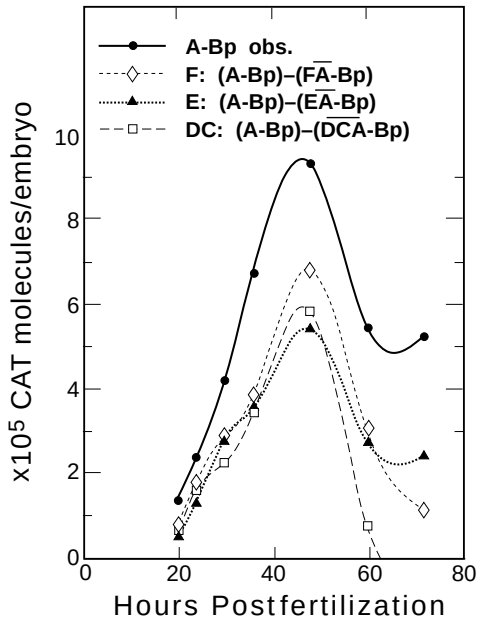

Fig. 6. Time course of repressive function in the *Endo16 cis*-regulatory system. The differences between the time-course data for *A-BpCAT* and *FA-BpCAT*, *EA-BpCAT* and *DCA-BpCAT*, respectively, are shown. Data are from Fig. 2E. Subtractions were carried out at each time point, and smooth curves were thereafter interpolated through the differences, as previously (interrupted lines). The time course for *A-BpCAT* is also reproduced (solid line) for comparison.

SVpCAT, *B-SVpCAT* and *GBA-SVpCAT*. Results are illustrated in Fig. 2H (except for *GBA-SVpCAT*, which was measured in a different experiment) and are listed in Table 3. These experiments also afforded an independent check, using entirely different measurements, on the behavior of *GBA-SVpCAT*. Table 3 shows that, almost exactly as in the experiment given in the first line of Table 2 and illustrated in Fig. 2D, the model $\overline{GBA}s=Bs\bullet As\bullet\lambda$ again provides the best fit, with λ (per function)=7.8 and ε (% max)=3%. The main conclusion from Fig. 2D is also evident by inspection. With respect to module A, when the CG sites are added to SVp, the activity of the newly created promoter is now almost identical to that of Bp; i.e., *A-BpCAT*≈*A(CG)₂-SVpCAT*, and both are much more active than is *A-SVpCAT*. Table 3 indicates that the linear amplification factor λ is, as expected, essentially the same for $A=A_s\bullet\lambda$ as for $A_s(CG)_2=A_s\bullet\lambda$ (where as above A represents *A-BpCAT* and A_s represents *A-SVpCAT*). These solutions are shown for the data set in Fig. 2H in Fig. 7A and B. A second data set, almost the same as that in Fig. 2H, gave the same results (data not shown). We conclude that the CG target sites suffice to endow the SV40 promoter with the same linear ability to amplify the output of module A, as observed with the endogenous Bp element. Therefore, the CG DNA-binding factor probably carries out this function.

Fig. 2H also shows that the CG target sites fail completely to amplify the activity of module B. *B(CG)₂-SVpCAT* is expressed almost identically with *B-SVpCAT*. The amplification function mediated by the CG sites thus depends on interaction specifically with elements of the contiguous region, module A. We wonder whether it is relevant that module A also contains CG sites, while module B lacks them (Fig. 1A); perhaps the CG-binding factor works by forming homomultimers, as does SpGCF1 (Zeller et al., 1995a). The source of the difference in activity between *B-BpCAT* and *B-SVpCAT* remains unresolved. This could depend on the SpGCF1 sites also present in Bp (Fig. 1) (although such an hypothesis would require that the Sp1 sites of SVp that resemble SpGCF1 sites [Yuh and Davidson, 1996] are for some reason inadequate).

DISCUSSION

Synergistic interactions amongst positively acting *Endo16 cis*-regulatory elements

The three positively acting modules of the *Endo16 cis*-regulatory system, G, B and A, respond to different transcription factors. We know this from direct determination of the minimum diversity of DNA-binding proteins that interact with target sites within these three regions (Yuh et al., 1994; see summary in Fig. 1). The only sites shared amongst modules G, B and A are those at which the ubiquitous SpGCF1 factor binds (Zeller et al., 1995a,b). This factor may act by multimerizing once it is bound, thus promoting or stabilizing physical interactions amongst distant regions of the regulatory system. When tested in isolation, i.e., in the constructs *G-BpCAT*, *B-BpCAT* and *A-BpCAT* each of these modules promotes vegetal plate expression as well as ectopic expression in the adjacent ectoderm and skeletogenic territories but not in the oral and

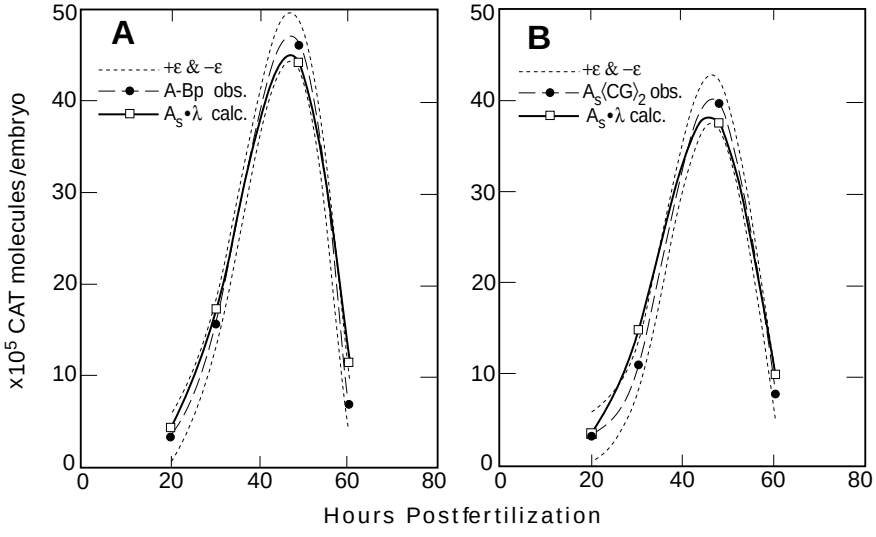

Fig. 7. Quantitative demonstration that *A(CG)₂-SVpCAT* and *A-BpCAT* produce the same linear amplification of *A-SVpCAT*. (A) Test of Model $A=A_s\bullet\lambda$. (B) Test of model $As(CG)_2=A_s\bullet\lambda$. As Table 3 shows, λ=2.1 for the solution shown in (A) and 1.8 for the solution shown in (B). Symbolism as in Figs 3-6.

aboral ectoderm territories that derive from cells above the horizontal 3rd cleavage plane (Yuh and Davidson, 1996). These three positive modules display distinct quantitative time courses of activity during embryonic development (Yuh and Davidson, 1996; Fig. 2 of this paper). We presume that their individual time courses reflect the effective occupancy of at least the key target sites within each module, and that these occupancies differ over time due to the changing availability of active forms of the respective transcription factors that bind at these sites. With respect to spatial expression, in unpublished experiments, we have shown that the key transcription factor in module A is that symbolized as 'J' in Fig. 1; and the key factor in module B is that symbolized as 'I' in Fig. 1. When combined with the *Endo16* Bp, oligonucleotide representations of these target sites alone suffice to reproduce the same spatial pattern of expression as generated by modules A and B, respectively. They also reproduce the time courses observed for *A-BpCAT* and *B-BpCAT*, though at significantly reduced quantitative levels. The experiment shown in Fig. 2G, however, shows that even when the factor J site is destroyed by mutation, the characteristic time course generated by module A is preserved. Therefore both factor J and at least one other of the transcription factors that bind within module A are presumably bound at peak concentrations at 40 hours.

A focus of this study is the synergism displayed when the positive modules are physically linked in expression constructs. In these constructs the 5′ to 3′ order in which modules G, B and A naturally occur has been preserved. We find that in none of the cases that we have analyzed is the output of the physically linked modules the sum of their separate outputs; rather the output of the combined modules is always significantly greater than the sum of the individual activities. Furthermore, we demonstrate (Figs 3, 4) that module A produces a linear amplification of the output of modules B and GB when linked together with these, rather than the rising and descending patterns displayed by module A when tested in isolation, with either Bp or SVp. The linear amplification also is observed when the target site for Factor J is mutated (Table 3), so this synergistic amplification function of module A depends on one of the other factors binding in module A. This result was already implied by the differences between the constant value of the synergism factor λ and the peaked time course produced by constructs driven by factor J.

When linked to the stripped down heterologous SV40 basal promoter, the linear synergism displayed by module A in Bp constructs is no longer observed. The output of *GBA-SVpCAT* is approximated instead by the *products* of the outputs of the constituent modules, and a much larger amplification factor as well. This is shown in Tables 2 and 3, and illustrated in Figs 2D and 5. The difference in behavior illuminates the function of sequences present in the endogenous *Endo16* Bp fragment. We think that the most likely interpretation is that, in the *GBA-SVpCAT* construct, outputs of the three modules are integrated at the basal transcription complex that binds within SVp (see legend to Fig. 8). That is, the individual upstream modules individually interact with the basal transcription apparatus, resulting in the relatively large amplification of activity observed, i.e., in comparison to *A-SVpCAT*, *B-SVpCAT* and *G-SVpCAT* individually.

As pointed out by Yuh and Davidson (1996), the term 'basal promoter' is probably a misnomer for the *Endo16 cis*-regula-

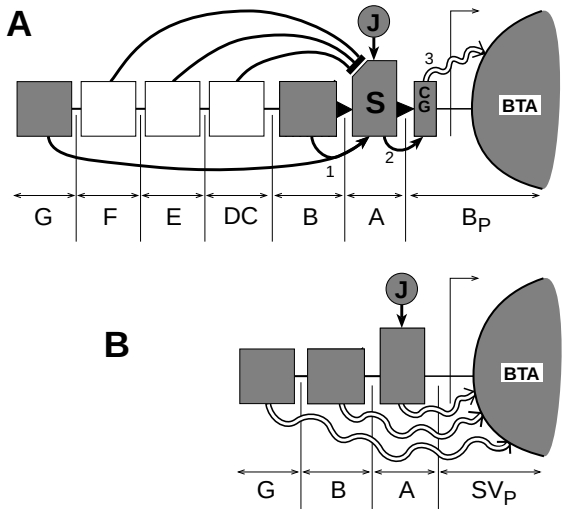

Fig. 8. Diagram summarizing internal interactions within the *Endo16 cis*-regulatory system. Positive modules or elements of modules are indicated as hatched forms and negatively active modules as open boxes. The spacing is not to scale (cf. Fig. 1A). For the complexity and identity of proteins that bind within each module see Yuh et al. (1994), and summary in Fig. 1 of this paper. Arrows indicate positive interactions; terminally barred lines indicate negative interactions. Modules (G-A) are named beneath the horizontal line representing the *cis*-regulatory DNA. Solid arrows denote intermodular interactions, and open wavy arrows indicate interactions (direct or via proteins that do not themselves bind DNA) with the basal transcription apparatus (BTA). The solid arrowheads separating module boxes denote presumed positive functional interactions amongst modules that are oriented functionally as shown. J represents 'factor J,' a unique DNA-binding protein of module A (lavender oval in Fig. 1A). An oligonucleotide that includes the specific target site for factor J alone confers the spatial and temporal behavior of module A on expression constructs when it is linked to Bp, though not its normal quantitative level of activity (unpublished data). S indicates the switch function of module A. At least four different proteins in addition to the J factor bind specific DNA target sites within module A, not considering proteins that may interact in turn with the DNA-binding proteins. The J factor functions may involve other proteins as well, and the identity of the proteins that execute the linear booster function of module A and its switch function (see text) is not known. As summarized in text, module A executes four different functions, and except for the one that can be assigned to the J factor, all of these module A interactions are symbolized as a single hatched box labeled S. (A) The complete endogenous regulatory system. Positive and negative functions were established by Yuh and Davidson (1996). Evidence for the specific features shown are as follows. Arrows (1 and 2) are implied by the successful models of Table 1 where (1) represents B or $\overline{GB}$ function; (2) is the modest step-up function (λ=3 to 4) of module A on the output of modules B or B and G. This arrow terminates on the CG box of the Bp region, as demonstrated in experiments of Fig. 2H, Fig. 7 and Table 3. The factor bound at the CG site interacts only with module A, as shown in Fig. 7. The final interaction of this linear, hierarchical system with BTA is indicated by the wavy arrow (3). (B) The *GBA-SVpCAT* system. The solution shown in Table 2 implies that all functional elements interact directly with BTA; multiplying the outputs of the individual functions is equivalent to multiplying the equilibrium constants for these interactions or to adding their free energies (i.e., output is proportional to transcriptional activity as discussed in Materials and Methods, and transcriptional activity should be proportional to modular transcription factor complex concentration, or occupancy).

Table 3. Models tested with $(CG)_2$ SV40 constructs

Model	ε (% max)	λ	λ/function
$\overline{GBA}_S=B_S\bullet A_S\bullet\lambda$	7.8 (3%)	7.8	2.8
$\overline{GBA}_S=B_S\bullet\lambda$	18.4 (7%)	26.6	26.6
$\overline{GBA}_S=A_S\bullet\lambda$	8.34 (34%)	23.4	23.4
$A=A_S\bullet\lambda$	2.7 (6%)	2.1	2.1
$A_S(CG)_2=A_S\bullet\lambda$	2.5 (6%)	1.8	1.8
$B=B_S\bullet\lambda$	3.3 (9%)	4.5	4.5
$B=B_S(CG)_2\bullet\lambda$	3.0 (8%)	4.4	4.4

tory fragment so denoted (see Fig. 1), since the Bp DNA fragment contains multiple target sites for the CG DNA-binding factor and for SpGCF1 (see Fig. 1), in addition to the sequences that presumably support assembly of the TBP-holoenzyme complex. These additional interactions account for the fact that the endogenous Bp is about 10-fold more active when combined with the single modules than is SVp (Yuh and Davidson, 1996; Fig. 2 of this paper). Insertion of CG target sites in the SV40 promoter converts its activity to that of Bp with respect to module A, as shown in Table 3 and Fig. 7.

Thus in the simplified context of the *GBA-SVpCAT* construct, we believe that G, B and A regulatory modules each interact with the basal transcription apparatus. In the natural system, the output of modules G, B and A depends instead on a sequence of interactions amongst upstream elements, including those that occur upstream of the start site in the endogenous 'basal' promoter. A similar theme obtains with respect to the negative spatial control functions modulated by the *Endo16 cis*-regulatory system, in that here again essential interactions occur amongst upstream regulatory elements.

Module A and the negative control functions mediated by DC, E and F modules

We summarized above the evidence of Yuh and Davidson (1996), showing that the repression of *Endo16* transcription in skeletogenic lineages by the DC module, and in ectoderm lineages by F and E modules, all operate by means of interactions with module A. None of the various functions identified for these three modules were found to be executed in constructs lacking module A, and the other positive modules, B and G, cannot substitute for A. An additional item of evidence is presented in Fig. 6 of this paper. Here we see that the time course of repressive DC, F and E function almost perfectly parallels the time course of module A positive function. It follows from this evidence, taken together with that of Yuh and Davidson (1996), that in the cells where they are responsible for shutting off *Endo16* expression, the negative modules function by means of interaction with module A. Since they do not work when placed with module B in the absence of A (Yuh and Davidson, 1996), they do not work by direct repressive interaction with the basal transcription apparatus. Yet modules B and G, as well as A, produce ectopic expression if these negative control functions are lacking (Yuh and Davidson, 1996). This means that the key spatial regulators that determine the positive function of modules G, B and A are all present and active not only in the vegetal plate but also in the surrounding domains of the early embryo as well. Therefore, the negative modules cannot function *exclusively* by interfering with the positive activity of module A per se. Instead module A seems to mediate the repressive output of the negative modules, by acting as a *switch* that permits or precludes expression of the other positive modules. When this switch is closed by the negative modules, none of the G-, B- or A-positive modules can act.

A qualitative model of the *Endo16 cis*-regulatory system

A hierarchical intermodular organization emerges from these studies of the *Endo16 cis*-regulatory system. In Fig. 8A, we combine the results obtained here with those of Yuh and Davidson (1996). The diagram represents the qualitative interrelations amongst the positive (hatched) and negative (open) modules that constitute the complete system. As described in more detail in the legend to Fig. 8, the modular functions presumed in this diagram are as determined by Yuh and Davidson (1996) with a number of additions from this work.

Module A serves as the integrating 'processor' of the whole system. As shown by Yuh and Davidson (1996), module A is required for function of the negative modules. We know from a series of experiments on module A mutations (unpublished data) that the module A site, which is required for the LiCl response of modules F, E and DC, is adjacent to that at which factor J binds, but is separate from that site. This adjacent site is therefore (Yuh and Davidson, 1996) probably also required for the module A function of mediating the spatial repression executed by the three negative modules. Another function of module A is to act as a 3- to 4-fold linear booster of the outputs of the other positive modules, i.e., B or GB (Table 1). By comparison with the results obtained with the SV40 constructs, modules B and G probably do not act by directly contacting the basal transcription apparatus individually, but instead interact with some element(s) of module A, which amplifies their output. However, this function does not require participation of the J factor of module A, as shown in the experiments with constructs bearing a J target site, summarized in Table 3. A third function of module A, established by Yuh and Davidson (1996), is to generate the early vegetal expression pattern, and for this factor J is required (unpublished data). A fourth function of module A is to communicate with the basal promoter. We show here that the CG factor target sites of the 'basal' promoter fragment are also required for the linear, modestly synergistic activity of the positive modules observed for the Bp constructs (Table 3; Fig. 7). In their absence, the SV40 promoter constructs containing single positive *Endo16* modules function poorly, and when the positive modules are combined, they synergize multiplicatively. We interpret this to mean that, in the SV40 constructs, each module interacts directly with the basal transcription apparatus, as indicated in Fig. 8B. However, the CG factors interact only with module A since, when linked instead to module B in the context of the SV40 promoter, they have no effect (Fig. 2B; Fig. 7). One possibility is that the CG factors that bind within module A mediate the interaction with the CG target sites in the basal promoter region, by means of a homotypic interaction.

In summary, we believe that the relationships uncovered in this work display interactions within the *Endo16 cis*-regulatory domain that are required for both its spatial and temporal developmental functions. The proximal region, module A, appears to play a special role in the integration of both positive and negative input from all the modules further upstream. Perhaps it alone communicates with the basal transcription

apparatus as suggested in Fig. 8A. An experimental caution, which illuminates the significance of the endogenous upstream interactions, is afforded by the SVp experiments. These show that replacement of the endogenous proximal sequence by the elemental SV40 heterologous promoter easily provokes a default set of interactions, that are different from the normal interactions and that obscure the subtle hierarchical organization of the endogenous regulatory system.

The authors gratefully acknowledge perspicacious and detailed reviews of the manuscript by Drs Roy J. Britten, Scott E. Fraser, James A. Coffman and Maria Arnone. It is a pleasure also to acknowledge the skillful technical assistance of Ms. Jessica Chang, then a Caltech undergraduate. This research was carried out with the support of the Stowers Institute for Medical Research, the Caltech Beckman Institute and with an NIH Grant (HD-05753).

REFERENCES

Cabrera, C. V., Lee, J. J., Ellison, J. W., Britten, R. J. and Davidson, E. H. (1984). Regulation of cytoplasmic mRNA prevalence in sea urchin embryos: Rates of appearance and turnover for specific sequences. *J. Mol. Biol.* **174**, 85-111.

Davidson, E. H. (1986). *Gene Activity in Early Development*, Third Edition. Orlando, Florida: Academic Press.

Davidson, E. H. (1994). Molecular biology of embryonic development: How far have we come in the last ten years? *BioEssays* **16**, 603-615.

Flytzanis, C. N., Britten, R. J. and Davidson, E. H. (1987). Ontogenic activation of a fusion gene introduced into sea urchin eggs. *Proc. Natl. Acad. Sci. USA* **84**, 151-155.

Franks, R. R., Anderson, R., Moore, J. G., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1990). Competitive titration in living sea urchin embryos of regulatory factors required for expression of the *CyIIIa* actin gene. *Development* **110**, 31-40.

Godin, R. E., Urry, L. A. and Ernst, S. G. (1996). Alternative splicing of the *Endo16* transcript produces differentially expressed mRNAs during sea urchin gastrulation. *Dev. Biol.*, in press.

Kirchhamer, C. V., Yuh, C.-H. and Davidson, E. H. (1996). Modular *cis*-regulatory organization of developmentally expressed genes: Two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proc. Natl. Acad. Sci. USA* **93**, in press.

Livant, D., Cutting, A., Britten, R. J. and Davidson, E. H. (1988). An *in vivo* titration of regulatory factors required for expression of a fusion gene in transgenic sea urchin embryos. *Proc. Natl. Acad. Sci. USA* **85**, 7607-7611.

Nocente-McGrath, C., Brenner, C. A. and Ernst, S. G. (1989). *Endo16*, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* **136**, 264-272.

Ransick, A. and Davidson, E. H. (1993). A complete second gut induced by transplanted micromeres in the sea urchin embryo. *Science* **259**, 1134-1138.

Soltysik-Espanola, M., Klinzing, D. C., Pfarr, K., Burke, R. D. and Ernst, S. G. (1994). *Endo16*, a large multidomain protein found on the surface and ECM of endodermal cells during sea urchin gastrulation, binds calcium. *Dev. Biol.* **165**, 73-85.

Yuh, C.-H. and Davidson, E. H. (1996). Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo. *Development* **122**, 1069-1082.

Yuh, C.-H., Ransick, A., Martinez, P., Britten, R. J. and Davidson, E. H. (1994). Complexity and organization of DNA-protein interactions in the 5′ regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mech. Dev.* **47**, 165-186.

Zeller, R. W., Griffith, J. D., Moore, J. G., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995a). A multimerizing transcription factor of sea urchin embryos capable of looping DNA. *Proc. Natl. Acad. Sci. USA* **92**, 2989-2993.

Zeller, R. W., Coffman, J. A., Harrington, M. G., Britten, R. J. and Davidson, E. H. (1995b). SpGCF1, a sea urchin embryo transcription factor, exists as five nested variants encoded by a single mRNA. *Dev. Biol.* **169**, 713-727.

(Accepted 13 September 1996)