

***Cis*-regulatory logic in the *endo16* gene: switching from a specification to a differentiation mode of control**

Chiou-Hwa Yuh^{1,2}, Hamid Bolouri³ and Eric H. Davidson^{2,*}

¹Stowers Institute for Medical Research, Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

²Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA

³Science and Technology Research Centre, University of Hertfordshire, UK

*Author for correspondence (e-mail: davidson@mirsrsky.caltech.edu)

Accepted 4 December 2000; published on WWW 7 February 2001

SUMMARY

The *endo16* gene of *Strongylocentrotus purpuratus* encodes a secreted protein of the embryonic and larval midgut. The overall functional organization of the spatial and temporal control system of this gene are relatively well known from a series of earlier *cis*-regulatory studies. Our recent computational model for the logic operations of the proximal region of the *endo16* control system (Module A) specifies the function of interactions at each transcription factor target site of Module A. Here, we extend sequence level functional analysis to the adjacent *cis*-regulatory region, Module B. The computational logic model is broadened to include B/A interactions as well as other Module B functions. Module B drives expression later in development and its major activator is responsible for a sharp, gut-specific increase in transcription after gastrulation. As shown earlier, Module B output undergoes a synergistic amplification that requires interactions within Module A. The interactions within Module B that are

required to generate and transmit its output to Module A are identified. Logic considerations predicted an internal *cis*-regulatory switch by which spatial control of *endo16* expression is shifted from Module A (early) to Module B (later). This prediction was confirmed experimentally and a distinct set of interactions in Module B that mediate the switch function was demonstrated. The *endo16* computational model now provides a detailed explanation of the information processing functions executed by the *cis*-regulatory system of this gene throughout embryogenesis. Early in development the gene participates in the specification events that define the endomesoderm; later it functions as a gut-specific differentiation gene. The *cis*-regulatory switch mediates this functional change.

Key words: Gene regulation, Computational model, Gene transfer, Sea urchin

INTRODUCTION

The information processing capacity of animal *cis*-regulatory systems is the basal feature that makes development possible. Each gene, in each cell of a developing animal, must read and respond to the presence or absence of multiple inputs. In effect, these inputs provide the gene with the regulatory information it requires to determine its own activity: this includes signaling inputs from adjacent cells and inputs that indicate what other relevant genes have been functioning in the cell in which the gene resides. These inputs are presented to the gene in terms of concentrations and activities of nuclear transcription factors. The heritable structural basis for *cis*-regulatory information processing functions consists of the target site sequences at which transcription factors bind to the DNA. The identity and disposition of these sites specify the regulatory activities that can be executed by the *cis*-regulatory system, depending on circumstances. This genetic hardwiring causally determines to which inputs each gene regulatory system will respond (Davidson, 1990; Davidson, 1999; Davidson, 2001).

The *cis*-regulatory system of the *endo16* gene of the sea

urchin *Strongylocentrotus purpuratus* provides a fascinating illustration of these principles. The major working parts of this 2300 bp *cis*-regulatory system have been defined in a series of earlier studies (Ransick et al., 1993; Yuh and Davidson, 1996; Yuh et al., 1994; Yuh et al., 1996). A key feature of our experimental approach has been the use of quantitative measurements of the kinetic outputs of various embryonic expression constructs that had been introduced into fertilized eggs. In this way, regulatory functions could be perceived that would otherwise be entirely invisible. An initial set of studies using this method (Yuh et al., 1996) indicated that the proximal region of the *endo16* regulatory system, the 185 bp cluster of sites known as Module A, plays an especially important role with respect to the operation of the remainder of the *cis*-regulatory system, and so we decided to determine the functional significance of each of the eight different sites of specific DNA/protein interaction that had been detected in vitro within this region (Yuh et al., 1994; Yuh et al., 1998). Remarkably, Module A was revealed to operate as a hardwired logic processor, the output of which is conditional on its inputs. These derive both from outside the gene, and from other

modular domains of the same *cis*-regulatory apparatus. Every one of the target sites previously observed was indeed shown to have a specific functional role. In order to describe these functions with respect to one another, a system-level logic model was required (Yuh et al., 1998). Module A is responsible for installing *endo16* expression in vegetal plate during its initial specification, and after this, a major function of Module A turned out to be receiving and processing the transcriptional regulatory output of the adjacent site cluster, Module B; this *cis*-regulatory element drives *endo16* expression as the gut differentiates. We now report the extension of our site-specific experimental analysis, and of the computational model, so as to include Module B. The complex interactions between Modules B and A over developmental time can now be treated explicitly. The resulting model provides satisfactory predictions of the behavior of the *endo16* *cis*-regulatory system for every mutational variant of the sequence tested. Most importantly, we now have an explanation of the multiple functions of this regulatory system that is couched directly in terms of its identified DNA target sites.

endo16 encodes a large, probably polyfunctional, secreted protein of the embryonic and larval midgut (Nocente-McGrath et al., 1989; Soltysik-Espanola et al., 1994). The gene begins to be expressed in mid-cleavage, long before there is a gut (Ransick et al., 1993; Godin et al., 1996). The initial domain of expression is the *veg2* lineage, from which most of the endomesoderm of the embryo derives (Davidson et al., 1998 for review). The major early activator of the gene, which binds in Module A, is an Otx regulator (SpOtx; Gan et al., 1995; Li et al., 1997; Li et al., 1999; Yuh et al., 1998). Early on, this activator is present ubiquitously. However, *endo16* expression is made specific to the *veg2* territory by (1) repressors that bind in an upstream region called Module DC and prevent expression in the adjoining skeletogenic precursors in the center of the vegetal plate; and (2) by other repressors binding in the adjacent regions, Modules F and E, which prevent expression in the overlying ectodermal domains (Yuh and Davidson, 1996; Yuh et al., 1996). However, in order for the upstream interactions with these spatial repressors to have any effect, a certain target site in Module A is required (Yuh et al., 1998). After the gut is formed, the level of *endo16* expression increases sharply, and the level of transcription now depends specifically on an activator (UI) that binds in Module B. Its activity rises late in development, as we inferred earlier (Yuh et al., 1998), and demonstrate anew here. But this Module B function also requires Module A, the late role of which is to perform a remarkable four- to fivefold amplification of the transcriptional stimulatory activity generated within Module B (Yuh et al., 1996; Yuh et al., 1998). A construct consisting only of Module B linked to Module A and its basal promoter produces almost the same temporal and quantitative output as does the whole 2300 bp *cis*-regulatory system (Yuh et al., 1996). However, since this construct, BA-Bp•CAT, lacks the repressor modules D, C, E and F, it is expressed in a few cells outside the gut as well as appropriately in the endoderm (CAT stands for chloramphenicol acetyl transferase, the reporter protein, whose enzymatic activities provide the kinetic output data). Two specific binding sites in Module A are required to link Module B to Module A and to mediate the amplification of Module B input (Yuh et al., 1998); below we demonstrate

the donor apparatus of Module B, which is required to provide the input to these sites. In addition we show the existence of a Module B to Module A switch that we refer to as the 'BA intermodule input switch'. The function of this switch is to turn off the input deriving from the Otx activator of Module A once the level of Module B activity rises above a threshold. This switch was predicted from measurements showing that from gastrula stage onwards, the kinetics of BA-Bp•CAT output are constituted entirely of the scalar magnification of the kinetic output of Module B, without any other kinetic contribution from Module A (Yuh et al., 1998). Here, we demonstrate the target site sequence of Module B that mediates this switch function and define the interrelationships by which it is controlled.

A simplified summary of these control relationships is shown in Fig. 1. We see the useful functions of the BA intermediate input switch: (1) It shifts control from a ubiquitous activator (Otx) to one which is evidently confined spatially to the archenteron, or later, subregions thereof. (2) The switch thereby relieves the system from further dependence on the cleavage- and blastula-stage spatial repression system, of which the necessary factors bind in

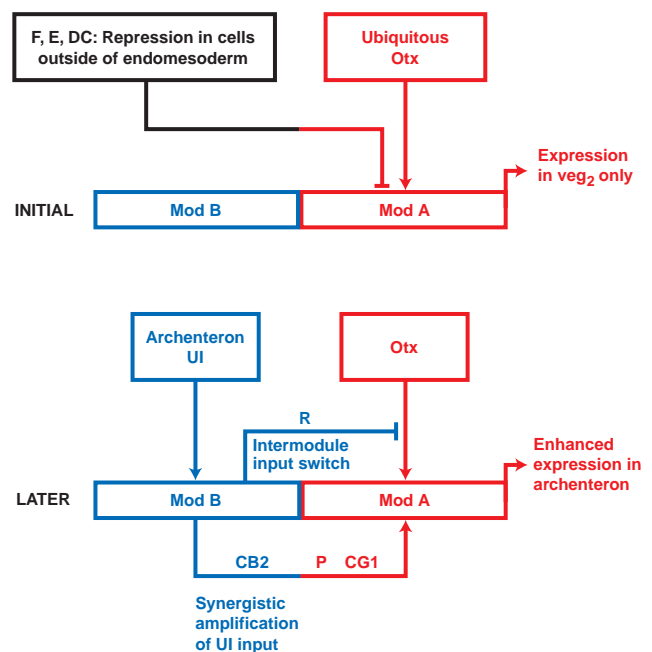


Fig. 1. Successive pathways of spatial and temporal control within the *endo16* *cis*-regulatory system. The diagram summarizes results from several previous studies (Yuh et al., 1996; Yuh et al., 1998; Yuh and Davidson, 1996). Module (Mod) A functions are shown in red; Module B functions in blue. Early in development the *endo16* gene responds to a ubiquitous activator (SpOtx1) binding in Module A, but in order to achieve accurate spatial expression, activity must be extinguished outside the *veg2* endomesodermal domain by repressors binding in the upstream modules (F, E and DC). Later in development, the activity of a transcriptional regulator (UI) binding in Module B rises and the internal BA intermodule input switch shuts off Otx input so that the system is now driven only by Module B input. This input is amplified in Module A, which provides the sole communication with the basal transcription apparatus: see text and references (Yuh et al., 1996; Yuh et al., 1998; Yuh and Davidson, 1996) for evidence and details.

Modules F, E and DC. This may be a necessary change, since these repressive interactions are likely to be in part signal mediated (C.-H. Y., unpublished observations), and all of the original *veg2* interfaces with other blastomeres are altered at gastrulation (see Davidson et al., 1998). (3) The switch is itself a temporal control subsystem, since it is activated only when the activity of the Module B regulator that we call UI rises. (4) The switch shunts control of *endo16* expression into a pathway capable of driving very high level expression in the differentiating gut, i.e. the BA amplification subsystem,

whereas at this time the activity of the SpOtx factor that drives Module A is in process of declining.

MATERIALS AND METHODS

All of the methods and procedures used to obtain the kinetic measurements of expression construct output referred to in this paper have been described earlier (Yuh et al., 1996; Yuh et al., 1998). Most of the expression constructs, however, were generated specifically for this work. A brief summary of their provenance follows; for constructs

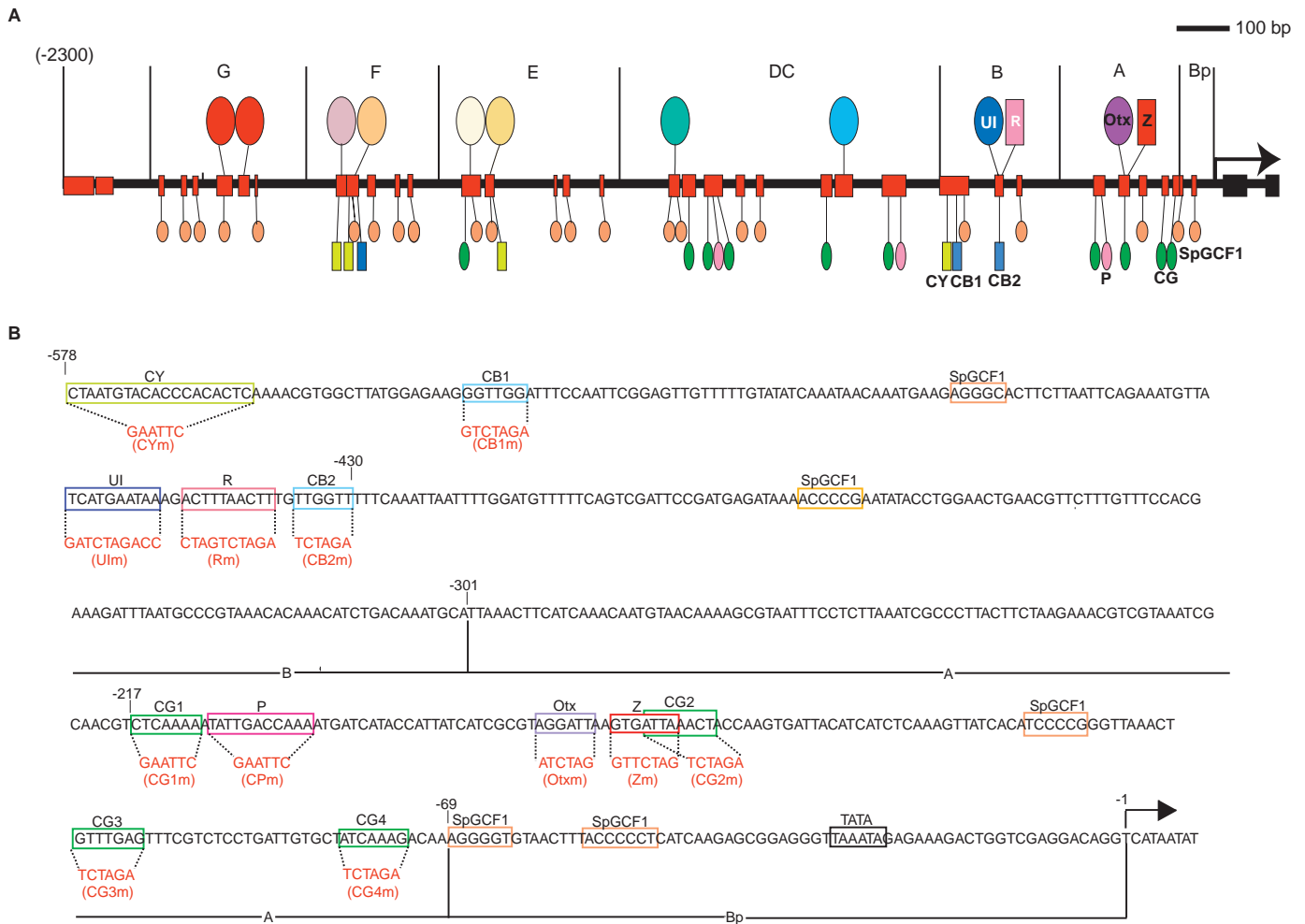


Fig. 2. Modules B and A of the *endo16* cis-regulatory system. (A) The protein binding maps of the 2300 bp sequence that is necessary and sufficient to generate an accurate spatial and temporal pattern of expression (Ransick et al., 1993; Yuh and Davidson, 1996). The map is modified slightly in accordance with current evidence from that derived by Yuh et al. (Yuh et al., 1994) by a three-step procedure. First, all sites of high specificity interaction were determined by a rapid gel shift mapping method in which embryo nuclear extract was reacted with nested sets of end-labeled probes (high specificity here denotes interactions for which $k_r \geq 5 \cdot 10^3$, where $k_r = k_s/k_n$, if k_s is the equilibrium constant for the interaction with a given site, and, k_n is the equilibrium constant for reaction of the factor with synthetic double-stranded DNA polynucleotide). Second, the location of the sites was further narrowed down by oligonucleotide gel shift competition mapping. Third, the binding factors were enriched by affinity chromatography and each challenged in turn for crossreaction with probes representing all of the identified binding sites. This permitted determination of the complexity and individuality of the binding factors (indicated by color in Fig. 1A), based both on the cross-reaction tests and on their molecular sizes, as estimated by DNA-protein interaction blots. Factors indicated above the line representing the DNA bind uniquely in a single region of the sequence; those indicated below interact in multiple regions. The factors with which this paper is concerned, i.e. those of Modules B and A, are indicated by labels: for Module A site functions see Yuh et al. (Yuh et al., 1998); for Module B, this paper. For overview of modular functions in this system see reviews by Davidson (Davidson 1999; Davidson, 2001); Module G is a general booster for the whole system; F, E and DC are repressor modules that permit ectopic expression. (Modified from Yuh et al. (Yuh et al., 1994)) (B) Sequence of *cis*-regulatory DNA of Modules B and A. Core target site sequences (Yuh et al., 1994; Yuh et al., 1998; Zeller et al., 1995a; Li et al., 1997) are boxed in the same respective colors as in (A), and beneath each, in red, is shown the target site mutations used to test function *in vivo* in the absence of that interaction.

referred to whose names do not appear below, see previous publications (Yuh et al., 1996; Yuh et al., 1998).

All mutations shown in red in Fig. 2 were generated by PCR. The basic method was as follows: primers were made for the expression constructs to be mutated (i.e., B-Bp•CAT, A-Bp•CAT, BA-Bp•CAT) such that the insert was divided into two fragments overlapping terminally in the region of the site to be changed. The upstream primer for the 5' of these two fragments was the m13 reverse primer of the vector, and within the downstream primer (indicated by the postscript '-1' in the following), the core target site sequence was changed to an *XbaI* restriction enzyme site (5'AGATCT). Near its 5' end, the upstream m13 reverse primer also included a *KpnI* site (5'GGTACC) for cloning back into the vector at the end of the construction. The sequence of this primer was 5'CGGGGTACCAACAGCTATG-ACCATG. For the 3' of the two fragments that represent the insert in the construct, the outside or downstream primer was made to represent an element of the CAT reporter sequence, i.e. 5'CTCCATT-TAGCTTCCTTAGCTCCTGAAAATCTCGCC (reverse orientation), while the upstream primer (indicated by the postscript '-2' in the following) overlapped the '-1' primer, including the *XbaI* site replacing the target sequence. The two PCR fragments were purified and digested with restriction enzymes: *KpnI* and *XbaI* for the upstream fragment; and *XbaI* and *SalI* for the downstream fragment (the *SalI* site joins the CAT reporter sequence to the basal promoter, which is located just upstream of the CAT sequence). Following this, the two PCR fragments were ligated together with the CAT reporter vector, which had been prepared by digestion with *KpnI* and *SalI*. The presence of an *XbaI* site within the insert provided an initial confirmation that the construct had been successful. The relevant regions of the constructs were then sequenced to confirm their structure.

The primers used for constructs not previously described are as follows (*XbaI* sites underlined): U1m-1, 5'CTGGTCTAGATCTAA-CATTTCTGAATTAG; U1m-2, 5'TAGATCTAGACCAGACTTTA-AACTTGTTGG; CB1m-1, 5'CTAGTCTAGACCTTCTCCATAAG3'; CB1m-2, 5'CTAGTCTAGAGATTTCCAATTTCGG3'; CB2m-1, 5'CTAGTCTAGACAAAGTTAAAG3'; CB2m-2, 5'CTAGTCTAGATTTCAAATTAATTTTGG3'; CYm-1, 5'CTAGGAATTCAGGA-GCCAGT3' (*EcoRI* is here the mutated site); CYm-2, 5'CTAGAATTCAAACGTGGCTTATGG3'; Rm-1, 5'CTGGTCTA-GACTTTATTCATGATAAC3'; Rm-2, 5'CTAGTCTAGAGTTGG-TTTTCAAATT3'.

RESULTS

Modules B and A

Nine different sequence-specific DNA binding factors interact within Modules B and A, as shown in the map in Fig. 2A (modified from Yuh et al., 1994). The sequence of the *cis*-regulatory DNA of Modules B and A is shown in Fig. 2B, including the target binding sites. These are indicated in the same colors (boxes) as in the protein binding map of Fig. 2A. The mutations used in the experiments of Yuh et al. (Yuh et al., 1998) and in this work are shown in red below each site: each of these was demonstrated by gel shift assays to eliminate DNA/protein interactions. The short vertical lines in Fig. 2B indicate the arbitrary divisions that separate the basal promoter (Bp) region from Module A, and Module A from Module B. These divisions were chosen a priori in our initial characterization of the *endo16* *cis*-regulatory system (Yuh et al., 1994; Yuh and Davidson, 1996). The Bp, as defined, is promiscuously able to respond to any upstream *cis*-regulatory elements so far tested, operating in any domain of the embryo (R. A. Cameron and E. H. D., unpublished observations); it has almost no activity on its own (Yuh and Davidson, 1996; Yuh

et al., 1996; Yuh et al., 2000) and in particular no detectable bias toward endoderm-specific activity (Yuh and Davidson, 1996; Yuh et al., 1996; Yuh et al., 2000; C.-H. Y., unpublished observations). Module B is separated from Module A by a relatively long gap in the distribution of specific target sites. As Fig. 2B indicates, upstream of the eight Module A target sites that are clustered between -69 and -217 in the sequence, no sites occur out to -380, and the only site out to -429 is an SpGCF1 site. SpGCF1 is an architectural DNA looping protein, sites for which occur ubiquitously in *S. purpuratus* *cis*-regulatory elements (Zeller et al., 1995a; Zeller et al., 1995b; orange ovals in Fig. 2A). SpGCF1 sites are unlikely to contribute significant specificity in terms of modular function, and the gap between Modules B and A might reasonably be considered to extend out to the cluster of Module B sites beginning at -430. The 'boundary' between Modules B and A might best be considered simply as this ~240 bp gap. The current definition of Modules B and A does not depend only on the clustered spatial distribution of their protein binding sites (though that indeed proved a useful initial guide) but also on a functional criterion. This is that either piece of DNA, i.e., Module B or Module A, if associated with Bp•CAT and injected into eggs, independently displays a specific and characteristic transcriptional activity (Yuh et al., 1996; Yuh et al., 1998; Yuh and Davidson, 1996; additional data shown below). Constructs driven by Module A alone are expressed early (at 30-48 hours) but hardly at all after this, while in contrast, constructs driven by Module B alone are better expressed late in development, though on an absolute scale rather weakly.

Identification of the kinetic driver of Module B and kinetic comparison with that of Module A

When Module A alone is associated with the basal promoter, its activity profile rises after about 20 hours postfertilization, attains a peak around 48 hours, and then rapidly falls (Yuh et al., 1996; Yuh et al., 1998; and below). As we showed earlier (and further substantiate in the following), this temporal characteristic is a direct reflection of the activity profile of the Otx factor that binds in Module A. Otx acts as the kinetic driver of Module A, in the sense that it provides the only time-varying input into this module; all the other interactions within it have other kinds of function (Yuh et al., 1998). Several independent pieces of evidence support the identification of the Otx factor as the driver of Module A: (1) The target site required for driver function in Module A constructs (i.e., A-Bp•CAT) includes the canonical core sequence recognized by Otx class homeodomain regulators, i.e. GATTA; and mutation of the ATTA element within A-Bp•CAT abolishes most of its transcriptional activity (except for the low residual activity of the basal promoter itself) (Yuh et al., 1998). (2) A construct consisting only of an oligonucleotide that includes as its essential sequence element the Otx target site, when linked to the basal promoter, recapitulates the temporal activity profile of Module A, though at a lower amplitude (Yuh et al., 1998). This timecourse is reproduced independently in Fig. 3B of this paper (construct [OtxZ]-Bp•CAT; blue curve). (3) Partial purification of the protein binding to this same oligonucleotide, and sequencing, demonstrates that it is identical to the SpOtx1 factor cloned by Li et al. (Li et al., 1997; X. Li., unpublished observations). (4) Introduction of mRNA encoding a negatively acting Otx-

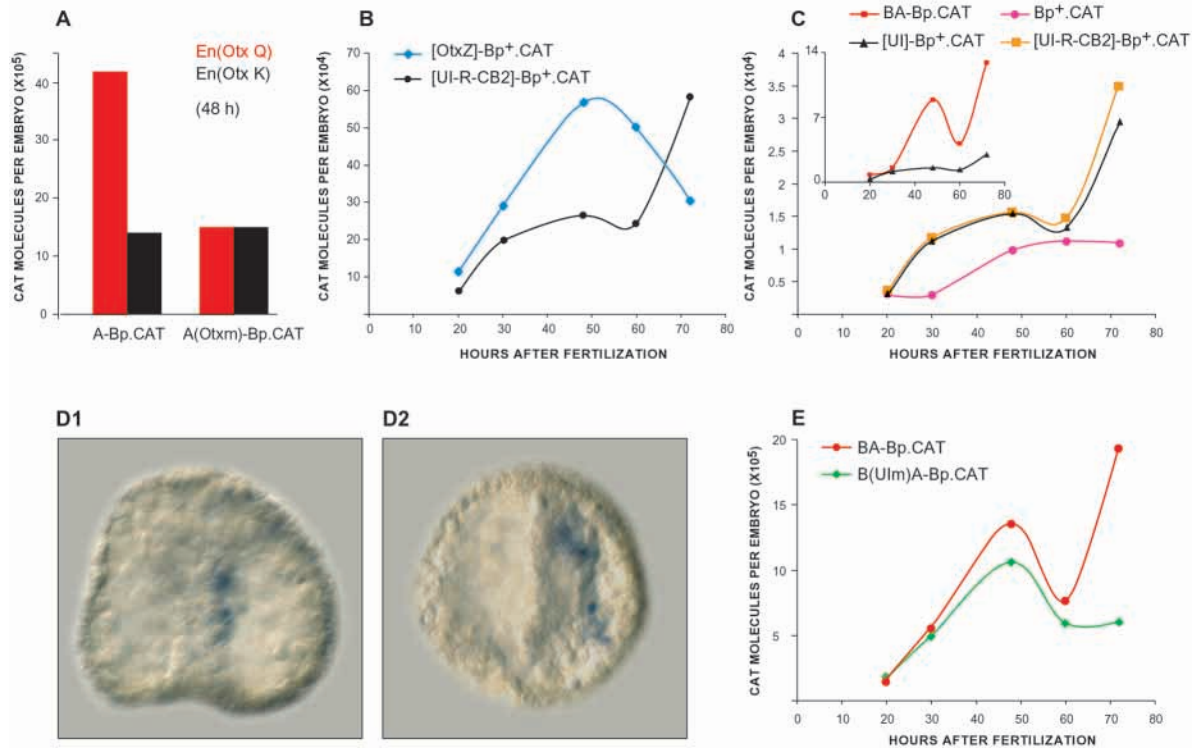


Fig. 3. Activities of Otx driver in Module A and UI driver in Module B. (A) *cis*- and *trans*-regulatory interference with Otx interactions in Module A have identical effects. An Otx-Engrailed (Otx-En) repressor fusion converts the Otx activator into a repressor of endogenous Otx target genes (Li et al., 1999). mRNA (0.025 pg/pl) encoding the chimeric protein (the construct was a kind gift from Prof. Wm. Klein and X.-t. Li) was injected into eggs also bearing the indicated *endo16* expression constructs (OtxK, black). As a control, mRNA encoding a variant which fails to bind DNA because a key lysine of the Otx homeodomain has been changed to a glutamine, was introduced into other eggs of the same batch (OtxQ, red). (Left) The amount of activity generated by control OtxQ embryos or embryos expressing OtxK in the A-Bp•CAT at 48 hours. (Right) Level of expression in eggs bearing either OtxQ or OtxK and the A(Otxm)-Bp•CAT construct (see Yuh et al., 1998 for details). (B) Timecourse of expression of regulatory constructs driven by oligonucleotides bearing Otx and UI target sites. The synthetic oligonucleotides are associated with the *endo16* basal promoter (Bp, i.e. -67 to +190 of the *endo16* gene), together with its adjacent CG3 and CG4 sites. These enhance Bp activity about twofold (Yuh et al., 1998). These CG sites have no endoderm-specific activity, neither does the Bp itself (Yuh et al., 1998; Yuh et al., 2001; Yuh and Davidson, 1996). In this and following figures, the enhanced Bp is denoted Bp⁺; where (in other experiments) Module A is present, the CG3 and CG4 sequences are included de facto. [OtxZ] is a 28 bp synthetic oligonucleotide including the *endo16* sequence from -157 to -175, which confers endoderm-specific expression. The timecourse of expression generated by [OtxZ]-Bp⁺•CAT (see also Yuh et al., 1998) is here compared with that of [UI-R-CB2]-Bp⁺•CAT, the sequence of which is given in Table 1. (C) The UI site alone produces the late rise in expression. Output kinetics are shown from an experiment in which only the UI site has been left intact (construct [UI]-Bp⁺•CAT of Table 1) so that it alone provides regulatory input into the enhanced basal promoter (black). This construct produces exactly the same output as does [UI-R-CB2]-Bp⁺•CAT (orange), again generating the late rise in expression. However, the output is of low magnitude relative to that of BA-Bp•CAT: the inset shows these data (i.e. for [UI]-Bp⁺•CAT) at reduced scale, to also indicate BA-Bp•CAT output in the same experiment. (D1,D2) Whole-mount in situ hybridization. These embryos illustrate expression in the archenteron, driven by the [UI-R-CB2]-Bp⁺•CAT construct (see Table 1 for quantitative data). (E) Requirement of UI site for late rise in expression of BA-Bp•CAT. BA-Bp•CAT kinetics (red) display the characteristic post-gastrular increase in CAT product; in the same experiment, embryos bearing a mutation of the UI site (Fig. 2) in an otherwise normal BA-Bp•CAT (construct B(UIm)A-Bp•CAT; green) fail to produce the late rise in expression. The data shown are the average of three replicate experiments, each carried out on one batch of eggs. The experiments were normalized to the 48 hour point on the BA-Bp•CAT control curve as earlier (Yuh et al., 1996). Standard deviations for the key 60 and 72 hour points were about 10% for the BA-Bp•CAT data and 15% for the B(UIm)A-Bp•CAT data, and were 30-40% elsewhere.

Engrailed fusion (Otx-En), which converts the Otx activation to a repressor for Otx target genes, severely depresses the endogenous *endo16* transcript levels (Li et al., 1999).

Additional evidence is shown in Fig. 3A. Here, Module A expression constructs have been introduced into eggs also expressing either the same Otx-En fusion (black bars) or a variant used as control, which cannot bind DNA (red bars; Li et al., 1999). The efficacy of the Otx-En fusion is shown on the left of Fig. 3A, where it can be seen to cut the level of expression of A-Bp•CAT down to about one third of that of the

control. On the right we see that this *trans* interference produces quantitatively just the same effect as does *cis* interference by mutation of the single Otx target site within A-Bp•CAT (construct A(Otxm)-Bp•CAT). The residual activity is due to the basal promoter and the adjacent CG3, CG4 and SpGCF1 sites, i.e. Bp⁺ (see legend to Fig. 3; Bp⁺•CAT expression can be seen in Fig. 3C; for additional experimental analyses of these CG elements see Yuh et al., 1998). Otx interactions should be and are irrelevant to this residuum (Fig. 3A, right pair of bars).

The major site of driver activity in Module B is that denoted 'UI' in Fig. 2, though as we show below, there is another less important time-varying driver input into Module B as well. Evidence for the key role of the UI target site is given in Fig. 3B-E and in Table 1. In Fig. 3B, the blue curve represents the timecourse of expression of synthetic constructs driven by an oligonucleotide containing the Otx site, here shown for comparison ([OtxZ]-Bp⁺•CAT; see legend and similar experiments by Yuh et al., 1998). A completely different timecourse is generated by a construct in which the regulatory element is an oligonucleotide that includes the UI target site, plus the two adjacent sites in the sequence of Module B (R and CB2 sites; construct [UI-R-CB2]-Bp⁺•CAT). The kinetics of expression of this construct are shown by the black curve in Fig. 3B. As shown in Fig. 3C (inset), and again in an independent set of experiments in Fig. 3E (red curves), the expression kinetics of BA-Bp⁺•CAT are characterized by a sharp increase in activity occurring after 60 hours, which is not generated by A-Bp⁺•CAT. The experiment illustrated in Fig. 3B demonstrates that a target site within the [UI-R-CB2] oligonucleotide suffices to confer this same kinetic character, i.e. the late rise in expression. It is the UI site which is responsible. This is shown explicitly in Fig. 3C, in which the R site and the CB2 site of the [UI-R-CB2] oligonucleotide have both been destroyed, leaving intact only the UI site (i.e. construct [UI]-Bp⁺•CAT). We see in Fig. 3C that this site alone suffices to provide the late rise in expression, though relatively weakly (Fig. 3C inset; note scale change). When tested together, [UI]-Bp⁺•CAT in fact functions indistinguishably from [UI-R-CB2]-Bp⁺•CAT (Fig. 3C). In addition, these oligonucleotides suffice to produce expression in endodermal cells, as illustrated by whole-mount in situ hybridization in Fig. 3D. Finally, mutation of the UI target site (construct B(UIm)A-Bp⁺•CAT) obliterates the late rise in expression. This is shown in Fig. 3E. When the same mutation is inserted

in a construct driven by Module B alone (construct B(UIm)-Bp⁺•CAT), it virtually eliminates endodermal activity (Table 1).

These experiments resolve the distinct character of the temporal outputs of Modules A and B to functions mediated by the target sites for the respective drivers of the Modules: the Otx site for Module A; the UI site for Module B. The protein binding at the UI site has not yet been cloned, but has been partially purified by affinity chromatography (C.-H. Y., unpublished observations). It is probably multimeric and in partially purified form it also binds with high affinity to the similar target sites that have been shown to confer gut-specific expression on the *cylla* cytoskeletal actin gene (Arnone and Davidson, 1997).

Synergism: the donor function of Module B

As reviewed above, Module A significantly amplifies the output of Module B (Fig. 1). Mutations of the P and CG1 sites of Module A in BA-Bp⁺•CAT constructs functionally unlink Module B from Module A. Thus, although Module B remains physically associated with Module A, the output of either mutation (i.e. BA(Pm)-Bp⁺•CAT or BA(CG1m)-Bp⁺•CAT) is identical to that of A-Bp⁺•CAT, and the late rise in activity characteristic of Module B is entirely lost (see Fig. 2 by Yuh et al., 1998). This and other results lead to the conclusion that the P and CG1 interactions are absolutely required for the transduction of Module B output into Module A, and for its synergistic amplification there. The implication is that Module B must provide a donor function into the synergistic B-A linkage. Conceivably, this input could be that of UI itself: however, the following experiment shows that the UI driver function is separable from a specific donor linkage function that is executed via the CB2 target site.

The key observation is illustrated in Fig. 4. With respect to both form and magnitude, the characteristic late-rising output

Table 1. Requirement for and sufficiency of UI target site for endoderm expression mediated by Module B (48 hours)

Constructs	Endoderm expression*	Sequence of relevant region						
		UI		R	CB2			
Endo16•CAT	95%	TGTTA ACAAT	TCATGAATAA AGTACTTATT	AG TC	ACTTTAACTT TGAAATTGAA	TG AC	TTGGTT AACCAA	TTTCAAATTAATT AAAGTTTAATTAA
B-Bp ⁺ •CAT	85%			''			''	
B(UIm)-Bp ⁺ •CAT	3%	TGTTA ACAAT	gatctAgacc ctagaTctgg	AG TC	ACTTTAACTT TGAAATTGAA	TG AC	TTGGTT AACCAA	TTTCAAATTAATT AAAGTTTAATTAA
[UI-R-CB2]-Bp ⁺ •CAT	65%	t cga	TCATGAATAA AGTACTTATT	AG TC	ACTTTAACTT TGAAATTGAA	TG AC	TTGGTT AACCAA	TTTCAAATa AAAGTTTAttcga
[UIm-R-CB2]-Bp ⁺ •CAT	16%	t cga	TCAcGt c g Ac AGTgCagcTg	AG TC	ACTTTAACTT TGAAATTGAA	TG AC	TTGGTT AACCAA	TTTCAAATa AAAGTTTAttcga
[UI-Rm-CB2]-Bp ⁺ •CAT	58%	t cga	TCATGAATAA AGTACTTATT	gt ca	gtccc g AaTT caggggTtAA	cG gC	TTGGTT AACCAA	TTTCAAATa AAAGTTTAttcga
[UI]-Bp ⁺ •CAT	Nil	t cga	TCATGAATAA AGTACTTATT	AG TC	ctagTctaga gatcAgatct	TG AC	tctaga agatct	TTTCAAATa AAAGTTTAttcga

*Observation by whole-mount in situ hybridization; see Yuh and Davidson (1996). Over 100 embryos were assayed for each entry. Percent values indicate the fraction of embryos displaying clones ≥ 2 cells expressing the CAT reporter in the archenteron, as normalized by whole-mount in situ hybridization.

Oligonucleotides are indicated in brackets in 'Constructs' column. Lower case letters denote bases different from wild-type sequence (Endo16•CAT, first line).

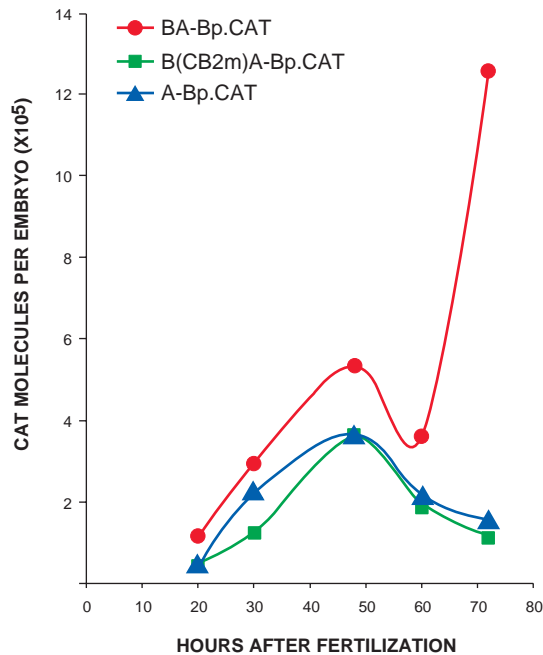


Fig. 4. Mutation of CB2 site in Module B severs the functional B-A linkage. Output kinetics for BA-Bp•CAT (red) display the typical late rise in CAT product observed repeatedly in these studies (Yuh et al., 1996; Yuh et al., 1998). Output of A-Bp•CAT (blue), in contrast, displays the early rise and then decline also reported previously (Yuh et al., 1996; Yuh et al., 1998; see also Fig. 3B, where it is shown that these Module A kinetics directly reflect the activity profile of the Otx driver of Module A). The sequence inserted in place of the CB2 site of Module B in B(CB2m)A-Bp•CAT is shown in Fig. 2; B(CB2m)A-Bp•CAT is the same as BA-Bp•CAT except for this mutation. The output (green) is, within observational variation, the same as that of A-Bp•CAT. This figure is a composite of five separate experiments, each carried out on a single batch of eggs (100 eggs per point) which were normalized to the peak value of the BA curve at 48 hours (Yuh et al., 1996). The 20 hour points, taken shortly after activity begins are near the limit of sensitivity and standard deviations are ~70% of the low values for this data point (this uncertainty has no effect on the sense of the experiment). Thereafter the standard deviations average ~25% for the BA-Bp•CAT curve and ~45% for the other two curves.

kinetics for the BA-Bp•CAT construct (red), and the early peaking kinetics of A-Bp•CAT (blue) shown in Fig. 4 reproduce those reported earlier (e.g. Fig. 2 by Yuh et al., 1996; Figs 2, 3 and 5 by Yuh et al., 1998). But when the CB2 site is mutated in BA-Bp•CAT (construct B(CB2m)A-Bp•CAT) we see that the same output (green) is generated as by A-Bp•CAT, within the attainable reproducibility. The strong late rise in expression, the signature of Module B, is completely absent. In other words, even though the UI site, and all the other sites of Module B remain physically linked in their normal positions with respect to Module A, the construct behaves as if none of the Module B sites were present at all: Module B input into the BA system requires the CB2 target site. Therefore a function of this site is to provide the donor linkage by which the regulatory output of Module B is passed to Module A. An interaction between proteins binding to CB2 in Module B and to the P and CG1 sites of Module A (direct or indirect) is implied, since all three respective mutations in the BA-Bp•CAT

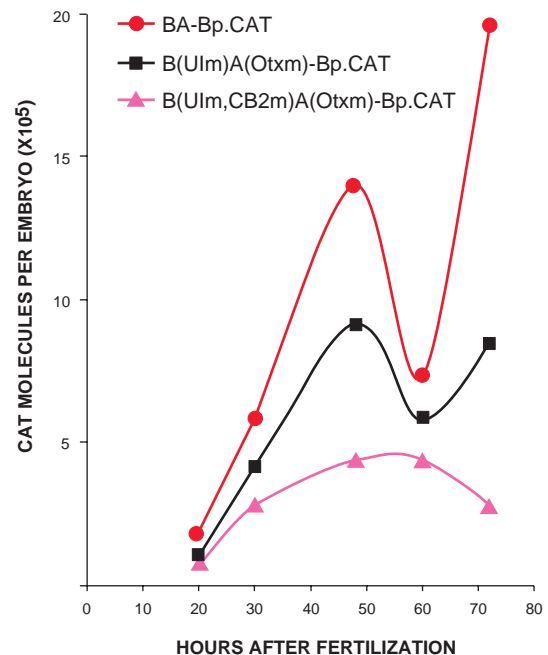


Fig. 5. Demonstration of low level CB2 driver function. Output kinetics are shown for embryos bearing BA-Bp•CAT (red), the double mutation B(UIm)A(Otxm)-Bp•CAT (black), and the triple mutation B(UIm,CB2m)A(Otxm)-Bp•CAT (magenta). A normalized average of two experiments each carried out on a single batch of eggs is shown (see legend to Fig. 4). For this data set standard deviations were for the BA curve about 35% of the values shown; about 17% for the double mutation; and about 50% for the triple mutation (owing to its low level of activity, which is difficult to measure accurately).

construct produce the identical phenotype, i.e. the output kinetics of A-Bp•CAT. No other mutation in Module A (Yuh et al., 1998) or of any other sites in Module B produces this result (below, and C.-H. Y., unpublished observations). Nor does the CB2 site affect the output in the absence of Module A (Fig. 3C, black and orange curves).

A further role of the CB2 interaction: additional Module B driver function

Were the Otx site of Module A and the UI site of Module B the only loci in these Modules where interactions that function as kinetic drivers occur within BA-Bp•CAT, then a double mutation of these sites would yield only the very low profile of activity generated by the enhanced basal promoter (Bp⁺) alone. However, the double mutation B(UIm)A(Otxm)-Bp•CAT (black curve) displays a 'residual' kinetic activity profile that is significantly higher than the basal level, though it lacks almost all of the late rise that the UI input confers on BA-Bp•CAT (red curve). The most prominent aspect of the timecourse displayed by the double mutation in Fig. 5 is the 'hump' at 40-50 hours, which parallels that of BA-Bp•CAT or the Otx input into A-Bp•CAT (Fig. 3B; Yuh et al., 1996; Yuh et al., 1998). But this residual activity is generated within Module B, since A(Otxm)-Bp•CAT has very much lower activity than B(UIm)A(Otxm)-Bp•CAT (Yuh et al., 1998; and data not shown). However, it also requires Module A, since the activity of B(UIm)-Bp•CAT is only barely measurable (below, and Yuh et al., 1998). The source of the residual driver activity

is revealed by a triple mutation, the output of which is illustrated in the magenta curve of Fig. 5. Here we see that in the construct B(UIm,CB2m)A(Otxm)-Bp•CAT, the residual driver function is lost, leaving only the background Bp⁺ level of expression (also seen in Fig. 3C; see legend). Therefore the source of the residual driver function, defined as the difference between the magenta and black curves of Fig. 5, is an interaction at the CB2 target site. Mutation of the CY, CB1, or R sites in B(UIm)A(Otxm)-Bp•CAT has no such effect (data not shown); these are the only additional relevant target sites in Module B. The CB2 interaction thus has a dual function: it serves to link the output of UI into the synergism apparatus of Module A (Fig. 4), and it also contributes a small driver function of its own (Fig. 5). The timecourse of this function is quite different from that of the UI output of Module B, as it lacks significant late activity. The CB2 driver output must feed into the synergism apparatus of Module A as well, since the scale of the residual activity of Fig. 5 is seen only in BA fusions. The residual driver function mediated by the CB2 interaction is entirely independent of the Otx input apparatus of Module A, since it occurs in the absence of the Otx target site (Fig. 5). Furthermore, CB2-mediated driver function is entirely insensitive to the presence of the Otx-En repressor (data not shown).

The role of the CB1 and CY sites

Of the seven specific target sites identified in Fig. 2 within the Module B region, we have thus far established specific

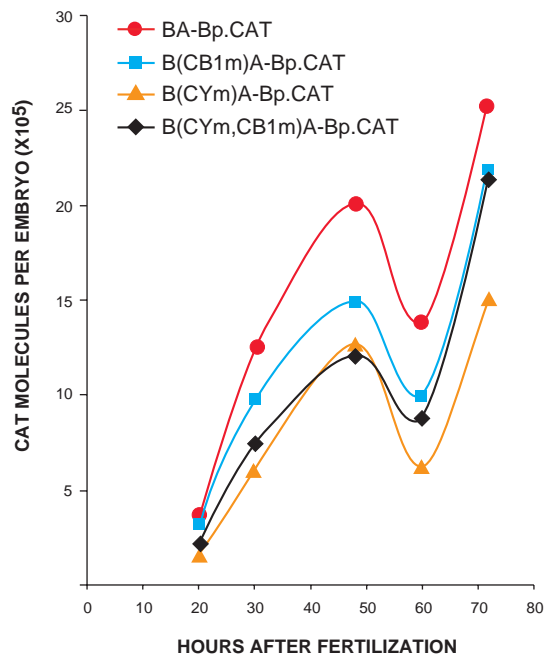


Fig. 6. Role of CB1 and CY interactions. The data shown are the normalized average of two experiments (see legend to Figs 3-5). The activity of the BA-Bp•CAT control (red) is compared with that of the same construct in which the CB1 site is mutated (turquoise); the CY site is mutated (orange); or both are mutated (black). Standard deviations at all points for each of the four constructs ranged from 24-33%. Thus, it is impossible to distinguish significantly the kinetics of any one of these mutations from that of the others, but all are clearly less active than is BA-Bp•CAT.

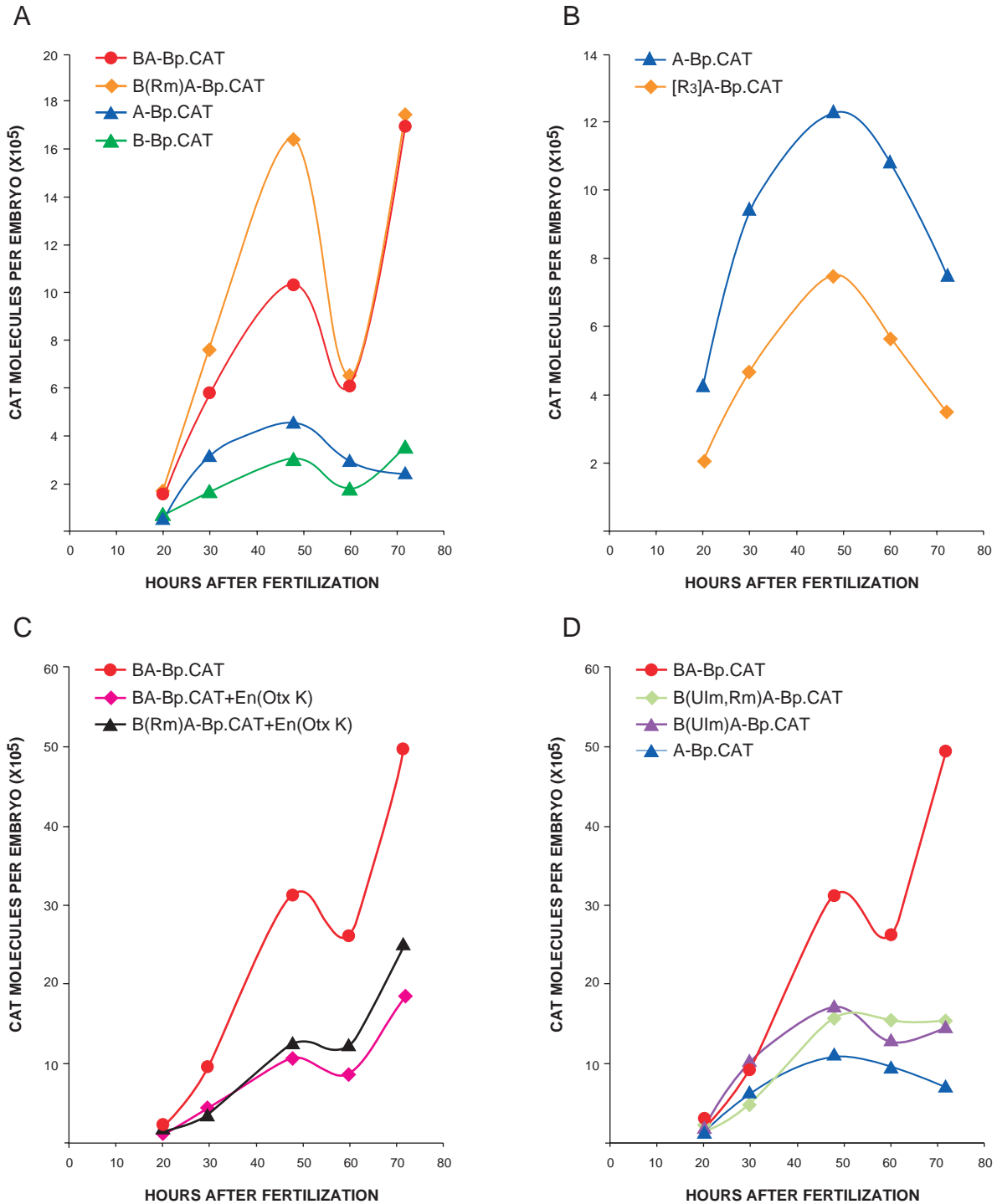
functions for the UI site and the CB2 site. The two SpGCF1 sites indicated in Fig. 2 are not considered in this paper, as SpGCF1 interactions, which occur commonly in *S. purpuratus* *cis*-regulatory elements, have been studied earlier. SpGCF1 is an architectural protein, that multimerizes after binding DNA, and thereby has the capacity to promote contact between noncontiguous *cis*-regulatory sites by stimulating DNA looping (Zeller et al., 1995b). There remain the CB1, CY and R sites of Module B. The experiments summarized in Fig. 6 concern the function of the CB1 and CY sites, and those that we take up in the following section, the R site. As in Module A (Yuh et al., 1998), a functional significance has been found for each one of the specific interactions that were originally detected in Module B by Yuh et al. (Yuh et al., 1994).

The CY and CB1 interactions contribute to the strength of the UI input into the BA system, and both are required for maximum UI effect. These target sites lie upstream of the UI site towards the 5' boundary of the BA system (Fig. 2). The effect of mutating one or both of these sites can be seen in Fig. 6. In this comparison, it is impossible to determine whether the effect of the CY mutation alone (orange curve) is greater than that of the CB1 mutation (turquoise curve) or of both together (black curve); nor could this detail be resolved in additional repeats of the experiment that we carried out. The outputs of all three of these mutated forms of BA-Bp•CAT are too similar, given the level of experimental variation (see

Fig. 7. The BA intermodule input switch function. (A) Effect of mutation of the R site in BA-Bp•CAT. The timecourse of expression of this construct (B(Rm)A-Bp•CAT, orange curve) is compared with that of the BA-Bp•CAT control (red), and to that of B-Bp•CAT (green) and A-Bp•CAT (blue). The normalized average of seven independent experiments is shown, each carried out on a single batch of eggs. The standard deviation at the crucial 48 hour time point of the B(Rm)A-Bp•CAT timecourse was about 30% for this data set, which was normalized as in earlier figures to the 48 hour time points of the BA-Bp•CAT curves (standard deviations for the remaining points were of about the same magnitude). (B) Direct demonstration of repression function mediated by the R target site. The timecourse generated by a construct consisting of three copies of an oligonucleotide that represents the R target site (see Fig. 2) ligated to A-Bp•CAT ([R₃]A-Bp•CAT, orange curve) is compared with the timecourse of A-Bp•CAT, blue. An average of two experiments, normalized at the 48 hour time point of the A-Bp•CAT curves is shown. From 30-72 hours, the standard deviations were 19% for A-Bp•CAT and 17% for [R₃]A-Bp•CAT. (C) Effect of R mutation is negated in presence of En-Otx fusion. An experiment is shown in which B(Rm)A-Bp•CAT (black) and BA-Bp•CAT (magenta) produce almost identical outputs in the presence of the En-Otx fusion (see legend to Fig. 3). The output of BA-Bp•CAT is shown in red, as a control carried out on the same batches of eggs. For this data set, from 48 to 72 hours, the standard deviations for all three curves ranged from 7-30% after the earliest stages. (D) Additional mutation of UI in a BA construct carrying the R mutation also negates the effect of the R mutation. An average of two experiments is shown in which the output of B(UIm,Rm)A-Bp•CAT (green) is seen to be very similar to that of B(UIm)A-Bp•CAT (purple). That is, though the R sites is intact in B(UIm)A-Bp•CAT, it fails to repress the activity of Module A in this context. As controls, BA-Bp•CAT (red) and A-Bp•CAT (blue) are shown in normal embryos of the same batch of eggs. For this set of data, the standard deviation for BA-Bp•CAT and B(UIm, Rm)A-Bp•CAT ranged from 3 to 30% at the later stages (48 to 72 hours). The B(UIm)A-Bp•CAT and A-Bp•CAT curves are from a single experiment.

legend to Fig. 6). In any case, what is clear is that CB1 and CY are both required for the level of Module B activity that is dependent on the UI driver observed in the intact BA-Bp•CAT construct. Possibly the proteins that bind to these two sites form a complex with the UI factor, and if either the CB1 or CY target site is destroyed (or, of course, if both are destroyed) this complex fails to form. The ancillary effect of the CB1 and CY interactions on UI output is on the average about twofold (i.e. subtracting the effect of CB2, which is not affected by this mutation). Fig. 6 shows that this moderate effect pertains

all across developmental time. Furthermore, as would be expected, the functions of CB1 and CY are independent of Module A: results similar in relative terms to those in Fig. 6 were obtained in a comparison of the activities of B-Bp⁺•CAT, B(CYm)-Bp⁺•CAT, B(CB1m)-Bp⁺•CAT and the double mutation B(CYm,CB1m)-Bp⁺•CAT, although of course here the outputs were of much lower amplitude in the absence of the synergistic amplification provided by Module A (data not shown): the levels of expression of all three mutations were significantly lower than that of the B-Bp⁺•CAT control across



the whole timecourse (data not shown). However, it was again impossible to distinguish the effect of the double mutation from that of either single mutation.

The BA intermodule input switch

One of the most interesting aspects of *endo16* cis-regulatory circuitry is the BA intermodule input switch (Fig. 1). The function of this switch is to cut off the input of the Otx driver of Module A when the output of UI (and its CY and CB1 facilitators) exceeds some threshold. This threshold is apparently crossed by late blastula stage, since for most of the timecourse, the level of the output of BA-BpCAT is essentially just that of B-Bp⁺CAT as linearly amplified by Module A (reviewed in Introduction). For example, except at very early times, the output of BA(Otxm)-BpCAT is the same as that of BA-BpCAT (Yuh et al., 1998). The BA intermodule input switch is very important. Its consequence is the transfer of control to UI, a dedicated gut-specific regulator (Fig. 1). As discussed earlier, it thereby relieves the system from dependence on the early spatial repression system that it relies on to confine activity to the vegetal plate while it is running off the ubiquitously active Otx driver. After the switch is thrown, the late rise in *endo16* expression can be produced simply by amplifying the late rise in driver activity of the Module B driver, UI (Fig. 3B).

The existence of the BA intermodule input switch was required by the results of the functional analysis of Module A (Yuh et al., 1998). But since its mechanism is rooted in Module B, further understanding of its workings awaited the present experiments. The predicted result of mutation of the target site(s) of Module B that mediate the operation of the switch would be an enhanced output of the BA-BpCAT construct, since now both the (unrepressed) output of Module A plus the amplified output of Module B would feed into the system. But such an enhancement should be observed only in the 30-50 hour period of the timecourse, since this period, when the Module A driver is active (Figs 3B, 4; Yuh et al., 1996; Yuh et al., 1998), should be the only part of the timecourse that the switch would affect when the R site of Module B is mutated. Exactly the predicted result occurs. These data are shown in Fig. 7A, which contains an average over seven different experiments all producing the same relative results. Here, the output of the mutation of the R site of Module B (construct B(Rm)A-BpCAT, orange curve) is compared in the same experiments with that of B-Bp⁺CAT (green), A-BpCAT (blue) and BA-BpCAT (red). The mutation, a 10 bp change within the R site, is shown in Fig. 2. The activity of B(Rm)A-BpCAT significantly exceeds that of the BA-BpCAT construct at the 30 and 48 hour points, but from 60 hours onwards, the control and the R mutation behave synonymously.

Quantitatively, the enhanced output of B(Rm)A-BpCAT is the sum of the normal amplified output of Module B in BA-BpCAT, plus that of A-BpCAT, plus a slightly increased input from the CB2 driver of Module B. The CB2 contribution increases when the R site is mutated, indicating that proteins binding to the adjacent R and CB2 sites probably interact, directly or indirectly, so that under normal conditions CB2 input is somewhat damped by the presence of the component(s) that bind at the R site. This detail aside, the experiments of Fig. 7A clearly implicate the R target site as the key element

in the BA intermodule input switch. No other mutation in Module B has any similar effect (Figs 3-5, and additional data not shown). Furthermore, as required, the derepression caused by the R mutation requires an intact Otx site in Module A. When this site is also mutated (construct B(Rm)A(Otxm)-BpCAT), the output is exactly the same as for BA(Otxm)-BpCAT, i.e. the affect of the R mutation is negated (data not shown). Similarly, there is no enhancement if B(Rm)A-BpCAT is introduced into embryos bearing the Otx-En fusion of Fig. 3A, as illustrated in Fig. 7C. These experiments prove that the enhanced expression of B(Rm)A-BpCAT versus BA-BpCAT is due largely to the normally repressed input from the Otx site in Module A.

To confirm independently that R has the capacity to bind a repressor of Otx input, synthetic constructs containing one or three copies of an oligonucleotide representing the R target site were linked to A-BpCAT. Fig. 7B shows the effect of the three-copy construct, [R3]A-BpCAT: the timecourse generated by this construct (orange curve) is similar to that of A-BpCAT (blue), but of much lower amplitude. Therefore, in vivo, the R oligonucleotide indeed mobilizes a repressor of Module A output, i.e. of its Otx driver. The effect of the one-copy construct was of the same nature, though weaker (data not shown).

Two other of the Module B interactions are required for the BA intermodule output switch to work. First, if there is no input into the system from UI, no repression of Module A Otx output occurs. Fig. 7D shows that the additional mutation of UI in the B(Rm)A-BpCAT (construct B(UIm, Rm)A-BpCAT) also negates the effect of the R mutation, so that the output is just that of B(UIm)A-BpCAT. That is, Module A input is not now repressed (compare Fig. 7A with Fig. 7D). Therefore, just as logic demands, activation of the R switch depends on UI activity. Second, the CB2 site is needed to link the intermodule switch function into Module A, just as it is needed to link UI function into Module A (Fig. 4). Thus the derepression of Module A observed in B(Rm)A-BpCAT experiments (Fig. 7A) is entirely lost in the double mutation B(Rm, CB2m)A-BpCAT. The output of this double mutation is just that of the single mutation B(CB2m)A-BpCAT illustrated in Fig. 4 (data not shown). To summarize, the BA intermodule input switch is mediated by a repressor binding directly or indirectly at the R target site (Fig. 7A,B); but activation of the switch function in BA-BpCAT depends on input from the UI site, and on the presence of the CB2 site as well.

DISCUSSION

Modules A and B of the *endo16* cis-regulatory system were first recognized simply as a proximal and an adjoining more-distal cluster of target sites for high specificity DNA/protein interactions (Yuh et al., 1994; Yuh and Davidson, 1996). Later, it became apparent that Module A is responsible for initiation of spatial expression in the early embryo, and Module B for later activity, but that, in addition, Module A is required in some ways for operation of the remainder of the whole system (Yuh et al., 1996; Yuh et al., 1998). As we described it earlier, Module A acts as the 'central processing unit' for the whole *endo16* regulatory apparatus. The study of Yuh et al. (Yuh et

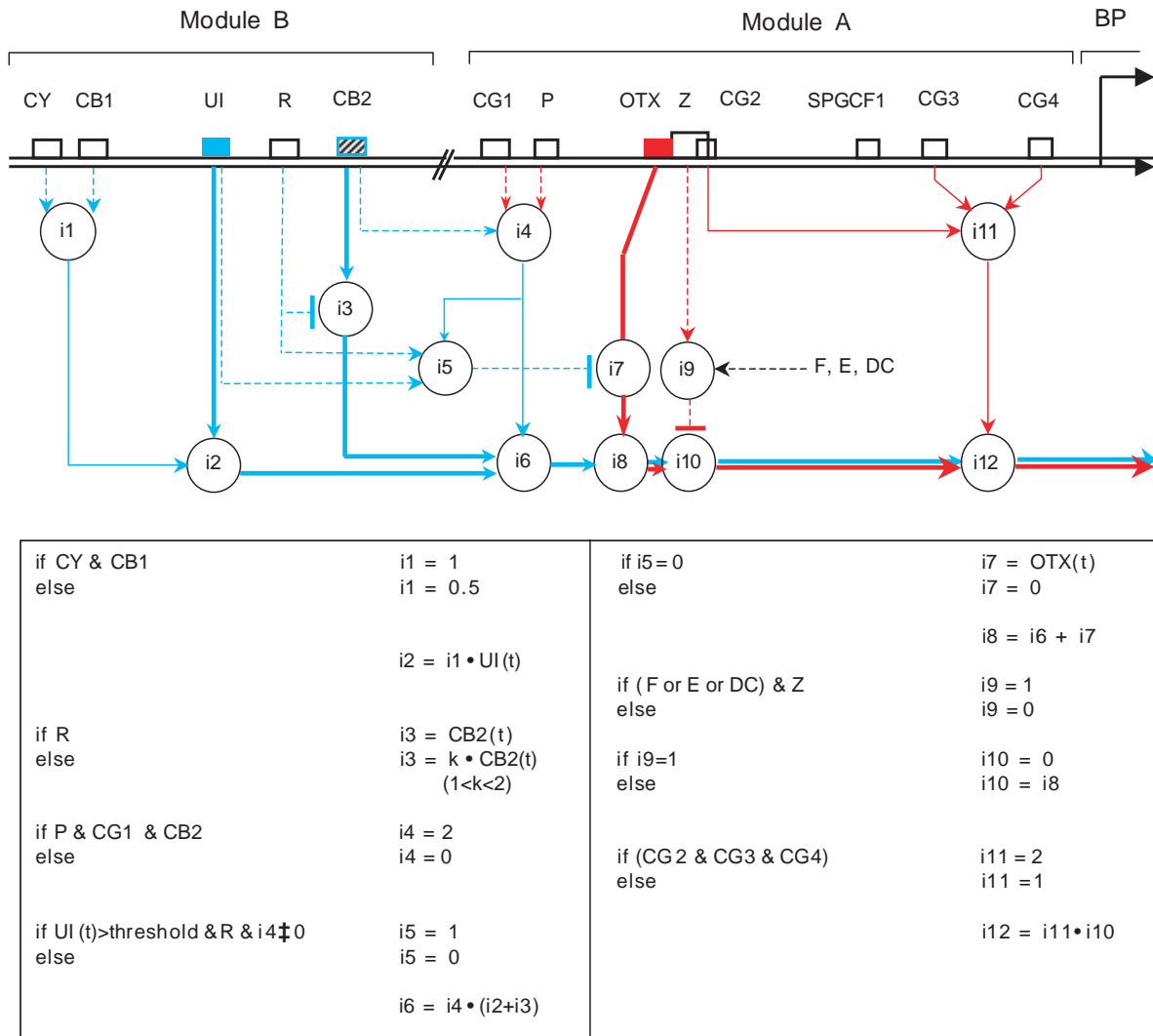


Fig. 8. Computational logic model for BA region of the *endo16* cis-regulatory system. The regulatory DNA of *endo16* is shown as a horizontal strip at the top of the diagram. The individual binding sites are indicated by labeled boxes. Module B and its effects are shown in blue; Module A and its effects are shown in red. Intermediate logic functions (i) are indicated by numbered circles. Each represents a specific regulatory interaction modeled as a logic operation. Driver functions (see text) are mediated by time-varying inputs, indicated by colored boxes, which determine the temporal and also spatial pattern of *endo16* expression. Time invariant interactions which affect the level of expression and control intra-system output and input traffic are shown as open boxes. Boolean functions in the model are indicated by broken lines; scalar functions by thin lines; functions of which the quantitative value varies with time, as thick lines. The individual logic functions are defined in the set of statements below the diagram. Here, statements of the form ‘If X’, where X is the name of a target site, mean that this site is present and occupied by the respective factor. If the site has been mutated (or were the factors inactivated or eliminated) this is denoted by a statement of the form ‘X=0’; or as the alternative (‘else’) to the site being present and occupied. The statements afford testable predictions of the output for any given mutation, alteration or subelement of the system (see examples in Table 2). The model was built using MATLAB (MathWorks) for programming, analysis and quantitative tests. Briefly, the BA system shown in the model works as follows: CB1 and CY interactions together synergistically increase by a factor of about two the output of the positive spatial and temporal driver of Module B, which binds at the UI site ($i1$, $i2$; Figs 3, 6). The output of the UI subsystem is at ($i2$). An additional, smaller, time varying driver input, which peaks at about 40 hours, is generated by the interaction at CB2 ($i3$; see Fig. 5). An interaction at R is required for the BA intermodule input switch, which shuts off Otx input ($i5$, $i7$; Fig. 7), but this switch operates only if there is input from UI, and the CB2 site is present and occupied ($i5$). Furthermore, the proteins binding at the adjacent R and CB2 sites apparently interact, in that if the R site is mutated CB2 input (at $i3$) is somewhat enhanced. The CG1 and P sites of Module A together with CB2 in Module B are all required for linkage of Module B to Module A ($i4$; Fig. 4 and Yuh et al., 1998), and for synergistic amplification (by a factor of about two) of the Module B input at $i6$. If the switch mediated by R does not function (i.e., in an R mutation) the summed input of Modules A and B at $i8$ is observed (Fig. 7A). If CB2, CG1 or P sites are mutated, Module B is unlinked from Module A (Fig. 4; Yuh et al., 1998). That is, now $i4=0$, so $i6=0$, and in this case $i8$ is just the output of the Otx interaction at $i7$. If the gene is in a location where the repressive interactions in any of the upstream F, E, or DC modules are operating, the system is shut off via an interaction at the Z site of Module A ($i9$, $i10$; Yuh et al., 1996; Yuh et al., 1998; Yuh and Davidson, 1996). Finally, the CG2, CG3 and CG4 sites combine (at $i11$) to boost whatever output is present (at $i12$) by an additional factor of two (Yuh et al., 1996; Yuh et al., 1998). The result at $i12$ is transmitted to the basal transcription apparatus. The output at each of the nodes (i) is calculated by the operations indicated in the Table below the Figure.

al., 1998) showed how Module A works, in terms of the functions mediated by each of the DNA target sites within it, and their interactions. This was formalized in a predictive logic model, in which the internal interrelations within Module A are explicit, while the inputs into Module A from Module B and other upstream regulatory elements (i.e. Modules F, E and D; Yuh et al., 1996) remained implicit. As evidence for essential B-A interactions emerged, the stage was set for the exploration of Module B at the DNA sequence level that is described in this paper. The new evidence has much deepened our understanding both of how different are the functions of Modules B and A (as summarized in Fig. 1 and Introduction), and also of how intimately the two regulatory units interact.

In a word, Module A starts off activity during vegetal plate specification, relying for its lineage-specific expression (Ransick et al., 1993) on a ubiquitous Otx driver, and on regional spatial repressors; but Module B is the primary operator once gut differentiation begins, relying on its own gut-specific driver. This is a partially purified but yet undefined sequence-specific regulatory protein, here referred to as UI. There are two key interactions between Modules B and A. These are the synergistic processing of Module B input within Module A, for which target sites in both modules is required; and what we have termed the 'BA intermodule input switch', for which the necessary target sites are located in Module B. The role of this switch is to transfer control of expression to Module B as UI driver activity rises. The threshold at which the switch is thrown is apparently crossed by the late blastula stage.

Extension of the computational logic model to encompass B-A interactions

The extended regulatory module for Modules B and A is shown in Fig. 8. The model is anchored in the DNA target sites of the B-A region, the core sequences of which are shown in Fig. 2. The specific functions that proteins binding to these sites mediate, as illustrated by the experiments described here and by Yuh et al. (Yuh et al., 1998) are summarized in the legend. The model in Fig. 8 is structured in terms of functional inputs into the system from each target site, which proceed to logic processing nodes, indicated by circles. Within each node the intermediate values of these inputs (denoted 'i') are processed according to the logic functions listed below the diagram. Those driver inputs that have time-varying activities, i.e. they change during development, are indicated as solid or hatched boxes at the respective target sites. These are the UI (blue box) and CB2 (blue hatched box) sites of Module B, and the Otx site (red box) in Module A. Thick lines connect those sites to the logic processing nodes, and the output sampled at any location along these lines will have a developmental or time-dependent magnitude. Other functions are indicated as scalar amplifications or enhancements of activity (thin lines) and still others as on or off (Boolean) functions (broken lines).

There are four Boolean subsystems included in this model. They are (1) the 'off' switch in Module A mediated by the Z-site interaction (i9), which turns the gene off in response to upstream repressor binding to sites in Modules F, E, or DC (Yuh et al., 1998); (2) the linkage system that connects Module B input into Module A, mediated by CB2, CG1 and P sites (i5; Yuh et al., 1998; Fig. 4 of this paper); (3) the synergism subsystem mediated by the CY and CB1 sites,

Table 2. Examples of quantitative predictions of model in Fig. 8: expected outputs produced by given module B mutations

B	= [CB2(t) + UI(t)]
B(CYm)	= 0.5 × UI(t) + CB2(t) = 0.5B + 0.5 × CB2(t)
B(CB1m)	= 0.5 × UI(t) + CB2(t) = 0.5B + 0.5 × CB2(t)
B(CYm,CB1m)	= 0.5 × UI(t) + CB2(t) = 0.5B + 0.5 × CB2(t)
B(Rm,CB2m)	= UI(t) = B - CB2(t)
B(UIm)	= CB2(t)
For UI(t) > threshold	
BA	= 4 × [CB2(t) + UI(t)]
BA(CYm)	= 4 × [CB2(t) + 0.5 × UI(t)] = 0.5BA + 2CB2(t)
BA(CB1m)	= 4 × [CB2(t) + 0.5 × UI(t)] = 0.5BA + 2CB2(t)
BA(CYm,CB1m)	= 4 × [CB2(t) + 0.5 × UI(t)] = 0.5BA + 2CB2(t)
BA(UIm)	= 4 × CB2(t) + A
BA(CB2m)	= A
BA(Rm)	= 4 × [UI(t) + k × CB2(t)] + A
	= BA + A + 4 × (k-1) × CB2(t)
BA(Rm,Otxm)	= 4 × [UI(t) + k × CB2(t)]
	= BA + 4 × (k-1) × CB2(t)
BA(UIm,CB2m)	= A
BA(Rm,CB2m)	= A
BA(UIm,Otxm)	= 4 × CB2(t) + A(Otxm)
BA(UIm,Otxm,CB2m)	= A(Otxm) = BP(t)
BA(Pm)	= A
BA(CG1m)	= A

Blue expressions consist of components that can be perceived directly from kinetic data (as illustrated in Figs 3-7).

which step up UI input if both sites are present but not otherwise (i1, i2; Fig. 6); and (4) the BA intermodule input switch (i5, i7; Fig. 7). The Boolean functions provide many of the testable logical constraints that constitute the model. Clearly, however, the key aspect of the system is that it processes time-varying inputs, i.e. it is a kinetic information processor. An obvious caution to be noted is that this system would never have been revealed by a conventional 'deletion analysis', nor by any experimental approach focused only on discovering its drivers (i.e. UI and Otx, and to a lesser extent the protein(s) binding at the CB2 site). Discovery of the regulatory organization of the *endo16* system depended on first establishing all the target sites, and then determining their functions and inter-relationships.

Why does *endo16* require the elaborate regulatory apparatus summarized in Fig. 8? In hindsight, the answer is clear: the gene needs to respond differentially to a great variety of different circumstances in developmental space and time. It uses the BA intermodule input switch to convert its regulatory function from a specification mode (i.e., the functions at i7, i8, i9 and i10) to a differentiation execution mode (i.e., the functions at i1-i7). It uses its synergistic amplification devices (i1 and i4) to run up its Module B output by a very considerable factor late in development, so as to convert the rather modest late increase in the basal UI driver activity seen in Fig. 3B,C into a strong midgut-specific transcriptional function. It runs all of its output through a further step-up and summing device through which it communicates to the basal transcriptional apparatus (i8, i11, i12). To carry out all these regulatory activities, a number of elemental subfunctions are executed: repression, enhancement, summing, sensing of threshold, intrasystem linkage and output transmission. These elemental subsystem functions must of course be carried out by protein

complexes mobilized at the target sites hardwired in the regulatory DNA sequence. The biochemical nature of the proteins and their interactions remains almost entirely unexplored, though the model provides specific predictions as to the various regulatory activities they mediate.

The model in Fig. 8 is specifically predictive in another sense. In Table 2 we see a list of predicted quantitative outputs for many different mutational alterations. These of course could also be used to predict the results of knockouts of the activities of the corresponding transcription factors. Such knockouts are now accessible in the sea urchin system by the TKO method (Bogarad et al., 1998); by use of Engrailed repressor domain fusions (e.g., Li et al., 1999); or by use of morpholino antisense oligonucleotides (Heasman et al., 2000; C.-H. Y., H. B. and E. H. O., unpublished observations). The predictions in Table 2 are kinetic, i.e. they indicate the quantitative timecourse that will be observed throughout development, for each case. They are experimentally testable, as shown in Figs 3-7 of this paper, and by Yuh et al. (Yuh et al., 1998), as are the many additional predictions that can be generated by the relationships specified in Fig. 8.

General implications

endo16 is a garden variety, downstream embryonic gene. It is unlikely that there is anything special about its spatial or temporal regulation. Its *cis*-regulatory system is probably typical in complexity (Arnone and Davidson, 1997). Few developmentally regulated genes have been analyzed at this level of detail. But enough is known to suggest strongly that information processing systems such as that of *endo16* are the rule, not the exception, for any gene that is differentially expressed in the process of development (Davidson, 2001). Indeed such systems constitute the fundamental genomic machinery that makes bilaterian development possible, for the basis of this process is the integration of diverse spatial and temporal inputs executed by the genomic regulatory apparatus that controls expression of each gene. The *endo16* system is an example that shows how developmental information processing works, and how what it does in each given case is hardwired into the regulatory DNA. We are left with an image of the animal genome as a network of tens of thousands of such information processing systems, one to several per gene. It is this genomic feature that underlies the basic evolutionary heritage of the Bilateria, that is, their amazing capacity to carry out development.

We are grateful to Prof. Ellen V. Rothenberg for her perspicacious and helpful review of a draft of the manuscript. We are very pleased to acknowledge the support of the Stowers Institute for Medical Research, and of the National Institute of Child Health and Human Development (Grant HD 37105).

REFERENCES

Arnone, M. I. and Davidson, E. H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**, 1851-1864.

- Arnone, M. I., Martin, E. L. and Davidson, E. H. (1998). *Cis*-regulation downstream of cell type specification: a single compact element controls the complex expression of the *Cy11a* gene in sea urchin embryos. *Development* **125**, 1381-1395.
- Bogarad, L. D., Arnone, M. I., Chang, C. and Davidson, E. H. (1998). Interference with gene regulation in living sea urchin embryos: Transcription factor Knock Out (TKO), a genetically controlled vector for blockade for specific transcription factors. *Proc. Natl. Acad. Sci. USA* **95**, 14827-14832.
- Davidson, E. H. (1990). How embryos work: a comparative view of diverse modes of cell fate specification. *Development* **108**, 365-389.
- Davidson, E. H. (1999). A view from the genome: spatial control of transcription in sea urchin development. *Curr. Opin. Genet. Dev.* **9**, 530-541.
- Davidson, E. H. (2001). *Genomic Regulatory Systems. Development and Evolution*. San Diego: Academic Press.
- Davidson, E. H., Cameron, R. A. and Ransick, A. (1998). Specification of cell fate in the sea urchin embryo: Summary and some proposed mechanisms. *Development* **125**, 3269-3290.
- Gan, L., Mao, C.-A., Wikramanyake, A., Angerer, L. M., Angerer, R. C. and Klein, W. H. (1995). An orthodenticle-related protein from *Strongylocentrotus purpuratus*. *Dev. Biol.* **167**, 517-528.
- 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.* **179**, 148-159.
- Heasman, J., Kofron, M. and Wylie, C. (2000). β -Catenin signaling activity dissected in the early *Xenopus* embryo: A novel antisense approach. *Dev. Biol.* **222**, 124-134.
- Li, X., Chuang, C. K., Mao, C. A., Angerer, L. M. and Klein, W. H. (1997). Two Otx proteins generated from multiple transcripts of a single gene in *Strongylocentrotus purpuratus*. *Dev. Biol.* **187**, 253-266.
- Li, X., Wikramanayake, A. H. and Klein, W. H. (1999). Requirement of SpOtx in cell fate decisions in the sea urchin embryo and possible role as a mediator of β -catenin signaling. *Dev. Biol.* **212**, 425-439.
- 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., Ernst, S., Britten, R. J. and Davidson, E. H. (1993). Whole mount *in situ* hybridization shows *Endo-16* to be a marker for the vegetal plate territory in sea urchin embryos. *Mech. Dev.* **42**, 117-124.
- 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.
- Yuh, C.-H., Moore, J. G. and Davidson, E. H. (1996). Quantitative functional interrelations within the *cis*-regulatory system of the *S. purpuratus Endo16* gene. *Development* **122**, 4045-4056.
- Yuh, C. H., Bolouri, H. and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: Functional analysis and computational model of a sea urchin gene control system. *Science* **279**, 1896-1902.
- Yuh, C. H., Li, X., Davidson, E. H. and Klein, W. H. (2001). Correct expression of *specza* in the sea urchin embryo requires both *Otx* and other *cis*-regulatory elements. *Dev. Biol.* (in press).
- Zeller, R. W., Coffman, J. A., Harrington, M. G., Britten, R. J. and Davidson, E. H. (1995a). SpGCF1, a sea urchin embryo transcription factor, exists as five nested variants encoded by a single mRNA. *Dev. Biol.* **169**, 713-727.
- Zeller, R. W., Griffith, J. D., Moore, J. G., Kirchhamer, C. V., Britten, R. J. and Davidson, E. H. (1995b). A multimerizing transcription factor of sea urchin embryos capable of looping DNA. *Proc. Natl. Acad. Sci. USA* **92**, 2989-2993.