

Genome-wide Mapping of *in Vivo* Targets of the *Drosophila* Transcription Factor Krüppel*[§]

Received for publication, March 25, 2004, and in revised form, May 5, 2004
Published, JBC Papers in Press, May 6, 2004, DOI 10.1074/jbc.M403345200

Alexey Matyash, Ho-Ryun Chung[‡], and Herbert Jäckle[§]

From the Max-Planck-Institut für biophysikalische Chemie, Abteilung Molekulare Entwicklungsbiologie, D-37070 Göttingen, Germany

Krüppel (Kr), a member of the gap class of *Drosophila* segmentation genes, encodes a DNA binding zinc finger-type transcription factor. In addition to its segmentation function at the blastoderm stage, Krüppel also plays a critical role in organ formation during later stages of embryogenesis. To systematically identify *in vivo* target genes of Krüppel, we isolated DNA fragments from the Krüppel-associated portion of chromatin and used them to find and map Krüppel-dependent cis-acting regulatory sites in the *Drosophila* genome. We show that Krüppel binding sites are not enriched in Krüppel-associated chromatin and that the clustering of Krüppel binding sites, as found in the cis-acting elements of Krüppel-dependent segmentation genes used for *in silico* searches of Krüppel target genes, is not a prerequisite for the *in vivo* binding of Krüppel to its regulatory elements. Results obtained with the newly identified target gene *ken* and *barbie* (*ken*) indicate that Krüppel represses transcription and thereby restricts the spatial expression pattern of *ken* during blastoderm and gastrulation.

The *Drosophila* segmentation gene *Krüppel* (*Kr*)¹ participates in the subdivision of the embryo into increasingly smaller segment equivalents along the anterior-posterior axis (1, 2). It encodes a transcription factor that contains a DNA binding domain composed of five C2H2-type zinc finger motifs (3). *Kr* activity controls the localized expression of other segmentation genes (4) that are required for the establishment of thoracic and anterior abdominal segments (for review, see Ref. 5). The *Kr* protein (Krüppel) acts as a *Drosophila* C terminus-binding protein-dependent transcriptional repressor (Refs. 6 and 7 and references therein) and as an activator (8–13), and it can maintain gene expression activated by other transcriptional regulators (14).

Kr is initially expressed at the syncytial blastoderm stage and, subsequently, in a number of spatially and temporally restricted patterns throughout embryogenesis. The *Kr* expression patterns are controlled by an 18-kb cis-acting upstream regulatory region composed of separable and partially redundant cis-acting modules (see Refs. 15 and 16). They are neces-

sary and sufficient for initial *Kr* expression in the anterior, central, and posterior regions of the blastoderm embryo, expression of *Kr* in distinct sets of muscle and neural precursor cells, in the developing kidney-like Malpighian tubules, the amnioserosa, and the larval light sensory system, called Bolwig's organ (15).

Studies concerning the biological function of Krüppel were focused on its role in segmentation and on target genes that are controlled by Krüppel during early body pattern formation (7, 17–26). In addition, a few Krüppel-regulated genes have been identified on the basis of altered gene expression patterns in *Kr* mutant embryos, by *in vitro* studies showing that Krüppel binds to the respective cis-acting control elements (14, 27, 28), and by genetic modifier screens involving the dominant *Kr* mutation *Irregular facets* (*If*) (29, 30).

To systematically assess target genes of Krüppel that are expressed between early and mid stages of embryogenesis, we isolated DNA fragments from the Krüppel-associated chromatin of embryos that were collected during early and mid-stages of embryogenesis (0–14 h after egg deposition). We used endogenously expressed FLAG-tagged Krüppel protein to isolate this chromatin fraction by immunoprecipitation (31), cloned the associated DNA fragments, and mapped them to the *Drosophila* genome. We present an initial screen in which we identified 82 putative Krüppel target DNA fragments of which more than half were examined with respect to enrichment in Krüppel-associated chromatin and Krüppel binding properties *in vitro*. We show that one of the Krüppel target genes that was identified in this screen, *ken* and *barbie* (*ken*) (32), is regulated in a *Kr*-dependent manner.

EXPERIMENTAL PROCEDURES

Fly Strains and Construction of Transgenes Expressing Tagged Krüppel—Flies were cultured under standard conditions using Oregon R as a wild type strain. Homozygous *Kr* mutant embryos were identified by the absence of lacZ activity in the *y,w*; *Sco*/*Cyo*, P[*hb-lacZ*] *Kr1*/*SM5* strain. The *y,w*; *Ly*/*TM3* line was used for balancing the *FX(10.7)-Kr-2F* transgene integrated into the 3rd chromosome. *hs-Kr*/*Cyo*, P[*hb-lacZ*] strain (13) was used to induce ectopic *Kr* expression in response to heat-shock treatment.

The double-tagged *FX(10.7)-Kr-2F* transgene included 10.7 kb of the *Kr* upstream region, 1.9 kb of the transcribed region, and 1.45 kb of downstream DNA. It was generated by fusing five separate DNA fragments. They were separately amplified from *Drosophila* Oregon R DNA by PCR (Stratagene, La Jolla, CA) using the primers listed in Supplemental Table 1 (for their location see Fig. 1). The amplified fragments were ligated in a way that the Krüppel wild type sequence was restored but extended by the FLAG octapeptide DYKDDDDK (58). In addition, a 34-bp module made of a pair of the self-annealing 30-mers, SphI_F_up/SphI_F_bot (Supplemental Table 1), was inserted to add a second FLAG epitope (position 470–477), resulting in a putative 520-amino acid protein with two FLAG-epitopes that are separated by 35 amino acid residues. White flies were transformed (33) with the 14-kb-long *FX(10.7)-Kr-2F* gene inserted into the pP(CaSpeR-4) vector (59).

* This work was supported by the Max Planck Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[§] The on-line version of this article (available at <http://www.jbc.org>) contains Supplemental Fig. 1 and Tables 1 and 2.

[‡] A predoctoral fellow of the Boehringer Ingelheim Fonds.

[§] To whom correspondence should be addressed. Tel.: 49-551-201-1482/3; Fax: 49-551-201-1755; E-mail: hjaeckl@gwdg.de.

¹ The abbreviations used are: *Kr*, *Krüppel*; kb, kilobase(s); GST, glutathione S-transferase; STAT, signal transducers and activators of transcription.

Production of Glutathione S-Transferase (GST)-tagged Krüppel and Antibodies—Full-length Krüppel cDNA (1509 bp) was obtained from the *Drosophila* early embryonic cDNA library by PCR using the primer pair 5'-CTTTAGAATTCCATATCAATGCTTCAAGAC-3' (forward primer) and 5'-TCACTCGAGCTAATGTTGTTGATGGCCAT-3' (backward primer). They contained EcoRI (forward primer) and XhoI (backward primer) sites to facilitate cloning into pGEX-4T-3 (Amersham Biosciences) so that the Krüppel sequence extends the N-terminally positioned GST via the thrombin-sensitive linker. GST-(N)Kr was expressed in BL21 bacteria cells (Invitrogen) induced with 0.4 mM isopropyl β -D-thiogalactopyranoside (4 h; 30 °C). The 83-kDa fusion protein was purified (60), solubilized, and bound to GSH-agarose beads (Sigma) (4 °C; 2 h). GST-(N)Kr was eluted from beads by 15 mM reduced glutathione.

For chromatin immunoprecipitation, Western blots, and *in situ* detection of protein, we used anti-FLAG M2 mouse monoclonal antibodies (Sigma), anti-GST rabbit polyclonal antibody (Amersham Biosciences), and alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche Applied Science). Horseradish peroxidase-conjugated anti-mouse goat polyclonal antiserum (Amersham Biosciences) was employed as a secondary antibody.

Chromatin Immunoprecipitation and Cloning of the Associated DNA—Chromatin was prepared from 5 g of 0–14 h *FX(10.7)-Kr-2F* embryo as described by Cavalli *et al.* (31). Cross-linking was performed with 2% (v/v) formaldehyde (15 min of incubation at room temperature with vigorous shaking). After stopping the reaction (adding 0.125 M glycine and 5 min of incubation) embryos were pelleted and extensively washed (phosphate-buffered saline), and chromatin was fragmented by sonication (Sonifier 250; Branson, Fürth, Germany) into 0.2–3-kb DNA fragments (average size of about 1 kb). DNA-protein complexes were purified by CsCl density gradient centrifugation and dialyzed (overnight in Tris-EDTA buffer, Slide-A-Lyzer[®] dialysis cassette, 6000–8000 molecular weight cut off; Pierce; yield, 50–70 μ g/g of embryos). Immunoprecipitation was carried out with 13-ml chromatin fractions that were incubated (1 h; 4 °C) with 200 μ l of protein G-agarose resin (Invitrogen). Agarose beads were removed, and chromatin was incubated overnight (4 °C) with 10 μ g/ml anti-FLAG M2 mouse monoclonal antibody (antibody-immunoprecipitation fraction) or without antibody (total chromatin fraction). 200 μ l of protein G-agarose beads were newly added to each fraction (3 h; 4 °C), harvested by centrifugation, and extensively washed. Resin-bound chromatin was treated with RNase, proteinase K, and 0.5% SDS followed by a phenol/chloroform (1:1, v/v) extraction. Alternatively, DNA-protein complexes were eluted by incubation with 150 μ g/ml FLAG peptide. After phenol/chloroform extraction, chromatin-associated DNA was ethanol-precipitated in the presence of 20 μ g of glycogen (10 ng of DNA/1 g of embryos). DNA was treated with Klenow enzyme and ligated with dephosphorylated pCR[®]4Blunt-TOPO[®] (Invitrogen) vector DNA. Recombinant plasmid DNA was isolated (Bio Robot 9600; Qiagen, Hilden, Germany) followed by size determination and sequencing of the inserts. 30% of the antibody-immunoprecipitated DNA and total chromatin DNA were used for Southern blot analysis. Linkers were added to the DNA fragment isolates, and they were amplified by PCR using a primer that covers the linker. The PCR-derived material (several μ g for each fraction) was affinity-purified by incubation (3 h, 4 °C) with GST-(N)Kr-containing resin (50 μ l of GSH-agarose resin bound to 10 μ g of the Krüppel fusion protein), extensive washing, and elution (1 M NaCl). Eluted DNA was precipitated (see above), PCR-amplified (linker primers), and used for ³²P-labeling (Rediprime[™] random primer labeling kit; Amersham Biosciences) to either obtain molecular probes for Krüppel-associated chromatin or total chromatin for Southern blot analysis.

Multiplex Semiquantitative PCR and Southern Blot Analysis—MatInspector V.2.2 software tool (61) was used to identify Krüppel binding sites within the cloned DNA using the Krüppel consensus matrix (62). Sequence matches were quality-based-filtered so that only matches scoring $\geq 80\%$ similarity to the Krüppel matrix and $\geq 75\%$ similarity to the Krüppel core binding sequence were left. Primers were designed (Oligo 4.0 software tool; Molecular Biology Insights, Inc.) to amplify 300–600-bp-long DNA fragments that contain the identified Krüppel binding sites.

Multiplex PCR was performed with the HotStarTaq[™] Master Mix kit (Qiagen) using primer sets for 23 different DNA fragments (1 ng each) of the immunoprecipitated chromatin. The following PCR conditions were used: 95 °C (14 min), 75 °C (2 min), 55 °C (1 min), 71 °C (100 s), an additional 32 cycles of 94 °C (45 s), 55 °C (1 min), 71 °C (100 s), followed by a final extension at 71 °C (10 min). The PCR fragments were size-fractionated on polyacrylamide gels and stained with EtBr

followed by image development and signal quantification (if necessary) with Lumi-imager[™] (Roche Applied Science).

For Southern blot analysis, 1 μ g of DNA from each clone was digested with EcoRI, separated on a 1% agarose gel (0.8 \times Tris-buffered EDTA), transferred onto the Hybond[™]-N+ membranes (Amersham Biosciences), hybridized in Rapid-hyb[™] buffer (Amersham Biosciences), and washed as per the manufacturer's instructions. Probes were prepared by [α -³²P]CTP labeling of PCR-amplified DNA (5 ng/ml) (see above). Signals were developed in PhosphorImager cassettes and quantified by PhosphorImager[™] (Molecular Dynamics, Krefeld, Germany).

In Vitro DNA Binding Assay and Gel Shift Assay—1 μ g of plasmid DNA was digested by EcoRI and separated on a 1% agarose gel. DNA fragments were extracted (agarose gel extraction kit; Qiagen). Fragments of different sizes were pooled and ³²P-labeled by T4 polynucleotide kinase (Fermentas, St. Leon-Rot, Germany). 500 ng of labeled DNA was mixed with 30 μ l of GSH-agarose beads coupled with 10 μ g of GST-(N)Kr and incubated for 20 min at 25 °C in the Zn²⁺-containing binding buffer (20 mM Tris, pH 7.5, 100 mM NaCl, 0.5 mM ZnSO₄, 0.1% Triton X-100, 1 mM dithiothreitol, 4% glycerol). Unbound fragments were washed off in 500 μ l of binding buffer. Fragments were eluted in a series of washes with 500 μ l of binding buffer (100–1000 mM NaCl). The ³²P-labeled DNA fragments were precipitated and separated by electrophoresis in 5% polyacrylamide gel, and the signals were quantified by PhosphorImager. Exponential regression graphs were made for 4 elution points between 200 and 800 mM NaCl.

Gel shift assays were performed with self-complementary DNA oligomers of defined sequences that were made double-stranded by heating (95 °C, 5 min) and subsequent cooling to room temperature (3 h) in RE reaction buffer 2 (New England Biolabs, Frankfurt am Main, Germany). They were ³²P-labeled by a Klenow fill-in reaction (Roche Applied Science). The binding reaction mix included the ³²P-labeled-specific DNA (0.5 nmol/ μ l), unlabeled nonspecific DNA competitor poly-(dI:dC) (10 ng/ μ l; Amersham Biosciences), and the GST-(N)Kr fusion protein (0.5 ng/ μ l). The binding reaction was performed in Zn²⁺-containing electrophoretic mobility shift assay buffer (20 mM HEPES, pH 7.9, 40 mM KCl, 1.4 mM MgCl₂, 0.3 mM ZnSO₄, 0.1 mM EGTA, 0.5 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 5% glycerol in a 30- μ l volume for 30 min at room temperature) before electrophoresis. For competition assay, a 200-fold molar excess of either specific or nonspecific competitor DNA was added to the reaction mix (30-min incubation at room temperature). DNA-protein complexes were resolved on 6% native polyacrylamide gels (0.8 \times Tris borate buffer, pH 8.5) lacking EDTA. After drying, gels were quantified by phosphorimaging.

In Silico Analysis of the DNA Fragments and RNA in Situ Hybridization—The positions of the isolated fragments were determined by blast searches (63) against the *Drosophila melanogaster* genome (Release 3 according to Ref. 37). The gene that had the smallest distance between the transcription start site and the midpoint of the fragment was assigned as a putative target gene. The positions were determined using Release 3.1 of the *D. melanogaster* genome. Krüppel binding sites were scored according to 38 using either the combined sequence information of the chromatin-associated DNA fragments or the euchromatic portion of *D. melanogaster* genome (Release 3).

RNA probes were prepared from plasmids containing *ken*- and *lacZ*-coding regions using the DIG-labeling kit (Roche Applied Science). *In situ* hybridizations to whole-mount preparations of embryos were performed as described (64).

RESULTS

Generation of a FLAG-tagged Krüppel Protein and Its Activity in Vivo—To isolate DNA fragments from Krüppel-associated chromatin, we marked this chromatin fraction by endogenous expression of a functional Krüppel protein that contains two FLAG tags. The *Kr*-FLAG fusion gene (*FX(10.7)-Kr-2F*; Fig. 1A) contains the *Kr* cDNA, 2 FLAG epitope sequences, 1.45 kb of *Kr* downstream, and 10.7 kb of *Kr* upstream DNA. It was cloned into the P-element vector pP(CaSpeR-4) (33) and used to transform Oregon R wild type flies. The pattern of FLAG-tagged Krüppel (Fig. 1B) shows that the transgene expresses a properly distributed protein in early embryos.

Individuals that are homozygous for the *FX(10.7)-Kr-2F* transgene-bearing chromosome develop normally. Furthermore, one copy of the *FX(10.7)-Kr-2F* transgene rescues the segmentation phenotype of homozygous *Kr^l* lack-of-function mutant embryos, indicating that FLAG-tagged Krüppel is func-

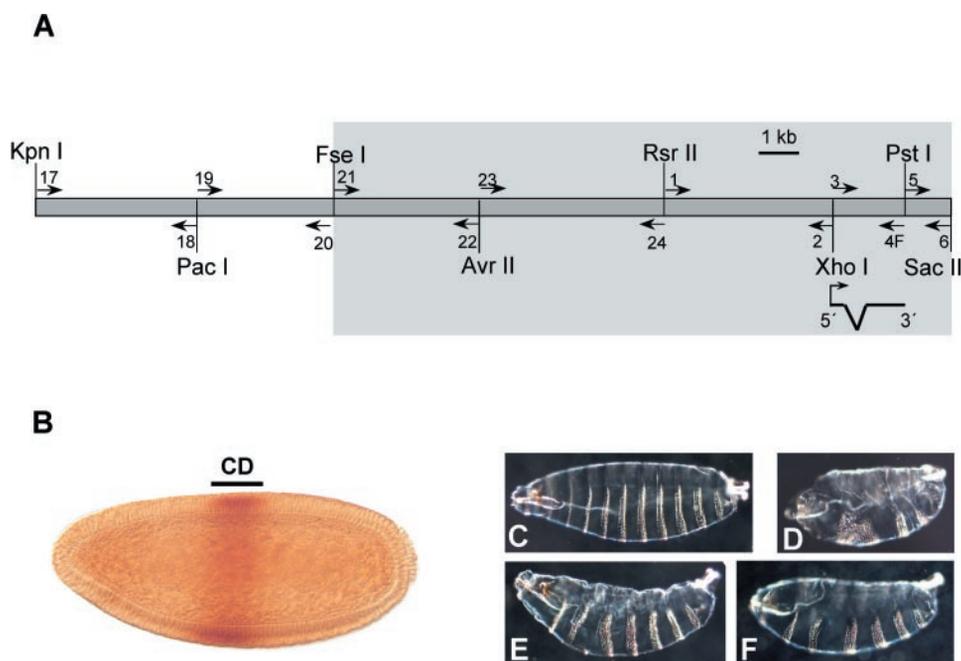


FIG. 1. A, schematic representation of the *Krüppel* locus including its regulatory region as described in Ref. 15. Arrows indicate the position of primers (see the supplemental material) used to amplify individual DNA fragments that were ligated to obtain the DNA of the *Krüppel* wild type gene (KpnI/SacI DNA fragment) including 17.6 kb of *Krüppel* upstream sequences (KpnI/XhoI DNA fragment) and the transcript. The transcription start site (arrow) and the *Krüppel* cDNA (note the 5'-3' orientation and the intron) are shown below the bar. Diagnostic restriction sites are indicated. The gray box highlights the *FX(10.7)-Kr-2F* transgene (see "Experimental Procedures"), which lacks a portion of the *Krüppel* control region required for expression in the posterior domain, the anterior pole, and the Malpighian tubules (for details see Ref. 15). B, *FX(10.7)-Kr-2F* transgene expression in the *Kr* central domain (CD) of a blastoderm embryo as revealed by anti-FLAG antibody staining. C, wild type larvae. D, homozygous *Kr*¹ mutant larvae lacking thorax and the anterior abdominal segments (for details, see Ref. 35). E and F, examples of homozygous *Kr*¹ mutant larvae containing one copy of the *FX(10.7)-Kr-2F* transgene. Note that the majority of the larvae (74%) are wild type, whereas the other larvae show mild defects in the thorax-abdomen formation (examples shown in E and F) as seen with heterozygous *Kr*¹ mutants. For details see "Experimental Procedures."

tional (34, 35) (Fig. 1, C–F). However, the rescued embryos fail to hatch since the transgene lacks cis-acting elements necessary for *Kr*-dependent Malpighian tubule development, which is essential for viability (16, 36).

Isolation of Krüppel-associated Chromatin—To isolate *in vivo* target DNA of FLAG-tagged Krüppel, we performed immunoprecipitation experiments using chromatin isolated from formaldehyde-treated *Kr-2F* transgene-expressing embryos (0–14 h after egg deposition) (Ref. 31; for details see "Experimental Procedures" and Supplemental Fig. 1) and monoclonal anti-FLAG M2 antibodies. We extracted and cloned the co-immunoprecipitated DNA, sequenced a total of 104 DNA fragments, and mapped them to the *Drosophila* genome sequence (37). We found 85 unique and 19 repetitive DNA fragments (see Supplemental Table 2). Three of the 85 non-repetitive DNA fragments were present twice, resulting in a total of 82 putative DNA targets of Krüppel.

Previous footprinting studies with Krüppel showed that the functional cis-acting regulatory elements of its target genes contain multiple binding sites (e.g. Ref. 38 and references therein). To test whether Krüppel binding sites are accordingly enriched in the DNA of the Krüppel-associated chromatin fraction, we employed the Cis-analyst program (38). A search for at least five binding sites within a 500-bp stretch of DNA yielded 46 clusters within the *Drosophila* genome; only one of them was found among the 82 DNA fragments (see Supplemental Table 2). Furthermore, a manual search using the Patser program (v3d; Ref. 39) and a position weight matrix of Krüppel binding sites (38) led to an average of 3.02 Krüppel sites per 1,000 bp of the isolated DNA (287 sites with scores of above 4 in 94,930 bp). This number is comparable with the 3.11 Krüppel sites per 1,000 bp (364,601 sites with scores above 4 in 116,914,271 bp) in the euchromatic portion of the *Drosophila*

genome. Thus, DNA fragments isolated from Krüppel-associated chromatin are not selected on the basis of clustered Krüppel binding sites. We next asked whether previously identified Krüppel-dependent cis-acting elements are enriched in Krüppel-associated chromatin.

Functional Krüppel Binding Site Regions Are Enriched—To assay for an enrichment of known Krüppel target DNA, we used multiplex semiquantitative PCR ("MQ-PCR," Ref. 40; for details, see "Experimental Procedures") to amplify DNA fragments containing the "stripe 2 enhancer" of the segmentation gene *even skipped* (*eve*; Ref. 21), a well established target of Krüppel, and of the gene coding for Sec23p, a component of the COP-II protein complex (41) that is not regulated by Krüppel. Both DNA fragments were found in DNA obtained from total chromatin, whereas in the Krüppel-associated chromatin fraction, only the *eve* stripe 2 enhancer DNA was found (Fig. 2A). This result indicates that Krüppel-dependent cis-acting sequences are indeed enriched in Krüppel-associated chromatin. The same result was obtained with 15 of a total of 23 DNA fragments examined (Table I; 6 examples are shown in Fig. 2B). In addition, we tested for a specific enrichment of DNA fragments in Krüppel-associated chromatin by Southern blot hybridization (31, 42) using template DNA from Krüppel-containing chromatin and control chromatin (for details see "Experimental Procedures"). With this technique we found that of 43 DNA fragments examined, 21 were highly enriched (Table I; 9 examples are shown in Fig. 2C).

We finally asked whether Krüppel-associated chromatin DNA contains *in vitro* binding sites for Krüppel. Because the average size of the cloned DNA fragments was around 1 kb, conventional electrophoretic mobility shift assay could not be applied. We, therefore, used an alternative *in vitro* approach suited to detect binding sites in large DNA fragments by spe-

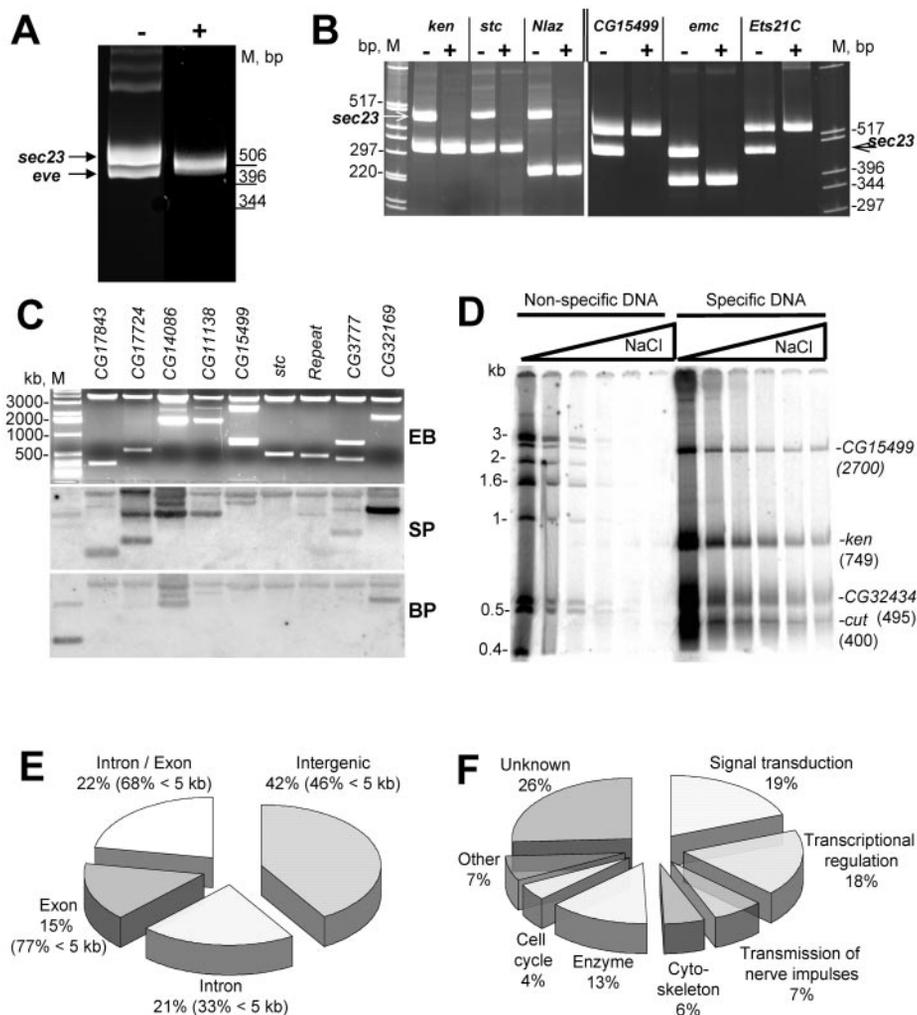


FIG. 2. *A*, amplification of DNA of Krüppel-associated chromatin (+) and total chromatin (-) showing that the *even-skipped* (*eve*) *stripe2* enhancer element (488 bp; Ref. 21) is highly enriched in Krüppel-associated chromatin. Note that due to the large amounts of the *sec23* DNA (-), the DNA band corresponding to the *eve stripe2* enhancer is slightly shifted as compared with its counterpart in the (+) lane. *B*, six PCR-amplified DNA fragments present in Krüppel-associated (+) chromatin and total (-) chromatin. Note that DNA fragments corresponding to the genes *ken*, *stc*, *Nlaz*, *CG15499*, *emc*, and *Ets21C* are enriched in Krüppel-associated chromatin. *C*, nine examples of Southern blot hybridization tests. *EB*, ethidium bromide-stained gel; *SP*, Southern blot probed with labeled DNA fragments obtained from Krüppel-associated chromatin; *BP*, Southern blot probed with labeled DNA fragments obtained from total chromatin. The 3.5-kb fragment visible in all lanes represents vector DNA since no pre-absorption of the probes was done. We used this background hybridization as a control for similar loads and transfer of DNA fragments to the Southern blots. Note the enrichment of DNA fragments in Krüppel-associated chromatin (for details see “Results” and Table I). *Repeat* denotes a fragment of transposon DNA that was isolated but not included as a Krüppel target in Supplemental Table II. *D*, *in vitro* DNA binding assay performed with resin-bound GST-(N)Kr fusion protein. Four differently sized DNA fragments obtained from Krüppel-associated chromatin (*Specific DNA*; 1 nM each) were labeled and incubated with GST-(N)Kr protein-loaded glutathione-agarose. A 1-kb DNA marker (Invitrogen) was used as control (*Nonspecific DNA*). DNA fragments were gradually eluted (the *triangle* refers to increasing salt concentrations ranging from 0.1 to 1 M NaCl) and separated by gel electrophoresis. Note that nonspecific DNA was almost entirely removed from the beads. *E*, localization of 85 isolated DNA fragments relative to the putative Krüppel target genes. *F*, classification of Krüppel target genes according to *Gene Ontology* (37). For details see text and “Experimental Procedures.”

sific retention to resin-bound GST-Krüppel fusion protein (for details see “Experimental Procedures”).² The DNA of 24 clones is able to bind Krüppel *in vitro* (see Fig. 2*D*). This set of clones includes DNA fragments that were enriched in Krüppel-associated chromatin (Table I), indicating that they contain Krüppel binding sites. Clustering of Krüppel *in vitro* binding sites, as observed in the cis-acting regions of Krüppel-dependent segmentation genes, is therefore not a necessity for the *in vivo* targeting of DNA by Krüppel.

Identification of Potential Krüppel Target Genes—Of the 85 isolated DNA fragments, 36 (42%) correspond to intergenic regions, 18 (21%) to introns, 19 (22%) to exon/intron boundaries, and 12 (15%) to exons (Fig. 2*E*) of the assigned Krüppel target genes listed in Supplemental Table 2. This assignment

rests on a linkage of the isolated DNA fragments to the closest transcription start sites. We are aware that due to this arbitrary assignment, the Krüppel binding DNA segment may participate in the cis-acting control of a different, nearby transcription unit. With this caution in mind, we found that the majority of the putative Krüppel target genes (18%) encode transcription factors, a class of genes that represents only 5% of the total *Drosophila* genes (37). Other putative Krüppel targets (see Fig. 2*F*, Supplemental Table 2) encode functionally diverse proteins such as components of cell-cell communication processes including signal transduction pathways (protein kinases, phosphatases, membrane receptors, and ion channels), cell adhesion proteins, and RNA binding factors. Of the 55 known genes that are potentially regulated by Krüppel, 10 participate in embryonic body pattern formation, 28 in neurogenesis and axon guidance, 13 in light sensory organ development, and 4 in

² P. Shaw, personal communications.

TABLE I
Enhancement and DNA binding

Forty-three candidate genes for Krüppel regulation were tested in at least one of the three target-validation assays. The genes are listed according to their molecular function as defined by Gene Ontology (37). Note that the transcription factors (TF) alone account for about 23% of all candidates showing positive response in at least one of the performed tests. Genes coding for cell adhesion and cell communication molecules constitute another class of relatively less abundant putative Kr targets. Positive and negative responses in the tests are designated with + and -, respectively; +/- refers to the experiments with inconclusive results; NA: data not available.

Target gene name	Molecular function or recognizable domain	Enrichment		<i>In vitro</i> binding to Kr
		Multiplex quantitative-PCR	Southern blot	
<i>emc</i>	Specific TF	+	NA	NA
<i>Ets21C</i>	Specific TF	+	+	NA
<i>Hr46</i>	Specific TF	+/-	+	NA
<i>kan</i>	Specific TF	+	NA	+
<i>stc</i>	Specific TF	+	NA	+
<i>tup</i>	Specific TF	+/-	NA	NA
<i>HLHm5</i>	Specific TF	+	+	NA
<i>cut</i>	Specific TF	NA	NA	+
<i>Taf110</i>	General TF	NA	NA	+
<i>osa</i>	DNA binding factor	NA	NA	+
<i>CG6905</i>	Pre-mRNA splicing factor	-	+	+
<i>CG10366</i>	C2H2-type zinc finger	-	NA	+
<i>CG12071</i>	C2H2-type zinc finger	NA	+	+
<i>CG17724</i>	Cell adhesion molecule	+	+	NA
<i>CG14521</i>	Cell adhesion molecule, IgC2	+	NA	NA
<i>sns</i>	Cell adhesion molecule	+	+	NA
<i>shot</i>	Actin binding molecule	+	NA	+
<i>CG31858</i>	Receptor activity	+	NA	NA
<i>CG17129</i>	DNA topoisomerase	-	NA	NA
<i>ast</i>	5'3'-Exonuclease N and I domains	+	+	NA
<i>CG7097</i>	Protein serine/threonine kinase	NA	NA	+
<i>CG12517</i>	Periplasmio-binding protein-like II domain	NA	+	+
<i>CG12007</i>	RAB-protein geranylgeranyltransferase	NA	NA	+
<i>CG9929</i>	Protein-arginine N-methyltransferase	NA	NA	+
<i>Nrv1</i>	Sodium/potassium ATPase	NA	NA	+
<i>ort</i>	Histamine-gated chloride channel	NA	+	NA
<i>CG32434</i>	Sec7 domain	NA	NA	+
<i>CG15429</i>	Cytochrome b5, lipid metabolism	-	NA	NA
<i>NLaz</i>	Lipid binding molecule	+	NA	+
<i>Fer1HCH</i>	Ferrous ion binding	+	NA	+
<i>CG32986</i>	Oxidoreductase activity	NA	+	+
<i>CG17843</i>	Flavin-linked sulfhydryl oxidase	NA	+	NA
<i>1(2)05510</i>	Interacts genetically with pnr	NA	+	NA
<i>CG14086</i>	Unknown	+	NA	+
<i>CG11138</i>	Unknown	-	+	+
<i>CG5868</i>	Unknown	+/-	+	+
<i>CG15499</i>	Unknown	+	+	+
<i>CG30044</i>	Unknown	NA	+	NA
<i>CG32111</i>	Unknown	NA	+	+
<i>CG3777</i>	Unknown	NA	+	NA
<i>CG32169</i>	Unknown	NA	+	NA
<i>CG30377</i>	Unknown	NA	+	NA
<i>CG13881</i>	Unknown	NA	NA	+

muscle development. Thus, the identified putative Krüppel target genes with known functions participate in processes and organs where *Kr* is known to act (15).

ken Is a Krüppel Target—To establish whether the newly identified candidate genes are indeed regulated in a Krüppel-dependent fashion, we focused on *ken*. The reason for this choice was that *ken*, which encodes a DNA binding zinc finger-type transcription factor (43), appears at a first glance unlikely to be a *Kr* target gene. This is because *Kr* activity is not required for male genitalia formation and adult eye development, the two processes in which *ken* is involved (32, 43, 44).³ Secondly, *ken* is expressed early in two stripes that do not overlap with the *Kr* expression domain during blastoderm stage and gastrulation (Refs. 32 and 45; see also below). On the other hand, we found that the isolated 749-bp DNA fragment (Fig. 3A) is highly enriched in the DNA of Krüppel-associated chromatin (Fig. 2B) and that it contains five Krüppel binding sites (see Fig 3B; see also the legend) confirmed by gel mobility shift assays (Fig. 3C).

To solve this apparent dilemma and to thereby demonstrate that our screen has indeed led to Krüppel target genes, we asked whether Krüppel does regulate *ken* expression *in vivo* by performing *in situ* hybridizations of *ken* probes to whole mount preparations of wild type and homozygous *Kr¹* lack-of-function mutant embryos (34, 35). In wild type, Krüppel is initially expressed in a broad band in the central region of the blastoderm (15). In contrast, *ken* is expressed in two distinct stripes that are anteriorly adjacent and posterior to the *Kr* central domain (32). In *Kr* mutant embryos, the two stripes of *ken* expression are not altered (Fig. 4), but we observed an additional expression domain where *Kr* is normally expressed at syncytial blastoderm stage (Fig. 4, A and B). This expression domain appears earlier than the normal stripes of *ken* expression, and it subsequently fades in a posterior to anterior direction, resulting in a third narrow stripe that remains separated from the anterior *ken* stripe (Fig. 4, C and D). These observations establish that in the absence of *Kr* activity, *ken* is activated in the central region of the embryo and that this aspect of *ken* activity is normally repressed in a Krüppel-dependent manner.

³ N. Arbouzova, personal communication.

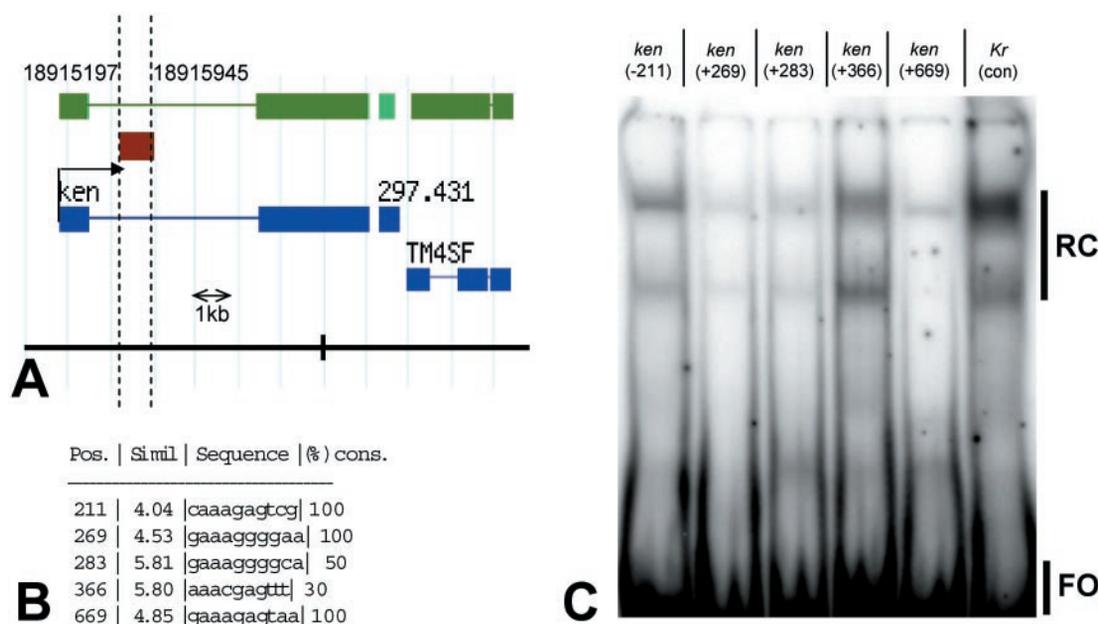


FIG. 3. **A**, the *ken* (749) fragment (red box) is located in the first intron of the *ken* transcription unit. Blue boxes represent the transcribed region of *ken*; green boxes refer to the ill-defined transcription unit, as annotated in Ref. 37, which corresponds to a fusion of *ken* and the unrelated transcription unit TM4SF. Numbers define the position within the *Drosophila* genome. **B**, computer searches revealed five potential Krüppel binding sites in the *ken* (749) fragment that are conserved in the corresponding sequences of the *Drosophila pseudoobscura* genome. Numbers refer to the positions within the 749-bp fragment; 1 is the 5' end. **C**, gel shift assays showing that specific 22-mer oligomers representing the Krüppel binding sites shown in **B** are able to bind Krüppel *in vitro* (RC, retarded complexes; FO, free oligos). Con refers to an oligo that contains the canonical high affinity binding site AAAAGGGGTAA. + and - indicate the upper and the lower strand of DNA, respectively.

Previous results have shown that the expression of the anterior stripe of *ken* is activated in response to the transcription factors encoded by *bicoid* and *hunchback*, whereas the posterior stripe is activated by the transcription factor of *tailless*, and its shape and size are due to repression by Hucklebein (32). To establish whether ectopic expression of Krüppel also causes the repression of *ken*, we used a heat shock-driven *Kr* transgene (13) to misexpress *Kr* uniformly in the blastoderm embryo. Fig. 4, *E-H*, shows that the posterior stripe of *ken* expression is not affected by ectopic *Kr* activity, whereas the anterior *ken* stripe is lacking. Collectively, the results demonstrate that Krüppel participates in early *ken* regulation by acting as a local repressor of the gene in wild type embryos.

DISCUSSION

We reported a pilot screen to identify genes which are regulated by the transcription factor Krüppel. This screen, which revealed 82 potential target genes of Krüppel (see Supplemental Table 2), is possibly far from saturation since only three genomic regions were represented twice among the isolated DNA, and the known early target genes that emerged from hypothesis-driven genetic approaches were not recovered. The latter result is not surprising in view of our experimental bias (using a 0–14 h embryo collection), which was directed at the identification of Krüppel-dependent genes involved in neurogenesis, muscle, and Bolwig organ development (Ref. 29 and references therein). In fact, 55 of the 82 isolated genes are known to participate in these developmental processes. Thus, we expect Krüppel to regulate possibly several hundreds of genes during the entire life cycle of the fly.

Two of the *Kr* target genes (*emc* and *osa*; Table I) were previously identified in a genetic modifier screen for gene products that mediate *Kr* activity (29, 30). In addition, a DNA fragment corresponds to the intron of the gene *CG7097* (46), a putative regulatory target of segmentation genes expressed during blastoderm formation (38). Microarray-based expression data and whole mount *in situ* hybridization of early embryos (www.fruitfly.org/cgi-bin/ex/basic.pl) showed that this

gene as well as additional 29 of the 43 candidate genes listed in Table I are expressed during the first 14 h of embryonic development. These observations and the results of the genetic studies with *ken* indicate that the DNA isolated from Krüppel-associated chromatin revealed *in vivo* target sites of the transcription factor.

Previous analysis has shown that during segmentation Krüppel controls the activity of other transcription factors that are part of a cell fate-determining gene network (5, 47). Our results suggest that this earlier finding is not restricted to *Kr* segmentation function since the majority of the Krüppel target genes identified in this study (18% of the total isolates) encode transcription factors as well. The more important notion is, however, that Krüppel not only participates in the regulation of transcription factor networks at the different levels of the segmentation gene cascade (48) but also assists signaling events by regulating various pathway components, as exemplified by target genes coding for components of the JAK/STAT-signaling pathway. Krüppel target DNA includes portions of the genes *ken*, *STAT92E*, and *stc*, which code for JAK/STAT-mediating transcription factors (Refs. 49 and 50)³ as well as factors known to participate in signaling by the epidermal growth factor receptor (Asteroid; Ref. 51) and Rho GTPases (Gef64C; Ref. 52). Moreover, the isolation of genes encoding lipid metabolism-related enzymes and the lipid carrier Neural Lazarillo (NLaz; Ref. 53) suggests that Krüppel not only takes part in embryonic fat body development (54) but also participates in metabolic functions (fat storage or fat consumption) of the organ.

The majority of the newly isolated Krüppel target sites lack Krüppel binding site clusters as revealed in cis-acting elements of the Krüppel-dependent segmentation genes (4, 7, 21, 25, 55). However, the isolated and subsequently tested set of DNA fragments is enriched in Krüppel-associated chromatin, as has been found with the *eve* stripe 2 element, which contains clustered Krüppel target sites (21). This finding suggests that the clustering of binding sites is not the sole biologically relevant marker for Krüppel-dependent cis-acting control elements.

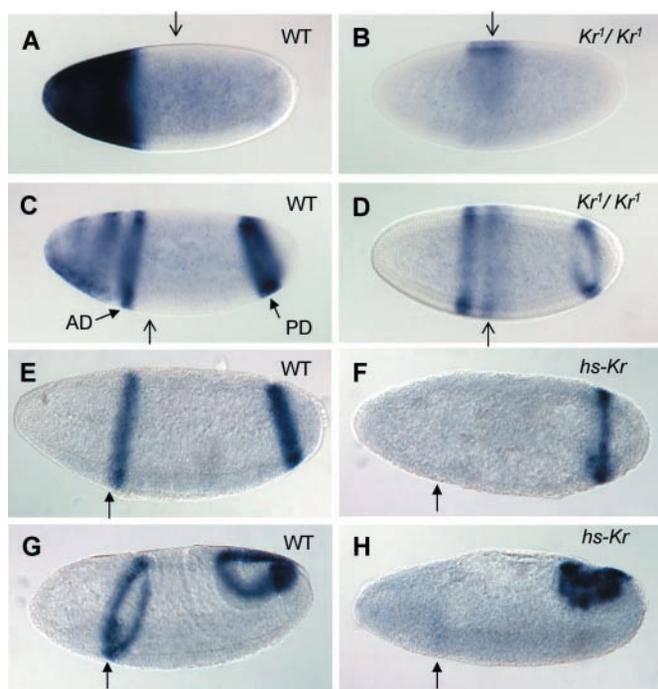


FIG. 4. Expression of *ken* in wild type (WT), homozygous *Kr¹/Kr¹* lack-of-function mutant embryos, and in response to ubiquitous Krüppel activity. The anterior pole is left, dorsal side up. Note that wild type and heterozygous *Kr¹* mutant embryos can be distinguished by a *lacZ* transgene that was placed under the control of the *hunchback* (*hb*) promoter. *A*, early stage wild type blastoderm embryo showing *hb*-dependent LacZ expression in the anterior region. *B*, homozygous *Kr¹* embryo of the corresponding stage showing that *ken* is expressed in the central region of the embryo (arrow). *C*, wild type embryo (late blastoderm) showing the remaining stripes of *hb* expression as well as the anterior (AD) and posterior (PD) of *ken* expression. *D*, homozygous *Kr¹* embryo of the corresponding stage showing that *ken* is expressed in the anterior and posterior domains and that a third stripe of *ken* expression is maintained in the central region of the embryo (arrow). *E* and *F*, *ken* expression in late blastoderm stage wild type embryos (*E*) and in an embryo of the corresponding stage that received ubiquitous Krüppel expression (*hs-Kr*) from a heat shock-dependent transgene (see "Experimental Procedures") (*F*). Note the absence of the *ken* AD (arrow in *E* and *F*). *G* and *H*, *ken* expression in wild type embryos at an early phase of gastrulation (*G*) and in an embryo that received ubiquitous Krüppel expression (*hs-Kr*) (*H*). Note the absence of the *ken* AD (arrow in *G* and *H*). Note that the embryo shown in *G* is slightly tilted to show that the stripe is actually a circumferential ring of *ken* expression.

Furthermore, the algorithm applied to detect Krüppel binding sites only counted matches of sequences to a weighted matrix that were arbitrarily set above a certain threshold. In consequence, functional low affinity binding sites or Krüppel-dependent DNA segments that contain only few and unclustered high affinity binding sites were left undetected (see the Discussion in Ref. 56).

Interestingly, more than half of the Krüppel target DNA fragments (68%) were located in introns and exon/intron overlap sequences or in exons (see Fig. 2E) and not at the canonical 5' termini of protein-coding genes. The location of these fragments downstream of the transcription start sites suggests that they may represent distal regulatory elements (e.g. enhancers or silencers) or promoters for non-coding RNAs, as implied by a most recent study on transcription factor binding along human chromosome 21 and 22 (57). Because noncoding transcripts within the *Drosophila* genome are not systematically annotated, we cannot decide whether Krüppel participates in the transcription of such transcripts.

A surprising result of our study was that *ken*, which is not expressed in the Krüppel domain of wild type blastoderm embryos (32), is in fact a target of Krüppel. In the absence of *Kr*

activity, *ken* is activated in the central region of the blastoderm. Thus, in addition to the regulation of *ken* expression in the anterior and posterior stripe domains, which involves the activities of *bicoid* in cooperation with the gap genes *hunchback*, *tailless*, and *huckebein* (32), Krüppel is needed to prevent ectopic *ken* activation in the blastoderm embryo. This finding and the notion that ubiquitous Krüppel expression abolishes *ken* activity in the anterior but not in the posterior stripe domain suggest that the two stripes of *ken* expression are under the control of separate cis-acting elements, of which only one mediates repression by Krüppel.

Acknowledgments—We thank G. Dowe for sequencing, N. Arbouzova for providing the *ken* RNA probe, and U. Jahns-Meyer for transformations.

REFERENCES

- Nüsslein-Volhard, C., and Wieschaus, E. (1980) *Nature* **287**, 795–801
- Pankratz, M. J., Busch, M., Hoch, M., Seifert, E., and Jäckle, H. (1992) *Science* **255**, 986–989
- Rosenberg, U. B., Schröder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I., and Jäckle, H. (1986) *Nature* **319**, 336–339
- Pankratz, M. J., Seifert, E., Gerwin, N., Billi, B., Nauber, U., and Jäckle, H. (1990) *Cell* **61**, 309–317
- Gaul, U., and Jäckle, H. (1990) *Adv. Genet.* **27**, 239–275
- Nibu, Y., Zhang, H., Bajor, E., Barolo, S., Small, S., and Levine, M. (1998) *EMBO J.* **17**, 7009–7020
- Shimell, M. J., Peterson, A. J., Burr, J., Simon, J. A., and O'Connor, M. B. (2000) *Dev. Biol.* **218**, 38–52
- Sauer, F., and Jäckle, H. (1991) *Nature* **353**, 563–566
- Sauer, F., and Jäckle, H. (1993) *Nature* **364**, 454–457
- Sauer, F., and Jäckle, H. (1995) *EMBO J.* **14**, 4773–4780
- Sauer, F., Fondell, J. D., Ohkuma, Y., Roeder, R. G., and Jäckle, H. (1995) *Nature* **375**, 162–164
- Licht, J. D., Ro, M., English, M. A., Gossel, M., and Hansen, U. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11361–11365
- La Rosee-Borggreve, A., Häder, T., Wainwright, D., Sauer, F., and Jäckle, H. (1999) *Mech. Dev.* **89**, 133–140
- Hartmann, C., Landgraf, M., Bate, M., and Jäckle, H. (1997) *EMBO J.* **16**, 5299–5309
- Hoch, M., Schröder, C., Seifert, E., and Jäckle, H. (1990) *EMBO J.* **9**, 2587–2595
- Hoch, M., Seifert, E., and Jäckle, H. (1991) *EMBO J.* **10**, 2267–2278
- Pankratz, M. J., Hoch, M., Seifert, E., and Jäckle, H. (1989) *Nature* **341**, 337–340
- Treisman, J., and Desplan, C. (1989) *Nature* **341**, 335–337
- Kraut, R., and Levine, M. (1991) *Development* **111**, 611–621
- Zuo, P., Stanojevic, D., Colgan, J., Han, K., Levine, M., and Manley, J. L. (1991) *Genes Dev.* **5**, 254–264
- Small, S., Blair, A., and Levine, M. (1992) *EMBO J.* **11**, 4047–4057
- Gutjahr, T., Frei, E., and Noll, M. (1993) *Development* **117**, 609–623
- Langeland, J. A., Attai, S. F., Vorwerk, K., and Carroll, S. B. (1994) *Development* **120**, 2945–2955
- Casares, F., and Sanchez-Herrero, E. (1995) *Development* **121**, 1855–1866
- La Rosee, A., Häder, T., Taubert, H., Rivera-Pomar, R., and Jäckle, H. (1997) *EMBO J.* **16**, 4403–4411
- Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T., and Jaynes, J. B. (1999) *Development* **126**, 2527–2538
- Hartmann, C., and Jäckle, H. (1997) *Dev. Genes Evol.* **207**, 186–193
- Liu, X., and Lengyel, J. A. (2000) *Dev. Biol.* **221**, 419–434
- Carrera, P., Abrell, S., Kerber, B., Walldorf, U., Preiss, A., Hoch, M., and Jäckle, H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10779–10784
- Abrell, S., Carrera, P., and Jäckle, H. (2000) *Chromosoma (Berl.)* **109**, 334–342
- Cavalli, G., Orlando, V., and Paro, R. (1999) in *Chromosome Structural Analysis: A Practical Approach* (Bickmore, W. A., ed) pp. 20–37, Oxford University Press, Oxford
- Kühnlein, R. P., Chen, C. K., and Schuh, R. (1998) *Mech. Dev.* **79**, 161–164
- Rubin, G. M., and Spradling, A. C. (1983) *Nucleic Acids Res.* **11**, 6341–6351
- Wieschaus, E., Nüsslein-Volhard, C., and Kluding, H. (1984) *Dev. Biol.* **104**, 172–186
- Preiss, A., Rosenberg, U. B., Kienlin, A., Seifert, E., and Jäckle, H. (1985) *Nature* **313**, 27–32
- Gaul, U., Redemann, N., and Jäckle, H. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 4599–4603
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., et al. (2000) *Science* **287**, 2185–2195
- Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M., and Eisen, M. B. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 757–762
- Hertz, G. Z., and Stormo, G. D. (1999) *Bioinformatics* **15**, 563–577
- Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H., and Vogt, P. H. (1997) *Biotechniques* **23**, 504–511
- Pagano, A., Letourneur, F., Garcia-Estefania, D., Carpentier, J. L., Orci, L., and Paccaud, J. P. (1999) *J. Biol. Chem.* **274**, 7833–7840
- Solano, P. J., Mugat, B., Martin, D., Girard, F., Huibant, J. M., Ferraz, C., Jacq, B., Demaille, J., and Maschat, F. (2003) *Development* **130**, 1243–1254
- Castrillon, D. H., Gonczy, P., Alexander, S., Rawson, R., Eberhart, C. G.,

- Viswanathan, S., DiNardo, S., and Wasserman, S. A. (1993) *Genetics* **135**, 489–505
44. Michaut, L., Flister, S., Neeb, M., White, K. P., Certa, U., and Gehring, W. J. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4024–4029
45. Knipple, D. C., Seifert, E., Rosenberg, U. B., Preiss, A., and Jäckle, H. (1985) *Nature* **317**, 40–44
46. Morrison, D. K., Murakami, M. S., and Cleghon, V. (2000) *J. Cell Biol.* **150**, 57–62
47. Sanson, B. (2001) *EMBO Rep.* **2**, 1083–1088
48. Sauer, F., Rivera-Pomar, R., Hoch, M., and Jäckle, H. (1996) *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **351**, 579–587
49. Hombria, J. C., and Brown, S. (2002) *Curr. Biol.* **12**, 569–575
50. Tolia, P. P., and Stroumbakis, N. D. (1998) *Dev. Genes Evol.* **208**, 274–282
51. Kotarski, M. A., Leonard, D. A., Bennett, S. A., Bishop, C. P., Wahn, S. D., Sedore, S. A., and Shrader, M. (1998) *Genome* **41**, 295–302
52. Bashaw, G. J., Hu, H., Nobes, C. D., and Goodman, C. S. (2001) *J. Cell Biol.* **155**, 1117–1122
53. Sanchez, D., Ganfornina, M. D., Torres-Schumann, S., Speese, S. D., Lora, J. M., and Bastiani, M. J. (2000) *Int. J. Dev. Biol.* **44**, 349–359
54. Hoshizaki, D. K., Lunz, R., Ghosh, M., and Johnson, W. (1995) *Genome* **38**, 497–506
55. Häder, T., La Rosee, A., Ziebold, U., Busch, M., Taubert, H., Jäckle, H., and Rivera-Pomar, R. (1998) *Mech. Dev.* **71**, 177–186
56. Rajewsky, N., Vergassola, M., Gaul, U., and Siggia, E. D. (2002) *BMC Bioinformatics* **3**, 30
57. Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., et al. (2004) *Cell* **116**, 499–509
58. Chiang, C. M., and Roeder, R. G. (1993) *Pept. Res.* **6**, 62–64
59. Thummel, C. S., and Pirrotta, V. (1991) *Dros. Info. Serv.* **71**, 150
60. Smith, D. B., Berger, L. C., and Wildeman, A. G. (1993) *Nucleic Acids Res.* **21**, 359–360
61. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) *Nucleic Acids Res.* **23**, 4878–4884
62. Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., and Schacherer, F. (2000) *Nucleic Acids Res.* **28**, 316–319
63. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
64. Lehmann, R., and Tautz, D. (1994) *Methods Cell Biol.* **44**, 575–598