Proteins that bind to *Drosophila* chorion *cis*-regulatory elements: A new C₂H₂ zinc finger protein and a C₂C₂ steroid receptor-like component

Martin J. Shea,¹ Dennis L. King,¹ Michael J. Conboy,² Brian D. Mariani,³ and Fotis C. Kafatos^{1,3}

¹Department of Cellular and Developmental Biology, Harvard University Biological Laboratories, Cambridge, Massachusetts 02138 USA; ²Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Philadelphia, Pennsylvania 19107 USA;

³Institute of Molecular Biology and Biotechnology, Research Center of Crete, and Department of Biology, University of Crete, Heraklion 711 10, Crete, Greece

Gel mobility-shift assays have been used to identify proteins that bind specifically to the promoter region of the *Drosophila* *s15* chorion gene. These proteins are present in nuclear extracts of ovarian follicles, the tissue where *s15* is expressed during development, and bind to specific elements of the promoter that have been shown by transformation analysis to be important for *in vivo* expression. The DNA binding specificity has been used for molecular cloning of two components from expression cDNA libraries and for their tentative identification with specific DNA-binding proteins of the nuclear extracts. The mRNAs for both of these components, CF1 and CF2, are differentially enriched in the follicles. DNA sequence analysis suggests that both CF1 and CF2 are novel *Drosophila* transcription factors. CF2 is a member of the C₂H₂ family of zinc finger proteins, whereas CF1 is a member of the family of steroid hormone receptors. The putative DNA-binding domain of CF1 is highly similar to the corresponding domains of certain vertebrate hormone receptors and recognizes a region of DNA with similar, hyphenated palindromic sequences. The nature of CF1 raises the possibility of hormonal control of choriogenesis in *Drosophila*.

[Key Words: Trans-acting factors; nuclear hormone receptors; chorion gene regulation; zinc finger protein; *Drosophila* transcription factors; DNA-binding protein]

Received February 9, 1990; revised version accepted April 30, 1990.

In recent years, the mechanisms responsible for developmental patterning in *Drosophila* have been elucidated to remarkable molecular detail. Starting with mutations that interfere with the overall polarity and segmentation of the embryo or with the pattern of adult structures, genetic regulatory hierarchies have been defined, and key members have been cloned. The proteins encoded by many of these regulatory genes show sequence features characteristic of DNA-binding proteins, consistent with the suggestion that they function as transcriptional regulatory factors. These proteins are thought to regulate each other's production during development by binding to promoter and enhancer elements and thus refine the spatial and temporal pattern of their expression. The same proteins are also thought to bind and thereby regulate downstream, as yet undefined "realizator" genes.

A complementary approach, predominant in the field of vertebrate regulation, is to start from a highly specialized tissue-specific gene and proceed "bottom up" in defining the regulatory interactions: Characterizing the *cis*-regulatory elements of the chosen gene in detail; bio-

chemically defining the proteins that interact with these elements; and then molecularly cloning the corresponding genes and subjecting them to reverse genetic functional assays. This approach in *Drosophila* has the attraction that ultimately it can be combined with classical genetic studies, thus intersecting with the "top-down" approach beginning with pattern formation.

We have been engaged in a bottom-up study of transcriptional regulation of a model developmental system in *Drosophila*, the *s15* chorion gene. This gene encodes an eggshell protein that is exclusively expressed in the last 2–3 hr of chorion formation (stages 13 and 14 of oogenesis), in the ~1000 follicular epithelial cells that surround each developing oocyte. In previous experiments, we have used various chorion DNA constructs in conjunction with germ line transformation analysis to delimit the DNA regions responsible for the developmentally regulated expression of *s15* (Mariani et al. 1988; Romano et al. 1988). The results indicated that sufficient regulatory elements exist in the proximal 5'-flanking DNA to permit developmentally correct expression. Internal deletions within that DNA suggested

the existence of at least three positive and negative *cis*-regulatory elements. The most definitive experiments identified TCACGT, a hexamer motif at -55 to -60 , as essential for *s15* expression: Elimination of the hexamer, either by deletions or by a linker-scanning (clustered substitution) mutation, abolished gene expression. The hexamer motif was initially pinpointed by comparative sequence analysis: It is present in comparable locations upstream of all chorion genes that have been characterized to date in four *Drosophila* species (Levine and Spradling 1985; Wong et al. 1985; Martinez-Cruzado et al. 1988; Fenerjian et al. 1989), as well as in nearly all silkworm chorion gene promoters (Spoerel et al. 1986; Mitsialis et al. 1989).

With this information as background, we have undertaken binding studies with ovarian nuclear extracts to identify proteins that recognize the known *s15* regulatory elements. We report the molecular cloning and characterization of two of these components, which proved to be novel members of the zinc finger protein superfamily: One has zinc fingers of the C_2H_2 type, and the other is a member of the nuclear hormone receptor family.

Results

A quantitative *cis*-activator in the *s15* chorion promoter

Our previous studies of the *s15* promoter (Mariani et al. 1988) had indicated the existence, in the vicinity of nucleotide position -69 , of a *cis*-activating element that is important for expression at the normal late period of choriogenesis. A 119-bp deletion mutation, $\Delta -189/-69$, substantially reduced late expression in transgenic *Drosophila*, whereas a deletion that was only 11 bp shorter, $\Delta -189/-80$, permitted normal late expression (both deletions resulted in abnormal early expression, because of the elimination of an upstream early repressing sequence). We have confirmed that even in the two most highly expressing transformant lines, deletion $\Delta -189/-69$ reduces by four- to ninefold the late expression of a reporter gene (Fig. 1). To further delimit this and other *cis*-regulatory elements of *s15*, we have constructed a series of linker-scanning mutations throughout the -13 to -148 region. The results revealed a complexity and redundancy of *cis*-regulatory elements greater than that suggested by the deletions alone and will be reported in full elsewhere. However, some of the results are pertinent to the present study. As shown in Figure 1, mutation *d*, which alters the -63 to -72 DNA, results in a threefold average decrease in *s15* expression in vivo (tested in four independent transformant lines). The effect is only quantitative: The temporal specificity of mRNA appearance is not altered (data not shown). In contrast to *d*, a similar mutation just upstream (mutation *de*, -72 to -80) has no significant effect on expression (data not shown). The effect of $\Delta -189/-69$ is clearly more severe than the effect of mutation *d*. This may be due to the existence of one or more functionally redundant elements within the de-

Figure 1. Effects of two mutations of the *s15* promoter on in vivo expression. Stage 14 follicles were dissected from the parental *cn;ry* *Drosophila* line (–) and from transformants bearing a wild-type (WT) *s15-P* gene [construct B/S of Mariani et al. 1988], a clustered substitution mutation (*d*; see Fig. 6), or a deletion mutation ($\Delta -189/-69$). Individual transformant lines for each construct are numbered. RNAs were extracted from each pool of follicles, electrophoresed, and blot-hybridized with probes that revealed the transcripts of the introduced *s15-P* gene (T) and the control, endogenous *s15-l* gene (E). For each RNA sample, the ratio of the two species of transcripts was evaluated more accurately by densitometry of quantitative dot-blots, hybridized with probes specific for each species; results from that independent experiment are shown below the respective lane, after normalization to the average ratio of the wild-type controls. Mutation *d* reduces average stage 14 expression to 0.31, and $\Delta -189/-69$ to 0.17.

leted DNA: Mutation *g* (-101 to -109), by itself, reduces in vivo late expression fourfold (data not shown). Although additional experiments are needed to pinpoint these regulatory elements accurately, the results to date suggest that a late activator maps in the -72 to -63 DNA region of *s15*.

An ovarian nuclear extract contains factors that recognize specific binding sites in the *s15* promoter

To characterize the *trans*-acting factors that might be responsible for the developmental regulation of the *s15* chorion gene, a procedure was developed to isolate nuclear proteins from ovaries. A previously reported method for mass isolation of developing follicles (Petri et al. 1977) was modified to recover purified nuclei (see Materials and methods). Salt extraction (Wu 1984) yielded nuclear extracts that showed highly specific DNA-binding activity.

Fragments of the *s15* promoter (-145 to $+32$ or shorter) were used in gel retardation assays (Fried and Crothers 1981) to detect specific DNA-binding proteins. Extracts from mixed stage follicles (stages 1–14 of King 1970, including early oogenesis, vitellogenesis, and chorion formation) yielded a series of protein–DNA complexes; the major ones were named I and X. These complexes were formed with extracts from both early (predominantly stage 1–10a) and late (choriogenic, stage 10b–14) follicles (Fig. 2a and b, respectively). An

Figure 2. Gel retardation assays of specific DNA-binding proteins in ovarian nuclear extracts. Labeled fragments of the *s15* gene, -145 to +33 (a) and -145 to -47 (b), as indicated in the diagrams, were mixed with nuclear extracts from early (a, predominantly stage 1–10a) or late (b, predominantly stage 10b–14) follicles. Major retarded complexes are labeled I, II, and X; free DNA is labeled F. The fragments were of wild-type sequence (WT), truncated (Δ -113), or mutated (a to g). Mutations were clustered (linker-scanning) substitutions at the positions indicated by black bars in the diagrams, except for mutation a, which included both a substitution and a small deletion. Mutation c eliminates the TCACGT hexamer (open bar), and mutation a eliminates the TATA box (stippled bar). Note that mutations c and d interfered with complex I formation, d interfered with complex II, and e–g interfered with complex X, thus revealing the approximate binding sites of the respective proteins. c is from a competition experiment, using nuclear extract from early follicles and a constant amount of labeled wild-type *s15* DNA (-113 to +33), in the absence (-) or presence of 50-fold molar excess of unlabeled competing fragments, from -381 to +599 of *s15*. Note that wild-type fragments or those that bear mutations a, b, and e compete equally well, the d mutant fragment competes partially, and the c mutant fragment shows essentially no competition. In this experiment, the gel conditions were slightly different, and complex X is not seen, presumably because of overlap with free DNA.

additional complex, which was named II, was most abundant in the late extract (Fig. 2b).

The binding specificities of the proteins responsible for these major complexes were confirmed by varying concentrations of NaCl and poly[d(I-C)]·poly[d(I-C)] (data not shown). The binding sites were localized more precisely within ~80 bp upstream of the TATA motif, by use of a series of linker-scanning mutations of the promoter. As can be observed in Figure 2, specific complexes disappeared when defined short sequences were altered. In particular, the binding site necessary for complex I formation is encompassed in the DNA segment between -55 and -72 (mutations c and d); the site necessary for complex II overlaps the -63 to -72 segment (mutation d); and that needed for complex X spans the -81 to -109 segment (mutations e–g). The specificity of complex I was also examined by competition experiments, using wild-type promoter fragments as the labeled probe and corresponding fragments derived from wild-type or mutated promoters as the unlabeled competitor. While confirming the previous results, these competition experiments revealed that complex I is more destabilized by mutation c than by mutation d (Fig. 2c). Furthermore, methylation interference showed that the guanine contact sites for complex I are within the -62 to -57 segment (see below). Thus, this protein is largely a TCACGT-binding factor, although its binding site extends farther upstream than the hexamer.

In preparation for the molecular cloning of these DNA-binding proteins, two double-stranded oligonucleotides spanning the putative binding sites were syn-

thesized. The downstream hexamer-centered oligonucleotide (H, -80 to -42) included the binding sites for complexes I and II, whereas the upstream oligonucleotide (U, -122 to -76) included the binding site of complex X. When H or U was mixed with ovarian extracts, the expected retarded complexes were observed (data not shown).

Recombinant clones producing proteins that specifically recognize *s15* promoter sequences

By use of the procedure of Vinson et al. (1988), the synthetic oligonucleotides were individually catenated, labeled, and used as binding probes for proteins produced by clones of λ gt11 expression cDNA libraries. The primary screens were performed with a mixture of the catenated H and U oligonucleotides, together with a control catenated 50-mer probe derived from the adult enhancer of the *Drosophila Adh* gene. Secondary screens of putative positives were performed with each of the three probes separately. In our experience, parallel secondary screens with individual control and experimental probes are essential, because primary screens yield a high proportion (~90%) of false positives, corresponding to proteins with indiscriminate DNA-binding activity. Figure 3 shows secondary screens of five positive clones, each of which recognized only one of our probes. Restriction analysis confirmed that each clone shown in Figure 3 is distinct, except for the clones named CF2, which are identical (data not shown, and see below). Clone AEF1 (adult enhancer factor 1; isolated by D. Falb) served as a

Figure 3. Secondary screens of five cDNA clones encoding specific DNA-binding proteins. The clones were isolated initially, from λ gt11 expression libraries (see Materials and methods), by use of a mixed binding probe consisting of the U and H *s15* oligonucleotides plus a 50-mer derived from the *Adh* gene adult enhancer. The two CF2 clones were isolated independently but proved to be identical. Each of these clones (CF1, CF2, CF3, and AEF1) was replated on triplicate plates, which were then screened with the three probes individually. Note that CF1 and CF2 only react with the H probe, CF3 only reacts with the U probe, and AEF1 only reacts with the *Adh* probe, confirming the specificity of binding.

control; it corresponds to a putative *trans*-acting factor for an adult *Adh* enhancer element.

With the chorion oligonucleotide H, we recovered clones corresponding to two proteins that bind specifically: CF1 (chorion factor 1) and CF2 (chorion factor 2). Most likely, these proteins correspond to complexes I and II, respectively (see below). Each sequence is unique in the genome. CF1 and CF2 were mapped by *in situ* hybridization to *Drosophila* polytene chromosomes: CF1 to cytogenetic location 2C4-8 (X chromosome), and CF2 to 25A5-8 (chromosome 2L). As shown in Figure 4, both CF1 and CF2 mRNAs are substantially more abundant in ovarian follicles than in the other tissues and developmental stages that were tested (embryos, pupae,

and adults of both sexes). Multiple transcripts are evident, suggestive of developmentally regulated alternative splicing. The major form of CF1 mRNA is 2.4 kb long. CF2 has two major mRNA forms, ~3.5 and 2.5 kb long, which are found in different ratios in prechoriogenic and choriogenic follicles (Fig. 4).

With the U oligonucleotide, we have not as yet obtained a clone corresponding to factor X, which we suspect to be a repressor, according to the results of *in vivo* deletion analysis (Mariani et al. 1988). However, the U oligonucleotide did select a clone that encodes a specific DNA-binding protein, CF3 (Fig. 3). The binding site of this factor coincides with the site of mutation *g*, which is important *in vivo* (see above). Our analysis of CF3 and its equivalent in follicular extracts is incomplete and will be presented elsewhere; however, the distinct binding specificity of CF3 provides an additional control, which emphasizes the specificity of CF1 and CF2 for the H oligonucleotide (Fig. 3).

CF1 and CF2 proteins bind to distinct elements of the s15 promoter

*Eco*RI fragments containing the *Drosophila* CF1 and CF2 inserts were subcloned in the translation vector pET-3b (Rosenberg et al. 1987). Overexpressed proteins

Figure 4. Distribution of CF1 and CF2 mRNA in selected tissues of *Drosophila*. Two micrograms each of poly(A)⁺ RNA was isolated from embryos [all stages (e); pupae (p); whole males (δ); females minus ovaries (♀-ov.); prechoriogenic follicles (1–10); choriogenic follicles (11–14)]. The RNA was gel-blot hybridized with RNA probes derived from the 0.9-kb CF1 cDNA (left) or the 1.6-kb CF2 cDNA (right). Two different exposures of the same filter, separated by a bar, are included in each panel. Major transcripts, as discussed in the text, are indicated by asterisks (*). These include a 2.4-kb mRNA for CF1 and both 2.5- and 3.5-kb mRNAs for CF2. Minor transcripts, denoted by dots, include both a 1.5- and a 1.1-kb species for CF1 and a 0.6-kb species for CF2.

present in the bacterial extracts were tested for binding to wild-type and mutated *s15* promoter fragments (Fig. 5). Control extracts were obtained from bacteria harboring each of the inserts in an inverted orientation. Each of the inserts, in the correct orientation only, yielded a protein that specifically recognized *s15* promoter sequences. The results showed that CF1 binding is abolished by mutation *c*; under some conditions, the binding appeared to be weakened by mutation *d*, in a manner resembling the results of competition experiments with follicular extracts (Fig. 2c). Taking into account the incomplete nature of the cloned protein (see below), these results are consistent with the interpretation that CF1 corresponds to follicular complex I. Similarly, both CF2 and complex II are uniquely sensitive to the *d* mutation, suggesting that they are identical.

We used DNase I footprinting analysis to define more precisely the boundaries of interaction between the bacterially produced factors and the *s15* promoter sequences. Examples of the footprints and a diagram of the results are shown in Figure 6. CF1 and CF2 yielded nuclease protection footprints of ~20 bp (for both strands combined) and some major nearby enhancements. The footprint of CF1 includes the TCACGT hexamer, 2 bases downstream and 10 bases upstream. A similar DNase I footprint was obtained with complex I from follicular extracts (data not shown). The major footprint of CF2 overlaps that of CF1 extensively but is shifted

slightly upstream; as expected from Figure 5, it encompasses the site of mutation *d*. The addition of 1 mM Zn^{2+} enhances the CF2 footprint and reveals two secondary binding sites: One is centered over the TATA box, and the other is centered over an upstream region, approximately -130 to -100. In 10–12 of 16 nucleotide positions, the upstream ends of the primary and secondary CF2 footprints match the consensus AATtG.GTA-TAta.g (invariant nucleotides are uppercase and those shared by two of three sites are lowercase).

Methylation interference was used to identify and compare the guanines that are contacted by the bacterially expressed proteins and the authentic DNA-binding proteins of follicular extracts. The results were perfectly consistent between CF1 and complex I (autoradiographs in Fig. 6) and also between CF2 and complex II (not shown). Furthermore, the results confirmed that the binding sites of CF1 and CF2 overlap extensively but are slightly offset; the results are summarized in Figure 6. CF1 and complex I require the same 4 guanines for binding, 2 within the hexamer in the negative strand (-57 and -59) and 2 immediately upstream of the hexamer in the plus strand (-61, -62); the contact with the guanine at -57 is slightly less important for binding than the rest (Fig. 6). Similarly, CF2 and complex II both have a strong requirement for guanine -71 and weaker requirements for guanines -69, -62, and -61, all in the positive strand; the latter 2 guanines are required much more strongly for binding of CF1 and complex I.

Figure 5. Gel retardation assays of DNA-binding proteins expressed in bacteria. (top autoradiograph) Results obtained with the CF1 clone; (bottom autoradiograph) results with the CF2 clone; retarded bands attributed to these factors are labeled, and free DNA is also indicated (F). The DNA was a -145 to -47 fragment of the *s15* promoter and was of wild-type sequence (WT) or carried a clustered substitution mutation (*c* to *g*). The diagram (top) shows the locations of the mutations on the DNA fragment used for each experiment. Note that CF1 cannot bind on DNA carrying the *c* mutation, whereas CF2 cannot bind on DNA carrying the *d* mutation.

CF1 and CF2 are new zinc finger proteins; CF1 is a member of the nuclear hormone receptor family

The CF1 and CF2 inserts were subcloned into M13 and subjected to DNA sequence analysis by the dideoxynucleotide method. Figure 7 presents the coding strand sequences and their conceptual translations.

Both clones are incomplete and encompass open reading frames of 280 (CF1) or 235 (CF2) codons. Although CF1 is incomplete at both ends, CF2 includes a termination codon, followed by 819 nucleotides of putative 3' untranslated sequence (including an AATAAA motif) and a remnant of the poly(A) tail (19 A residues).

In each case, the open reading frame shows a putative DNA-binding domain consisting of zinc finger motifs. The domain of CF2 (113 amino acid residues) consists of four contiguous zinc fingers of the C_2H_2 type (Cys-Cys-His-His). As seen in Figure 8a, in addition to the conserved C and H residues, the CF2 fingers show characteristic, properly spaced hydrophobic residues (F and L) typical of this type of motif. They most closely resemble the developmentally regulated *Drosophila* Krüppel, mouse *mkr1* and *mkr2*, and *Xenopus* *Xfin* and *XLcOF* zinc finger sequences, although they also share some similarities with the constitutive mammalian transcription factor Sp1 (for detailed comparisons and references, see Discussion).

The CF1 sequence is especially intriguing. It includes a central segment of 68 amino acid residues, with unmistakable homology to the DNA-binding domain of

binding domain of CF1 and the vertebrate sequences (41%, 41%, and 40% identity to CF1), whereas *svp*, which appears to be the *Drosophila* homolog of ear3/COUP, is 65% identical to the DNA-binding domain of CF1. Outside that domain, CF1 shows no substantial sequence identity to any previously described protein.

Cloning of two chorion promoter binding factors

GENES & DEVELOPMENT 1133

Figure 7. DNA sequence of the coding strand and conceptual translation of the CF1 and CF2 cDNA clones. The putative DNA-binding domains are boxed, and conserved residues are in boldface type (cf. Fig. 8). A polyadenylation signal, AATAAA, is highlighted in the 3'-untranslated region of CF2.

now culminated in the molecular cloning of cDNAs for proteins that recognize *cis*-regulatory elements of *s15*. Three lines of evidence strongly suggest that the cloned CF1 and CF2 components correspond to the follicular DNA-binding proteins responsible for complexes I and II, respectively. First, CF1 and CF2 RNAs are highly enriched in follicles, relative to other tested tissues and developmental stages. Second, the cloned components are indistinguishable from the authentic follicular proteins in their methylation interference patterns. Third, the binding of cloned and authentic components is similarly affected by the linker-scanning mutations used: Binding is abolished by mutation *d* in the case of CF2/complex II, and by mutation *c* in the case of CF1/complex I. Mutation *d* appears to weaken the binding of CF1/complex I; the apparent difference between Figures

2 and 5 is attributable to differences in protein concentration and possibly the incomplete nature of the CF1 clone. Definitive identification will require additional studies, such as the use of anti-CF1 and anti-CF2 antibodies to interfere with or alter the mobility of complexes observed with follicular extracts.

Binding alone does not establish *in vivo* function: Direct *in vivo* and *in vitro* functional assays will be needed to document the importance of CF1 and CF2. However, we note that the *s15*-binding sites for these proteins are important *in vivo* and that the proteins belong to two well-known families of transcription factors (see below). Both CF1 and CF2 also occur outside the ovary, and detailed studies on their tissue and temporal distribution during *Drosophila* development will be of interest; however, it is highly suggestive that their mRNAs are excep-

Figure 8. (a) Comparison of the DNA-binding domains of CF2 and two typical zinc finger proteins. For each protein the sequences shown are contiguous from line to line. Each line includes one zinc finger, conforming to the C_2H_2 consensus ($CX_2CX_3FX_5LX_2HX_3H$); the conserved residues are highlighted with vertical solid lines. Note that in two Sp1 fingers the first 2 cysteines are separated by 4 rather than 2 residues. Sequences linking adjacent fingers are also shown; they frequently conform to the H/C link consensus (Berg 1988), TGE $\overline{K}P\overline{F}$ (vertical dashed lines). (b) Comparison of the DNA-binding domains of CF1 and 12 additional members of the nuclear hormone receptor family. This domain encompasses two zinc fingers of the C_2C_2 type. Residues that are invariant in all of the sequences shown are highlighted with vertical solid lines. Percent sequence identities relative to the CF1 domain are listed at right. The closest known homolog of CF1 is the ear3/COUP transcription factor and the product of the corresponding *Drosophila* gene *svp*. Asterisks show that these two proteins share the P-box sequence characteristic of the estrogen receptor/thyroid receptor subfamily (Umesono and Evans 1989) and are also unusually similar in the D box, which is thought to aid recognition of the DNA-binding site (Umesono and Evans 1989). Unusually high similarity between CF1 and ear3 or *svp* (vertical dashed lines) is seen in two additional segments: The link between the two zinc fingers and a trailer segment following the second finger. Additional sequence matches between CF1 and other proteins are indicated by a dot beneath the respective residue. Sequences other than CF1 were obtained from the following publications: (ear3 and ear2) Miyajima et al. (1988), (*svp*) Mlodzik et al. (1990); [human retinoic acid receptor α (hRAR α)] Giguere et al. (1987); [human estrogen hormone receptor (hER)] Greene et al. (1986); [human thyroid hormone receptor β (hTR β)] Weinberger et al. (1986); [oncogene product of avian erythroblastosis virus (AEV; v-erbA)] Debuire et al. (1984); [human vitamin D₃ receptor (hVDR)] Baker et al. (1988); [*kni* protein] Nauber et al. (1988); [*knrl* protein] Oro et al. (1988); (*egon* protein) Rothe et al. (1989); [human glucocorticoid hormone receptor (hGR)] Hollenberg et al. (1985).

tionally abundant in follicles. Although additional, follicle-specific factors for regulating choriogenesis may well exist, there are many precedents for combinatorial regulation of development-specific genes by factors that are not unique to the tissue of interest.

Although the binding sites of CF1 and CF2 are distinguishable, they overlap extensively. Methylation of the same 2 guanines (-62 and -61) prevents binding of CF1 and weakens the CF2 binding, whereas the converse is true for mutation *d*. These observations raise the possibility that the two factors might bind in a mutually exclusive manner. It is interesting that complex I is at least as abundant in early as compared to late ovarian extracts, whereas the abundance of complex II is substantially greater at the late stages.

Nature and significance of CF2

CF2 has the hallmarks of transcription factors. At the beginning of the available sequence, a segment of 37 residues includes 17 glutamines interspersed with 7 acidic residues; such clusters are frequently found in activation domains of transcription factors (Mitchell and Tjian 1989). As shown in Figure 8a, the zinc finger domain of CF2 is reminiscent of the protein encoded by the *Drosophila* regulatory gene *Krüppel* (Rosenberg et al. 1986); it also resembles the finger domains encoded by *hunchback* (Tautz et al. 1987), *serendipity* β and δ (Vincent et al. 1985), and *glass* (Moses et al. 1989). All of these proteins have a small number (five to seven) of clustered zinc fingers of the C_2H_2 class, each with 2 amino acid

residues between the first 2 cysteines (in a different sub-type of C_2H_2 zinc finger, 4 residues are present between the cysteines, resulting in an extra β -bend in the expected structure; Berg 1988). Zinc fingers, of the same type as in CF2, sometimes in very long arrays, are found in certain vertebrate proteins such as those encoded by the *Xenopus Xfin* (Ruiz i Altaba et al. 1987) and *XLcOF* (Köster et al. 1988) sequences, and the mouse *mkr1* and *mkr2* genes (Chowdhury et al. 1987). CF2 also shares with many of the previously mentioned proteins a characteristic short sequence between the zinc fingers (H/C link; Berg 1988), conforming to the consensus TGEKPPX (Fig. 8a).

CF2 shows enhanced DNA binding in the presence of added Zn^{2+} ; a similar property has been documented for Sp1 (Kadonaga et al. 1987). With added Zn^{2+} , CF2 binds not only to its primary recognition site, just upstream of the TCACGT hexamer, but also to two flanking, secondary sites of related sequence: the TATA box and a site at the upstream end of the previously defined minimal regulatory region (Mariani et al. 1988).

Although CF2 and CF1 show extensively overlapping footprints, some of the mutations tested prevent the binding of only one of these factors, thus permitting separate inference of their possible *in vivo* functions. Mutation *c*, which eliminates the TCACGT hexamer and abolishes expression (Mariani et al. 1988), prevents the binding of CF1 but not of CF2; this suggests that CF1 may be an essential hexamer-binding positive factor. The possible effect of CF2 cannot be inferred from mutation *d* alone, because this mutation also weakens CF1 binding. However, the effect of this mutation, coupled with the observation that $\Delta-189/-69$ suppresses late expression (Fig. 1; Mariani et al. 1988) and that $\Delta-189/-69$ hinders CF2 binding and complex II formation but not the binding of CF1 or complex I formation (data not shown), all argue that CF2 may act as a late activator (Mariani et al. 1988). If CF2 binds *in vivo* to two secondary sites, as it does in the presence of added Zn^{2+} *in vitro*, the phenotype of mutation *d* may grossly underestimate the importance of CF2.

CF1 and the control of *s15* expression

The nature of CF1, a factor that binds strongly to the conserved chorion hexamer (TCACGT) and adjacent DNA, is particularly intriguing. Again, direct functional tests are needed to establish whether CF1 is a transcription factor involved in the regulation of choriogenesis. Genetics could provide such a test: CF1 maps to a favorable location in the polytene chromosomes in a region that has been genetically well studied, because of proximity to the pattern formation gene *polyhomeotic* (Dura et al. 1987). However, CF1 is not *polyhomeotic*, nor is it likely to correspond to any of the nearby complementation groups that are uncovered by deficiency *Df(1)A52*: Our CF1 cDNA maps between 43 and 51 kb on the chromosomal walk of Dura et al. (1987), ~25 kb outside the nearest breakpoint of *Df(1)A52* (data not shown).

On the basis of its putative DNA-binding domain, CF1 is clearly a member of the family of proteins that have two zinc fingers of a special class (C_2C_2) and function as nuclear hormone receptors (for review, see Evans 1988; Berg 1989). This factor has been cloned independently by two additional groups: by A. Oro and R. Evans (pers. comm.), by use of the DNA-binding domain of the human retinoic acid receptor as probe; and by V. Henrich and L. Gilbert (pers. comm.), using as probe a consensus oligonucleotide from a portion of the DNA-binding region of human steroid receptors. At the protein level (Fig. 8b) the DNA-binding domain of CF1 is most similar (68% identity) to the binding domain of the human ear3 receptor, a ubiquitous component of fetal tissues (Miyajima et al. 1988). ear3 corresponds to the COUP transcription factor which, in conjunction with a second protein, stimulates *in vitro* transcription from the promoter of the chicken ovalbumin gene (Wang et al. 1989). The apparent ear3/COUP homolog of *Drosophila*, *svp*, is an essential gene that is involved in the control of cell fate during the generation of neuronal diversity in the eye (Mlodzik et al. 1990). CF1 also shows identities in the 43–57% range with several additional vertebrate members of the nuclear receptor family, although only 40–41% with three *Drosophila* members that have been described thus far (*kni*, *knrl*, and *egon*).

The presence of a zinc finger structure characteristic of the steroid receptor superfamily suggests that CF1 may be a ligand-dependent transcription factor. This raises the intriguing possibility of hormonal control of choriogenesis in *Drosophila*. Two lines of indirect evidence favor that possibility. First is the isolation of a somatic cell-specific mutation, *fs(1)de12*, which results in drastically underdeveloped chorionic appendages (Orr et al. 1989). That mutation is allelic to *l(1)npr-1* and, thus, part of the *Broad-Complex* locus, which has been implicated in the mediation of ecdysone effects (Belyaeva et al. 1981; Meyerowitz et al. 1985). Second, oogenesis in *Drosophila* is feedback-inhibited by failure of oviposition; the mechanism of that inhibition is unknown but could be hormonal, as is the regulation of other steps in oogenesis such as vitellogenesis in many insects (Postlethwait et al. 1980; Racciopi et al. 1986).

Direct identification of the putative CF1 ligand would be of major interest. Pending that, closer scrutiny of the CF1 DNA-binding domain and the DNA site that it recognizes is informative. The nuclear hormone receptors have been classified into two subfamilies, GR and ER/TR, typified by the human glucocorticoid and estrogen or thyroid hormone receptors, respectively (Umesono and Evans 1989). The discriminating characteristic is the sequence of the P box at the end of the first zinc finger, where the GR subfamily shows a GSKV motif and the ER/TR subfamily shows an EGCK motif. Interchange of these motifs is sufficient to change the specificity of a glucocorticoid receptor so that it recognizes an estrogen response DNA element (ERE), rather than a glucocorticoid response element (GRE; Klock et al. 1987; Martinez et al. 1987; Green et al. 1988; Umesono and Evans 1989). CF1 is clearly a member of the ER/TR

subfamily of receptors by the criterion of P box sequence, as are *ear3*, *svp*, *kni*, *knrl*, and *egon* (Fig. 8b).

Another interesting feature of nuclear hormone receptors is the D box between the first 2 cysteines of the second zinc finger (Umesono and Evans 1989). That region is moderately variable in sequence, and it is also thought to be involved in some aspect of DNA recognition (Umesono and Evans 1989). In the D box, CF1 differs from most of the hormone receptors but is nearly identical to *ear3*, the *svp* product, and *ear2* (Fig. 8b). It also shows unusually strong sequence similarity with *ear3* and the *svp* product in two additional regions: a DNA segment that links the two zinc fingers, between the P and D boxes, and a trailer segment that follows the second zinc finger (Fig. 8b).

Identification of CF1 as a member of the nuclear hormone receptor family led us to re-examine the sequence of its binding site. Members of this family are known to recognize DNA palindromes, perfect or imperfect, with two half-sites that are separated by variable DNA spacing, specific for each class of receptors (for review, see Evans 1988; Beato 1989; Berg 1989). As shown in Figure 9, the binding site of CF1 in the *s15* promoter is the first half-site of an imperfect palindrome, with substantial similarity to vertebrate hormone response ele-

ments. The major difference is the spacing of the half-sites, which are 6 nucleotides apart for the CF1-binding site, directly contiguous for TRE (thyroid), and 3 nucleotides apart for ERE (estrogen) or GRE (glucocorticoid), as well as the binding site of the COUP transcription factor. The invariant chorion hexamer TCACGT overlaps the first half-site and the beginning of the spacer; the rest of this putative CF1 response element is also extensively conserved in the *s15* promoter of four different *Drosophila* species (Martinez-Cruzado et al. 1988), and in silkworm chorion genes (Mitsialis et al. 1989).

Our working hypothesis, based on these sequence similarities and the known mode of binding of vertebrate hormone receptors, is that full-length CF1 would bind avidly to the first half-site of the motif shown in Figure 9 and, upon dimerization, also bind to the second half-site. The CF1 factor currently in our possession is incomplete and has not shown dimerization. The prediction can be tested by future binding studies involving full-length CF1 protein synthesized in bacteria, as well as the authentic factor responsible for complex I formation, purified from follicular extracts.

Materials and methods

In vitro mutagenesis

Promoter mutants of the *s15* gene were obtained by site-directed mutagenesis (Carter et al. 1985) of a 990-bp *Xba*I-*Kpn*I DNA fragment. Selected target sites were mutagenized by use of synthetic oligonucleotide primers of varying lengths (36–44 nucleotides). Each primer contained a centrally located 10-nucleotide mutagenic sequence (5'-CCGAATTCC-3' or 5'-CGGAATTCC-3'), which included an *Eco*RI recognition site (underlined), flanked by sequences wild type for *s15*. Thus, mutagenized fragments contained *Eco*RI sites in their target sites; the number of mismatches to wild-type *s15* sequence ranged from 6 to 9 bp. Mutant clones were verified by dideoxynucleotide sequencing.

Gel-blot and dot-blot hybridization analysis

For *in vivo* analysis of *s15* promoter mutations, the *s15* gene was marked by an insertion of silkworm DNA and used as a reporter (*s15-P*; Mariani et al. 1988). Ovaries from conditioned transformed females were dissected, and stage 14 follicles were harvested and total RNA extracted, and gel-blot hybridization was performed as in Mariani et al. (1988). For dot-blot hybridizations, 5 µg of total RNA (for detection of both endogenous *s15-l* transcripts and transformant-specific *s15-P* transcripts) from the above samples was incubated in 200 µl of a solution of 5 mM Tris-HCl (pH 8.0), 0.5 mM EDTA (pH 8.0), 3 × SSC, and 7% (wt/vol) formaldehyde for 15 min at 65°C prior to dot-blotting to Biotrans (ICN) nylon membranes with a suction manifold. The RNA samples for the *s15-l* probeings were supplemented with 4.5 µg yeast tRNA. RNA was fixed to membranes and hybridized as described in Mariani et al. (1988), with the addition of 0.5 mg/ml sheared and denatured herring sperm DNA. To detect *s15-P* or *s15-l* transcripts, the RNA probes were derived from the T3 promoter of a Bluescribe plasmid (Stratagene, San Diego, CA) that contained either a 202-bp *Sac*I fragment from the silkworm insert of the *s15-P* gene or the full-length wild-type cDNA (*s15-l*), respectively.

Densitometry values were determined by use of an LKB

Figure 9. Comparison of binding sites for the *Drosophila* CF1 protein and for vertebrate hormone receptors. The sites are imperfect palindromes; the commonly recognized palindromic half-sites are indicated by horizontal black bars, and a more extensive dyad symmetry in the *Xenopus* vitellogenin A₂ ERE (ERE-Xvit.A₂) is emphasized by arrows. Other vertebrate hormone response elements are shown as consensus derived from data in the literature (GRE cons., TRE cons., and ERE cons.). The COUP transcription factor-binding site is shown as it occurs in the chicken ovalbumin promoter (COUP-Covalb.). The insect chorion sequences that contain the TCACGT hexamer (asterisks) are shown for the *s15* gene in four *Drosophila* species (*D. melanogaster*, *Dm*; *D. simulans*, *Ds*; *D. virilis*, *Dv*; and *D. grimshawi*, *Dg*), and for the *L12* gene pair in *Bombyx mori* (*Bm*). Nucleotides that show matches between the vertebrate and insect elements are boxed; dashes signify absence of a nucleotide, and *n* signifies an unspecified nucleotide. Note that the half-sites are separated by zero or 3 nucleotides in the vertebrate elements and by 6 nucleotides in the insect elements. Essential contact sites for CF1 are indicated by arrowheads. The reported sequences were obtained as referenced at right: (1, 4, and 5) Evans (1988); (2) Beato (1989); (3) Wang et al. (1989); (6) Martinez-Cruzado et al. (1988); (7) Mitsialis et al. (1989).

Bromma Ultrosan XL laser densitometer, from autoradiographs obtained without an intensifying screen and exposed within the linear range of the film. The values for *s15-P* were normalized relative to the *s15-I* values in the same sample; and for each line, the normalized *s15-P* value was listed in Figure 1 relative to the corresponding value of the wild-type control (full-length B/S construct; see Mariani et al. 1988).

For *in vivo* analysis of CF1 and CF2 expression, total RNA was extracted from tissues of various life stages, as indicated in Figure 4, and poly(A)⁺ RNA was selected by use of oligo(dT)-cellulose type 7 (Pharmacia). Samples of poly(A)⁺ RNA (2 µg each) were gel-blot analyzed as described above. Transcripts were detected with RNA probes derived from the T₃ promoter of Bluescribe plasmids containing either a 0.9-kb CF1 cDNA or a 1.6-kb CF2 cDNA fragment.

Follicle isolation and nuclear extract preparation

Follicles were isolated as in Petri et al. (1977), with the following modifications. Canton S flies were conditioned for 36–48 hr in one-half pint, yeast-sprinkled food bottles (~200–300 flies per bottle), with one bottle change after 18–24 hr. Step C3, stirring of the brei with a large magnetic stirring bar, was omitted. To isolate a late stage fraction, the total pool of follicles was shaken vigorously through a Nitex 202 screen, and follicles that were retained—stages 10b–14 but predominant stage 14—were collected. To obtain early stage follicles, the flowthrough fraction in the above fractionation was lightly shaken through Nitex 143, and the subsequent flowthrough fraction, consisting predominantly of stage 1–10a follicles, was collected.

Isolation and salt extraction of nuclei was similar to that of Wu (1984), with several modifications. Briefly, follicles were allowed to settle, and an equal volume of 2× Robb's saline buffer [RSB: 100 mM Tris-HCl (pH 7.4), 100 mM NaCl, 30 mM MgCl₂] plus 1% NP-40 and 1 mM PMSF was added. This mixture was homogenized (Thomas homogenizer 3431-D76), placed on top of a sucrose cushion (1 M sucrose, 1× RSB, 0.5 mM PMSF), and spun for 15 min at 3000 rpm in an HB4 rotor. The nuclear pellet was resuspended in a wash solution (1× RSB, 0.5 mM PMSF) and repelleted as above. The nuclei were then resuspended in a small volume of a salt extraction wash buffer [10 mM HEPES (pH 7.9), 10 mM KCl, 1.5 mM MgCl₂, 0.1 mM EGTA, 0.5 mM DTT, 0.5 mM PMSF], transferred to Eppendorf tubes, and repelleted at 3000 rpm for 15 min. After estimating the volume of the pellet, solution II (Wu 1984), at 3.5× the pellet volume, was added and rotated gently for 30 min. This was followed by a 50-min spin at 34,000 rpm in a TFT80.4 rotor. All of the above steps were performed at 4°C. The supernatant (nuclear extract) was then aliquotted and stored in liquid N₂.

Retardation gels

The following binding conditions were used for gel retardation experiments (Fried and Crothers 1981). Nuclear or bacterial extracts (10–20 µg protein) were incubated with ~0.2–1.0 ng of ³²P-end-labeled DNA fragment in the presence of 3 µg of poly[d(I-C)]·poly[d(I-C)] (Pharmacia), in buffer that was 4 mM Tris (pH 7.5), 75 mM NaCl (final), 5% glycerol, 0.5 mM DTT, and 1 mM EDTA. The incubation was carried out in 25 µl at 22°C for 30 min and was followed by loading on a 6% 0.25× Tris-borate-EDTA buffer, nondenaturing polyacrylamide gel (containing 5% glycerol) that was prerun and run at 4°C. For the competition experiment shown in Figure 2c, a 4% gel with 1× Tris-acetate-EDTA buffer and without glycerol was run at 22°C.

DNase I and methylation interference footprints

Direct DNase I footprinting was performed as in Heberlein et al. (1985), with the following modifications (K. Zinn, pers. comm.). Binding reactions of 15 µl, consisting of 10 µl of bacterial extract, probe (~20,000 cpm), 2 µl of polyvinyl alcohol (13.3%), 3 µg of poly[d(I-C)] and 75 mM NaCl (final), were incubated for 30 min at room temperature (~22°C). A 35-µl aliquot of a 5 mM CaCl₂, 10 mM MgCl₂ solution was added, followed by the addition of 2.5 µl DNase I stock (2 µg/ml). After 1 min at room temperature, 75 µl of a stop mix (1.7 M ammonium acetate, 2 µg glycogen) was added, followed by phenol, phenol-chloroform, and chloroform extractions, and ethanol precipitation. The DNAs were subsequently analyzed on a 7% polyacrylamide sequencing gel.

Methylation interference footprinting was performed in the following manner. Radiolabeled DNA was pretreated with dimethyl sulfate (Hendrickson and Schleif 1985). A preparative retardation gel was run, the wet gel was placed on film, and free and bound complexes were subsequently excised from the gel. The DNA was electroeluted, and the solution was brought to 0.6 M NaCl, followed by phenol, phenol-chloroform, and chloroform extractions. The DNA solution was then diluted to 0.1 M NaCl and passed over Elutip-d columns (Schleicher & Schuell). Glycogen (2 µg; Boehringer-Mannheim) was added prior to ethanol precipitation to serve as a carrier. The DNA was resuspended in 20 µl of water, and 20 µl of 2 M piperidine was added, followed by incubation at 90°C for 30 min. The solution was then frozen and vacuum-dried. The residue was resuspended repeatedly in water (200, 100, 50, 25, 15, 10, 10 µl) and vacuum-dried, before being resuspended in loading buffer and analyzed on a 7% polyacrylamide sequencing gel.

Screening cDNA libraries

Ovarian (Steinhauer et al. 1989) and 9- to 12-hr embryonic (Zinn et al. 1988) cDNA libraries in λgt11 were screened as in Vinson et al. (1988), with the following modifications. Plates were incubated overnight after overlaying with the IPTG-impregnated nitrocellulose filters. BSA was used in place of Carnation nonfat dry milk. The filters were exposed to the DNA probe overnight with the addition of salmon sperm DNA (10 µg/ml) to the binding buffer. The filters were washed over a period of 30 min with six changes of binding buffer plus 0.25% BSA. CF1 was isolated from the ovarian cDNA library. CF2 and CF3 were obtained from the embryonic library.

Sequences for the H and U oligonucleotides were as follows (dots denote wild-type *s15* sequence):

```
H:      -80                               -42
          *                               *
      GATCTGAACA.....AGGCTG
      ACTTGT.....TCCGACCTAG
```

```
U:      -120                             -76
          *                             *
      GATCCAGCCG.....AGAACA
      GTCGGC.....TCTTGTCTAG
```

Preparation of bacterial extract

EcoRI fragments (consisting of the cDNA insert) were obtained from plaque-purified phage and subcloned into the translation vector pET-3b (Rosenberg et al. 1987). *Escherichia coli* BL21 cells were then infected with phage and induced to produce protein off the T7 promoter, and total bacterial protein was subsequently isolated as in Desplan et al. (1985).

Sequencing analysis

Restriction fragments were cloned into M13 vectors and sequenced by the chain-termination method (Sanger et al. 1977) by using ³⁵S-labeled dATP. Data were analyzed by computer programs [Pustell and Kafatos 1982, 1984].

Acknowledgments

We are grateful to Michael Ashburner for in situ mapping of CF1, CF2, and CF3. Athanassia Economou for oligonucleotide synthesis, Laura Kalfayan and Kai Zinn for gifts of expression libraries, Janet Deatrick and Neel Randsholt for a gift of fly stocks and clones from the chromosome walk of J.M. Dura et al., Marta Frisardi for localizing CF1 on that walk, and Dean Falb for jointly screening the λgt11 expression libraries. We thank Tony Oro, Ron Evans, Vince Henrich, and Larry Gilbert for exchange of information, Xi He for pointing out the similarity between ERE and the CF1-binding site, and Paul Young, Dean Falb, Ted Abel, Peter Tolias, Michael Koelle, Lucy Cherbas, and Peter Cherbas for helpful discussions. We acknowledge the photography of Bianca Klumpa and the secretarial assistance of Esther Fenerjian. This work was supported by grants from the National Institutes of Health and the National Science Foundation.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Note added in proof

The sequence of full-length, independently isolated CF1 cDNA clone, and comparison with its closest homolog, the mouse H-2RIIBP component [Hamada et al. 1989, *Proc. Natl. Acad. Sci.* **86**: 8289–8293] are presented in Henrich et al. [*Nucleic Acids Res.*, in press]. Analysis of the recently published sequence of the *Drosophila* E75 protein [Segraves and Hogness 1990, *Genes Dev.* **4**: 204–219] reveals a 56% sequence identity to CF1 in the DNA-binding domain. Sequence data reported in this paper have been submitted to the EMBL/GenBank Data Libraries.

References

- Baker, A.R., D.P. McDonnell, M. Hughes, T.M. Crisp, D.J. Mangelsdorf, M.R. Haussler, J.W. Pike, J. Shine, and B. O'Malley. 1988. Cloning and expression of full-length cDNA encoding human vitamin D receptor. *Proc. Natl. Acad. Sci.* **85**: 3294–3298.
- Beato, M. 1989. Gene regulation by steroid hormones. *Cell* **56**: 335–344.
- Belyaeva, E.S., I.E. Vlassova, Z.M. Biyasheava, V.T. Kakpakov, G. Richards, and I.F. Zhimulev. 1981. Cytogenetic analysis of the 2B3-4–2B11 region of the X chromosome of *Drosophila melanogaster*. II. Changes in 20-OH ecdysone puffing caused by genetic defects of puff 2B5. *Chromosoma* **84**: 207–219.
- Berg, J. 1988. Proposed structure for the zinc-binding domains from transcriptional factor IIIA and related proteins. *Proc. Natl. Acad. Sci.* **85**: 99–102.
- . 1989. DNA binding specificity of steroid receptors. *Cell* **57**: 1065–1068.
- Carter, P., H. Bedouelle, and G. Winter. 1985. Improved oligonucleotide site-directed mutagenesis using M13 vectors. *Nucleic Acids Res.* **13**: 4431–4443.
- Chowdhury, K., U. Deutsch, and P. Gruss. 1987. A multigene family encoding several finger structures is present and differentially active in mammalian genomes. *Cell* **48**: 771–778.
- Debuire, B., C. Henry, M. Benaisse, and G. Biserte. 1984. Sequencing the *erb A* gene of avian erythroblastosis virus reveals a new type of oncogene. *Science* **224**: 1456–1459.
- Desplan, C., J. Theis, and P.H. O'Farrell. 1985. The *Drosophila* developmental gene, *engrailed*, encodes a sequence-specific DNA binding activity. *Nature* **318**: 630–635.
- Dura, J.M., N.B. Randsholt, J. Deatrick, I. Erk, P. Santamaria, S.J. Freeman, D. Weddell, and H.W. Brock. 1987. A complex genetic locus, *polyhomeotic*, is required for segmental specification and epidermal development in *D. melanogaster*. *Cell* **51**: 829–839.
- Evans, R.M. 1988. The steroid and thyroid hormone receptor superfamily. *Science* **240**: 889–895.
- Fenerjian, M.G., J.C. Martinez-Cruzado, C. Swimmer, D. King, and F.C. Kafatos. 1989. Evolution of the autosomal chorion locus in *Drosophila*. II. Chorion gene expression and sequence comparisons of the *s16* and *s19* genes in evolutionarily distant species. *J. Mol. Evol.* **29**: 108–125.
- Fried, M. and D. Crothers. 1981. Equilibrium and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **8**: 6505–6525.
- Giguere, V., E.S. Ong, P. Segui, and R.M. Evans. 1987. Identification of a receptor for the morphogen retinoic acid. *Nature* **330**: 624–629.
- Green, S., V. Kumar, I. Theulaz, W. Wahli, and P. Chambon. 1988. The N-terminal DNA-binding zinc finger of the oestrogen and glucocorticoid receptors determines target gene specificity. *EMBO J.* **7**: 3037–3044.
- Greene, G.L., P. Gilna, M. Waterfield, A. Baker, Y. Hort, and J. Shine. 1986. Sequence and expression of human estrogen receptor complementary DNA. *Science* **231**: 1150–1154.
- Heberlein, U., B. England, and R. Tjian. 1985. Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. *Cell* **41**: 965–977.
- Hendrickson, W. and R. Schleif. 1985. A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. *Proc. Natl. Acad. Sci.* **82**: 3129–3133.
- Hollenberg, S.M., C. Weinberger, E.S. Ong, G. Cerelli, A. Oro, R. Lebo, E.B. Thompson, M.G. Rosenfeld, and R.M. Evans. 1985. Primary structure and expression of a functional glucocorticoid receptor cDNA. *Nature* **318**: 635–641.
- Kadonaga, J.T., K.R. Carner, F.R. Masiarz, and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**: 1079–1090.
- King, R.C. 1970. *Ovarian development in Drosophila melanogaster*. Academic Press, New York.
- Klock, G., U. Strahle, and G. Schutz. 1987. Oestrogen and glucocorticoid responsive elements are closely related but distinct. *Nature* **329**: 734–736.
- Köster, M., T. Pieler, A. Poting, and W. Knochel. 1988. The finger motif defines a multigene family represented in the maternal mRNA of *Xenopus laevis* oocytes. *EMBO J.* **7**: 1735–1741.
- Levine, J. and A. Spradling. 1985. DNA sequence of a 3.8 kilobase pair region controlling *Drosophila* chorion gene amplification. *Chromosoma* **92**: 136–142.
- Mariani, B.D., J.R. Lingappa, and F.C. Kafatos. 1988. Temporal regulation in development: Negative and positive *cis*-regulators dictate the precise timing of expression of a *Drosophila* chorion gene. *Proc. Natl. Acad. Sci.* **85**: 3029–3033.
- Martinez, E., F. Givel, and W. Wahli. 1987. The estrogen-re-

- sponsive element as an inducible enhancer: DNA sequence requirements and conversion to a glucocorticoid-responsive element. *EMBO J.* **6**: 3719–3727.
- Martinez-Cruzado, J.C., C. Swimmer, M.G. Fenerjian, and F.C. Kafatos. 1988. Evolution of the autosomal chorion locus in *Drosophila* I General organization of the locus and sequence comparisons of genes *s15* and *s19* in evolutionarily distant species. *Genetics* **119**: 663–667.
- Meyerowitz, E.M., M.A. Crosby, M.D. Garfinkel, C.H. Martin, P.H. Mathers, and K. Vijayraghavan. 1985. The 68 glue puff of *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 347–353.
- Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**: 371–378.
- Mitsialis, S.A., S. Veletzka, and F.C. Kafatos. 1989. Transgenic regulation of moth chorion gene promoters in *Drosophila*: Tissue, temporal and quantitative control of four bidirectional promoters. *J. Mol. Evol.* **29**: 486–495.
- Miyajima, N., Y. Kadowaki, S. Fukushige, S. Shimizu, K. Semba, Y. Yamanashi, K. Matsubara, K. Toyoshima, and T. Yamamoto. 1988. Identification of two novel members of the *erbA* superfamily by molecular cloning: The gene products of two are highly related to each other. *Nucleic Acids Res.* **16**: 11057–11074.
- Mlodzik, M., Y. Hiromi, U. Weber, C.S. Goodman, and G.M. Rubin. 1990. The *Drosophila* *seven-up* gene, a member of the steroid receptor gene superfamily, controls photoreceptor cell fates. *Cell* **60**: 211–224.
- Moses, K., M.C. Ellis, and G.M. Rubin. 1989. The *glass* gene encodes a zinc-finger protein required by *Drosophila* photoreceptor cells. *Nature* **340**: 531–534.
- Nauber, U., M.J. Pankratz, A. Kienlin, E. Seifert, U. Klemm, and H. Jäckle. 1988. Abdominal segmentation of *Drosophila* embryo requires a hormone receptor-like protein encoded by the gap gene *knirps*. *Nature* **336**: 489–491.
- Oro, A.E., E.S. Ong, J.S. Margolis, J.W. Posakony, M. McKeown, and R.M. Evans. 1988. The *Drosophila* gene *knirps-related* is a member of the steroid-receptor gene superfamily. *Nature* **336**: 493–496.
- Orr, W.C., V.K. Galanopoulos, C.P. Romano, and F.C. Kafatos. 1989. A female sterile screen of the *Drosophila melanogaster* X chromosome using hybrid dysgenesis: Identification and characterization of egg morphology mutants. *Genetics* **122**: 847–858.
- Petri, W.H., A.R. Wyman, and F.C. Kafatos. 1977. Mass isolation of *Drosophila* follicles and eggshells. *The genetics and biology of Drosophila*, vol. 2 (ed. M. Ashburner and T. Wright) pp. 136–140. Academic Press, London.
- Postlethwait, J.H., M. Bownes, and T. Jowett. 1980. Sexual phenotype and vitellogenin synthesis in *Drosophila melanogaster*. *Dev. Biol.* **79**: 379–387.
- Pustell, J. and F.C. Kafatos. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. *Nucleic Acids Res.* **12**: 643–655.
- . 1982. A high speed, high capacity homology matrix: Zooming through SV40 and polyoma. *Nucleic Acids Res.* **10**: 4765–4782.
- Racioppi, J.V., R.M. Gemmill, P.H. Kogan, J.M. Calvo, and H.H. Hagedorn. 1986. Expression and regulation of vitellogenin messenger RNA in the mosquito, *Aedes aegypti*. *Insect Biochem.* **16**: 255–262.
- Romano, C.P., B. Bienz-Tadmor, B.D. Mariani, and F.C. Kafatos. 1988. Both early and late *Drosophila* chorion gene promoters confer correct temporal, tissue and sex specificity on a reporter *Adh* gene. *EMBO J.* **7**: 783–790.
- Rosenberg, A.H., B.N. Lade, D. Chui, S. Lin, J.J. Dunn, and F.W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**: 125–135.
- Rosenberg, U.B., C. Schroder, A. Preiss, A. Kienlin, S. Cote, I. Riede, and H. Jäckle. 1986. Structural homology of the product of the *Drosophila* Krüppel gene with *Xenopus* transcription factor IIIA. *Nature* **319**: 336–339.
- Rothe, M., U. Nauber, and H. Jäckle. 1989. Three hormone receptor-like *Drosophila* genes encode an identical DNA-binding finger. *EMBO J.* **8**: 3087–3094.
- Ruiz i Altaba, A., H. Perry-O'Keefe, and D.A. Melton. 1987. *Xfin*: An embryonic gene encoding a multifingered protein in *Xenopus*. *EMBO J.* **6**: 3065–3070.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
- Spoerel, N., H.T. Nguyen, and F.C. Kafatos. 1986. Gene regulation and evolution in the chorion locus of *B. mori*: Structural and developmental characterization of four eggshell genes and their flanking DNA regions. *J. Mol. Biol.* **190**: 23–35.
- Steinhauer, W., R. Walsh, and L. Kalfayan. 1989. Sequence and structure of the *Drosophila melanogaster* ovarian tumor gene and generation of an antibody specific for the ovarian tumor protein. *Mol. Cell. Biol.* **12**: 5726–5732.
- Tautz, D., R. Lehmann, H. Schnurch, R. Schuh, E. Seifert, A. Kienlin, K. Jones, and H. Jäckle. 1987. Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. *Nature* **327**: 383–389.
- Umesono, K. and R.M. Evans. 1989. Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* **57**: 1139–1146.
- Vincent, A., H.V. Colot, and M. Rosbash. 1985. Sequence and structure of the *Serendipity* locus of *Drosophila melanogaster*: A densely transcribed region including a blastoderm-specific gene. *J. Mol. Biol.* **186**: 149–166.
- Vinson, C.R., K.L. LaMarco, P.F. Johnson, W.H. Landschulz, and S.L. McKnight. 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* **2**: 801–806.
- Wang, L., S.Y. Tsai, R.G. Cook, W.G. Beattie, M. Tsai, and B.W. O'Malley. 1989. COUP transcription factor is a member of the steroid receptor superfamily. *Nature* **40**: 163–166.
- Weinberger, C., C.C. Thompson, E.S. Ong, R. Lebo, D.J. Gruol, and R.M. Evans. 1986. Three *c-erb-A* gene encodes a thyroid hormone receptor. *Nature* **324**: 641–646.
- Wong, Y.-C., J. Pustell, N. Spoerel, and F.C. Kafatos. 1985. Coding and potential regulatory sequences of a cluster of chorion genes in *Drosophila melanogaster*. *Chromosoma* **92**: 124–135.
- Wu, C. 1984. Activating protein factor binds *in vitro* to upstream control sequences in heat shock gene chromatin. *Nature* **311**: 81–84.
- Zinn, K., L. McAllister, and C. Goodman. 1988. Sequence analysis and neuronal expression of fasciclin in grasshopper and *Drosophila*. *Cell* **53**: 577–587.

Proteins that bind to *Drosophila* chorion cis-regulatory elements: a new C2H2 zinc finger protein and a C2C2 steroid receptor-like component.

M J Shea, D L King, M J Conboy, et al.

Genes Dev. 1990, **4**:

Access the most recent version at doi:[10.1101/gad.4.7.1128](https://doi.org/10.1101/gad.4.7.1128)

References

This article cites 56 articles, 13 of which can be accessed free at:
<http://genesdev.cshlp.org/content/4/7/1128.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).