Regulation of segment polarity genes in the head region of different arthropods

Dissertation submitted in partial fulfillment of the requirements for the degree of
"doctor rerum naturalium (Dr. rer. nat.)"
of the Georg-August-University Göttingen

by

Evgenia Ntini

from

Athens,
Greece

Göttingen, 2009

Contents

1. Introduction

Drosophila melanogaster has been extensively used as an excellent developmental model system to reveal genetic and molecular mechanisms that underlie early embryonic pattern formation. Almost twenty years after the first use of the *Drosophila* cuticle in screening for patterning mutants (Nüsslein-Volhard and Wieschaus, 1980), most of the elements of the hierarchical gene cascade governing the early developmental processes were identified and have then been analyzed in detail (reviewed in Ingham and Martinez Arias, 1992; St Johnston and Nüsslein-Volhard, 1992). Most of the studies however primarily focused on the clearly segmented trunk region. Understanding the segmented nature of the embryonic head and the mechanisms leading to patterning and establishment of the head structures advanced much slower. This is due to the inherent complexity of the embryonic head morphological aspects which makes them hard to understand (reviewed in Finkelstein and Perrimon, 1991). Particularly in *Drosophila* the highly evolved process of head involution makes the situation even more complicated and the effort to elucidate head patterning mechanisms even harder.

1.1 Principles of early embryonic patterning in the trunk

Early development and pattern formation in the central trunk region is maternally triggered by the deposition of maternal determinants which form functional gradients in the zygotic syncytium before cellularization. The integrated maternal systems specify the broad domains of gap gene expression along the antero/posterior (A/P) axis. Gap genes then activate the repetitive patterns of transcription of the pair-rule genes, which establish the initial parasegmental subdivisions of the trunk. The pair-rule genes in turn define in a combinatorial manner the initial metameric expression domains of segment-polarity genes which further subdivide each parasegmental unit. More specifically, the pair-rule genes initiate the transcription of *wingless* and *engrailed/hedgehog* in adjacent domains which subsequently regulate each-other to stabilize their expression in two adjacent but not overlapping stripes, thereby defining the parasegmental (PS) boundary 'organizer' (reviewed in Dinardo et al., 1994). Further, the gap genes and the pair-rule genes in combination define the spatial domains of the homeotic selector genes which assign segment identity (reviewed in Akam, 1987). Gene activities of the segment polarity class mediate intra- and inter-parasegmental cell-signaling (reviewed in Martinez Arias, 1993; Perrimon, 1995) which is thought of crucial function in organizing the patterning within the presumptive segmental units leading to distinct cell-fate specification (reviewed in Sanson, 2001). Very recently however, functional role of segment polarity gene-mediated cell-cell interactions was restricted only to – still of substantial importance – stabilizing pre-specified cell identities (Vincent et al., 2008). This is performed by Wg and Hh patterning signals that prevent switch of already specified cell identities towards different ones; thus,

although incapable of inducing cell fates, stabilizing segment-polarity gene signaling allows for the differentiation process to act on the cells according to their current, already specified identity.

1.2 Early embryonic head development – a comparative approach

In contrast to the trunk, mechanisms underlying establishment of metamerization of the head region and anterior head patterning have been elucidated to a less extent (reviewed in Finkelstein and Perrimon, 1991). In principle, the embryonic head region is finally metamerized into the pre-gnathal segments (ocular, antennal and intercalary) and the gnathal segments (mandibular, maxillary and labial). The pregnathal segments constitute the procephalic or anterior head region. All segments are posterior-marked by the ectodermal expression pattern of the segment polarity gene *engrailed*. All six segments, marked by the expression of *engrailed* (*en*) and *hedgehog* (*hh*) at their posterior edge, also display the respective anterior adjacent expression domains of *wingless (wg)* (Schmidt-Ott and Technau, 1992). The clypeolabral lobe at the anterior-most region of the embryo corresponds to a parasegmental unit rather than to a segment (Mohler et al., 1995; Seecomar et al., 2000). Expression domains of the segment polarity genes in the anterior head region provide an important molecular marker for the anterior head segments, helpful to analyze the otherwise difficult to assess effects of mutations on anterior head morphogenesis.

Embryonic expression patterns of segment polarity genes in the head region are extensively conserved across different arthropod species (Rogers and Kaufman 1996, 1997 (Review)) providing a clear indication that some mechanisms of head development may also be phylogenetically conserved. Still, no systematic analyses and comparisons of the regulatory networks that underlie head development in diverse arthropod species have been undertaken so far. Thus, using a model organism other than *Drosophila*, namely the short-germ band insect *Tribolium castaneum* would make comparative analyses possible, which in turn would help to understand whether the so far revealed genetic network underlying *Drosophila* head development is conserved to some extent in other species or whether it rather reflects the highly evolved nature of higher Dipteran morphogenesis. In addition, *Tribolium* is a perfect system to study head development mechanisms as it lacks the complex process of head involution and the cuticular structures directly reflect the embryonic segmental organization of the head region. It therefore offers the advantage of relatively easy phenotypic analyses after reverse-genetics through the effective RNA interference procedure. Transgenesis has also been established in *Tribolium* allowing for the analysis of enhancers and *cis*-regulatory elements (Bucher and Wimmer, 2005).

a					
Procephalon			Gnathocephalon		
Oc	An	Ic	Mn	Mx	Lb
non-pair-rule			pair-rule		
non-Hox		Hox			
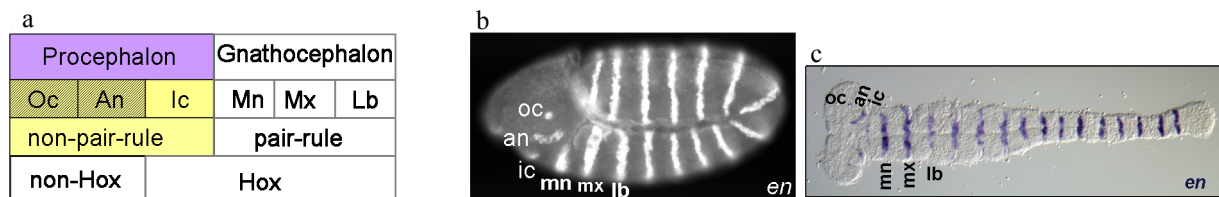
Fig. 1_1. a) Summary of the gene cascades governing anterior head and gnathal head morphogenesis. b, c) Segmental organization of the *Drosophila* and *Tribolium* embryonic head region revealed by the conserved expression pattern of the segment polarity gene *engrailed*. oc, ocular; an, antennal; ic, intercalary; mn, mandibular; mx, maxillary; lb, labial segment.

1.3 Head segmentation – role of the head gap genes

The first debate among arthropodists about the segmented or not nature of the embryonic head region, and the exact number of segmental units that constitute it, was overcome after revealing the segmental ectodermal expression patterns of *en* and *wg* (Schmidt-Ott and Technau, 1992; Schmidt-Ott et al., 1994a) from which all of them later during development contribute cells to neuromeres of the central nervous system. In fact the ectodermal segmental pattern of the anterior (procephalic) head region is well correlated to the segmental organization of the brain (Urbach and Technau, 2003a), with the ocular segment corresponding to the protocerebrum, the antennal to the deutocerebrum and the intercalary to the tritocerebrum. Revealing the – distinct from the trunk – underlying gene networks, cascades and regulatory interactions at the molecular level that govern anterior head regionalization and segmentation remains in progress and comprises a fruitful ground for discovering and establishing novel principles of patterning in the field of Developmental Biology.

It has been well argued so far that the two most posterior gnathal segments, namely the maxillary and the labial, develop like the rest of the trunk segments under the well established hierarchical gene segmentation cascade mentioned above. Development of the mandibular segment integrates inputs from both the head and the trunk patterning systems (Cohen and Jurgens, 1990; Vincent et al., 1997; Grossniklaus et al., 1994). In principle, development of the anterior head region is under the control of the head gap-like genes *orthodenticle* (*otd*), *empty-spiracles* (*ems*), *buttonhead* (*btd*) and *sloppy-paired* (*slp*). In particular, *otd, ems,* and *btd* control the formation of the cephalic *en* and *wg* segmental patterns at germband extension stage (Cohen and Jürgens, 1990). However, due to the effect on segment polarity gene expression observed rather late in the developmental time window, it remained unclear if these genes act to integrate maternal information into blastodermal embryonic pattern in a way similar to the gap genes in the trunk, therefore these gene activities were referred to as anterior head *gap-like* (Mohler, 1995). Nevertheless, the deletion effects of mutations in the anterior head structures (Finkelstein and Perrimon, 1990) as well as the correlation between the combinatorial overlapping expression domains of the head gap genes and the functional domain-requirements (reviewed in Finkelstein and Perrimon, 1991; Cohen and Jürgens, 1991) lead to the

proposal of a combinatorial model controlling anterior head morphogenesis; a combinatorial controlling input from the head gap-like genes can subdivide the anterior head region into a defined number segments and in the absence of pair-rule gene patterning activity in the cephalic region may also directly activate the expression of the segment polarity genes, while the respective level of second-order regulation acting in the trunk being omitted. In addition, in account for the absence of homeotic selector gene activity in the anterior head region, the combinatorial controlling input directed from the head-gap genes was thought to define anterior head segment identities as well.

Indeed, the anterior most cephalic segment expressing a gene of the ANT-C Hox cluster is the intercalary segment where *labial* is expressed (Abzhanov and Kaufman, 1999). *labial* however does not provide a homeotic selector gene activity. Rather, head phenotypic defects are secondary effects of impaired head involution (Merrill et al., 1989; Diederich et al., 1989). At least one of the head gap genes in *Drosophila* seems to have a functional role in anterior head-segment specification; ectopic expression of the homeodomain encoding gene *ems* causes homeotic transformation of mandibular derived structures into intercalary ones (Schöck et al., 2000). This kind of homeotic selector function of *ems* however requires *btd* (zinc finger; Wimmer et al., 1993; Schöck et al., 1999) to overcome the phenotypic suppression of *ems* function by other Hox gene activities (Macias and Morata, 1996). Moreover it acts in a way opposite to the posterior prevalence principle governing trunk homeotic segment transformations. In agreement with the combinatorial model, development of the intercalary segment requires *btd* and *ems* activities, but not *otd*.

Grossniklaus and colleagues in 1994 further supported the combinatorial model while additionally introducing *sloppy paired* (*slp*) gene activity to participate in the functional combinatorial domains of overlapping head gap gene expression. *slp* is required for the establishment of the ocular, antennal and mandibular segments. More specifically, the existence of seven combinatorial expression domains was proposed, however not all of them could be correlated precisely with the actual functional requirements for segment establishment, nor was it possible to specify in detail the precise extent of some of these domains at blastoderm stage (Fig. 1_2).
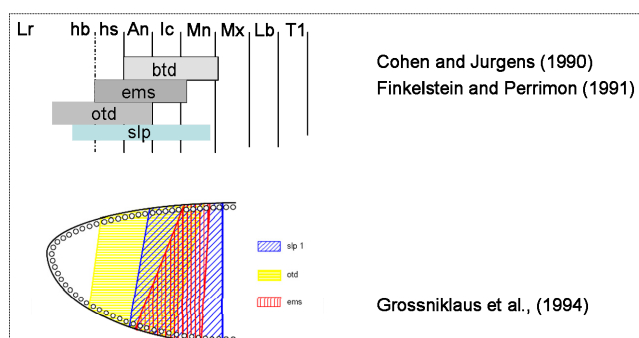
Fig. 1_2. Combinatorial mode of function of the head gap-like genes. Lower scheme is from Grossniklaus et al., 1994. Table summarizes the combinatorial model revised from Grossniklaus et al., 1994.

Combinatorial domains	otd	otd and slp	otd, slp and ems	otd, slp, ems and btd	slp ems and btd	slp and btd	btd
Segment anlage	-	ocular (wg head blob)	ocular (en head spot)	antennal	intercalary	mandibular	-

1.4 Unsolved questions regarding anterior head morphogenesis

Subsequent studies however opposed the proposed function of the combinatorial model in assigning anterior head segment identities. Ectopic expression of *buttonhead* (under the control of *hunchback* promoter) did not affect anterior head segment identities, nor are the spatial limits of *btd* expression instructive for metamerization of the anterior head region (Wimmer et al., 1997). Therefore, *btd* activity was excluded from the proposed combinatorial code for the anterior head segment identity specification. In addition, heatshock-inducible ubiquitous expression of *otd* did not alter anterior head segment identities and only marginally affected expression patterns of the segment polarity genes (Gallitano-Mendel and Finkelstein, 1998). A further interpretation of the combinatorial model leading to metamerization of the anterior head region involved the blastodermal overlap in the early broad expression domains of the segment polarity genes *wg* and *hh*, as being defined by a presumptively direct head gap gene input (Mohler, 1995). Due to the mutual exclusiveness in the successive establishment of the expression domains and in the activities of these two genes, interference might generate their segmentally iterated expression pattern and subsequent metamerization of the anterior head region (Fig. 1_3). However, embryos mutant for either *wg* or *hh* develop the wild type number of anterior head segments (Gallitano-Mendel and Finkelstein, 1999).
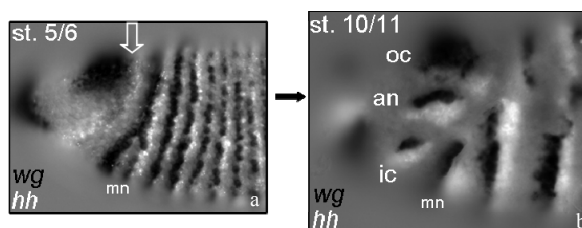
Fig. 1_3. Early blastodemal overlap (arrow in a) in the expression domains of *wg* and *hh* in the anterior head region and subsequent marking of the procephalic segments by the adjacent, mutually abutting stripes of segment polarity gene expression (b).

In addition, apart from a distinct initial activation mode (other than pair rule activity) also a distinct network of cross-regulatory interactions among segment polarity genes was revealed for each of the anterior head segments, differing from the typical maintenance cross-regulation in the trunk (Gallitano-Mendel and Finkelstein, 1999). This mechanism indicates a unique mode of establishment for each of the anterior head segments and it was suggested that it evolved to specify the high degree of structural diversity required for the anterior head morphogenesis.

Further, data presented by Crozatier and colleagues (1996, 1999) support neither a direct activation of procephalic segment polarity gene expression by the head gap genes nor a simple combinatorial input for metamerization of the anterior head region and the assignment of the anterior head segment identities. Rather, their data argued for the functional role of second order regulators acting in the anterior head region. The helix-loop-helix COE factor Collier was the first introduced to act as a crucial transcription factor that is required for the establishment of segment polarity gene

expression in the intercalary segment. *collier* expression is under the strict control by the head gap gene *btd* and functions to establish the PS0/PS(-1) parasegmental boundary and intercalary derived structures. Still, the proposed functions of Collier were indicated only in the context of the analyzed genetic interactions and the issue if regulation of segment polarity gene expression is under the direct control by the transcription factor activity in the intercalary segment remained open (Fig. 1_4).
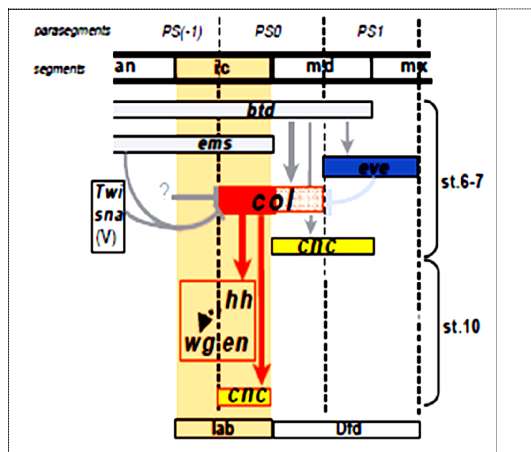
Fig. 1_4. Function of the second-order regulator *collier* in patterning of the intercalary segment. Scheme is from Crozatier et al., 1999. *col* expression is controlled by the head gap gene *btd*. *col* regulates expression of *cap `n` collar* (*cnc*) in the anterior part of the mandibular segment and later during germ band extension it initiates a secondary site of *cnc* expression in the posterior part (in the *hh* expressing cells) of the intercalary segment. *col* is required for the establishment of segment polarity gene expression in the intercalary segment; since *col* expression does not overlap the *wg* expressing cells, positive regulation of *wg* is thought to be mediated by a *hh*-dependent mechanism (Crozatier et al., 1999).

1.5 Aim of the thesis: A bottom-up approach identifies the role of second-level regulators

Overall, it still remains quite obscure how metamerization of the anterior head region succeeds, how the expression domains of the segment polarity genes are established in the absence of pair-rule input activity and how the procephalic segmental unit identities are specified. Further, what is the exact functional role and requirement of the head-gap and segment polarity genes in specifying naïve fields of cells into segmental unit identities and which are the additional 'key-players'. Thus, analyzing and identifying key components of the underlying gene network will help to understand the genetic cascade and the functional interactions at the molecular level that govern anterior head morphogenesis. It seems that continuing research in the field of insect and arthropod (anterior) head morphogenesis will reveal additional key factors and functional molecular interactions involved in the patterning of the anterior head region and the establishment of the anterior head structures.

Taking a bottom-up approach by revealing mechanisms that govern transcriptional regulation of segment polarity gene expression in the procephalic region has been exactly the aim of this thesis. Such an approach leads to the identification of transcription factors directly involved in the gene network(s) patterning each of the anterior head segments, since by directly controlling establishment of segment polarity gene expression they are consequently functional for metamerization of the anterior head region at a level corresponding to the second-order regulation mediated by the pair-rule gene activity in the trunk. During this thesis it could be shown that activity of Collier in the gene

network underlying establishment of the intercalary segment actually involves a direct transcriptional interaction with the downstream segment polarity gene target *hh*, thereby further supporting the critical function of second-order regulators in the patterning process of the anterior head region. In addition, since *collier* has an early parasegmental register of embryonic expression, means for restricting the positive function of the Collier activator to the posterior part of the intercalary segment were examined. Further research in that direction identified a physical protein interaction between Collier and the bZIP factor Cap `n` Collar, the biological relevance of which is addressed in this thesis. Working in that direction resulted from the functional dissection of segment polarity gene *cis*-regulatory regions which led to the isolation of an intercalary-specific *cis*-regulatory element of *hh*. Thus, such a bottom-up approach seems quite promising for the identification of transcriptional control mechanisms underlying segment-polarity gene expression for the rest of the anterior head segments as well.

2. Results

Developmental staging of *Drosophila* embryos was after Campos-Ortega and Hartenstein, 1997 and Hartenstein, 1993 (Appendix).

2.1 Functional dissection of *cis*-regulatory regions of segment polarity genes *wg, en* and *hh* in *Drosophila melanogaster* and *Tribolium castaneum.*

In order to localize *cis*-regulatory information governing anterior-head segment-specific pattern of segment polarity gene expression in *Drosophila melanogaster* and *Tribolium castaneum* functional dissection of upstream or intronic genomic regions was performed. Results are presented starting from *wg, en* and *hh* in the two insect species and ending with dissection of the *hh* upstream region in *Drosophila*.

2.1.1 Functional dissection of *wg* upstream enhancer in *Drosophila*

In *Drosophila melanogaster* transcription of the *wingless* gene is triggered from two alternative transcription start sites, or two alternative promoters, producing two transcript variants. Transcript variant A, 2907 bp, found at genomic locus NM_078788 (NCBI Reference Sequence) and variant B, 2656 bp, locus NM_164746. Transcription start site of transcript B is located within the first intron of transcript A (Fig. 2_1). Thus, the two generated polypeptides differ in their N-terminus, in particular isoform A extends by 53 aa at the N-terminus.

A total of 10.216 kb of upstream sequence was cloned by Long Range PCR (§5.1.5) spanning region -8.094 kb upstream of tsA (transcription start site A) to +195 bp downstream of tsB. The full-length upstream sequence was subcloned in front of a *lacZ* reporter and via *piggyBac* mediated transgenesis tested *in vivo.* At the same time a few subfragment-constructs were also prepared and assayed (Fig. 2_2).

The full-length upstream sequence tested drives expression in the trunk stripes, also in the clypeolabral region and foregut primordium, but lacks *cis*-regulatory information for the cephalic head stripes (Fig. 2_3). Therefore, *cis*-regulatory information governing expression of *wg* in the anterior head region is not included in the 10 kb upstream enhancer that contains *cis*-regulatory elements sufficient for the trunk and the anterior most terminal expression pattern during germ band extension. At stage 6 the reporter trunk stripes appear in an odd pair-rule mode but have fully developed by stage 8.

The construct spanning 4.8 kb upstream of tsA, was not re-injected after a fist unsuccessful round, since it had already been tested by Lessing and Nusse, 1998. It contains elements sufficient for trunk expression, including the gnathal stripes, but not for the anterior cephalic region. A 5' overlapping fragment (-6.7_-3.8 kb) was tested in combination with a *hs43* basal promoter and shows only weak expression after T1 at stage 11 while it leads to minimal ectopic expression in the anterior head region at blastoderm stage (Fig. 2_4). Note that *cis*-elements controlling maintenance of gnathal segment-specific expression of *wg* are distinct from those for the rest of the trunk.

Construct spanning (-8.1_-3.9 kb) ensures clypeolabrum expression and since fragment (-6.7_-3.8 kb) does not mediate expression in that region, the 1.4 kb fragment (-8.1_-6.7 kb) contains *cis*-elements necessary for clypeolabrum-specific expression of *wg* (Fig. 2_5).

In search of *cis*-regulatory information governing expression of *wg* in the anterior head region, 8.4 kb of further upstream enhancer sequence was isolated by long-range PCR, spanning region [-16.212 kb_-7.813 kb] relative to tsA (using primers wg_upF2/wg_upR2; Table 5-1) and subcloned in front of 280 bp endogenous promoter region surrounding tsA [-159 bp_+121 bp] (Fig 2_2). This 8.4 kb sequence contains *cis*-elements also driving expression in the trunk stripes, overlapping to some degree the expression pattern produced by the (-8.1 kb) enhancer fragment. This result implies a degree of redundancy shared by the regions (-16 kb_-8.1 kb) and (-4.8 kb_tsA) (§3.1.3). It also contains *cis*-regulatory information for the antennal segment at stage 10 and for the ventral most part of the ocular segment ('head blob'). A nascent conclusion from this observation is that regulation of the ocular-specific expression pattern of *wg* involves dorso-ventral differences (§3.1.2). Finally, intercalary-specific expression comes up rather late, at st. 11, missing early onset which is during stage 10 for *wg*. Therefore, region [-16.212 kb_-7.813 kb] lacks *cis*-regulatory information for the early temporal control of intercalary-specific expression of *wg*, i.e. it lacks elements required for activation and the establishment of *wg* expression in the intercalary segment, while *cis*-elements found in this upstream sequence are rather involved in maintenance of intercalary-specific *wg* expression (§3.1.4).
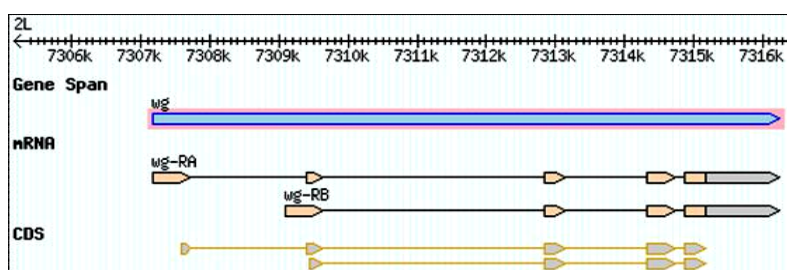
Fig. 2_1.Transcriptional units of *wg* and the respective coding sequences
(CDS) as depicted in Flybase.org.
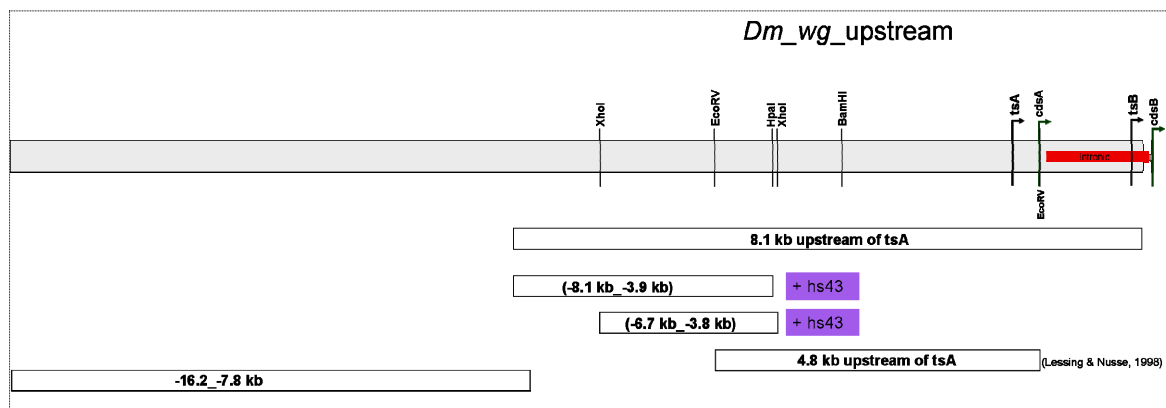
Fig. 2_2. Functional dissection of the *wg* upstream region. Fragments (-8.1 kb_-3.9 kb) and (-6.7 kb_-3.8 kb) were tested in combination with the *hs43* basal promoter. Fragment (-16.2 kb_-7.8 kb) was tested in combination with the endogenous promoter (-128_+121 bp) around the tsA.
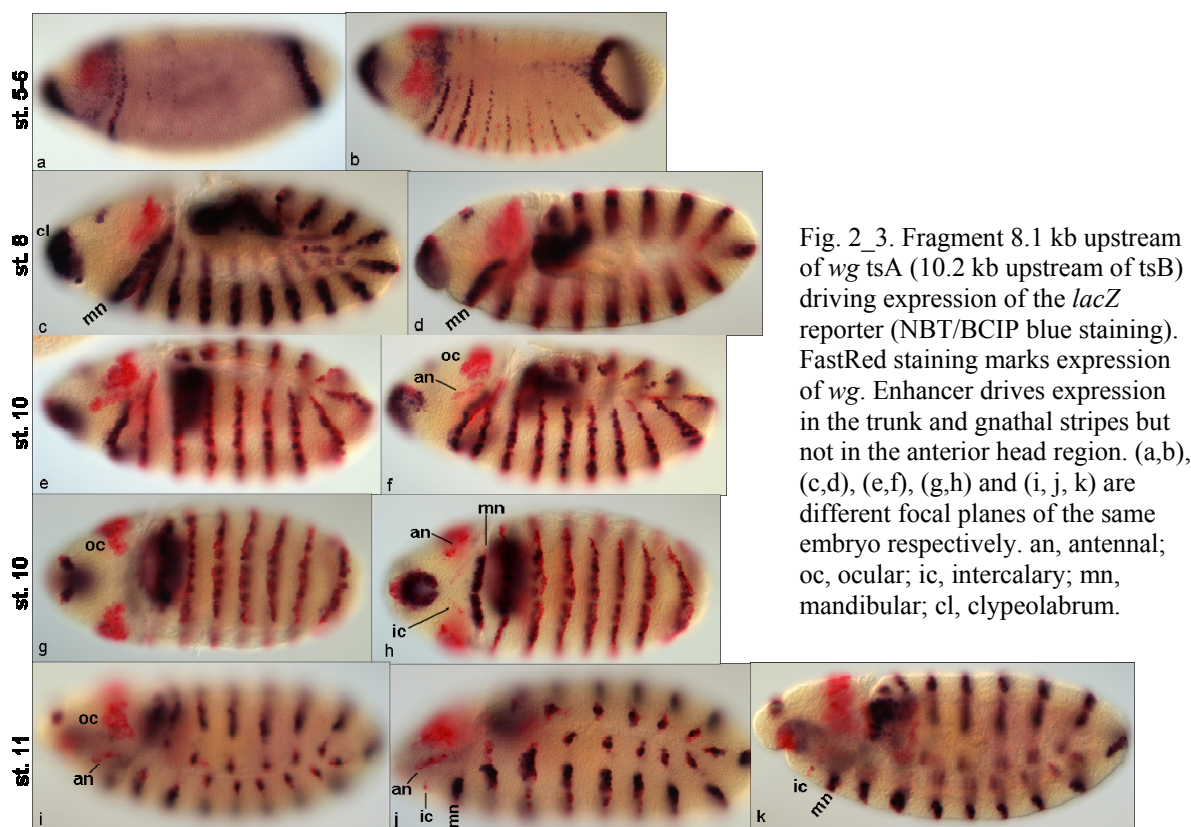
Fig. 2_3. Fragment 8.1 kb upstream of *wg* tsA (10.2 kb upstream of tsB) driving expression of the *lacZ* reporter (NBT/BCIP blue staining). FastRed staining marks expression of *wg*. Enhancer drives expression in the trunk and gnathal stripes but not in the anterior head region. (a,b), (c,d), (e,f), (g,h) and (i, j, k) are different focal planes of the same embryo respectively. an, antennal; oc, ocular; ic, intercalary; mn, mandibular; cl, clypeolabrum.
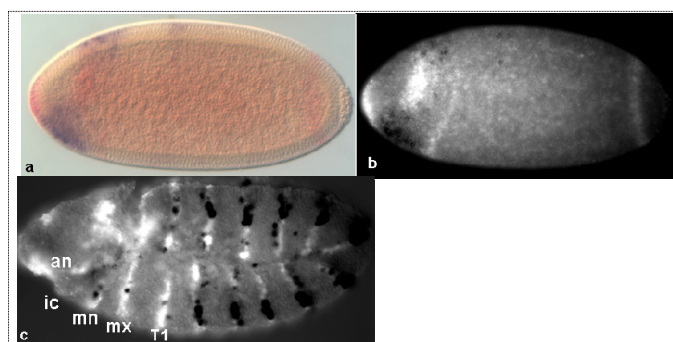
Fig. 2_4. Fragment (-6.7 kb_-3.8 kb) upstream of *wg* tsA dives ectopic expression of the reporter at blastoderm stage (a,b) (blue in the DIC picture, black in the fluorescent). FastRed staining (red in the DIC picture (a) and white in the fluorescent (b,c)) marks expression of *wg*. At stage 11 (c) it drives partial expression in the trunk stripes but not in the gnathal, indicating the function of distinct maintenance regulatory elements.
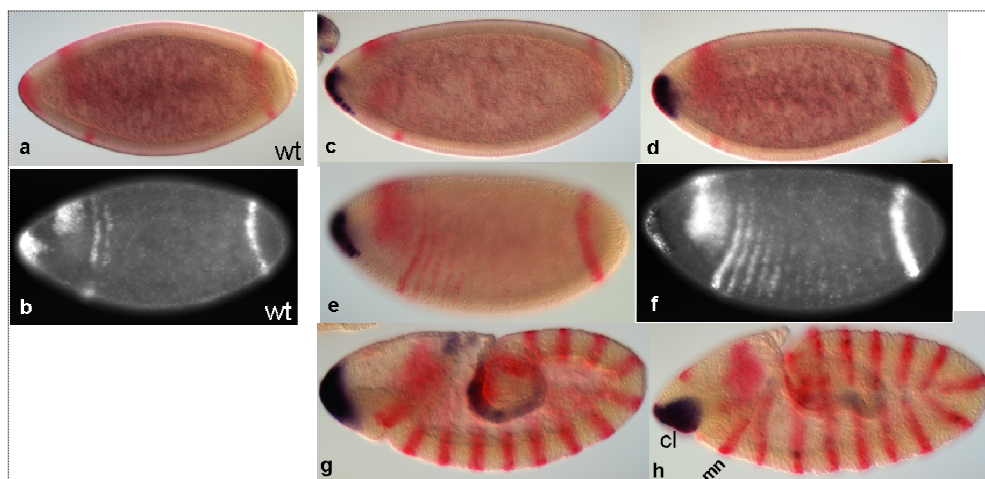
Fig. 2_5. Fragment (-8.1_-3.9 kb) in combination with the *hs43* promoter drives expression in the anterior terminal region corresponding to the clypeolabrum (cl). a,b wild-type (wt) embryos at blastoderm stage. FastRed (white in the fluorescent pictures b, f) marks expression of *wg*. c, d, e, f blastoderm stage. c, e different focal planes of the same embryo. g, stage 8. h, stage 9.
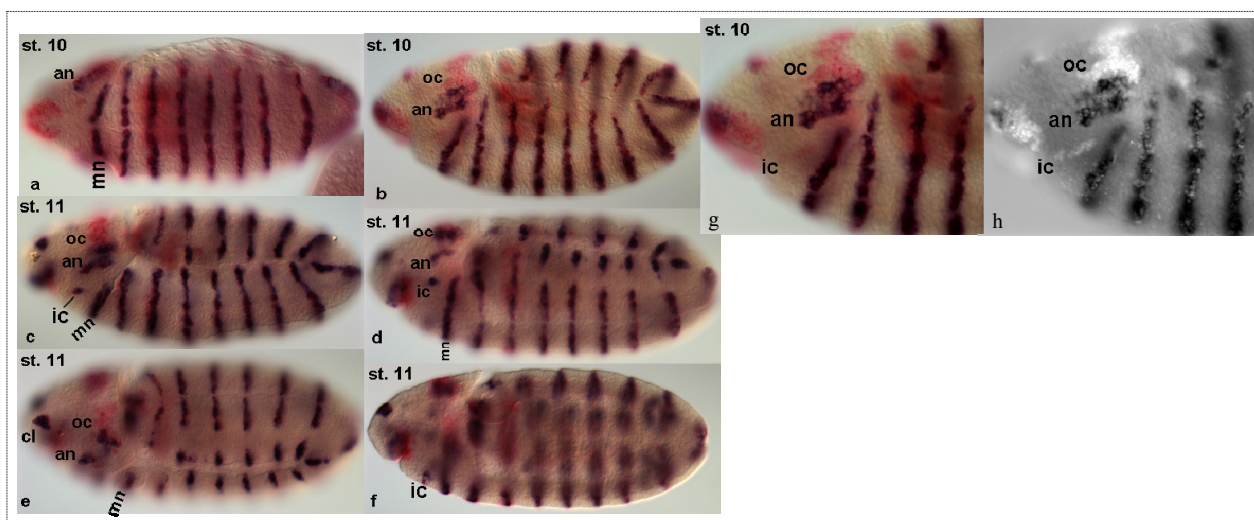
Fig. 2_6. Fragment (-16.2_-7.8 kb) combined with the endogenous promoter region around tsA drives expression in the gnathal and trunk stripes during stage 10 (a, b, g, h); it also mediates expression in the antennal segment and the ventral part of the ocular segment, but not in the intercalary segment (g, h). g, h 40x magnification of the embryo depicted in b (stage 10). Intercalary expression is evident at stage 11 (c-f). c, e and d, f are different focal planes of the same embryo respectively (stage 11). an, antennal; oc, ocular; ic, intercalary.

2.1.2 Functional dissection of *en cis*-regulatory region

The first intronic sequence of Dm_*en* had already been tested (Kassis, 1990); it functions as an enhancer which in combination with the endogenous promoter region drives expression in the trunk stripes. This was confirmed by subcloning and testing the first intron of *en* combined with the endogenous promoter sequence (-75_+55 bp) which also contains *cis*-regulatory elements that ensure or enhance spatial-specific expression (Orihara et al., 1999 and §3.3). In particular two GAGA elements found at -33 and -45 bp (Fig. 2_7) are thought to augment transcriptional

efficiency. Indeed, via *piggyBac* mediated transgenesis, in two independent lines, trunk expression pattern could be generated with this construct, which lacks however *cis*-regulatory information specific for the anterior head region (Fig. 2_8). It also seems to contain *cis*-elements for establishment of expression of the trunk stripes, but not for the gnathal stripes; reporter is expressed in the gnathal segments later, at stage 11, missing stage 10. Perhaps it is then driven by the same maintenance *cis*-elements that also control late expression in the rest of the trunk.

In search of *cis*-elements controlling anterior head expression of *en*, an 8.8 kb upstream sequence was cloned (-8812_+165 bp) and tested *in vivo* (Fig. 2_9). This upstream region drives some expression in the trunk stripes, and starting from blastoderm stage it also drives ectopic expression in the anterior head region. Also later, during germ band elongation, reporter expression exhibits a loss of anterior parasegmental boundary in the trunk, which is reminiscent of anterior expansion of the endogenous *en* expression pattern in the case of mutants lacking anterior parasegmental repressor activity, such as *slp* (Cadigan et al., 1994). At stage 10 ectopic expression in the anterior parts of the mandibular and the maxillary segments is also detected (Fig. 2_9 i, j.).

If the ectopic expression driven by the 8.8 kb upstream enhancer sequence is not due to a position effect caused by the insertion site, then it is possible that the 8.8 kb upstream sequence lacks specific negative-repression elements which, otherwise functional in the wild-type locus, ensure silencing of the endogenous *engrailed* expression in these ectopically driven reporter regions. However, this was not further examined and clarified since only one transgenic line was generated for that construct. Regarding the anterior head stripes, expression of the reporter could be observed only in the antennal segment. Therefore, *cis*-regulatory information driving the rest of the anterior head-segment expression should be searched elsewhere in the ~50 kb *en* locus.
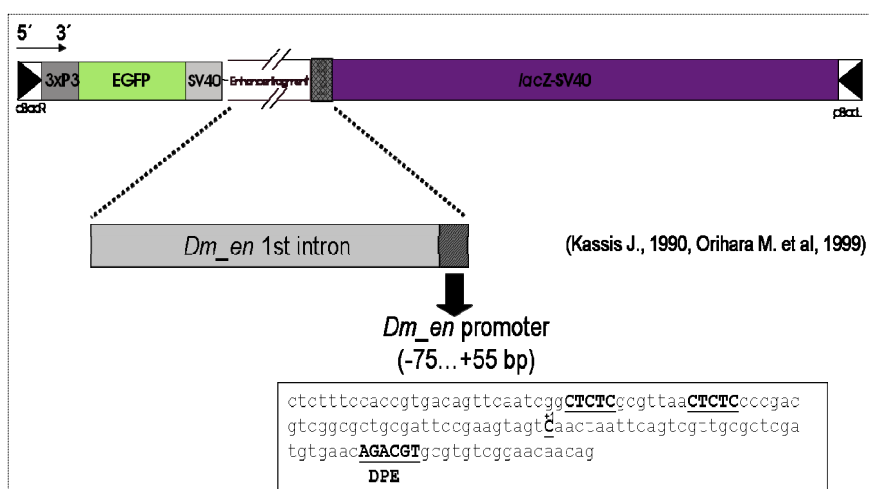
Fig. 2_7. Construct assaying the first intron of *engrailed* in combination with the endogenous promoter region (-75_+55 bp) driving expression of the *lacZ* reporter. In the promoter region two GAGA elements (CTCTC) and the Downstream Promoter Element (DPE) are depicted. Transcription start site C (+1).
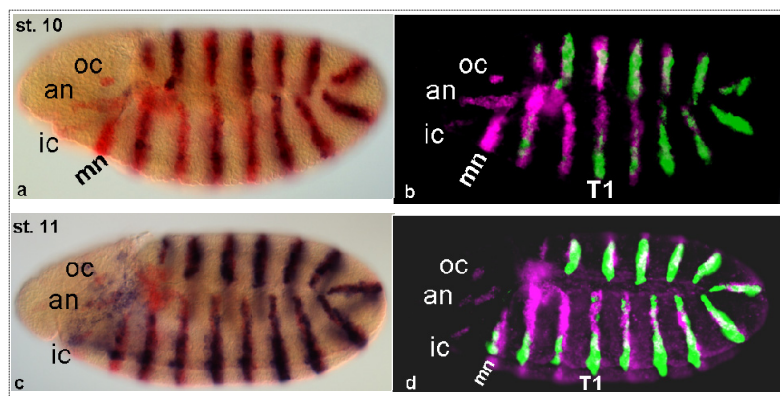
Fig. 2_8. Embryos in a (st. 10) and c (st. 11) are depicted in false-colour in b and d respectively. First intron of *en* drives expression in the trunk stripes by late stage 10 (a, b) but not in the head region. Reporter is stained blue in the DIC picture (a,c) and green in false-colour (b,d). At stage 11 partial expression in the gnathal stripes is also detected (c,d).
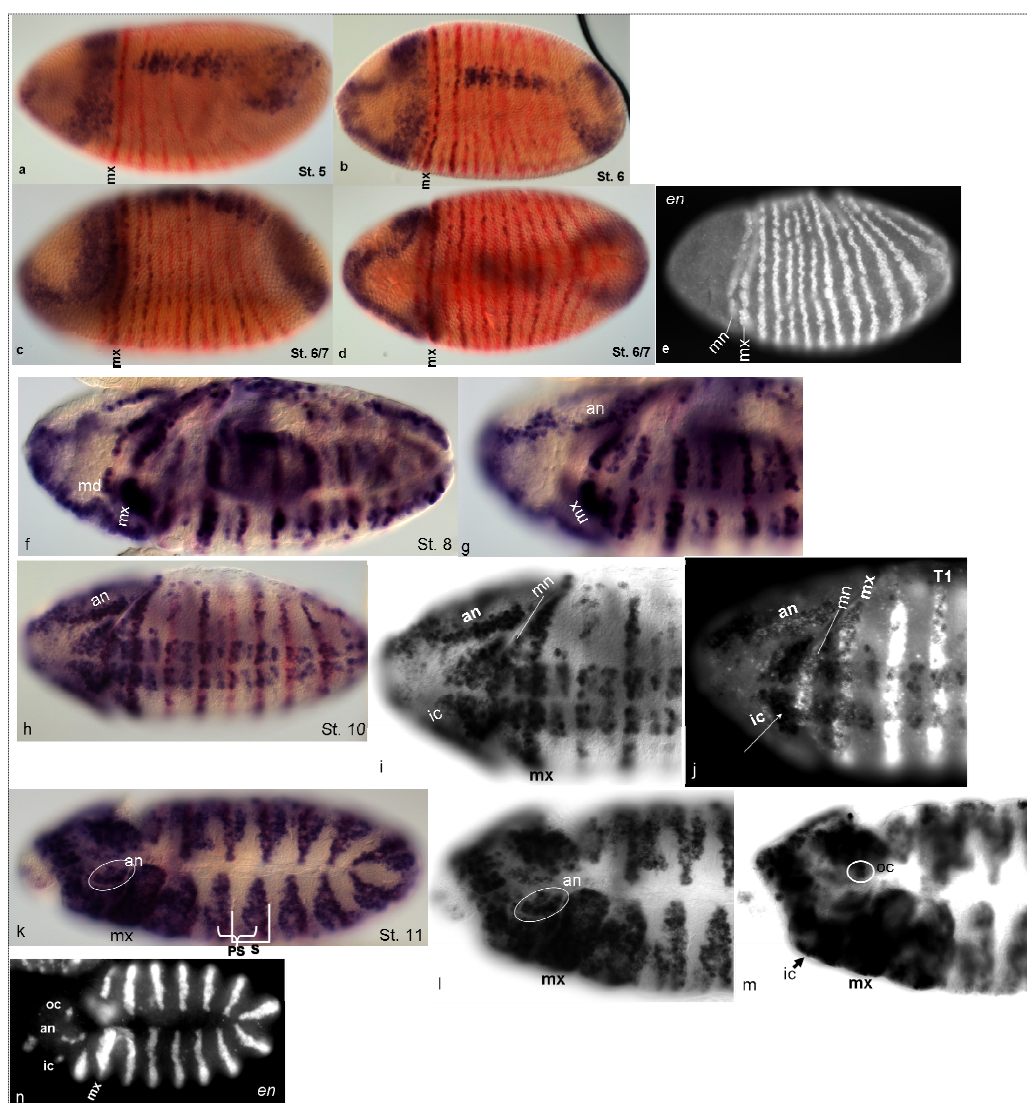
Fig. 2_9. 8.8 kb upstream of *engrailed* drives ectopic expression in the prospective procephalic region of early embryos (a-d). e: fluorescent (FastRed) staining of the wt *engrailed* expression pattern; there is no expression of *en* in the prospective procephalic region of st. 6/7 embryos. The most anterior stripe of wt *en* expression detected (e) corresponds to the mandibular segment (mn); the maxillary stripe (mx) is also depicted. Ectopic expression in the trunk and the head region is also observed at later stages (f-m). g is 40x magnification of the embryo depicted in f (st. 8). i and j (fluorescent) are 40x magnification of the embryo depicted in h (DIC). Arrow in j depicts ectopic reporter expression in the anterior part of the mandibular segment. l, m are different focal planes of 40x magnification of the embryo depicted in k; reporter pattern exhibits loss of anterior parasegmental boundary of expression (PS: parasegment, S: segment). Fluorescent staining in n marks the wt expression pattern of *en* at st. 11 embryos.

16

2.1.3 Determination of transcription-start sites and functional assay of segment polarity gene (*wg*, *en*, and *hh*) *cis*-regulatory regions in *Tribolium castaneum*.

A similar isolation of *cis*-regulatory regions was performed in *Tribolium castaneum*. Transcription start sites were determined by 5' RACE PCR. For *wg* and *en* there was one transcription start site determined, matching for both cases the insect transcription start site consensus 5' TC**A**GT 3' (Hultmark et al., 1986). For Tc_*wg*, by similarity to the *Drosophila* genomic organization, one more alternative transcription start site within the first intron of Tc_*wg* transcript A was predicted as putative (tsB). A 7.9 kb sequence upstream of Tc_*wg*_tsA was isolated by Long Range PCR, and the fragment XbaI (-7.4 kb)_SacII (+ 271 bp) was subcloned in an orientation-directed way driving expression of the *lacZ* reporter (2_10 A). This upstream sequence contains minimal *cis*-regulatory information driving only very weak expression in the gnathal stripes, ocular and antennal segments and in the posterior growth zone (Fig. 2_10 B).

For Tc_*hh*, two transcription start sites were identified by 5' RACE PCR (tsA, tsB). Only tsB matches the insect transcription start site consensus. Transcript B has a 5' UTR of 62 bp, while transcript A has a 5' UTR of 251 bp after removing by splicing a 3.4 kb intron (Fig. 2_11). A 5.2 kb EcoRI fragment spanning (-4941_+272 bp) relative to tsA (Fig. 2_11) did not contain sufficient *cis*-regulatory information to drive specific expression of the reporter.

For Tc_*en*, in comparison to the *Drosophila* situation, the first intron was tested in combination with the endogenous promoter region, determined by 5' RACE PCR (Fig. 2_10). Dm_*en* and Tc_*en* have a similar genomic organization (two introns), and *in silico* analysis performed on the first 3.036 kb intron of Tc_*en* predicts several putative homeodomain binding sites (TAATTA) which have been implicated as functional in the case of Dm_*en* 1st intron (Kassis, 1990). However first intronic sequence of Tc_*en* does not mediate specific expression and thus other regions of the locus should be tested.
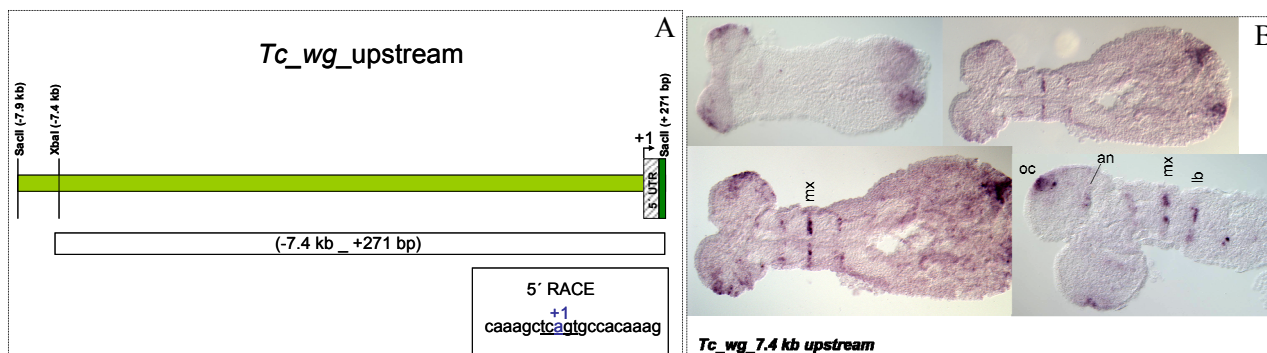
Fig. 2_10. Transcription start site of Tc_*wg* was determined by 5' RACE PCR (A). Construct (-7.4 kb_+271 bp) was tested via *piggyBac* transgenesis; it drives very weak expression in the gnathal stripes, ocular and antennal segments (B).
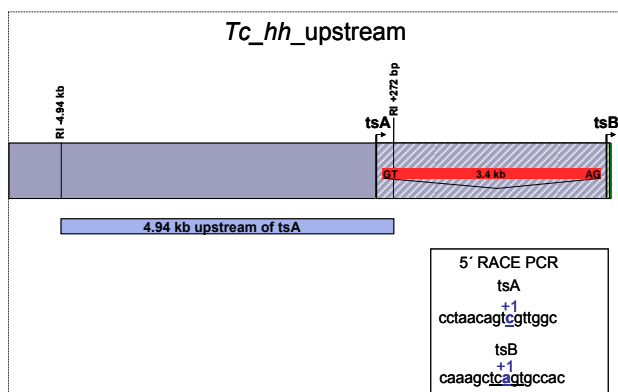
Fig. 2_11. Two transcription start sites or two alternative promoters were identified by 5' RACE PCR for Tc_hh. Only tsB matches the insect transcription start site consensus TCAGT. 4.9 kb upstream of tsA does not mediate specific expression pattern.
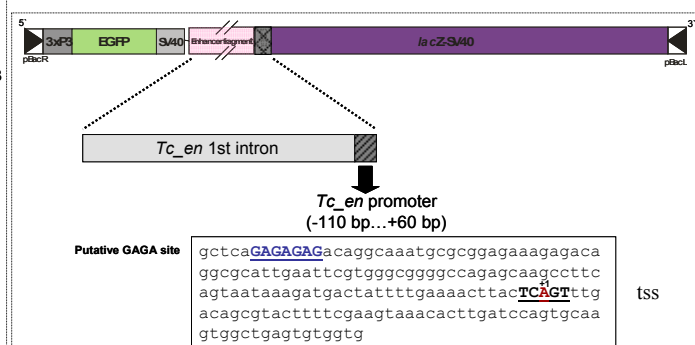
Fig. 2_12. Transcription start site of Tc_en was identified by 5' RACE PCR and matches the insect consensus. A GAGA element (blue and underlined) can be found. tss: transcription start site determined by 5' RACE PCR.

2.1.4 Functional dissection of Dm_*hh cis*-regulatory region

Functional dissection of the *hh* upstream region in *Drosophila* is presented last in this section since further analysis ensued.

Transcription start site (tss) of *hh* was identified by 5' RACE PCR to be located +33 bp downstream of the tss reported in Lee et al., 1992. This may be due to a nucleotide polymorphism (T>C) that the used strain carries at position [-2] relative to the published tss. [+1] of the annotated EST (EK111112.5prime) is also affected by a polymorphism (G>C) found at -21 bp relative to the transcription start site identified by 5' RACE PCR. *hh* transcript produced from the identified tss has a 5' UTR of 353 nucleotides (Fig. 2_13).

```
>5' RACE PCR        -392 CGTACCTATACTCGCATTCGTATTCGCTGCCGTACTCGTACTCGCACTCG -343
>Lee et al., 1992   -392 CGTACTTATACTCGCATTCGTATTCGCTGCCGTACTCGTACTCGCACTCG -343
>EK111112.5prime    -392 CGTACTTATACTCGCATTGGTATTCGCTGCCGTACTCGTACTCGCACTCG -343
```

Fig. 2_13. Nucleotide polymorphisms (red) may affect transcription start site (blue). (-392_-343 bp) relative to translation initiation codon.

As a first step a 6.43 kb upstream sequence spanning region [-6.43 kb (NsiI site)_+265 bp] was tested. The full-length upstream enhancer drives expression early at blastoderm stage in a broad procephalic domain overlapping the endogenous anterior expression domain of *hh*. At st. 8 it drives trunk expression in an even pair-rule mode. During germ-band elongation the reporter gene is also expressed in the procephalic head stripes, meaning that the essential *cis*-regulatory elements controlling expression in the anterior head region are included in this 6.43 kb enhancer region (Fig. 2_14). The 6.43 kb upstream enhancer was further dissected into 5' shorter fragments (Fig. 2_15 A) which were tested via *piggyBac*-mediated transgenesis.

Fragment -4.08 kb (PvuII restriction site) drives expression in the intercalary segment and the dorsal most part of the antennal segment (Fig. 2_15 II a-e). Fragment -3.17 kb (NheI site) retains expression only in the dorsal most part of the antennal segment (Fig. 2_15 II f). Therefore, enhancer sequence [-4.08 kb_3.17 kb] contains *cis*-regulatory information <u>essential</u> for controlling expression of *hh* in the intercalary segment.

To test whether this enhancer element is also <u>sufficient</u> to ensure intercalary-specific expression of *hh*, fragment [-4.08 kb_3.077 kb] ('α') was subcloned and tested in combination with a *hs43* basal promoter or with the endogenous promoter region (-120_+99 bp), respectively. Expression of the reporter was specifically driven in the intercalary segment, essentially when the construct contained the endogenous promoter region (Fig. 2_16). Therefore, enhancer fragment [-4.08 kb_3.077 kb] is <u>sufficient</u> and <u>essential</u> for controlling intercalary-specific expression of *hh*. This sequence and its functional subfragments will be referred from now on as the <u>intercalary-specific *cis*-regulatory element</u> (**ic-CRE**). It is also preliminary concluded that intercalary-specific expression of *hh* requires an enhancer-promoter specific interaction or that the *hs43* TATA-box minimal promoter does not contain core elements required for the ic-CRE function (§3.3).
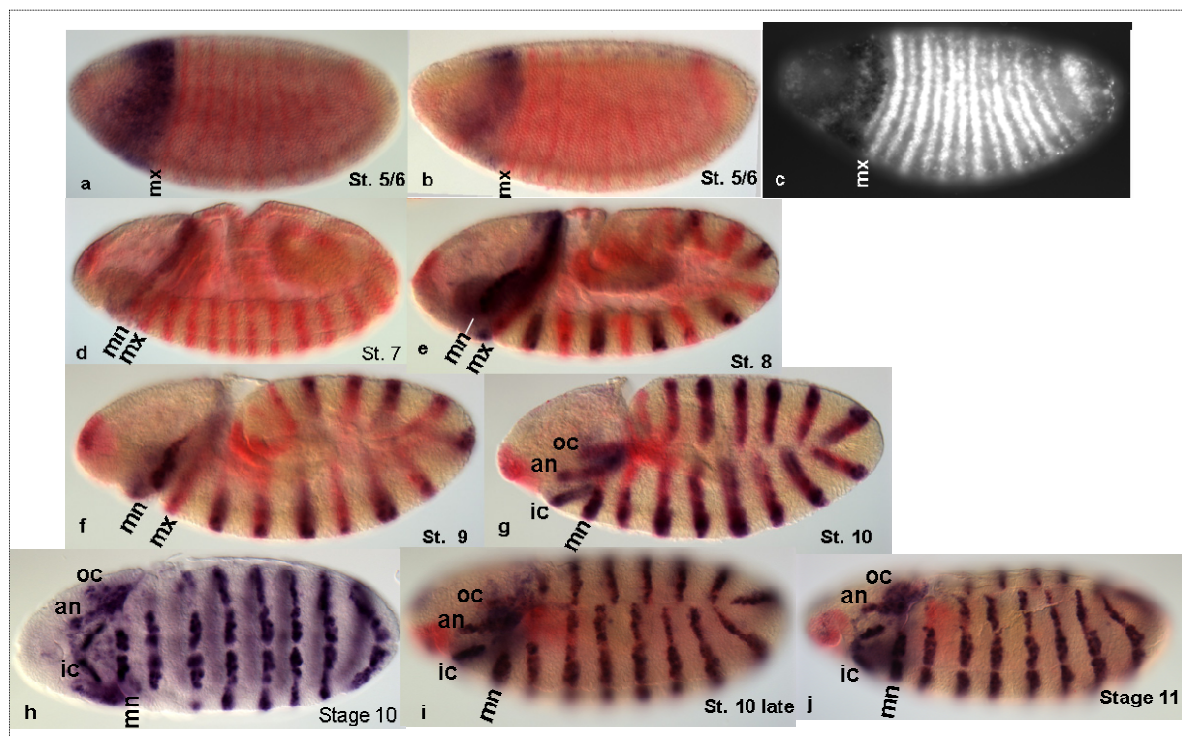
Fig. 2_14. 6.43 kb upstream of *hh* drives expression in the early broad anterior domain corresponding to the presumptive procephalic region (a-c), and later it drives complete expression in the trunk and the anterior head region. Blue (NBT) staining marks the reporter expression. FastRed staining (fluorescent in c) marks expression of *hh*. c, fluorescent picture of b (DIC). a, b, c blastodermal stages. d, stage 7; e, stage 8; f, stage 9; g, h stage 10. i, j late 10-11.
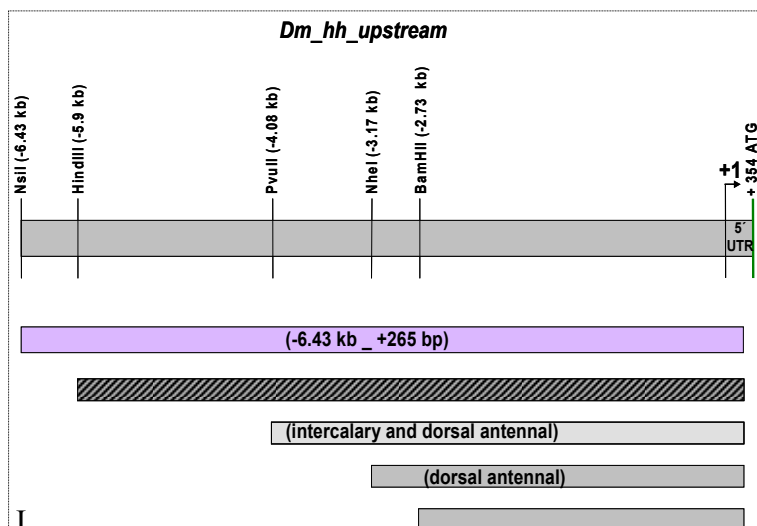
Fig. 2_15 (I) Functional dissection of the -6.43 kb upstream region of *hh*. Fragment -5.9 kb is functionally similar to the -6.43 kb upstream region. -4.08 kb drives expression in the intercalary segment, -3.17 kb does not. Dissection was performed using the depicted restrictions sites.
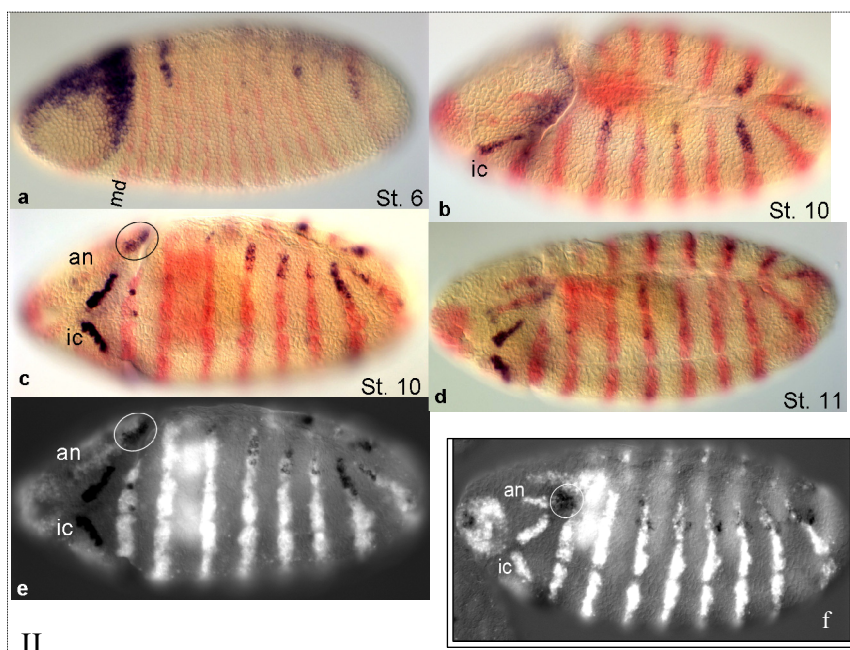
Fig. 2_15 (II) a-e: 4.08 kb subconstruct. f: 3.17 kb subconstruct. a: ectopic expression in the procephalic region at st. 6 driven by the 4.08 kb upstream *hh* enhancer sequence. During germ band extension, 4.08 kb upstream of *hh* drives expression in the intercalary segment (b-e) and the anterior most portion of the antennal (circle in c, e) while 3.17 kb upstream region (f) does not retain expression in the intercalary segment.
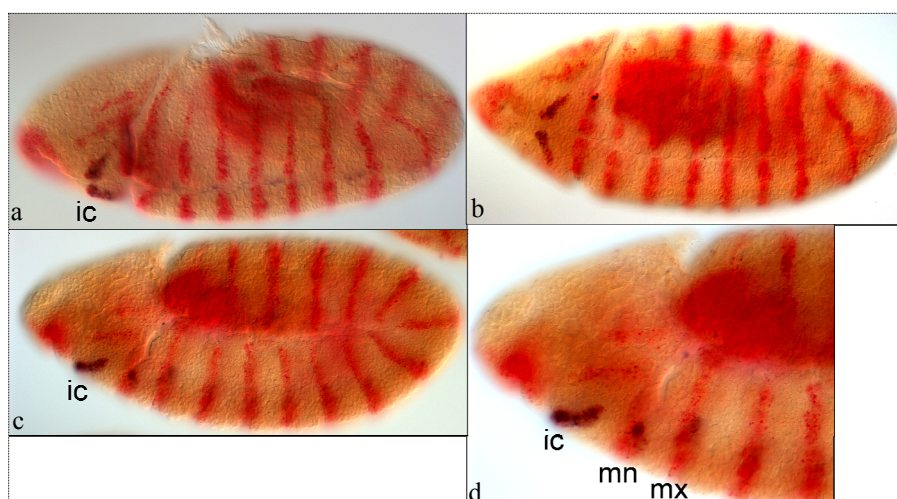
Fig. 2_16. The 'α' fragment (-4085_-3077 kb) in combination with the endogenous promoter (-120_+99 bp) specifically drives expression in the intercalary segment and 2-3 cells of the mandibular and maxillary segment. a: st. 9 (ventral-lateral view); b-d; st. 10. b, ventral view; c, lateral – anterior to the left. d is 40x magnification of c.

2.2 Analysis of the intercalary-specific *cis*-regulatory element of *hh*

2.2.1 Dissection of the 1 kb ic-CRE

The 1kb intercalary-specific *cis*-regulatory element [-4.08 kb_3.077 kb] was further dissected by testing overlapping subfragments in an effort to further localize crucial *cis*-regulatory information in a more defined context (Fig. 2_17).
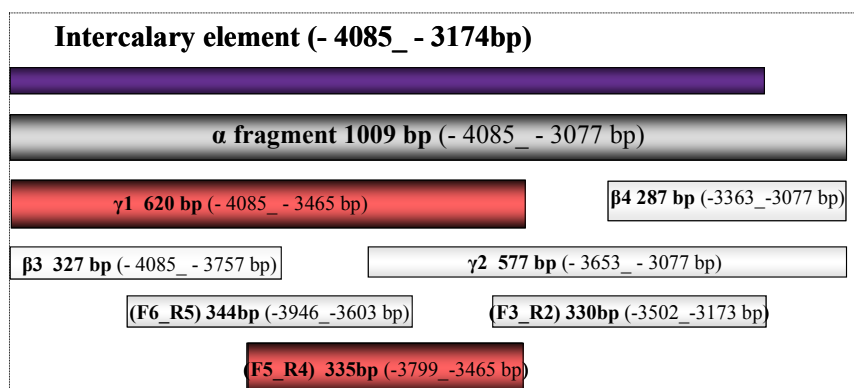
Fig. 2_17. Functional dissection of the intercalary-specific *cis*-regulatory element of *hh* by testing smaller overlapping subfragments of the 'α' fragment. Red are the ones that retain expression in the intercalary segment.

Two of the overlapping sub-constructs retain intercalary-expression (Fig. 2_18, 2_19); construct *'γ1'* 620 bp (-4085_-3465 bp) is expressed in the intercalary segment displaying an intercalary-specific early onset at stage 8. At stage 10 it is specifically expressed in the intercalary segment, plus a few cells of the mandibular and maxillary which also exhibit reporter expression. Later, at stage 11 it is de-repressed in the trunk (§3.4). Construct *'F5_R4'* (-3799_-3464 bp) was the minimum 335 bp fragment to be tested that retains expression in the intercalary segment during germ-band extension stage 10 while later it is partially derepressed in the trunk, mainly in the gnathal stripes. It lacks however early onset at stage 8.

Since two constructs, 5' and 3' overlapping the *'F5_R4'*, do not mediate intercalary-expression (Fig. 2_17), namely the fragment 'F6_R5' (-3946_3603 bp) and 'γ2' (-3653_-3077 bp), this allows for a preliminary conclusion; either sequences (-3799_-3654 bp) and (-3602_-3464 bp) contain *cis*-regulatory elements that direct a synergistic positive input required to specify intercalary-specific expression, or the net outcome of positive (activator) and negative (repressor) binding sites is positive (activating) only in the context of the 'F5_R4' sequence. This kind of organization and function of enhancer elements or *cis*-regulatory modules is described in Segal et al., 2008.

γ1 620 bp (- 4085_ - 3465 bp)_hh prom
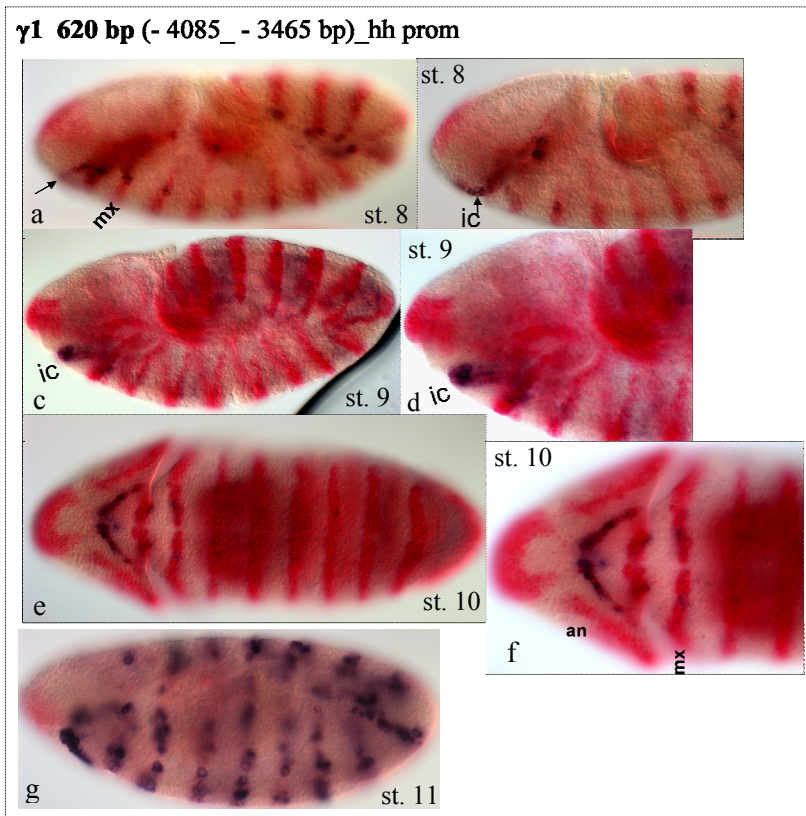
Fig. 2_18. The γ1 fragment retains expression in the intercalary segment while ensuring the early onset of expression at stage 8. Later at stage 11 reporter expression is derepressed in the trunk and gnathal segments (g). a, b different focal planes of the same embryo (st. 8). c,d st. 9; d is 40x magnificatioon of c. e,f st. 10; f is 40x magnification of e. g st. 11.
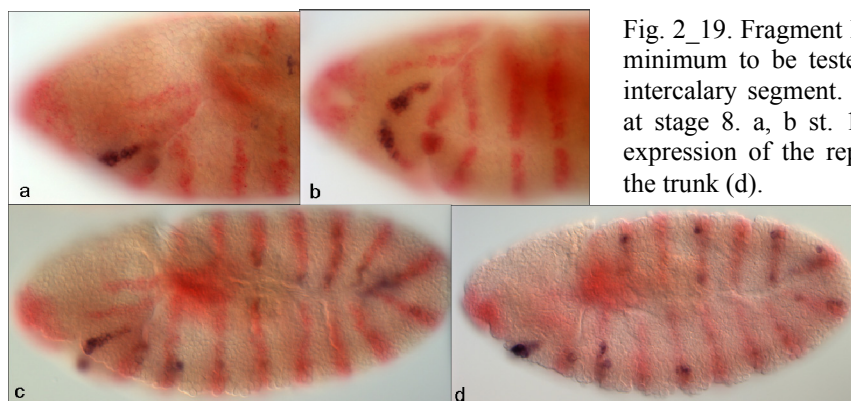
Fig. 2_19. Fragment F5_R4 (-3799_-3465 bp) was the minimum to be tested that retains expression in the intercalary segment. It lacks however the early onset at stage 8. a, b st. 10, c late st. 10. Later at st. 11 expression of the reporter is partially derepressed in the trunk (d).

2.2.2 Phylogenetic conservation and deletion mutant analysis of the ic-CRE

The 335 bp minimum ic-CRE was submitted to phylogenetic analysis (12 *Drosophilidae*) using the UCSC Genome Browser (http://www.genome.ucsc.edu/cgi-bin/hgGateway) or the DrosophilaEvoprinterHD (http://evoprinter.ninds.nih.gov/). The sequence consists of six highly conserved blocks separated by less conserved DNA stretches (Fig. 2_20).

To define functional *cis*-binding sites found within the ic-CRE, deletion mutant analysis was performed based on the phylogenetic footprint. Highly conserved sequences found within non-coding regions usually have a functional role including specific recognition and binding by transcription factors; thus they potentially function as transcriptional elements (Bejerano et al., 2005). As a first approach a series of deletion mutant constructs lacking each of the highly conserved blocks of the ic–CRE were cloned and tested *in vivo*. The deletion mutants versions of the ic-CRE sequence were constructed in the context of the minimum 335 bp 'F5_R4' (-3799_-3465 bp) ic-CRE fragment. Construction of the deletion mutant constructs is described in §5.1.3. From that point of the analysis and on the entire reporter constructs series were tested integrated at the same genomic site using the *attP-attB* site-specific integration system (§5.1.2). Reason for that was to avoid misinterpretation of results caused by integration site position effects; differences in transcriptional outcome driven by slightly divergent *cis*-regulatory sequences is significantly prone to position effects which can overcome the true transcriptional potential of the tested *cis*-driving sequence. Tested at the same *attP* landing site, the wild type sequence 'γ1mF6_hhR4' (-3888_-3465 bp) retained expression in the intercalary segment (§2.5.1) and also a point mutant version of the minimum 335 bp 'F5_R4' ic-CRE (bearing a point mutation ccattag>cca**gg**ag that affects the putative homeodomain site found within the fourth conservation block; §2.2.3) was driving expression in the intercalary segment.

However, none of the deletion mutant constructs retained expression in the intercalary segment, either because functional *cis*-binding elements essential for intercalary-specific expression of *hh* lie within each of the highly conserved blocks, or due to disturbing inter-motif distances crucial for individual transcription factor binding and/or protein-protein interactions, or even for both reasons. Position and distance specificity are important determinants of *cis*-regulatory motifs in addition to evolutionary conservation (Vardhahabhuti et al 2007). Therefore, the performed deletion mutant analysis was not informative regarding precise location of intercalary-specific *cis*-regulatory information. Thus, another approach was followed, namely a comprehensive point mutant analysis.

Fig. 2_20. (on the next site): 12 species alignment of the minimum ic-CRE sequence representing 12 species phylogenetic conservation. The 12 species conservation blocks (I-VI) are depicted.

Fig. 2_20

Phylogenetic footprint - 12 Drosophila species

Multiple sequence alignment of the upstream regulatory region of 12 Drosophila species (pseudoobscura, persimilis, melanogaster, sechellia, simulans, yakuba, erecta, ananassae, willistoni, virilis, mojavensis, grimshawi). Conserved blocks (phylogenetic footprints) are highlighted and labelled I–VI.

Region I / II (positions to 51):

```
pseudoobscura  ACGTCCTCCCGTCCATTTGCCTAATTTC-------TAATCTCTAA----AGCAATTCCCCAAT  51
persimilis     ACGTCCTCCCGTCCATTTGCCTAATTTC-------AAATCTCTAA----AGCAATTCCCCAAT  51
melanogaster   ---ACATCCATCCATTTGCCTAATTTC--------TATTT-CGG-----CAGCAATTCCCCAAT  47
sechellia      ---ACATCCATCCATTTGCCTAATTTC--------TATCT-CGG-----CAGCAATTCCCCAAT  47
simulans       ---ACATCCATCCATTTGCCTAATTTC--------TATCT-CGG-----CAGCAATTCCCCAAT  47
yakuba         ---ACATCCATCCATTTGCCTAATTTC--------TATCT-CGG-----CAGCAATTCCCCAAT  47
erecta         ---ACATCCATCCATTTGCCTAATTTC--------TATCT-CGG-----CAGCAATTCCCCAAT  47
ananassae      --ACAGCCATTTGCCTAATTTC-------------TATCTGCAGGCCCAGCAATTCCCCAAT   47
willistoni     --GCAACCATTTGCCTAATTTC-------------AAACGTCTAA----GACCAATTCCCCAAA  49
virilis        ----GCAACCATTTGCCTAATTTC-----------TATCTCTGT--GAGACAATTCCCCAAT   45
mojavensis     --GCTGCCCATTTGCCTAATTTCATCTCTAGCTCGT-----------GAGACAATTCCCCAAT  50
grimshawi      ----GCATCCATTTGCCTAATTTC-----------TATCACTGCC-AAGACAATTCCCCAAT   46
                *       *********                   *    *  *******
```

Region III (positions to 104):

```
pseudoobscura  GGATTTATGTGAATTAAAC-----AGCCTCTCACTCTCTTTCGTTC  104
persimilis     GGATTTATGTGAATTAAAC-----AGCCTCTCACTCTCTTTCGTTC  104
melanogaster   GGATTTATGTGAATTAAAC----GAACACAGCCACTT            97
sechellia      GGATTTATGTGAATTAAAC----GGACACAGCCACTT            97
simulans       GGATTTATGTGAATTAAAC----GGACACAGCCACTT            97
yakuba         GGATTTATGTGAATTAAAC----GGACACAGACACTT            97
erecta         GGATTTATGTGAATTAAAC----A------TAGCCACTT           88
ananassae      GGATTTATGTGAATTAAAC----AGCC-----CCTTGCCGCCT       93
willistoni     GGATTTAACTTATTGTTAGATTTGTTAAGCC----A--------GCCACCC  94
virilis        GGATTTTTACTCA----------GGACACAGCCACCT-GCTCGCACACCT  93
mojavensis     GGCCGTTTACTCA----------AGCCT--------ACGAGTCCGC-T   97
grimshawi      GGCATTTATGTGAATTAAAC----AGATTTATGTGAATTAAACCAGACCT--ACGAATCCGTTT  98
                ***  ** *** *
```

Region IV (positions to 146):

```
pseudoobscura  ---CGTTCCGTTCCGCCTTTGAGCCAATACGAATCCACATCCAAA  146
persimilis     ---CGTTCCGTTCCGCCTTTGAGCCAATACGAATCCACATCCAAA  146
melanogaster   ---CCCCTTCCCCTTT-----TCCCTTCCACAACCCAATCCCAA   139
sechellia      ---CCCCTTCCCCTTT-----TCCCTTCCACAACCCAATCCCAA   132
simulans       ---CCCCTTCCCCTTCCC--CTTTCCCTTCCACGACCCAATCCCAA 138
yakuba         ---CCCCTTCCCCTTCC----TTCCTTTCCCTTCCCAATCCGAA   133
erecta         ---CCCCTTCCCCTTCC----TTCCTTTCCCTTCCCAATCCGAA   126
ananassae      ---CGCGGCTCGAGCCT---CGGAGCCTCAAGCGGCACA-TCCAAA 132
willistoni     ---CTGCCCCATC-------CCATTCCAATGGCAATTCAA        124
virilis        ----CCTCCCTGG--------CCTCCCTGG                  102
mojavensis     ---CCTCCCT----------CCTCCCT                     104
grimshawi      GTTCTGCTCGCCTCGTACACCTCCTTCCACTCCTACTCCTACTCCTA 158
                *****  **  **
```

Region II / V (positions to 234):

```
pseudoobscura  GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCGCAGCAGCATGA----  234
persimilis     GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCGCAGCAGCATGA----  234
melanogaster   GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCTATGAGGCA------  215
sechellia      GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCTATGAGGCA------  208
simulans       GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCTATGAGGCA------  214
yakuba         GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCTATGAGGCA------  209
erecta         GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCTATGAGGCA------  202
ananassae      GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGGCCAAAAAGAAAAAG---  220
willistoni     GACCTCAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGGCGCCGTGCGGCGTTTG--  213
virilis        GACCTTAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGCGGCCGCGCCAAC----CT  194
mojavensis     GACCTTAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGCGCCGCTGCGCGCTGCT--GCT 190
grimshawi      GACCTTAAACCGTTAGTGG-CATTTGTCGCACGCTATGAGCGGCCGCTGCGGCTTTTCAGTT 277
                ****              *  *************** ******           **
```

Region (positions to 294):

```
pseudoobscura  GCAGCAGCAGCCGCAGTGGCCAGTGGCAGTCGGAGGCCAGTCGGAGAC  294
persimilis     GCAGCA----GCCGCAGTGGCCAGTGGCAGTCGGAGGCCAGTCGGAGAC  291
melanogaster   -------------------GGCCGAAAGCAACCG--CCGCAGTCA-     239
sechellia      -------------------GGCCGAAAGCAACCG--CCGCAGTCA-     232
simulans       -------------------GGCCGAAAGCAACCG--CCGCAGTCA-     238
yakuba         -------------------GGCCGAAAGCAACCG--CCGCAGTCA-     233
erecta         -------------------GGCCGAAAGCAACCG--CCGCAGTCA-     226
ananassae      ACC----------CGATCGAAGGAAAATGAACCCAC--AACAGCAGTTC- 257
willistoni     CAAGCGTG----ATGCTGATGCCCCAGTTCCAGTCA--CAACAGCCGTCCCA 257
virilis        CCAGTTACCAGTTGC-CAGTTGCCA--GTTGCCAGTCACCAGCACCAGA--C 248
mojavensis     GCCGCTGCGTCCTCGTCTGCTGCTGGTCAGTTACCAGTTACCAGTCACCGATCAC 250
grimshawi      ACCAATTGCCAGTGCCAGCAGTCACCG--ACCACCAGTCACTGTTCACTGTTCAC  335
                *                  *
```

Region VI (positions to 347):

```
pseudoobscura  AGTCA-CAACAAGAGGC-CAACAGCT----CGGGGCATGATTGAGTGCCTGCTTCTTGG  347
persimilis     AGTCA-CAACAAGAGGC-CAACAGCT----CGGGGCATGATTGAGTGCCTGCTTCTTGG  344
melanogaster   ---CA-ACACAAGAGTC-CCACAACT----CGCAGCAATGATTGAGTGCCTGATTCTCGG  289
sechellia      ---CA-ACACAAGAGTC-CCACAACT----CGCAGCAATGATTGAGTGCCTGATTCTCGG  282
simulans       ---CA-ACACAAGAGTC-CCACAACT----CGCAGCAATGATTGAGTGCCTGATTCTCGG  288
yakuba         ---CA-ACACAAGAGTC-CCACAACT----CGCAGCAATGATTGAGTGCCTGATTCTCGG  283
erecta         ---CA-ACACAAGAGTC-CCACAACT----CGCAGCAATGATTGAGTGCCTGATTCTCGG  276
ananassae      ---CA-ACACAAGAGTC-CCACAACT----GGCTACAATGATTGAGTGCCTGCTTCTCGG  307
willistoni     AAACA-CACACACAGTC-C--CAGTC----CAGTACATGATTGAGTGCCTGTTTCTTGG  308
virilis        AGACAGCGGCAACAGTCACAACAGCA----GTCGCAGCAATGATTGAGTGGCCCGCTTCTCAG  305
mojavensis     AGACAGCGGCAGCAGTCACAACAGCA----GTCGCAGCAATGATTGAGTGGCCGCTTCTCGG  307
grimshawi      AGACAGCGGCAACAGTCACAACAAGTCGCACCATGATTGAGTGGTGTTTCTCAG  395
                          *     *           *********** *  *  **  ****
```

Region (positions to 395):

```
pseudoobscura  AATC--AAATCGAAAT------------------CAGCATTGGC-CTGATATCGAGAAT--TGGGACTGTGA  395
persimilis     AATC--AAATCGAAAT------------------CAGCATTGGC-CTGATATCGAGAAT--TGGGACTGTGA  392
melanogaster   AATC--AAATCGAAAT------------------CAGCATTGAC-TTGATATGAGA-----ACCGTC-      330
sechellia      AATC--AAATCGAAAT------------------CAGCATTGAC-TTGATATGAGA-----ACTGTG-      323
simulans       AATC--AAATCGAAAT------------------CAGCATTCTCATTGATATAGA-----ACTGTG-      330
yakuba         AATC--AAATCGAAAATATCGAAATACTGAAAATCAGCATTGAAAATCAGCATTGATATAAGAGAACTGAACTGTG-- 340
erecta         AATC--AAATCGATTT------------------CAGCATTGAC-TTGATATGAGA-----ACTGTG-      317
ananassae      AA----AAATCGATTT------------------GAGAACTGGGATCG-TAATTGGAT---CGAAGTG-      348
willistoni     AACTCAAAACTCAAAAT-----------------CTCAACATACAA-CTGAATATAGGGA----GTTAAAT     353
virilis        AATT--GCTACGAATT------------------CAGTAT-----TTGTTTTTGTT--------GTTATT-   342
mojavensis     AACT--GATACGAATT------------------CGGTAT-----TTGTTACTGTT--------GCCGCT-   344
grimshawi      AACT--GGTACGAGTT------------------CAGTAT-----TTGTTTTCGTC--------GCTATT-   432
                **                                 *         *    *  **
```

Fig. 2_21 A: Schematic representation of the highly conserved blocks (I-VI) and the deletion mutant constructs which were tested via site-specific recombination. None of them retained intercalary-expression. B: Sequence phylogenetic conservation (nucleotides in capitals) as depicted by the *cis*-Decoder. Nucleotides that disturb the 12 species conservation blocks may be filtered through the 7 species phylogenetic footprint (the red ones). The purple conservation block in the lower sequence encompasses the Collier binding site (§2.2.3).

2.2.3 *In silico* prediction of putative binding sites and site-directed point mutagenesis

In silico analysis was performed on the ic-CRE to identify putative binding sites. A set of online publically available prediction tools were employed, namely the rVISTA (http://genome.lbl.gov/vista/rvista/submit.shtml), Match and P-Match (www.gene-regulation.com/pub/programs.html) and MatInspector (http://www.genomatix.de/cgi-bin/matinspector_prof/mat_fam.pl). These programs generate predictions on novel DNA sequences based on scoring against known position weight matrices (PWMs) of transcription factors which can be found in the TransFac Database (http://www.genome.jp/dbget-bin/www_bfind?transfac) or in the MatBase of Genomatrix (http://www.genomatrix.de). Position weight matrices used were of insect and vertebrate transcription factors, taking into consideration that high degree of percentage identity in protein primary sequence and more significantly in the DNA binding domain shared by homologues is permissive for similar DNA binding specificity. Results were further inspected manually, checked in correlation with literature reports and filtered through the phylogenetic footprint.

In this paragraph predictions falling within each of the highly conserved blocks are depicted, their biological relevance is discussed and insertion of independent point mutations in the context of the 335 bp minimum ic-CRE ('F5_R4') or the 450 bp (γ1mF5_hhR4) ic-CRE (§2.5.1) and their effects are also presented. Point mutations were all tested at the same genomic position (landing site *96E*) using the *attP-attB* site-specific integration system (§5.1.2)

Tables presented in this section with predictions and [core similarity; matrix similarity] scores are from MatInspector. Capitals used in the 'Sequence' column of the prediction tables depict the 4 nucleotide core sequence.

As depicted in Fig. 2_22, for the first conservation block three main predictions are generated; the first regards a putative POU domain-factor recognition site, the second a putative C/EBP site on the opposite strand, and the third a homeodomain site (core CTAAT). All three predicted sites partially overlap which would imply differential recognition by distinct transcription factors in distinct group of cells or during different developmental windows, or even competition for binding. Notably, a deletion mutant construct lacking the first conservation block (Fig. 2_21 A) is deficient of reporter expression, implying possible involvement of the predicted binding activities in functionality of the ic-CRE.

Fig. 2_22. *In silico* predictions within the first conservation block and the respective position weight matrices.

Fig. 2_23. *In silico* predictions in the second conservation block.

Submitting the second conservation block to MatInspector with core similarity cut-off (0.95) and matrix similarity cut-off optimized by the program generates predictions listed in the table of Fig. 2_23.
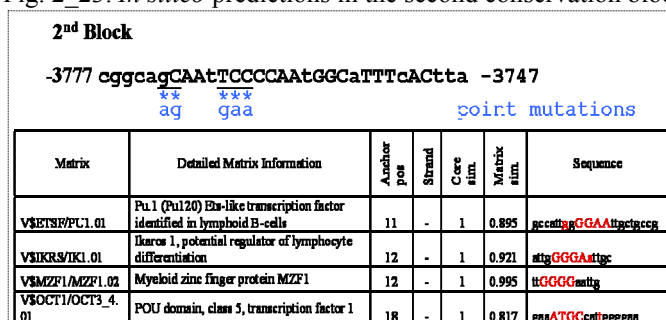
Setting the matrix similarity cut-off value to **0.8** generates one more prediction which scores the binding matrix of Olf-1 (Olfactory factor 1) with a core similarity value **1** and matrix similarity value **0.819**. This prediction does not come up when the matrix threshold is set to program-optimized, as the optimal matrix-similarity cut-off value for Olf-1 matrix is **0.82**. The predicted sequence displays two mismatches to the matrix (positions 14 and 18) which are actually nucleotide positions that disturb the 12 species conservation block.
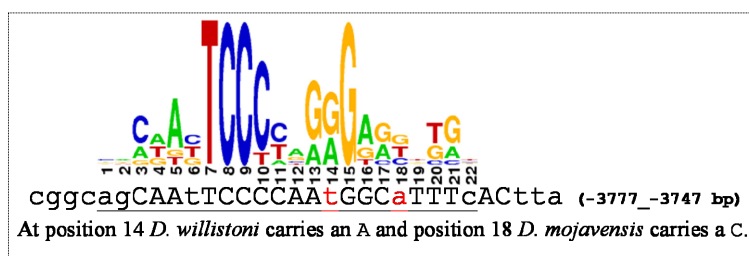
Fig. 2_24. Alignment of the sequence of the second conservation block to the binding matrix of Olf-1. Interestingly, nucleotides that do not match the respective positions of the matrix (small red) are not conserved across 12 species.

Olf-1/ Ebf (mouse Early B-cell factor) is the mammalian homolog of Dm_Collier. Factors share 86% identities in their DNA binding domain that would allow for a similar DNA-binding specificity. Thus, the predicted Olf-1 site in the second conservation block was regarded as a putative Collier recognition site. Further analysis ensued to elucidate this putative direct DNA-protein interaction as it shows quite an interest from a functional aspect (§2.7). Two point mutations inserted in that block, depicted in Fig. 2_23, abolish reporter expression, indicating that the site is necessary *in vivo* for the ic-CRE transcriptional outcome.
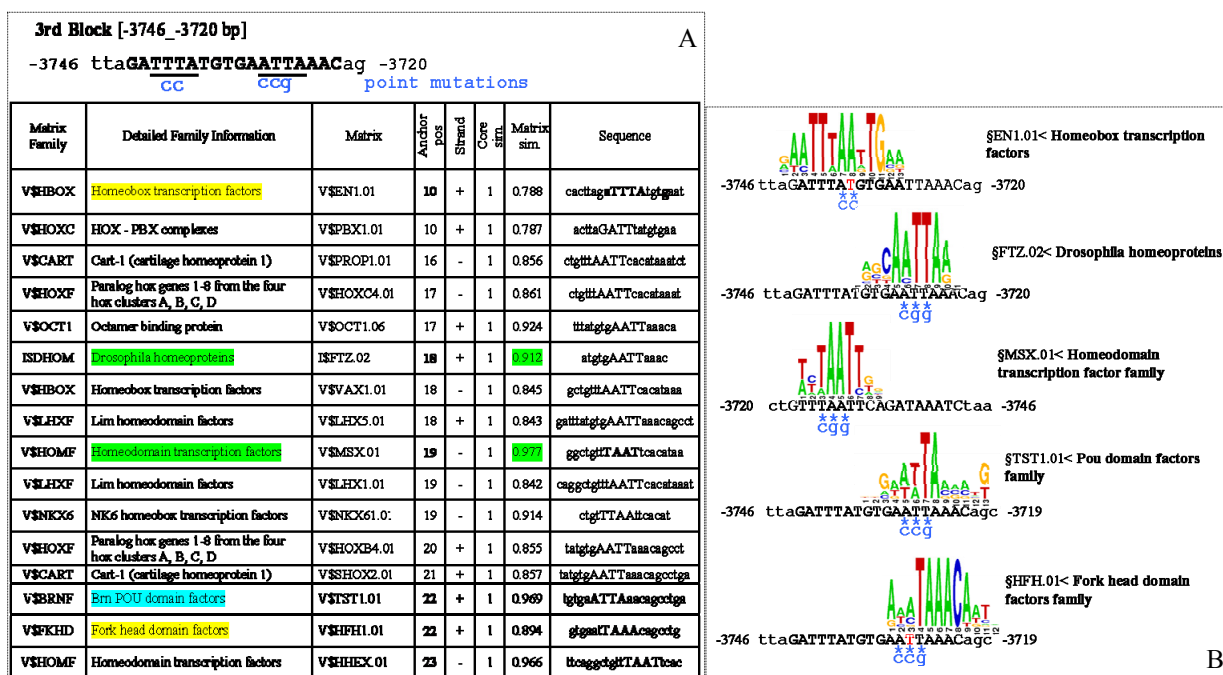
Fig. 2_25. Predictions of MatInspector within the third conservation block (Table in A) and alignment of the sequence of the third conservation block to some of the scoring PWMs (B).

Submitting the 33 bp sequence of the third conservation block to MatInspector (Core similarity cut-off: 0.9; Matrix similarity cut-off: optimized to minimize false-positives, specified for each individual matrix by the program) mainly predicts two putative homeodomain binding sites (two cores underlined) as well as a putative POU domain recognition site and a putative forkhead-domain (winged-helix) recognition site overlapping the second homeodomain site (Fig. 2_25). Both core sequences of the putative recognition sites were simultaneously mutated in the context of the 450 bp ic-CRE enhancer fragment (§2.5.1). In spite of one mismatch to the binding matrix, possibility that the second site may as well be recognized by a fork head domain factor *in vivo* cannot be excluded. The second recognition site also scores the binding matrix of Pou domain family factors which have both activation and repression transcriptional potential. However, the mutation did not alter expression outcome of the ic-CRE (Fig. 2_26). Since several putative homeodomain binding sites are found within the ic-CRE as well as throughout the full-length upstream enhancer, potential roles

of homeodomain proteins in development and means of determining DNA binding specificity and transcriptional outcome are further discussed in §3.5.
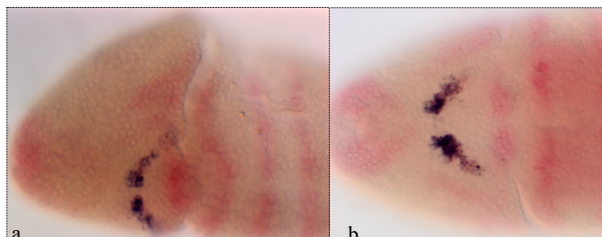
Fig. 2_26. *450 bp ic-CRE_homeo mut.* Point mutations disrupting the core sequences of the putative homeodomain sites within the third conservation block do not affect expression outcome of the ic-CRE. a, b two embryos at st. 10. a, ventral-lateral view; b, ventral.
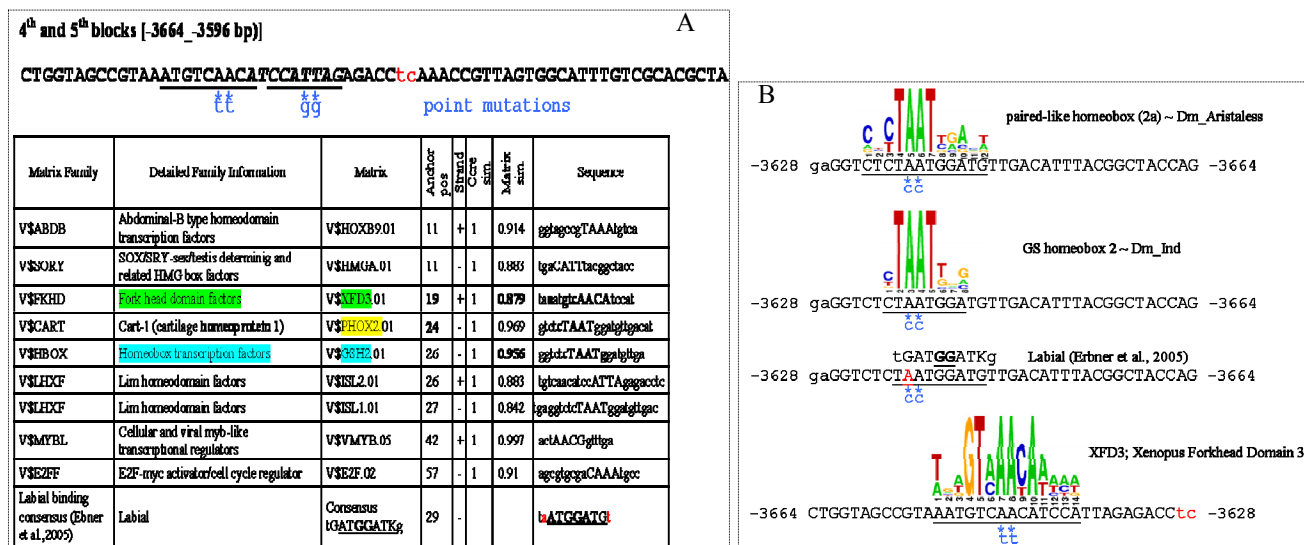
Fig. 2_27. *In silico* prediction results in the 4th and 5th blocks (-3664_-3596 bp) (A) and alignment of the sequence of the 4th conservation block to some of the PWMs (B).

Sequence [-3664_-3596 bp] was considered to consist of two highly conserved blocks (12 *Drosophila* species) separated by 2 bp. Submitting this 69 bp sequence to MatInspector generates predictions listed in the table of Fig. 2_27. (Core similarity cut-off: 0.9; Matrix similarity cut-off optimized). In the 4th sequence block interesting predictions include a putative fork-head domain (winged helix*)* recognition site `ATGTCAACAT` and a homeodomain binding site (core `TAAT`). The putative fork-head domain recognition site scores the binding matrix of XFD3. This factor is member of the Xenopus fork-head domain related multigene family. DNA binding specificity has been studied for three representative members of the family. XD-1 and -3 bind consensus `ANNGT(C/A)AACA` while XFD-2 binds consensus `ANNRT(C/A)AACA` (Kaufmann et al., 1995).

Homologues of fork head domain (winged-helix) factors in *Drosophila* include Forkhead (*fkh*), Crocodile (*croc*), Sloppy paired 1 and 2 (*slp1, slp2*) (Grossniklaus et al., 1994), FoxK/Mnf (*Mnf*) (Casas-Tinto et al., 2008) and Biniou/FoxF (*biniou*) (Popichenco et al., 2007). The predicted sequence resembles a site protected in DNaseI footprint assay by Slp1 found in the proximal promoter of *ftz* (`TCTTCGATGTCAACACACC`) (Yu et al., 1999). Crocodile binding consensus is quite different (`arTAAAtAtnn`) (Häcker et al., 1995) and more similar to a site footprinted by Fkh found

28

in a *krüppel* enhancer (ATCCTTTT<u>GTAAATATTA</u>ATGT) (Hoch and Jäckle, 1998). Consensus binding sequence recognized by Dm_Biniou is RYAAAYA (Popichenco et al., 2007; Peterson et al., 1997), thus it conforms more to binding specificities of Fkh and Croc. A putative fork head domain binding site is also found within the single conservation block of (-3914_-3888 bp) (Fig. 2_37) and might be responsible for restriction of the (*γ1mF5_hhR4*) 450 bp ic-CRE expression outcome strikingly in the intercalary segment, in contrast to the (*γ1_mF6_hhR4*) which is also expressed in some cells of the mandibular and maxillary. Since *croc* is not expressed in the gnathal segments, it is more likely that the site is recognized by a distinct winged helix factor with a repression effect outside the intercalary segment.

The point mutation disrupting the putative fork head domain recognition site of the fourth conservation block (caaca>cttca, Fig. 2_27 A) both in the context of the minimum 335 'F5_R4' ic-CRE sequence as well as in the context of the 450 bp 'γ1mF5_hhR4' ic-CRE abolished expression of the reporter (§2.3).

Furthermore, the putative homeodomain site predicted in the fourth conservation block scores the binding matrix of the vertebrate factor PHOX2A (Paired-like homeobox 2a; Fig. 2_27 A) which contains a paired-like box homeodomain most similar to that of Dm_Aristaless. Other homologues are encoded by genes including *PHDP (putative homeodomain protein), paired, homeobrain* and *gooseberry*.

The predicted homeodomain site also scores the binding matrix of the vertebrate homeodomain transcription factor GSH2 ('Genomic Screen' GS homeobox 2; Hsieh-Li et al, 1995). Gsh-2 is the vertebrate homolog of *Drosophila* homeodomain transcription factor Ind (*ind*) (intermediate neuroblasts defective) which is involved in embryonic brain development, neuroblast fate determination, procephalic ectoderm development and during stages 9-10 is expressed in the neuroectoderm of the intercalary segment (tritocerebral neuroblast subset) (Urbach and Technau, 2004).

The predicted site (cTAAT**GG**ATGt rev. compl.) also resembles binding consensus sequence of Labial (Ebner et al., 2005) which is the homolog of the mammalian HoxA1 (Mann and Chan, 1996) (§2.4). Introducing a point mutation in the putative homeodomain site of the fourth conservation block (ccattag>cca**gg**ag) in the context of the 335 bp minimum ic-CRE did not affect reporter expression. That might also be due to functional redundancy shared with the putative homeodomain binding sites of the third and first block.

The fourth conservation block consisting of one putative fork-head domain binding site and one homeodomain site is reminiscent of a {*HNF1 _HOXC* } *cis*-regulatory module where Pbx2/Prep1 homeodomain-heterodimer by binding to the homeodomain site interferes with binding of HNF1alpha (fork-head factor) to its adjacent recognition site. This results in downregulation of HNF1alpha-mediated activation (Gregory and Mackenzie, 2002).

Matrix Family	Detailed Family Information	Matrix	Anchor pos	Strand	Core sim.	Matrix sim.	Sequence
V$PAX8	PAX-2/5/8 binding sites	V$PAX8.01	13	-	1	0.891	cacTCAAtcatgc
V$HOXC	HOX - PBX complexes	V$PBX1.01	14	+	1	0.832	agcatGATTgagtgcct
V$PBXC	PBX1 - MEIS1 complexes	V$PBX1_MEIS1.32	14	+	1	0.796	agcaTGATTgagtgcct
I$ZEST	Zeste transvection gene product	I$ZESTE.01	16	+	1	0.975	gattGACTTgcc

Fig. 2_28. *In silico* predictions in the sixth conservation block.

Submitting the sixth conservation block to MatInspector predicts a putative Zeste recognition site (sequence matches binding consensus CRCTCR in the reverse-complement orientation). The overlapping seven nucleotide sequence TGATTGA conforms to the eight nucleotide PCE (Pbx consensus element) sequence [TGATTGAT] (Knoepfler and Kamps, 1997) and it is also predicted by the program scoring the binding matrix of Hox-Pbx complexes (Lu et al., 1994). Interestingly, this '7 bp element' is also present in the endogenous *hh* promoter region (position +60_+54) (§3.3), while it is encountered only once in the 6.43 kb upstream *hh* enhancer, exactly at this position falling within the sixth conservation block of the ic-CRE (position -3539_-3533). Two conserved putative Zeste sites can be found also in the promoter as well as at other positions of the enhancer (§3.3, 3.4).

Pbx is the mammalian homolog of Dm_Exd (*extradenticle*) which is involved in embryonic brain development (tritocerebral commisure is missing in mutants; Nagao et al., 2000). Pbx proteins are members of the TALE superclass (*three amino-acids loop extension*) of atypical homeodomain proteins. TALE members are characterized by a three-residue insertion in the first helix of the homeodomain being involved in interaction with Hox homologs partners, resulting in cooperative DNA binding to target sites. In *Drosophila* such a functional interaction between Exd and Hth (*homothorax*) has been reported (Ebner et al., 2005). In addition, Pbx homologues share a PBC domain (for *Pbx* and *ceh-20)* directly upstream adjacent to the TALE homeodomain (Bürglin, 1997). Finally, a C-terminal tail conserved in Exd improves affinity of monomeric Pbx1 DNA binding (Green et al., 1998). This C-terminal tail is termed the 'Hox cooperativity motif' (HCM) as it also enhances cooperative DNA binding by the minimal Pbx1-TALE homeodomain with a Hox homolog partner (Chang et al, 1995; Jabet et al., 1999).

Point mutation inserted in the sixth conservation block in the context of the 335 bp minimum ic-CRE sequence (**TGATTGA**GTG>**TGATTCA**TTG) abolished reporter expression. However, the same point mutation in the context of the 450 bp ic-CRE causes late derepression of reporter expression both in ectopic cells of the intercalary and the rest of the trunk (Fig. 2_29; §3.4). Ectopic expression in dorsal epidermal cells (and cells of the amnioserosa) in early embryos is also observed.
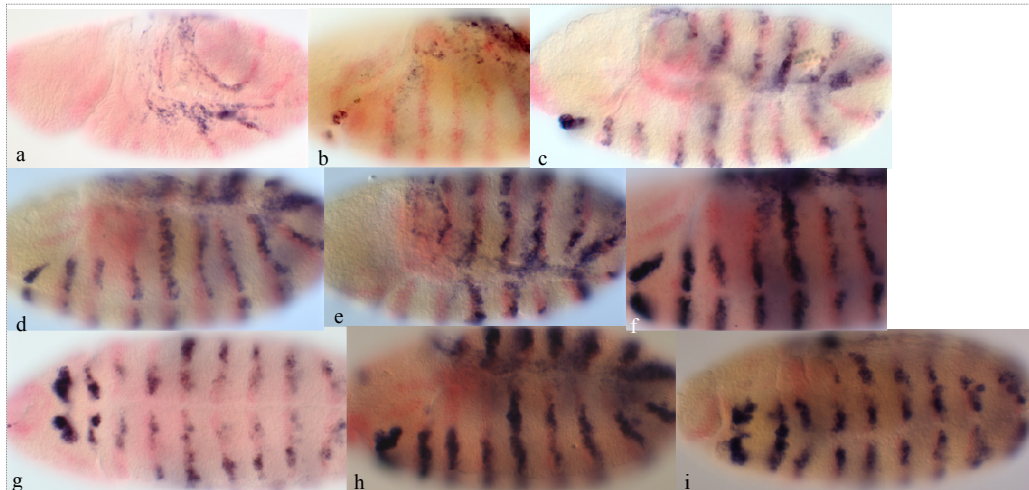
Fig. 2_29. *450_bp ic-CRE_zeste point mut.* Progress of the derepression effect of the point mutation on expression of the reporter during development. a, st. 8; b, st. 9; c-f, st. 10. g-i, st. 11. d, e ventral-lateral and dorso-lateral views of the same embryo, respectively. At early embryos (a, b) reporter is ectopically expressed in some dorsal epidermal cells (b) and cells of the amnioserosa. Expression of the reporter in intercalary segment is increased in intensity (c) in comparison to the pattern driven by the wild-type sequence (Fig. 2.35 III). At later stages, apart from the gradual trunk derepression effect (c-i) reporter is also ectopically expressed in cells of the intercalary lobes that do not express *hh* (g-i).

2.3 Transcriptional response of the ic-CRE expression to Slp1 activity

Due to the great similarity to the Slp1 footprinted site (Yu et al., 1999), the predicted fork-head domain binding sequence found within the fourth conservation block of the 335 bp ic-CRE (§2.2.3) was tested in mobility shift assays using *in vitro* expressed Slp1 protein. This did not however lead to specific complex formation, meaning that the particular DNA sequence is not *in vitro* recognized by Slp1. Still, site-directed point mutagenesis of this site (TCAACA>TCTTCA) in the context of the 335 bp ic-CRE caused loss of reporter expression. The mutation effect was also verified in the context of the 450 bp ic-CRE. Therefore, the identified site is bound *in vivo*, possibly by a member of the *Drosophila* winged-helix factors super-family, and this interaction may as well be involved in specifying expression outcome of the ic-CRE. Possibility that the site is recognized *in vivo* by Slp1, even with small affinity, and that this interaction contributes to positive transcriptional outcome of the ic-CRE, cannot be excluded for two reasons: a) the 450 bp ic-CRE was tested in *slp1* loss-of-function background using maternal {*pnos_GAL4/GCN4_bcd 3' utr*} input driving expression of a UAS-*slp1* hairpin (VDRC). Detection of the reporter transcripts in the intercalary segment is reduced (Fig. 2_30) (but not abolished as in heatshock-induced *collier* antisense RNA production; §2.7.1) in response to downregulation of Slp1 expression and activity after RNAi (still transcript

remnants are detectable by *in situ* hybridization Fig. 2_30). b) Although it has been shown to function as a transcriptional repressor of the *ftz* promoter (Yu et al, 1999), Slp1 has been postulated to exert a direct activating function on *wingless* expression (Cadican et al., 1994), meaning that an as yet unidentified intrinsic activation potential cannot be excluded.

  slp1 expression in the intercalary segment of stage 10-11 embryos overlaps only the *wg* expressing cells of the anterior part of the intercalary segment but not the *hh* expressing cells of the posterior part of the segment. However, the ic-CRE positive response to Slp1 activity does not seem likely to be mediated by *wg* signaling activity for two reasons: i) *wg* expression in the intercalary segment comes up later (st.10) than the relatively early *hh* expression onset (st. 8), therefore *wg* activity should not be involved, at least not in the establishment of the intercalary *hh* expression pattern ii) *wg* signaling activity has actually been implicated to restrict rather than to maintain *hh* expression in the cells of the posterior part of the intercalary segment, in the context of an anterior head-segment-specific cross-regulatory interaction (Gallitano-Mendel and Finkelstein, 1999). However, in early embryos (6-8) *slp1* RNA transcription is detected in the procephalic domain overlapping the presumptive intercalary anlage (Grossniklaus et al., 1994). Thus, it is possible that Slp1 protein still present in cells of the intercalary segment is involved in the establishment of *hh* expression in the posterior part of the segment. Nevertheless, in the case of *en* posterior segment polarity gene expression, Slp1 activity has been implicated in repressing it (while required for anterior *wg* activation). However, it cannot be excluded that *hh* and *en* are independently and differentially regulated by Slp1 in the intercalary segment. In any case, further analysis should be performed to verify or to exclude transcriptional responses of the ic-CRE and endogenous *hh* expression to Slp1 activity in the intercalary segment.
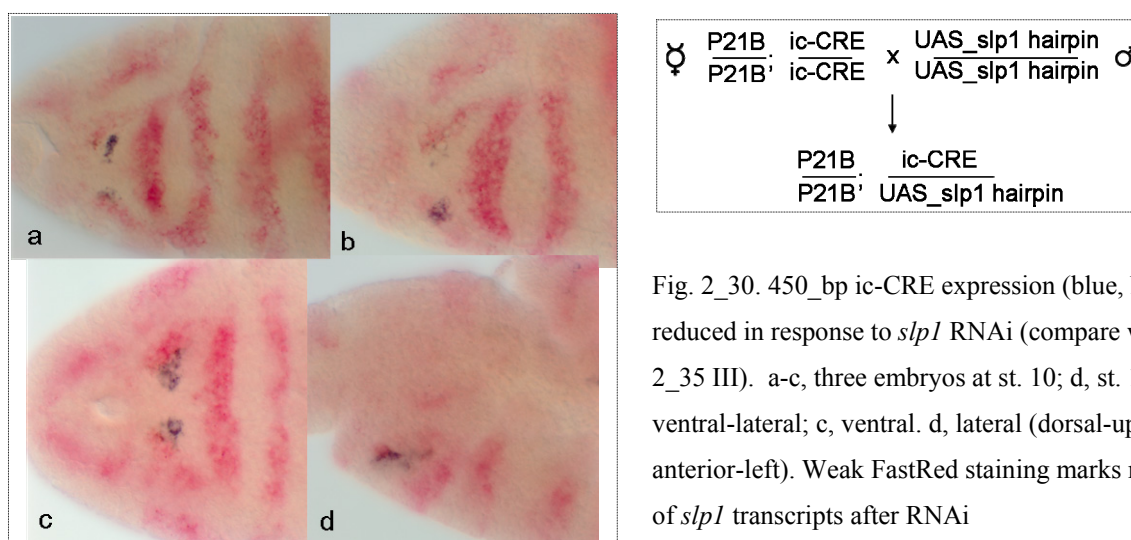
Fig. 2_30. 450_bp ic-CRE expression (blue, NBT) is reduced in response to *slp1* RNAi (compare with Fig. 2_35 III). a-c, three embryos at st. 10; d, st. 11. a, b ventral-lateral; c, ventral. d, lateral (dorsal-up, anterior-left). Weak FastRed staining marks remnants of *slp1* transcripts after RNAi

2.4 The ic-CRE does not respond to Labial activity

As mentioned in §2.2.3, the putative homeodomain binding site found within the fourth conservation block of the 335 bp ic-CRE (TAATGGATGT rev. compl.) resembles Labial binding consensus (TGATGGATKT; Ebner et al., 2005). One more site resembling binding specificity of Labial is found at position (-3797_-3806 bp reverse complement) GGATGGATGT just 5' of the first conservation block (Fig. 2_22, 2_20). *labial* belongs to the ANTC-Hox cluster and marks the intercalary segment across arthropod species (Abzhanof and Kaufman, 1999). Therefore, it was interesting to test if expression outcome of the ic-CRE would respond to Labial activity, although affecting the putative homeodomain site by point mutagenesis (§2.2.3) did not alter intercalary-expression of the ic-CRE. Expression of the ic-CRE was not affected either when tested in a *labial* loss-of-function mutant background (*lab^14*, Diederich et al., 1989). Also the endogenous *hh* expression pattern seems unaltered in those mutants. *labial* is expressed anterior to the PS0/PS(-1) parasegmental border, actually overlapping the *wg* expressing cells of the intercalary segment and abutting the intercalary *hh* stripe (Fig. 2_31 a,b). By *in situ* hybridization it was not possible to detect anterior derepression of *hh* or the ic-CRE expression pattern in *labial* null mutants. This however has to be repeated with a more critical examination for anterior expansion of only one or two cells. This is exactly the overlap between expression of the Collier activator (§2.7.1) in the posterior part of the intercalary segment and Labial in the anterior (Fig. 2_31, c, d).
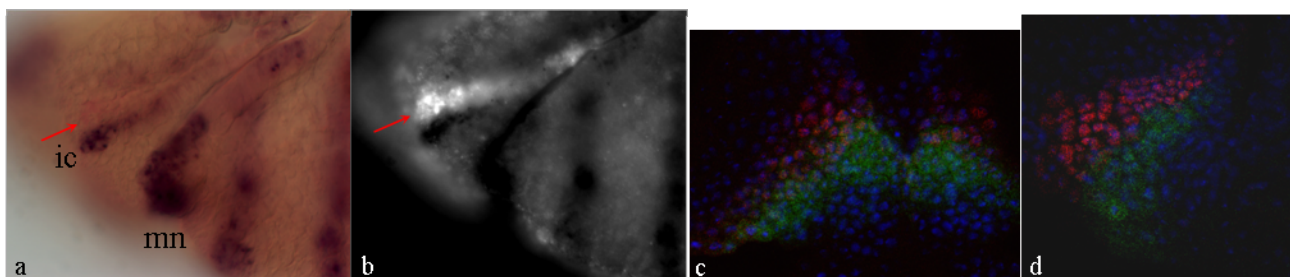
Fig. 2_31. *labial* (Fastred staining – fluorescent, red arrow) abuts the *hh* cells in the intercalary segment (a, b). c, d: Fluorescent immunostaining Alexa 488 (green) Collier, Cy3 (red) Labial. c anterior-up, d anterior-left, dorsal-up. Collier and Labial overlap by 1-2 cells.

2.5 Temporal control of the ic-CRE early onset of expression

During dissection analysis of the 1 kb ic-CRE, it became apparent that the minimum 335 bp element 'F5_R4' (-3799_3465 bp) lacks the early onset of *hh* expression at stage 8 (Fig. 2_32) while the *'γ1'* 620 bp (-4085_-3465 bp) construct ensures it (§2.2.1). Therefore, intercalary-specific expression of *hh* is not only under the control of spatial *cis*-regulatory information, but also involves temporal control.
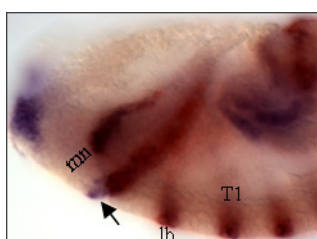
Fig. 2_32. Early procephalic detection of *hh* expression in the intercalary anlage (arrow). *hh*-NBT, *en*-FastRed).

2.5.1   5' dissection of the enhancer in order to localize temporal control *cis*-elements

In order to localize *cis*-regulatory elements conferring the early onset of the ic-CRE expression, being presumptively also involved in the initial activation process of the endogenous *hh* expression in the intercalary segment, a series of 5' shorter constructs were tested, spanning region (-4085 kb_-3799 bp), while ending at the same 3' point (-3465 bp) (Fig. 2_35 I). These constructs were designed based on the phylogenetic footprint of that region, in a means that during 'cutting-off' 5' fragments, highly conserved blocks were not disrupted (Fig. 2_37). Two conclusions could be made from this analysis: a) Positive and negative elements controlling late trunk derepression are in a way dispersed throughout the dissected enhancer fragment (§3.4) b) Sequence (-4014_-3985 bp) is required to ensure early onset of the reporter expression in the intercalary segment (Fig. 2_35 II).
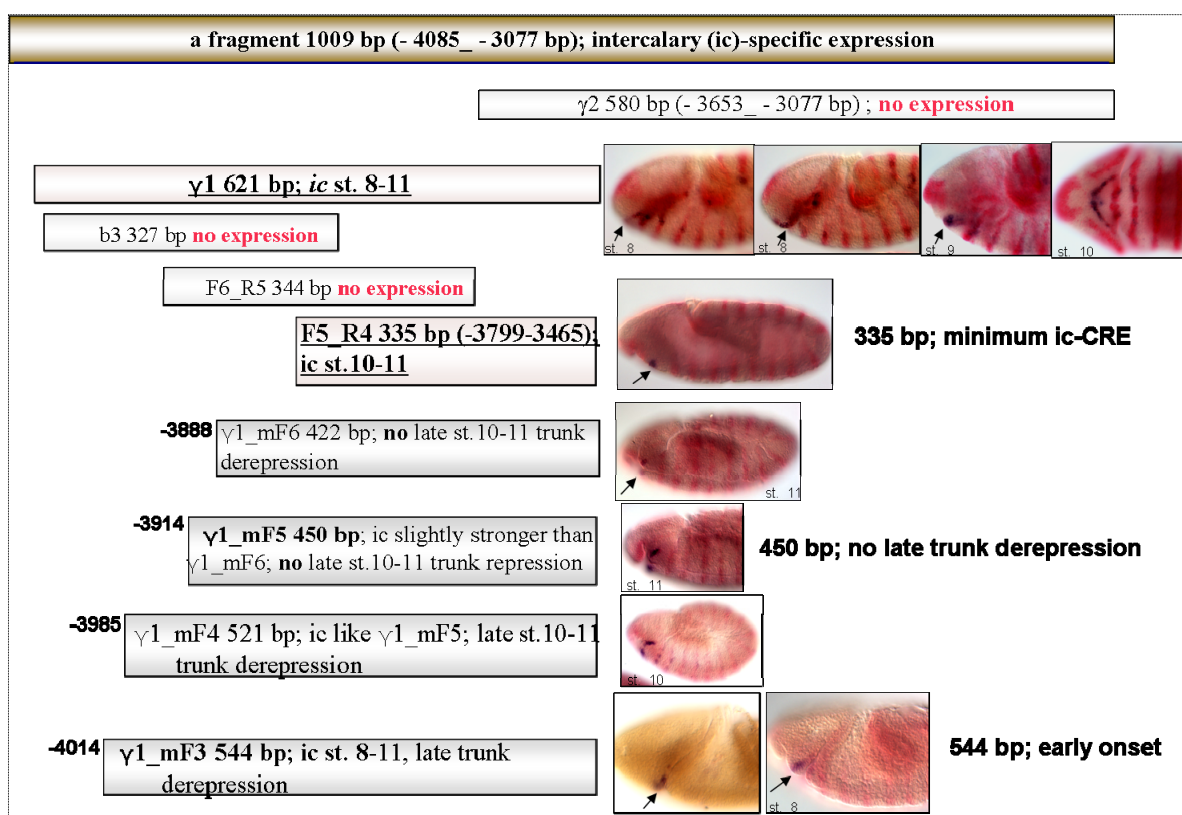
Fig. 2_35 I. 5' dissection of the γ1 fragment in order to localize *cis*-regulatory elements that generate the early onset of intercalary-expression (arrow) at stage 8.
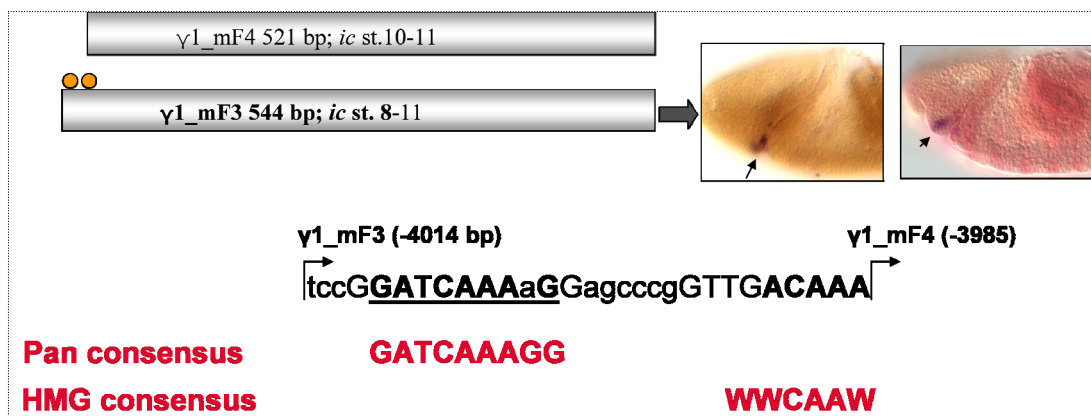
Fig. 2_35 II. Sequence (-4014_-3985 bp) ensures the early onset of expression.
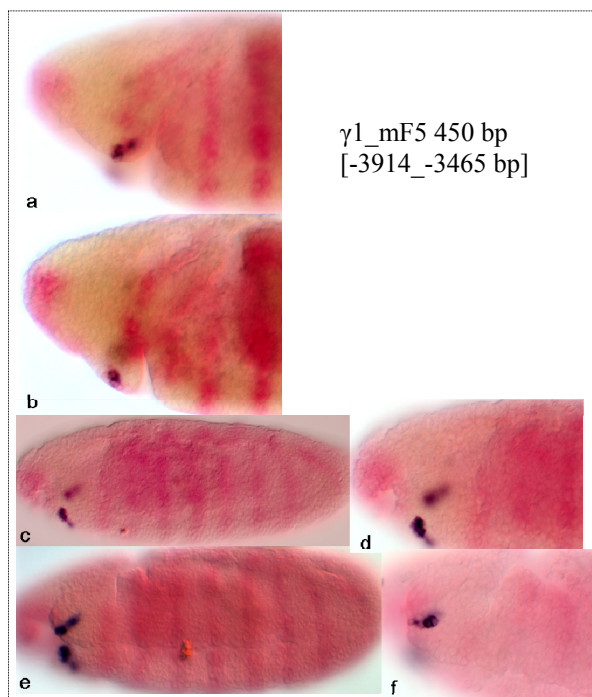
Fig. 2_35 III. 'γ1_mF5' 450 bp fragment drives expression specifically in the intercalary segment during st. 9-11. a, b different focal planes of the same embryo (st. 9). d is 40x magnification of embryo depicted in c (st. 10). e late 10. f stage 11.

2.5.2 HMG activity possibly involved in specifying the early onset of *hh* expression in the intercalary segment

Fragment (-4014_-3985 bp) that ensures early onset of *hh* expression in the intercalary segment consists of two highly conserved blocks GGAT**C**AAAaGG and GTTGA**C**AAAt, separated by a 6 bp stretch. Both sequences resemble the binding motif of HMG-box protein factors [W**C**AAAS] (reported in NCBI CDD Database; cd01388 'SOX-TCF_HMG-box'; Love et al., 1995; Werner et al., 1995). In addition, they both conform to the consensus binding sequence of HMG-box proteins of the SOX subclass [WW**C**AAW] (Churchill et al., 1995; Lefebvre et al., 2007).

In silico prediction on the 50 bp DNA sequence (-4019_-3970 bp) using MatInspector (http://www.genomatix.de/cgi-bin/matinspector_prof/mat_fam.pl) also generates a hit in the first block [GGA**TC**AAAaGG] scoring the binding matrix of dTCF (*Drosophila T-cell factor homolog or Pangolin*) which is [W**TC**AAAS] (MatInspector; Lee and Frasch 2000) (Core 4 nucleotide-sequence used by MatInspector underlined). The non-conserved 'a' nucleotide that disturbs the conservation block does not match the matrix at the corresponding position (S) (G/C). Still, the site strongly resembles consensus binding sequence of dTCF as determined by PCR-based binding site selection [GA**TC**AAA**G**G] (van de Wetering et al., 1997) which matches well the canonical Lef1/TCF binding motif [WW**TC**AAA**G**G]; (van de Wetering et al., 1991; van de Wetering et al., 1993) . Only this first block, and not the second one, scores *in silico* the binding matrix of dTCF, as it seems that a T

residue filling the (second) `W` position of the general HMG-box consensus binding sequence `WCAAAS` (or `WWCAAW`) is a prerequisite for specific recognition by the HMG-box of dTCF.

Notably, juxtaposed to the second highly conserved block, one more putative HMG binding site (`TACAAAC`) is found in the reverse-complement orientation, matching the `WCAAAS` consensus. This sequence is filtered through 11 species phylogenetic conservation, with *D. yakuba* species sequence being divergent (i.e. not in capitals in Fig. 2_37).

Based on the above *in silico* analysis performed, recognition of the 30 bp sequence conferring early expression onset of the ic-CRE by dTCF as a first candidate was tested *in vitro* in mobility shift assay. dTCF was expressed *in vitro* in a cell-free expression system (§5.7). Recognition of the DNA sequence of the putative site and binding resulting in mobility shift was tested for the wild-type 35 oligo spanning region (-4016_-3982 bp) and for a mutated version removing an 'A' from the first block sequence so that it perfectly matches the dTCF binding consensus (Fig. 2_36). The wt oligo generates only a very weak shift complex. The optimal sequence is recognized by dTCF producing a shift complex which is merely competed by 100x molar excess of wt oligos corresponding to each of the highly conserved blocks ('21 bp TCF' site and '18 bp HMG site' competitors). Thus, dTCF *in vitro* does not efficiently bind the oligo sequence of the first conservation block, despite the *in silico* prediction scoring the binding matrix of dTCF. This is most probably because at the corresponding position 7 of the dTCF binding consensus (`WTCAAAS`) the wt oligo sequence bears an A instead of a 'strong' `S(G/C)` nucleotide. This is also supported by recognition and binding *in vitro* of the optimal mutant probe (`GGATCAAAaGG`> `GGATCAAAGG`). Therefore, occurrence of a 'strong' (`G/C`) nucleotide at position 7 of the matrix seems to be a prerequisite for specific recognition and binding by the HMG-box domain of dTCF.

In vitro dTCF does not recognize either the second conservation block sequence (`GTTGACAAAtg`) as it can be concluded by the competition experiment. This is most probably because the second conserved sequence resembles dTCF binding consensus [`WTCAAAS`] only at positions 3, 4, 5, 6 (`CAAA`) but not at positions 1 (`W>G`), 2 (`T>A`) and 7 (`S>A`). Nevertheless, the second conservation block sequence could be recognized *in vivo* by another member of the HMG domain family, as it still conforms to the consensus binding sequences `WCAAAS` or `WWCAAW`, with one mismatch to the last or to the first nucleotide residue, respectively (§3.2).

In conclusion, region (-4014_-3975 bp) consists of three highly conserved blocks, the first two (-4014_-3985 bp) filtered through 12 species conservation, the third through 11 species conservation. All conform to the consensus binding sequence of HMG-domain proteins. Specifically, the first block scores with one mismatch the binding matrix of dTCF HMG-box protein (Pangolin), the wt

oligo sequence however is not shifted in EMSA by *in vitro* expressed dTCF. Since from the 5' dissection analysis it is concluded that fragment (-4014_-3985 bp) can confer early onset of the ic-CRE expression, and also supported from the *in silico* analysis, involvement of HMG activity in the early temporal control of intercalary-specific expression of *hh* and proposal of another candidate distinct from dTCF are issues further discussed in §3.2.

Fig. 2_36. Only the mutated version of the probe (optimal) that perfectly matches the dTCF binding consensus generates substantial shift complex formation (lane 2) which is merely competed by the wild-type sequences (lanes 3, 4 competition at 100x molar excess).

#1 wt probe (36 bp)   GGTCCGGATCAAA**A**GGAGCCCGGTTGACAAATGTT

#2 optimal probe (3 A) (35 bp)   GGTCCGGATCAAAGGAGCCCGGTTGACAAATGTT

#3 optimal probe; 100x TCF site competitor (21 bp)   GGTCCGGATCAAAAGGAGCCC

#4 *optimal probe;100x HMG competitor* (18 bp)   GCCCGGTTGACAAATGTT

A summary of the ic-CRE components presented in §2.2-2.5 is depicted in Fig. 2_37.

Fig. 2_37. A summary of the ic-CRE sequence (-4014_-3465 bp) and its components mentioned in §2.2-2.5. Short blue arrows mark 5' ends of Forward and Reverse primers spanning the Collier site used in qPCR (§ 2.7.3).

2.6 Further dissection of the *hh* upstream enhancer

As mentioned in §2.1.4, a 6.43 kb upstream enhancer sequence of *hh* confers expression of the reporter also in the other two procephalic segments, namely the ocular and the antennal. In an effort to localize *cis*-regulatory information governing *hh* expression in these two segments, enhancer region [-6.43 kb_-4 kb] was further dissected by testing shorter overlapping subfragments of ~600 bp size (Fig. 2_38). However, none of these fragments could mediate segment-specific embryonic expression pattern which raised the question if this is due to disrupting individual *cis*-binding elements during the dissection analysis, or because a high degree of synergy between individual *cis*-elements dispersed along different regions the enhancer is required to determine anterior head segment-specific expression in the ocular and the antennal segments. If this second possibility is indeed a prerequisite to be fulfilled, then a molecular mechanism mediating functional interactions among distant enhancer elements (and/or with core promoter elements) has to be considered (§3.1.5).
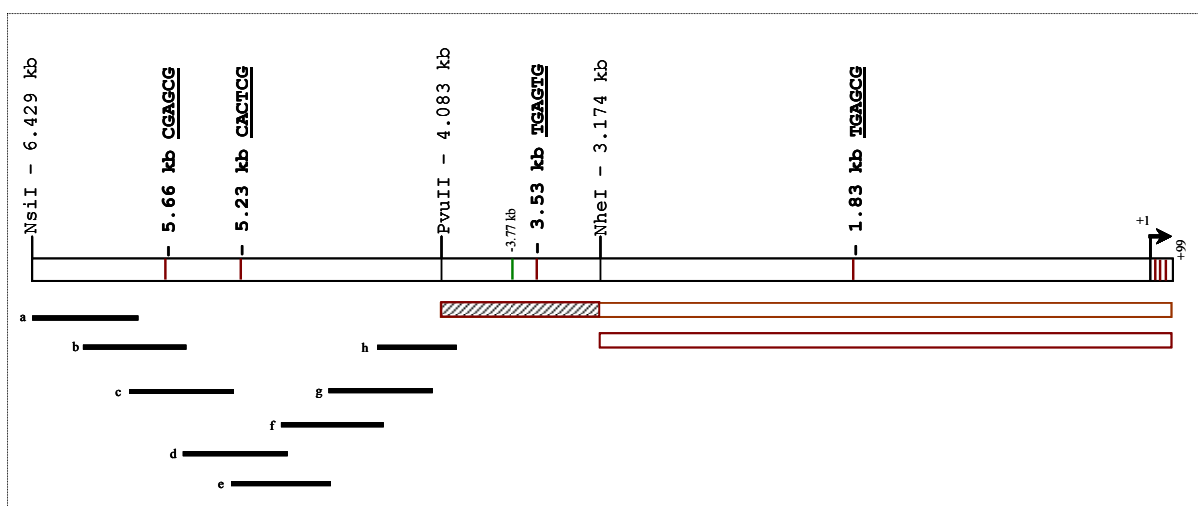
Fig. 2_38. Full-length upstream enhancer of *hh* (-6.43 kb) sufficiently drives head and trunk segmented expression pattern of the reporter (Fig. 2_14). Overlapping fragments (a-h) of ~600 bp spanning *hh* upstream enhancer (-6.43_-4.08 kb) do not mediate specific expression pattern. The ic-CRE (-4.083_-3.174 kb) is marked. Green bar at -3.77 kb represents the functional identified Collier binding site (§2.7). In total four conserved putative zeste sites can be found in the 6.43 kb upstream enhancer (CRCTCR).

2.7 Analysis of Collier DNA-binding activity on the ic-CRE

In silico analysis performed on the ic-CRE sequence revealed a site within the second conservation block that scores the binding matrix of the COE-HLH factor Olf-1 (Olfactory factor 1) (§2.2.3).

Olf-1/ Ebf (Early B cell factor) is the mammalian homolog of Dm_Collier (Crozatier et al., 1996). Because of the high degree of percentage identity in the DNA binding domain shared by the homologues which reaches up to 86% implying a similar DNA binding specificity, the predicted sequence was regarded as a putative Collier recognition site. Further analysis was performed to elucidate this putative DNA-protein interaction as it shows quite an interest from a functional aspect; Collier was proposed to function as an activator of segment polarity gene expression in the intercalary segment and was the first candidate factor proposed to act as a second-level regulator in patterning of the anterior head region, while being under the strict control of *btd* (Crozatier et al., 1999; Crozatier et al., 1996). Intercalary-specific expression pattern of segment polarity genes is lost in *collier* null mutants. Since Collier expression does not overlap with *wg* expressing cells of the intercalary segment, dependence of *wg* expression upon Collier is most probably mediated by the *hh* pathway. Also the question whether Collier-dependent intercalary-expression of *hh* depends on a direct interaction at protein-DNA level that triggers transcriptional activation of *hh* in the posterior cells of the intercalary segment was left open.

Identifying a Collier DNA-binding site, functional within the ic-CRE, would explain functional dependence of *hh* upon Collier activity specifically in the intercalary segment. Analysis further proceeded to prove that the underlying molecular mechanism involves a direct protein DNA-interaction leading to transcriptional control of segment polarity gene expression. That would further support the role of Collier as a second level regulator in the anterior head region, without of course excluding the possibility that head gap-like genes may also contribute a transcriptional control input in procephalic segment polarity gene expression.

2.7.1. Intercalary-specific expression of *hh* mediated by the ic-CRE is under positive regulatory control exerted by Collier

To determine whether the ic-CRE expression outcome is under positive control by Collier like the endogenous *hh* intercalary-pattern, the 450 bp ic-CRE was brought into a *collier* loss-of-function mutant background (*col[1]*; Crozatier et al., 1999; Fig 2_38). *col[1]* bears a nucleotide substitution (G>A) that eliminates a splice acceptor site resulting in the non-removal of intron 6 and the production of a truncated protein ending at aminoacid position 228.

The 450 bp (γ1mF5_hhR4) ic-CRE driving tGFP reporter expression was also crossed to a generated transgenic line that produces heatshock-inducible ColA_CDS_antisense RNA (line HS.ColA_antisense #M3, §5.3). Remnants of *collier* transcripts are still detectable by FastRed staining but ic-CRE expression is abolished (Fig. 2_40).

Abolishment of the ic-CRE-mediated reporter expression, as well as the intercalary expression pattern of *hh,* in *col* loss-of-function mutant background supports that intercalary-specific expression of *hh* is under positive regulatory function of Collier, mediated by the ic-CRE transcriptional activating input.
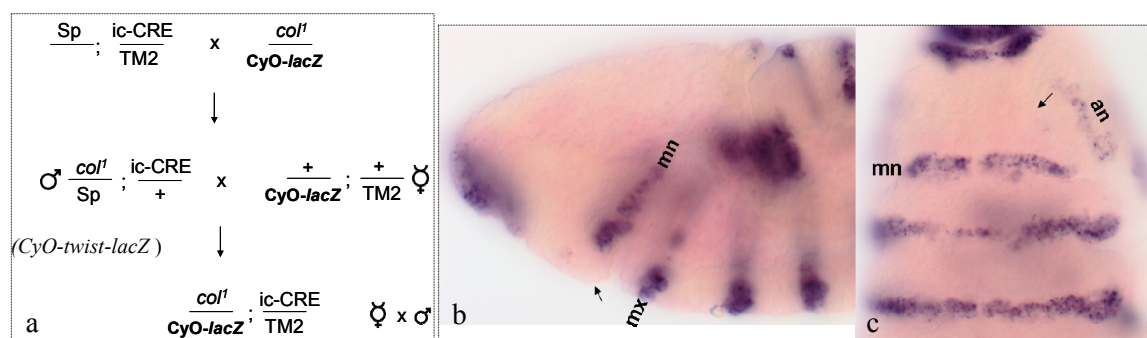
Fig. 2_39. Probes; *hh, lacZ* (NBT); *tgfp* (Fred). *collier* null embryos specifically lack intercalary expression pattern of *hh* and ic-CRE expression is abolished. Dig-labeled probe of *lacZ* is also added in the hybridization to distinguish the nulls, as the second chromosome balancer carries a *twist-lacZ* insertion. b (st. 9), anterior-left; c (st. late 10), anterior-up.

Fig. 2_40. Heat-shock inducible transgenic RNAi to knock-down *collier* activity. b: The ic-CRE mediated expression is abolished (no NBT staining against DIG-labeled probe of the *tgfp* reporter) while remnants of *collier* transcripts are still detectable (weak FastRed staining against the Fluo-labeled probe of *collier*). c: ic-CRE mediated expression is retained in embryos that have been heat-shocked but do not bear the HS.ColA_antisense insertion (genotype from the crossing scheme is ic-CRE/TM2)

2.7.2 *In vitro* analysis of Collier-DNA binding interaction

Recognition of the Collier-binding site identified *in silico* within the ic-CRE was tested *in vitro* in mobility shift assays. The first mobility shift assays were performed using crude nuclear embryonic extracts (0-10.5 h, §5.8). A 31 bp oligo encompassing the putative Collier binding site was used as a probe (cggcagCAAtTCCCCAAtGGCaTTTcACtta) (-3777_-3747 bp). The oligo is recognized by the extracts resulting in formation of three complexes (two major and one minor – slower migrating – one; Fig. 2_41) which are competed by including increasing amounts of unlabeled

competitor at 50x, 100x, and 150x molar excess. Complex formation was downregulated after preincubating the extract with an antibody against Collier, without however leading to supershift band detection (not shown). That observation triggered hypothesis that Collier activity present in the crude extracts might be involved in formation of the detected complexes, which was further explored and supported by chromatin immunoprecipitation experiments using the anti-Col and cross-linked chromatin extracted from (~2-10.5 h ael) embryos (§5.4). Further EMSA experiments were performed using proteins expressed *in vitro* in a cell-free transcription and translation system (§5.7). For that purpose the open reading frame of Collier protein was cloned.
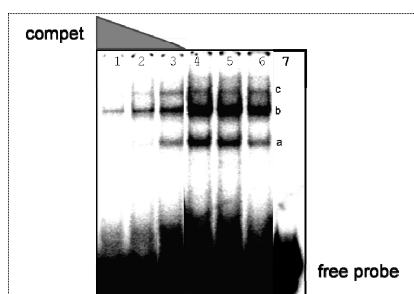
Fig. 2_41. Mobility shift assay with crude nuclear extracts and the 31bp Collier site – oligo probe. Two major complexes are generated (a, b) and one minor (c). Lane 1, 2, 3: Competition at 150x, 100x, and 50x molar excess, respectively.


```
ColC   MVTEQNADRRLLGKIIPSRFERFIIWPSQLGACPGIPGTINHPSEDPSPSPSGGQAALQL  60
ColA   ------------------------------------------------------------
ColB   ------------------------------------------------------------

ColC   VDILEWKKLLRAGLGGDVAAYADLHQHSHSHHPNPHPHPHPHPFATDGFKLRSEEPVPGF  120
ColA   ------------------------------------------------------------
ColB   ------------------------------------------------------------

ColC   SSASPWPTLELGMEWGRKLYPSAVSGPRSAGGLMFGLPPTAAVDMNQPRGPMTSLKEEPL  180
ColA   -------------MEWGRKLYPSAVSGPRSAGGLMFGLPPTAAVDMNQPRGPMTSLKEEPL   48
ColB   -------------MEWGRKLYPSAVSGPRSAGGLMFGLPPTAAVDMNQPRGPMTSLKEEPL   48
                    ********************************************

ColC   GSRWAMQFPVVDQSNLGIGRAHFEKQPPSNLRKSNFFHFVIALYDRAGQPIEIERTAFIGF  240
ColA   GSRWAMQFPVVDQSNLGIGRAHFEKQPPSNLRKSNFFHFVIALYDRAGQPIEIERTAFIGF  108
ColB   GSRWAMQFPVVDQSNLGIGRAHFEKQPPSNLRKSNFFHFVIALYDRAGQPIEIERTAFIGF  108
       ************************************************************

ColC   IEKDSESDATKTNNGIQYRLQLLYANGARQEQDIFVRLIDSVTKQAIIYEGQDKNPEMCR  300
ColA   IEKDSESDATKTNNGIQYRLQLLYANGARQEQDIFVRLIDSVTKQAIIYEGQDKNPEMCR  168
ColB   IEKDSESDATKTNNGIQYRLQLLYANGARQEQDIFVRLIDSVTKQAIIYEGQDKNPEMCR  168
       ***********************************************************

ColC   VLLTHEVMCSRCCDKKSCGNRNETPSDPVIIDRFFLKFFLKCNQNCLKNAGNPRDMRRFQ  360
ColA   VLLTHEVMCSRCCDKKSCGNRNETPSDPVIIDRFFLKFFLKCNQNCLKNAGNPRDMRRFQ  228
ColB   VLLTHEVMCSRCCDKKSCGNRNETPSDPVIIDRFFLKFFLKCNQNCLKNAGNPRDMRRFQ  228
       ***********************************************************

ColC   VVISTQVAVDGPLLAISDNMFVHNNSKHGRRAKRLDTTEGTGNTSLSISGHPLAPDSTYD  420
ColA   VVISTQVAVDGPLLAISDNMFVHNNSKHGRRAKRLDTTEGTGNTSLSISGHPLAPDSTYD  288
ColB   VVISTQVAVDGPLLAISDNMFVHNNSKHGRRAKRLDTTEGTGNTSLSISGHPLAPDSTYD  288
       ***********************************************************

ColC   GLYPPLPVATPCIKAISPSEGWTTGGATVIIVGDNFFDGLQVVFGTMLVWSELITSHAIR  480
ColA   GLYPPLPVATPCIKAISPSEGWTTGGATVIIVGDNFFDGLQVVFGTMLVWSELITSHAIR  348
ColB   GLYPPLPVATPCIKAISPSEGWTTGGATVIIVGDNFFDGLQVVFGTMLVWSELITSHAIR  348
       ***********************************************************

ColC   VQTPPRHIPGVVEVTLSYKSKQFCKGSPGRFVYVSALNEPTIDYGFQRLQKLIPRHPGDP  540
ColA   VQTPPRHIPGVVEVTLSYKSKQFCKGSPGRFVYVSALNEPTIDYGFQRLQKLIPRHPGDP  408
ColB   VQTPPRHIPGVVEVTLSYKSKQFCKGSPGRFVYVSALNEPTIDYGFQRLQKLIPRHPGDP  408
       ***********************************************************

ColC   EKLQKEIILKRAADLVEALYSMPRSPGGSTGFNSYAGQLAVSVQDGSGQWTEDDYQRAQS  600
ColA   EKLQKEIILKRAADLVEALYSMPRSPGGSTGFNSYAGQLAVSVQDGSGQWTEDDYQRAQS  468
ColB   EKLQKEIILKRAADLVEALYSMPRSPGGSTGFNSYAGQLAVSVQDGSGQWTEDDYQRAQS  468
       ***********************************************************

ColC   SSVSPRGGYCSSASTPHSSGGSYGATAASAAVAATANGYAPAPNMGTLSSSPGSVFNSTS  660
ColA   SSVSPRGGYCSSASTPHSSGGSYGATAASAAVAATANGYAPAPNMGTLSSSPGSVFNSTS  528
ColB   SSVSPRGGYCSSASTPHSSGGSYGATAASAAVAATANGYAPAPNMGTLSSSPGSVFNSTS  528
       ***********************************************************

ColC   R--VSSLSFNPFALPTCNT-----------QGYSTQLVTSTK----  689
ColA   R--VSSLSFNPFALPTCNT-----------QGYSTQLVTSTK----  557
ColB   MSAVSSTWHQAFVQHHHAATAHPHHHYPHPHQPWHNPAVSAATAAAV  575
       ***  .:.*.    :              * : .  *:::.
```

Fig. 2_42. Alignment of the annotated Collier isoforms. Collier A and Collier B share the same N-terminal 528 aa. Collier B has a His-rich (11/29 aa) C-terminus.

In *D. melanogaster* two Collier isoforms are present resulting from a developmentally regulated alternative splicing event (Crozatier et al., 1996). Specifically, the two protein isoforms share the same first 528 aa and differ in the last 29 aa for Collier A (Col2; Crozatier et al., 1996) and 47 aa for Collier B (Col1) which constitute a His-rich C-terminus specific for the B isoform. Calculated MW for the two proteins are 60.5 and 62.5 kD respectively. Collier C isoform included in the alignment of Fig. 2_42 is encoded from a weakly supported annotated transcript in Flybase. However, by using

a forward primer specific for the 5' of the coding sequence of Col_C open reading frame, it was not possible to isolate it by PCR from *Drosophila* embryonic cDNA pool and only the open reading frames of A (Col2) and B (Col1) isoforms were cloned. As reported in Crozatier et al. (1996), two major transcripts of length 3.9 and 3.4 kb, containing the respective open reading frames of ColB 1725 nucleotides and ColA 1671 nucleotides, were detected on developmental northern blots. ColB transcript is first detected at 4 h after egg lay, while both transcripts display a peak between 8-16 hours of development. The ColA and ColB cDNAs differ from each other by 465 nucleotides which are removed by a developmentally regulated splicing event (Crozatier et al., 1996). This results in two isoform-specific C-terminal regions (depicted in the alignment of Fig. 2_42; 29 aa C-terminus for ColA and 47 aa His-rich C-terminus for ColB). The intronic for Collier A, still coding for Collier B, 465 bp sequence (Fig. 2_43) was cloned (with primers ColB_specific_F/Rev; Table 5-1.1) in order to generate a labeled antisense RNA probe that specifically hybridizes with transcripts of Collier B in *in situ* whole-mount embryonic hybridization. Cytoplasmic transcripts detected with this probe are only of ColB while detectable nuclear dots correspond to total nascent transcripts (nuclear-dot staining is not distinguishable between ColA and ColB cell-group-specific expression). No tissue-specific or cell-group specific difference in the expression of the two splicing isoforms could be indicated from double *in situ* hybridization assays using the ColB-specific probe and a Fluo-probe that detects both transcripts (ColA_CDS_antisense). Therefore, both proteins were considered to be potentially involved in regulation of the ic-CRE and the endogenous *hh* expression in the intercalary segment, and they were both expressed *in vitro* to be used in EMSAs.
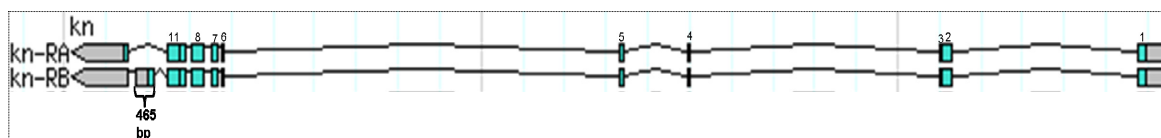
Fig. 2_43. Organization of ColA and ColB transcriptional units (*col* is referred as *kn; knot* in Flybase). Light blue and black bars correspond to coding sequence (exon 11 is the last common exon of the two splicing isoforms). Grey bars are untranslated regions of the transcribed mRNA. Lines are intronic regions. The 465 bp sequence depicted was subloned from the ColB cDNA to be used as a template for a ColB-specific antisense RNA probe.

EMSAs were performed testing two oligo sequences; the wild-type and an 'optimal' mutant version of probe that bears two nucleotide exchanges so that it perfectly matches the binding matrix of Olf-1 (§2.2.3) (Fig. 2_44 a.) Both probes generate formation of a complex shift with the two Collier splicing protein isoforms (Fig. 2_44 b). The optimal probe is recognized with slightly higher affinity. One major complex is formed (C1) which is most probably due to Collier homodimer and/or heterodimer binding when both proteins are present in the reaction. Dimerization upon binding is mediated by the Helix-Loop-Helix (HLH) motif present in all members of the COE transcription factor family (Daburon et al., 2008; §3.6). In the case of Collier B one faster migrating

complex (C2) is detectable which is most probably due to Collier B binding also as a monomer. Modest monomeric binding activity has also been reported for the mammalian homolog EBF (Early B-cell factor) (Hagman et al., 1995).

a)

Col site 31 bp wt   cggcag**CAAtTCCCCAAtGGCaTTT**c**AC**tta

Col site 31 bp optimal   cggcag**CAAtTCCCCAAGGGCTTTT**c**AC**tta

b)

mock	++	+	–	+	–	–	++	+	–	+	–	–
ColA	–	–	–	+	++	+	–	–	–	+	++	+
ColB	–	+	++	–	–	+	–	+	++	–	–	+
Lanes (#)	1	2	3	4	5	6	7	8	9	10	11	12

(+) 2 ul

← C₁ homodimer/ heterodimer

← C₂ ColB monomer

CD**ABTCCCYNRRGDVB**NBVN   Olf-1 consensus

1-6 Col site wt 31 bp   cggcag**CAAtTCCCCAAtGGCaTTT**c**AC**tta

7-12 Col site optimal   cggcag**CAAtTCCCCAAGGGCTTTT**c**AC**tta
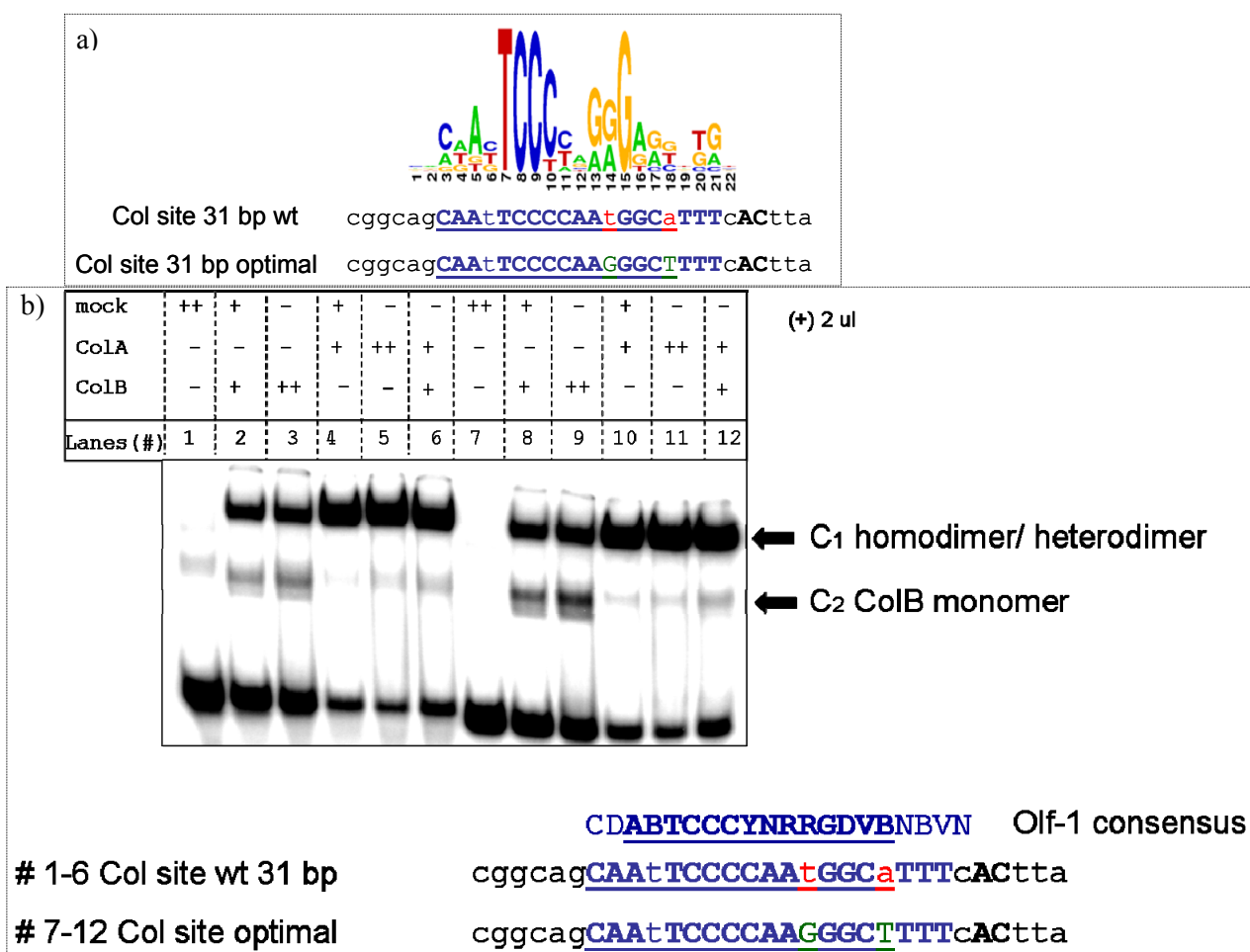
Fig. 2_44. a) Alignment of the wild-type and optimal Collier recognition sequences (-3777_-3747 bp) to the Olf-1 binding matrix. Capitals represent 12 species conservation. b) Mobility shift assay with the two *in vitro* expressed Collier isoforms and the wild type and optimal sequence probes. Both probes lead to complex shift formation with both proteins. The optimal probe is recognized and bound with slightly higher affinity. In the case of Collier B one faster migrating complex can be detected (C2).

It is also apparent that the homodimer of Collier A binds the oligo with more affinity than the homodimer of Collier B and that the AB heterodimer binds with intermediate affinity (compare complex C1 intensity in lanes 2 and 4 with lane 6 and the amount of free unbound probe in each case in Fig. 2_44 b).

Differential binding affinities of the homodimers and their heterodimer situation was further examined in titration experiments. While keeping steady the amount of Collier B, adding additional amounts of Collier A leads to increase of C1 complex formation, while C2 formation decreases. This is presumptively due to heterodimerization of Collier B to Collier A protein molecules, with the heterodimer situation binding the oligo sequence with less affinity than the homodimer of Collier A, since while keeping steady the amount of this factor, increasing the presence of Collier B attenuates C1 complex formation (Fig. 2_45 a).
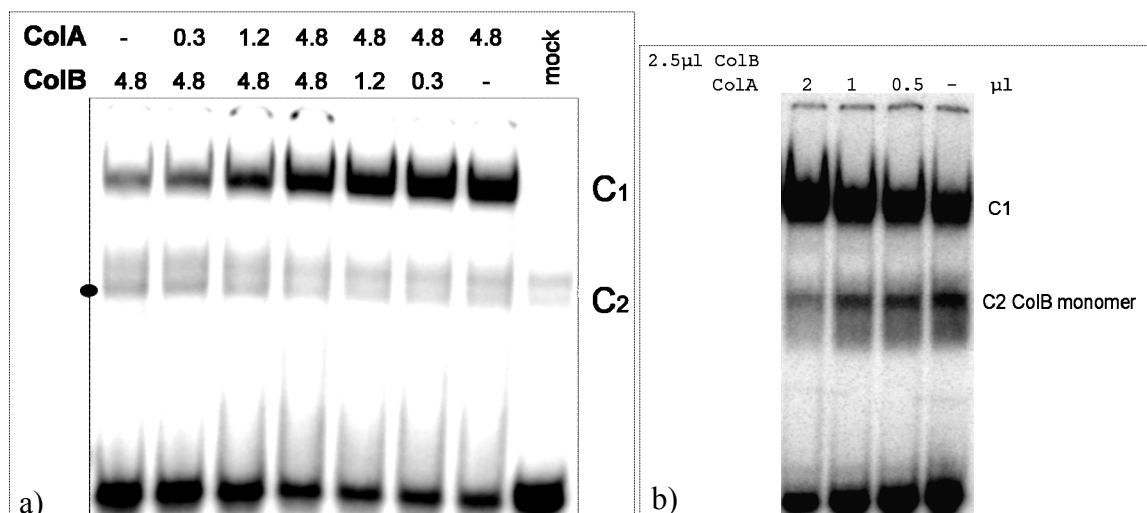
Fig. 2_45. a) Titration experiment. C1 homodimer/heterodimer, C2 ColB monomer (at the height of the black spot). The AB heterodimer binds the oligo with intermediate affinity, lesser than the A homodimer and greater than the B homodimer. b) Overexposed gel to show that heterodimer binding prevails over Collier B monomer binding.

Specificity of Collier-DNA complex formation was assessed in competition experiments using unlabeled wild type or point mutant sequence versions of oligo competitors at 100x molar excess (Fig 2_46). Although presence of the wild type competitor can reduce probe-complex formation (lanes 2, 8), an oligo bearing mutation in the core of the binding site is not recognized, thus incapable of competition at 100x molar excess (lanes 4, 10). A sequence mutated 5' of the core competes shift formation, but not as efficiently as the wild type version, implying that the affected nucleotides are recognized and participate in strengthening DNA binding as well (lanes 3, 9). Notably, the same point mutations *in vivo*, both tested in the context of the 335 bp minimum ic-CRE abolish reporter expression (§2.2.3). Competition events affect likewise formation of the C2 complex, implying that specific DNA interaction *in vitro* can be mediated by Collier B monomers as well.

Ability of a monoclonal anti-Col antibody (gift from Michelle Crozatier and Alain Vincent, Toulouse, France) to super-shift Collier-DNA complex was also tested (§5.8). Preincubation with anti-Col leads to partial supershift of the C1 complex (Fig. 2_47). This means that the anti-Col used is, at least *in vitro*, able to recognize and stably bind to the protein factor while the latter is in binding complex with a DNA oligo sequence. Since the anti-Col recognizes the transcription factor in the DNA-binding conformation this suggests that the same antibody can be used in chromatin immunoprecipitation experiments after *in vivo* crosslinking.

1-6; ColA 2 ul

7-8; ColB 2ul

2, 8; 100x wt compet

3, 9;100x mut1 compet

4, 10; 100x mut2 compet

5, 11; 100x oligo3 compet

6, 12; 100x oligo3 mut compet

Olf-1 consensus CDABTCCCYNRRGDVBNBVN

tcggCAGCAATTCCCCAATGGCATTTCAC

tcggCAagcATTCCCCAATGGCATTTCAC

tcggCAGCAATgaaCCAATGGCATTTCAC

ATCCCAATCCCTGGTAGCCGTAAAT

ATCCCAACGGCTTTTAGCCGTAAAT

Fig. 2_47 Supershift reaction. Arrow depicts the supershift band.

Fig. 2_46 Competition experiment using wild type and mutant versions of oligo sequences. Competitor 'mut2' which bears a mutation in the core of the binding site does not compete; competitor 'mut1' bearing a mutation at the 5' of the binding sequence competes but not as efficiently as the wild-type version. Sequence of oligo3 competitor is found at position (-3674_-3650 bp) within the ic-CRE; this sequence (or its mutant version) does not compete, thus it is not recognized by Collier.

2.7.3 *In vivo* occupancy of the identified Collier binding site within the ic-CRE

To address the question whether the identified Collier binding site is occupied by Collier *in vivo,* chromatin immunoprecipitation followed by quantitative real-time PCR was performed (§5.4, 5.5). As described in §5.5, enrichment ratios of the identified Collier binding site in the anti-Col sample over the negative control region (amplicon within the second exon of *caudal*) were assessed and compared to the enrichment ratios obtained from the mock IP in two independent experiments (A and B, using independently isolated starting material; Fig. 2_48).

Antibodies	Enrichment ratio Experiment A	Enrichment ratio Experiment B
Mock IP (m_a-BP102)	1.46	1.47
a-Col IP 0.5 μl	1.55	
a-Col IP 2.2 μl	10.97	
a-Col IP 1.9 μl		7.15

	A	B
a-BP102	1.46	1.47
a-Col	10.97	7.15
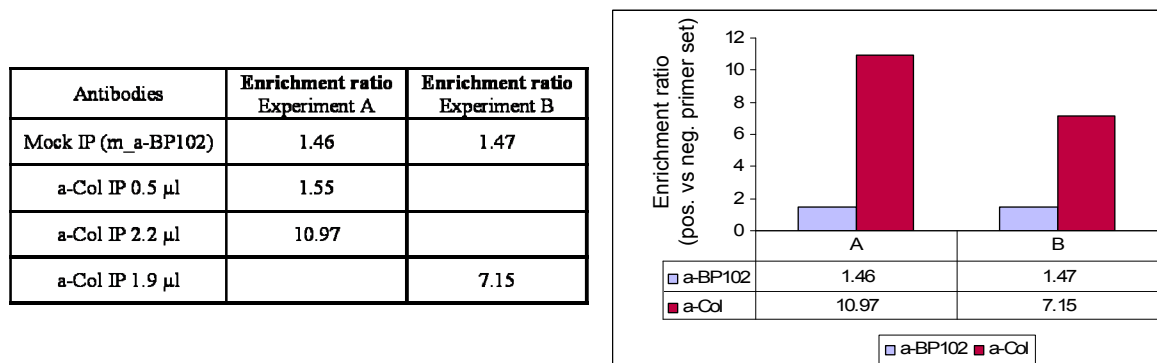
Fig. 2_48. ChIP results. In Experiment A 0.5 μl and 2.2 μl of monoclonal anti-Col (~4.5 μg/μl), 1 and 3 μl of anti-BP102 (~10 μg) were used respectively. In Experiment B 1.9 μl of a-Col (~8.6 μg) and 2.6 μl of a-BP102 (~8.6 μg) were used.

Occupancy of the identified binding site is enriched in the anti-Col IP samples, in contrast to the mock IP's where the ratios are around 1. The identified, as occupied by Collier *in vivo*, binding site could be further used as a proven endogenous positive site for data normalization in ChIP experiments followed by qPCR from transgenic embryos bearing either the wild-type ic-CRE sequence or the point-mutant versions affecting the Collier site. It could also be used for normalization when testing immunoprecipitated chromatin sample enrichment for novel Collier binding sites in a genome-wide distribution search.

2.7.4 *In vitro* analysis of Collier interaction with other *in silico* predicted putative recognition sites

It is expected that *cis*-regulatory modules controlling sub-patterns of the whole complex expression patterns of developmental genes are enriched in stronger binding sites for key-input transcription factors, recognized and bound with high affinity by the factor protein molecules (Segal et al., 2008). Nevertheless, weaker binding sites also contribute to the overall net expression outcome of a spatial-controlling module. Characteristic of these weaker binding sites is short-range homotypic clustering, usually within ~200 bp, which facilitates cooperative binding (Segal et al., 2008).

In silico analysis was performed to detect additional novel putative Collier binding sites within the *hh* upstream enhancer, in the vicinity of the identified functional binding site (-3777_-3747 bp), which would suggest an additional contributing input to the ic-CRE transcriptional outcome. Recognition of the novel – predicted *in silico* – putative sites by Collier protein was tested *in vitro* in competition EMSAs (Fig. 2_54 b).

The 6.43 kb upstream enhancer of *hh* was submitted to rVISTA analysis (http://genome.lbl.gov/vista/rvista/submit.shtml) scanning for sites matching the binding matrix of Olf-1. By setting the highest possible combination of values of core-similarity cut-off and matrix-similarity cut-off (0.95; 0.85), so that at least one prediction is generated, only one site is predicted which is the identified functional Collier binding site CAATTCCCCAATGGCAT (-3771_-3755 bp) found within the ic-CRE. Lowering the matrix similarity threshold only by 0.05, that is using combination similarity cut-off (core; matrix) (0.95; **0.8),** instead of (0.95; 0.85), then three additional predictions are generated, including two distant binding sites: GAGACACTTGGGATGAG at (-3963_-3947 bp) and CACACCACGGGGAAGCG at (-2872_-2856bp), and one promoter-proximal site CACTTCCCTTGCGCATA at (-212_-196 bp). These three predicted sequences were re-aligned with the already identified functional site using MEME (http://meme.nbcr.net/meme4_1/cgi-bin/meme.cgi) to generate a novel motif (Motif A) (Fig. 2_49).
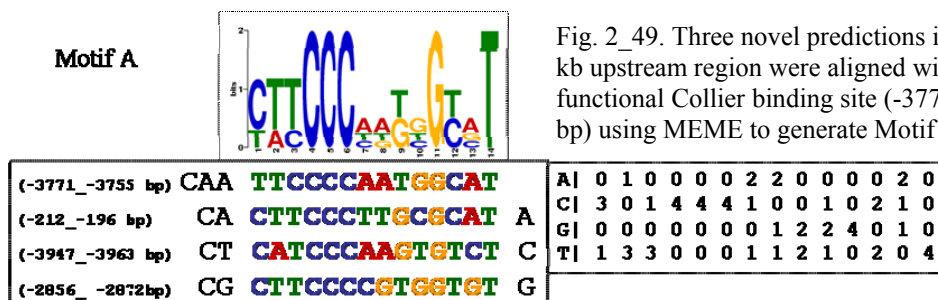
Motif A

Fig. 2_49. Three novel predictions in the 6.43 kb upstream region were aligned with the functional Collier binding site (-3771_-3755 bp) using MEME to generate Motif A.

(-3771_-3755 bp)	CAA TTCCCCAATGGCAT	
(-212_-196 bp)	CA CTTCCCTTGCGCAT A	
(-3947_-3963 bp)	CT CATCCCAAGTGTCT C	
(-2856_ -2872bp)	CG CTTCCCCGTGGTGT G	

```
A| 0 1 0 0 0 0 2 2 0 0 0 0 2 0
C| 3 0 1 4 4 4 1 0 0 1 0 2 1 0
G| 0 0 0 0 0 0 0 1 2 2 4 0 1 0
T| 1 3 3 0 0 0 1 1 2 1 0 2 0 4
```

Using the Motif A to scan the 6.43 kb upstream enhancer by rVISTA predicts sites (-2856_-2872 bp) and (-212_-196 bp) with core cut-off value 0.95. Threshold has to be lowered to 0.9 to generate the functional Collier binding site (-3771_-3755 bp) prediction. This is because core of the binding matrix has been repositioned. Therefore, the above alignment was manually corrected to generate a slightly distinct position weight matrix (Motif B) (Fig. 2_50) which if used to scan *in silico* the enhancer, the functional Collier binding site (-3771_-3755 bp) comes up with the highest core similarity cut-off.

Motif B

(-3771_-3755 bp)	CAATTCCCCAATGGCA T	
(-212_-196 bp)	CACTTCCCTTGCGCAT A	
(-3947_-3963 bp)	CTCATCCCAAGTGTCT C	
(-2856_-2872 bp)	CGCTTCCCCGTGGTGT G	

```
A| 0 2 1 1 0 0 0 0 1 2 1 0 0 0 1 1
C| 4 0 3 0 0 4 4 4 2 0 0 1 0 1 2 0
G| 0 1 0 0 0 0 0 0 0 1 2 1 4 1 1 0
T| 0 1 0 3 4 0 0 0 1 1 1 2 0 2 0 3
```

<=After small sample correction. (http://weblogo.berkeley.edu/logo.cgi)
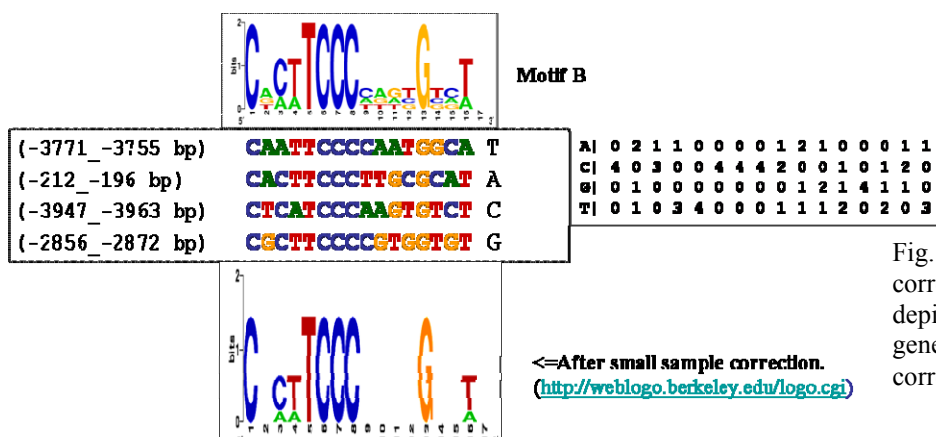
Fig. 2_50. After manual correction of the alignment depicted in Fig. 2_49 Motif B is generated. This can be further corrected for small sample input.

Notably, if in the alignment apart from the 17 nucleotide long sequences (which are the output of rVISTA prediction using the Olf-1 binding matrix) also surrounding nucleotides are included, in the respective context of the Olf-1 22 bp binding matrix (Fig. 2_24), then one strong consensus motif is generated by MEME shared by all set sequences, which perfectly matches the 5' half of Motif B (nucleotide positions 1-8; Fig. 2_50). This is depicted in Fig. 2_51.
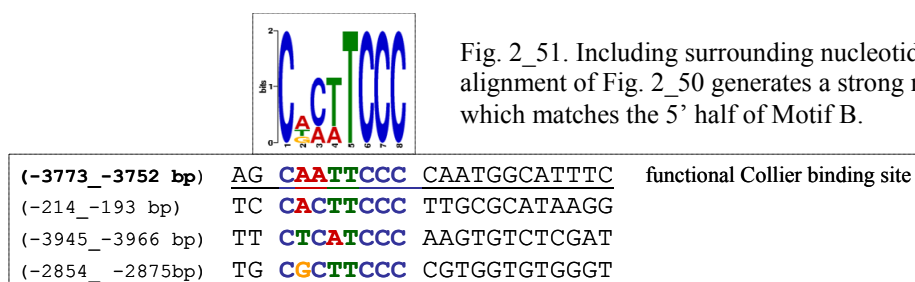
Fig. 2_51. Including surrounding nucleotides in the alignment of Fig. 2_50 generates a strong motif by MEME which matches the 5' half of Motif B.

(-3773_-3752 bp)	AG CAATTCCC CAATGGCATTTC	functional Collier binding site
(-214_-193 bp)	TC CACTTCCC TTGCGCATAAGG	
(-3945_-3966 bp)	TT CTCATCCC AAGTGTCTCGAT	
(-2854_ -2875bp)	TG CGCTTCCC CGTGGTGTGGGT	

If finally the vertebrate Olf-1 binding matrix and the new *in silico* generated 'Drosophila Collier Motif B' are compared, they can be finely aligned around an anchor position nucleotide **G** (position 15 of the Olf-1 matrix) spaced with 4 nucleotides from the core **TCCC** (Fig. 2_52).

Fig. 2_52. Alignment of the functional Collier binding site (-3777_-3747 bp) to the mammalian homolog binding matrix and to the novel *in silico* generated *Drosophila* Collier motif. The novel motif was generated from the alignment of four binding sites, the functional one plus three sites predicted from scanning *Drosophila hh* upstream region with rVISTA using the Olf-1 matrix and (0.95; 0.8) (core; matrix) similarity cut-off.

If the novel 'Drosophila Collier motif' is used to scan the upstream enhancer then one additional site is predicted which consists of two tandem repeats of the 5' half of the binding motif (overlapping by one nucleotide). Interestingly, this prediction found at position (-3714_-3701 bp) also falls within the ic-CRE in a distance of 35 nucleotide bases downstream of the functional Collier binding site. Although not phylogenetically conserved, clustering with the strong Collier binding site would imply that these weaker predicted sites might contribute as well to Collier positive input on the ic-CRE transcriptional outcome (if they are indeed bound by Collier protein molecules *in vivo*, perhaps by monomers, since they conform to the 5' half of the *Drosophila* binding motif; the second one with a mismatch) (Fig. 2_53).

Fig. 2_53. Alignment of the novel prediction, generated by scanning the full-length upstream enhancer using the *in silico* generated Dm_Col_motif, with the MEME motif of Fig. 2_51.
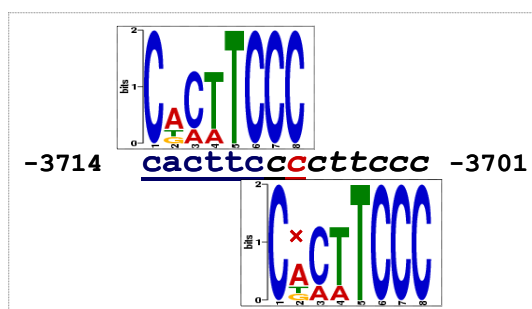
In conclusion four additional putative Collier recognition sites were predicted, two of them falling within the ic-CRE, [sites (-3947_-3964 bp); 'compet1' and (-3714_-3701 bp); compet2] in a distance 170 bp upstream and 35 bp downstream of the functional Collier binding site, respectively (Fig. 2_54_a). Notably, all three sites found within the ic-CRE are within the typical range of homotypic-clustering (~200 bp). The third additional predicted site is a distant element positioned downstream of the ic-CRE (- 2856_ -2872 bp; 'compet3') and the fourth is a promoter-proximal element (-212_-196 bp; 'compet4').

The four novel predicted binding sites were used as competitors at 100x molar excess in mobility shift assay (Fig. 2_54 b) with the functional binding site-probe (Fig. 2_44, 2_46). However these sequences were not as efficient in competition as the (100x) wt competitor (Fig. 2_46) (only a very weak competition effect can be observed in the case of competitor 1 – compare free probe and C1 intensity, lane #3 vs. #4, Fig. 2_54 b), meaning that the sites are not efficiently recognized by Collier *in vitro*. Still, if the sites are recognized even with minor affinity *in vivo*, homotypic clustering within the ic-CRE would suggest that they may contribute to the overall net transcriptional outcome of the ic-CRE.

On the other hand, from the total five sites predicted *in silico* within the 6.43 kb upstream enhancer only the one (the functional Collier binding site at position (-3773_-3752 bp) within the ic-CRE) that is predicted with the highest matrix similarity cut-off is specifically and with high affinity recognized by the factor *in vitro* (Fig. 2_44, 2_46).
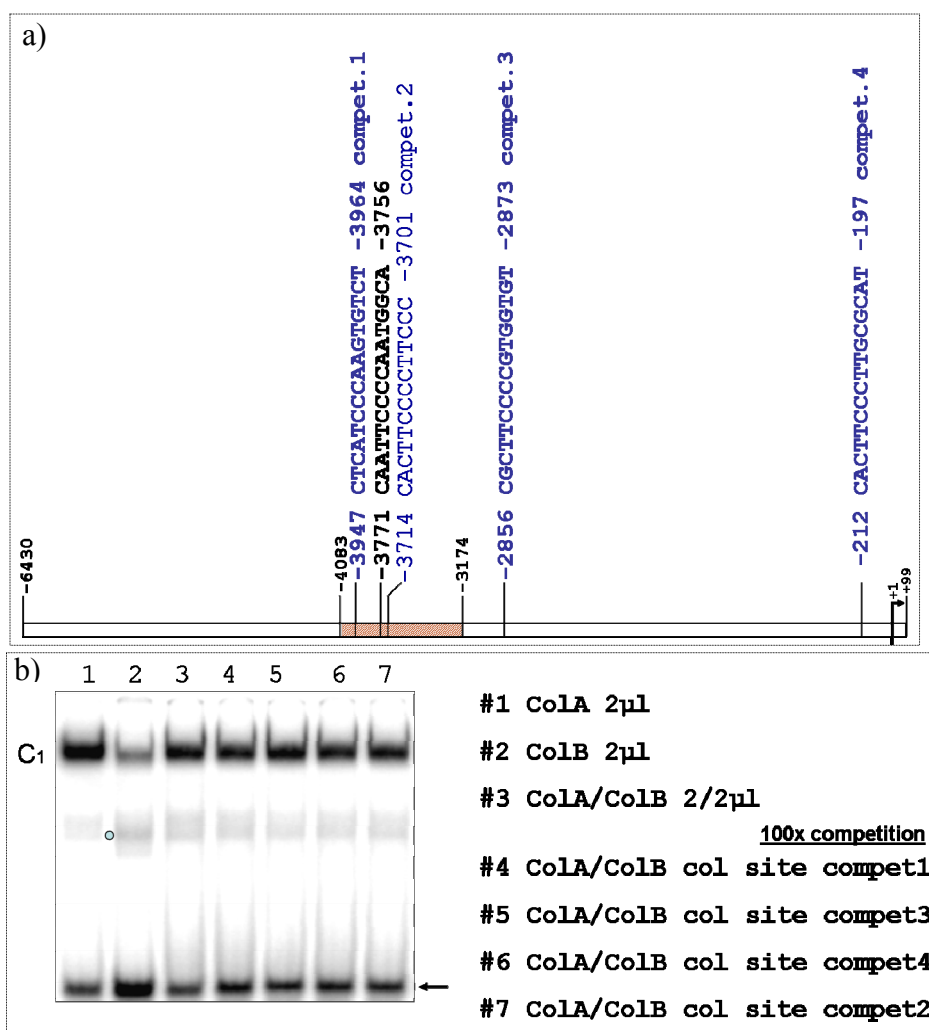
Fig. 2_54 a) Summary of the *in silico* predictions of putative Collier recognition sites within the -6.43 kb *hh* upstream enhancer. Black is the functional Collier binding site found within the ic-CRE. b) Competition assay with the novel *in silico* predicted oligo sequences. Blue spot marks the Col B monomer complex, arrow depicts the free probe. Only a minor competition effect can be observed (compare C1 complex intensity and free probe between lane 3 and lane 4).

2.8 Restriction of the activator function of Collier to the posterior part of the intercalary segment

Up to now a line of results has been presented supporting that Collier-dependent intercalary-specific expression of *hh* is succeeded through direct involvement of Collier in transcriptional activation of *hh* in the intercalary segment:

- Presence of a Collier binding site in the intercalary-specific *cis*-regulatory element of *hh* which is functional *in vitro* (generating specific Collier-DNA complex formation in EMSAs), as well as *in vivo,* including enrichment in anti-Col ChIP sample and functionality in the point mutagenesis screen; insertion of point mutations in that site in the context of the ic-CRE sequence causes loss of reporter expression. This is due to abolishment of recognition and binding of the mutated site by Collier *in vivo*, as it can be inferred *in vitro* by competition EMSAs (Fig. 2_46).

- Expression of the ic-CRE, as well as the endogenous intercalary-specific *hh* expression pattern, is abolished in *col* loss-of-function mutant backgrounds (§2.7.1).

Supporting a role of Collier functioning as a direct transcriptional activator of *hh* in the intercalary segment triggers the question of how positive transcriptional regulatory input is restricted to the posterior part of the intercalary segment. *In situ* hybridization and immunostaining reveals that Collier is not expressed solely in the *hh* expressing cells of the posterior part of the intercalary segment but its expression pattern exceeds also into the anterior most part of the mandibular segment. Therefore, an underlying molecular mechanism ensures that positive regulatory input directed by Collier on the ic-CRE expression outcome is restricted in the *hh* expressing cells of the posterior part of the intercalary segment. Presumptively the same mechanism may be involved in defining posterior segmental boundary of intercalary-specific expression of *hh*. Considering a synergistic positive regulatory input from a factor(s) absent from the anterior most part of the mandibular segment, required for Collier to conduct in a synergistic fashion its positive regulatory function, is one possibility. The second is to consider the presence of a negative regulatory factor in the anterior most part of the mandibular segment which inhibits or competes positive regulatory function of Collier in that region.

2.8.1 ic-CRE expression under negative regulation of CncB in the anterior part of the mandibular segment

Among genes being expressed in the embryonic head region, *cap `n` collar* (*cnc*) displays a well characterized expression pattern (Crozatier et al., 1999; Veraksa et al, 2000; Seecoomar et al., 2000) that would finely serve the second possibility; *cnc* is required for the development of the hypopharyngeal lobes, structures that derive from progenitor cells of the anterior part of the mandibular segment (Seecoomar et al., 2000), and by stage 10 it is expressed throughout the

mandibular compartment precisely abutting the *hh* expressing cells of the posterior part of the intercalary segment. Therefore, by stage 10 expression domains of *collier* and *cnc* overlap only in the anterior most part of the mandibular segment which is exactly the domain where Collier fails to activate the ic-CRE (Fig. 2_67).

There is one more reason why *cnc* could be involved in a negative regulatory aspect of the ic-CRE expression, preventing it from being expressed in the anterior most part of mandibular segment; after initiation of *cnc* expression in the posterior part of the intercalary segment with onset at stage 10, triggered by Collier activity (Crozatier et al., 1999), reporter expression driven by the ic-CRE is greatly reduced to 2-3 cells in the intercalary segment by the end of stage 11 (Fig. 2_55). Expression of the endogenous intercalary *hh* stripe also gradually diminishes by late stage 11 (Fig. 2_56).
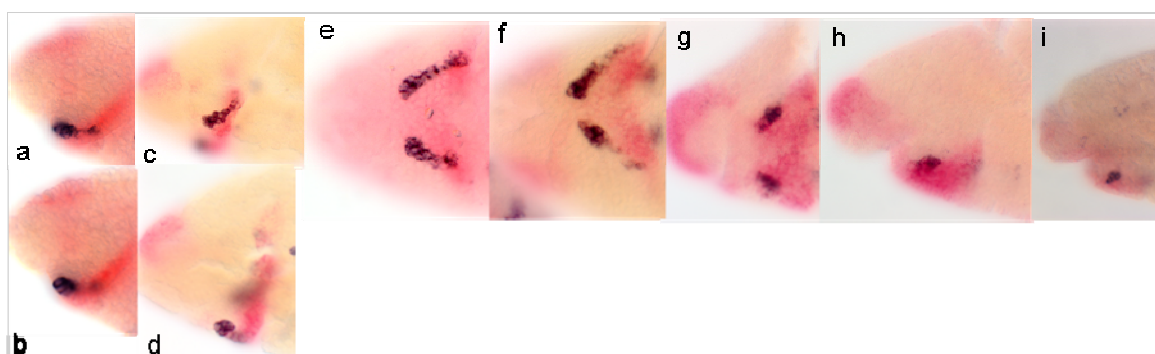
Fig. 2_55. ic-CRE_*tgfp* (NBT), *cnc* (Red). Progressive reduction in the detection of the ic-CRE-mediated expression with onset during st. 10. a, b (st. 9 lateral view) and c, d (early st. 10 ventral-lateral view) are different focal planes of the same embryo respectively. e early st. 10 ventral view, f st. 10; g, h, i different embryos at st. 11. g, ventral view; h, i lateral views, anterior to the left, dorsal – up.
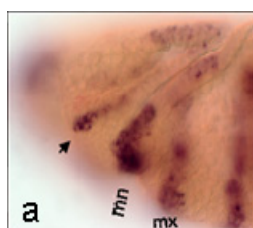
Fig.2_56. Panel I (a-d): Gradual diminishing in the intercalary ectodermal stripe of *hh* expression (arrow) in wild type embryos. a, b same embryo stage 10. c, late 10. d, st. 11. Decrease can also be observed for the antennal and ocular stripes (focal plane allows to note the difference in the intensity of the ocular stripe between b and c).
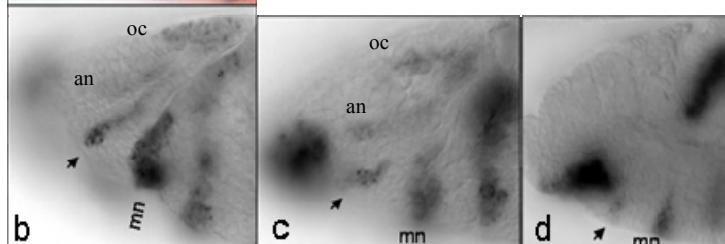
Fig.2_56 Panel II (a-h). Observation of the gradual decrease in intensity of the procephalic ectodermal stripes. Double staining *hh-cnc*. a, c, h ventral views. b, d, e, f ventral-lateral. g, lateral. a, b 'snapshots' during stage 10, c, d, e, late 10. f, g st. 11. h end of st. 11. Arrow depicts the reduced in intensity intercalary stripe (intercalary spot) at stage 11. The antennal and ocular ectodermal stripes also progressivelly decrease in intensity (note the difference in the intensity of the antennal stripe between e and f).
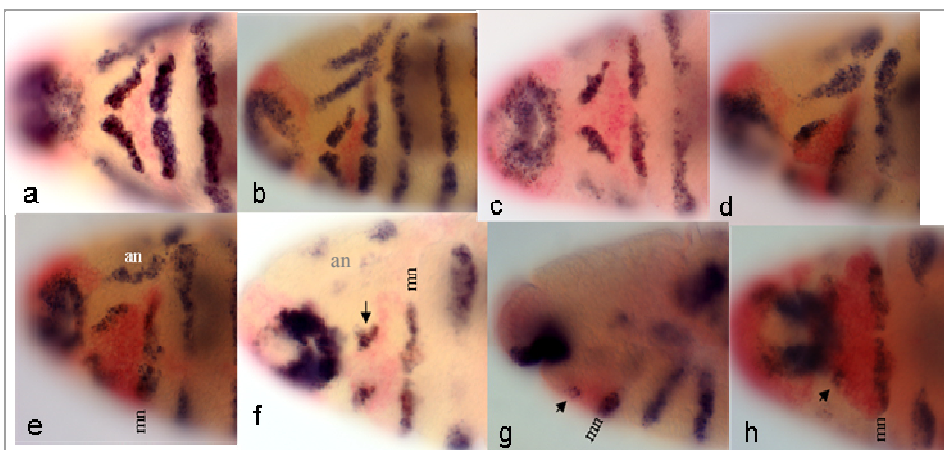
Moreover, it has been reported that *cnc* is also involved in negative regulation of another gene, namely *Deformed*; CncB inhibits both the function of Deformed protein and *Dfd* expression in the anterior part of the mandibular segment although this effect is likely to be indirect rather than resulting from direct CncB-dependent transcriptional repression of *Dfd* transcription (McGinnis et al., 1998; Veraksa et al., 2000). Nevertheless, an observation of a CncB-Dfd protein-protein interaction as it could be inferred from GST-pull down immunoprecipitation was reported in Veraksa et al., (2000).

In *Drosophila melanogaster* three Cnc isoforms are expressed throughout development (McGinnis et al., 1998). *cncA* transcript is ubiquitously present in 0-2 h embryos due to maternal contribution; *cncB* is absent from 0-2 h but present in all other embryonic stages, detected throughout the mandibular segment and labral region; *cncC* is present in 0-2 h embryos, barely detected in 2-12 and later low level ubiquitous expression is detected in 12-24 h embryos. The three produced protein isoforms share the same C-terminus bZIP domain (Fig. 2_57_a). Only CncB displays a key-function in head development (McGinnis et al., 1998).
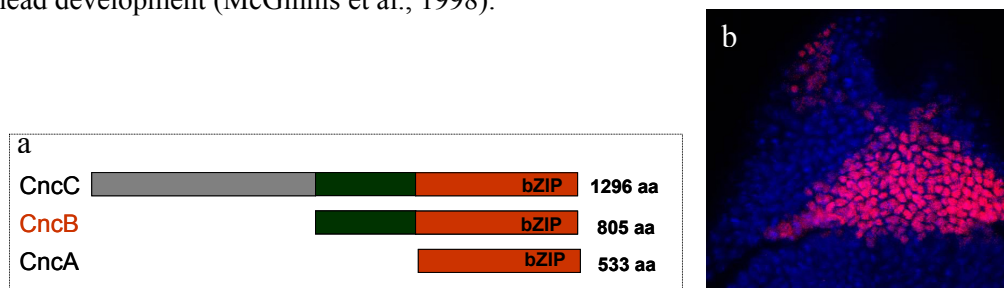
Fig. 2_57. a) Schematic representation of the three Cnc isoforms. b) Fluorescent immunostaining detects CncB expression in the mandibular compartment and around the stomodeum at stage 10 (anterior up).

Question if CncB could indeed be involved in a negative regulation aspect of the ic-CRE-mediated expression in the anterior part of the mandibular segment was addressed by following approaches *in vitro* and *in vivo*.

2.8.2 *In vitro* analysis of Collier-CncB interaction and effect on Collier-DNA binding complex formation

As a first approach, the 450 bp sequence of the intercalary-specific *cis*-regulatory element [-4.08 kb_3.077 kb] was examined for the presence of putative CncB recognition sites, conditionally overlooking transcriptional activation potential as an intrinsic characteristic of CncB (Veraksa et al., 2000). CncB was shown to bind in heterodimers with small Maf protein (Maf-S) consensus sequence `TGCTGA(G/T)TCAT` with the `TGCTGAG` half site being contacted by Maf-S and the `GTCAT` half-site by Cnc homologues (Veraksa et al., 2000). Binding is highly cooperative and results in transcriptional activation of target sequences. Although no site matching Maf-S contacting

consensus sequence could be found within the 450 bp ic-CRE, still an imperfect putative site for CncB, TTCAC, is juxtaposed 3' of the Collier binding site (Fig. 2_58).

CncB as well as the shortest CncA isoforms were both expressed *in vitro* to be used in mobility shift assays using as a probe a 35 bp oligo sequence containing both the identified functional Collier binding site and the putative Cnc recognition site. CncB (or CncA) could not specifically recognize and bind the probe when alone included in the reaction, in agreement with Veraksa et al. (2000) reporting that CncB binding to its cognate site requires an adjacent recognition sequence cooperatively bound by Maf-S. When both Collier and Cnc factors are included in the reaction, apart from the characterized Collier-DNA shift complex, no additional complex could be detected that would imply synergistic binding with Collier. On the contrary, including CncB (or CncA) in the binding reaction caused decrease in the Collier-DNA shift complex formation (compare C1 complex intensity of line 8 with lanes 3, 4 and lane 7 with lanes 1, 2 in Fig. 2_58).

ColA	–	–	2	2	–	–	–	2	–	–
ColB	2	2	–	–	–	–	2	–	–	–
CncA	–	2	–	2	–	2	–	–	–	–
CncB	2	–	2	–	2	–	–	–	–	–
	1	2	3	4	5	6	7	8	9	10

Col site probe (-3778_3744)   tcggc**AGCAATTCCCCAATGGCATTTCAC**TTagat
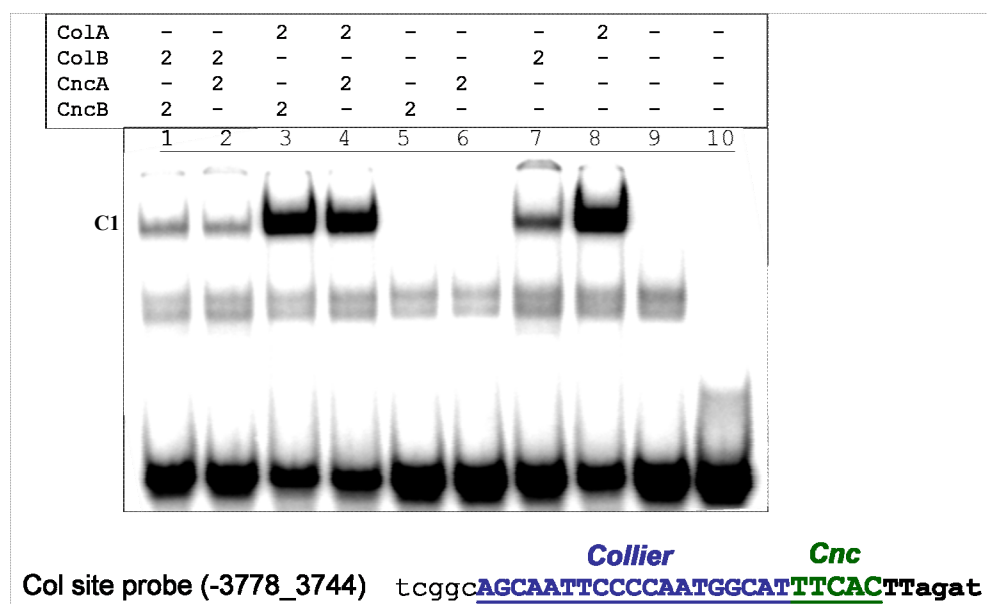
Fig. 2_58. Cnc homologues prevent Collier-DNA complex formation. Line 9 contains 4 µl of TNT mock.

Because this small effect indicates that, at least *in vitro* and on that specific DNA sequence, Cnc homologues seem to interfere with Collier binding to its cognate site, additional analysis was performed. Firstly, another well characterized member of the *Drosphila* bZIP superfamily, namely Dm_C/EBP (Slbo) was tested for generating a similar interference effect with Collier DNA binding activity. DNA binding specificity of C/EBP is TKNNGCAAT and two putative 5' and 3' degenerated bZIP sites are found upstream of the Collier binding site, lying on opposite strands and separated by a 10 bp sequence (position (-3798_-3765 bp); Fig. 2_59). The 34 bp sequence (-3798_-3765 bp) containing the two bZIP sites was labeled to be used as a positive target binding sequence-probe for *in vitro* expressed C/EBP in order to test for functionality of the expressed protein. Indeed, C/EBP

binding reaction leads to the formation of a specific complex formation mediated by homodimers (lane 1, Fig. 2_59), but in contrast to CncB or CncA, including C/EBP factor in the binding reaction of Collier to its cognate site (using Collier probe [-3778_3744 bp]) does not efficiently interfere with Collier complex formation (Fig. 2_59). This result implies that ability of CncB for interaction with Collier, resulting in interference of Collier DNA binding to its cognate site, may be an intrinsic property of Cnc homologues not shared by all members of Drosophila bZIP superfamily (§3.7.1). Moreover, and in correlation with the study reported by Bengal et al., 1992 (§3.7.1), Cnc homologues are also capable of interacting with C/EBP and causing reduction of its binding to the positive target sequence (Fig. 2_60).
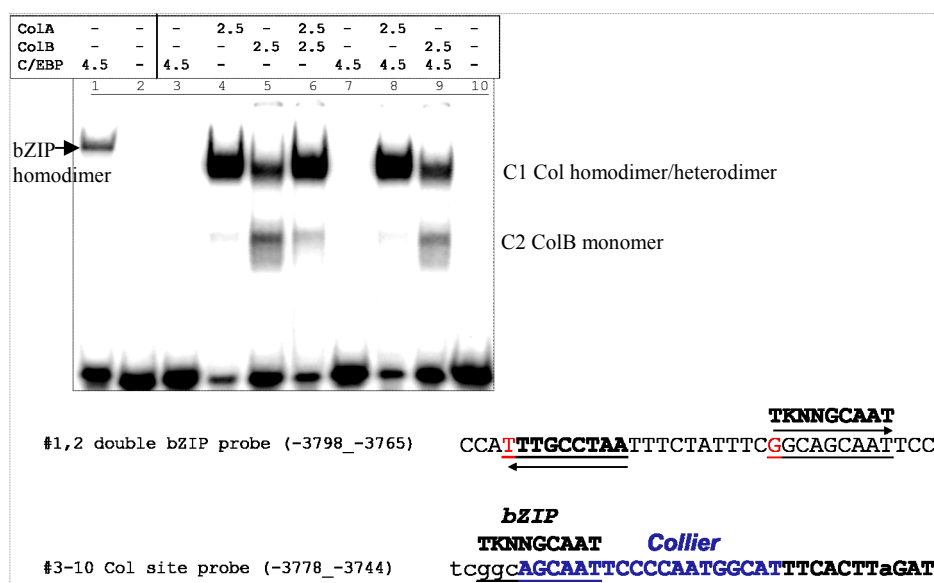
Fig. 2_59. C/EBP does not interfere with Collier-DNA complex formation.
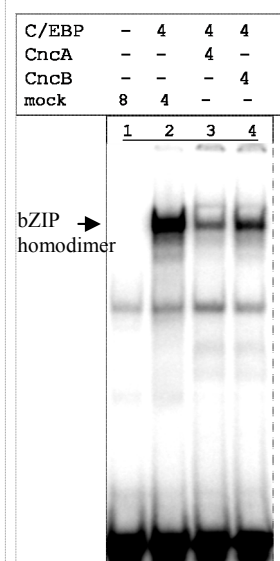
Fig. 2_60. CncA and CncB reduce C/EBP complex shift.

Interference of Collier DNA-binding to its cognate site caused by interaction with CncB was further titrated showing that increasing amounts of CncB factor included in the binding reaction attenuates complex formation (Fig. 2_61). To further examine if CncB can act *in vitro* as an inhibitor of Collier DNA-binding in a dominant way, 'competition' shift assays were performed; Collier DNA-binding complex formation was allowed to reach the dynamic equilibrium and then challenged with addition of the same amount of CncB for gradually longer incubation time periods prior to immediate gel loading. As depicted in Figure 2_62, *in vitro* interaction of CncB with Collier resembles a kind of 'sequestering' function of CncB partially preventing free unbound Collier protein molecules of the dynamic DNA-binding state from re-associating with DNA. Thus, CncB seems to function as a dominant inhibitor of Collier DNA-binding activity *in vitro,* presumptively as an antagonist of its transcriptional activation potential *in vivo* (§3.7.1, §3.7.2).
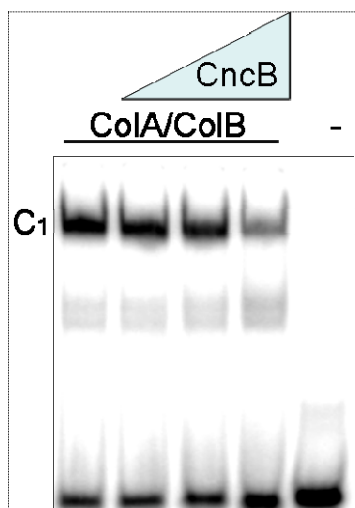
Fig. 2_61. Increasing concentration of CncB attenuates Collier-DNA complex formation.
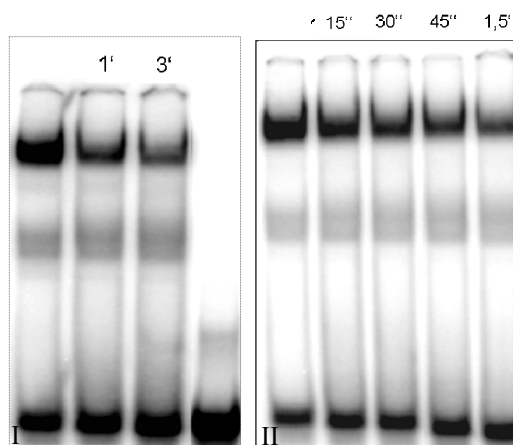
Fig. 2_62. Collier-DNA complex formation was challenged with CncB for gradual longer incubation times before loading the gels. I and II are from two independent experiments.

2.8.3 *In vivo* analysis of Collier-CncB interaction; negative regulatory effect of CncB function on the ic-CRE transcriptional outcome

A physical interaction between Collier and CncB inferred in *vitro* from the electrophoretic mobility shift assays could be further supported by *in vivo* co-immunoprecipitation from crude nuclear extracts (§5.6). After *in vivo* anti-Col IP and immunoblotting, CncB specific band detection on the blot corresponding to co-immunoprecipitated CncB suggests that stable interactions between Collier and CncB can exist (§3.7.1). The result was reproduced in three independent experiments. Two independent a-Cnc blots are depicted in Fig. 2_63.
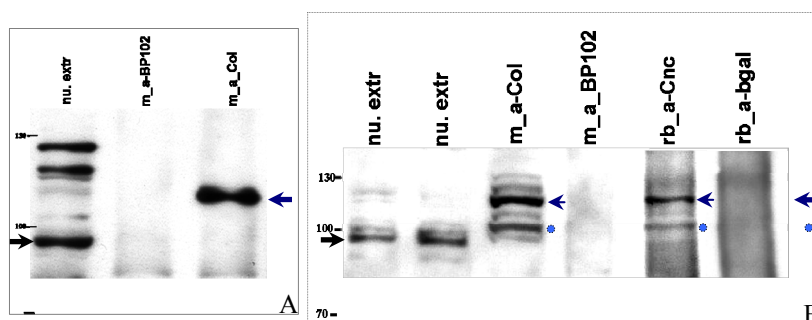
Fig. 2_63. a-Cnc blots from two independent experiments (A, B). Blue arrow marks the co-immmunopreicipited CncB isoform in the a-Col IP from crude embryonic nuclear extracts. The black arrow marks the CnCB ~90 kDa isoform detected in the nuclear extracts.

In both cases the co-immunoprecipitated in the a-Col IP and the immunoprecipitated in the a-CncB IP band (blue arrow) is detected at ~112 KDa, higher than the ~90 kDa CncB band detected in the extracts (black arrow), probably due to secondary modifications during the IP procedure. The blue spot marks an additional band (~105 kDa) immunoprecipitated with a-CncB and co-immunoprecipitated with a-Col during the second experiment. In the nuclear extract lane of a-Cnc blot A two additional higher bands (~123, ~126 kDa) are detected presumably because of aggregate formation after boiling or secondary modifications not detected in B.

To examine if the observed physical interaction between Collier and CncB is involved *in vivo* in a negative regulatory aspect of ic-CRE expression in the anterior part of the mandibular segment, the element was tested in an ectopic CncB background. Background was created by crossing the maternal driver {*pnos_GAL4/GCN4_ bcd 3' UTR*} (insertion 'P21B'; Janody et al., 2000) to a UAS-CncB line (Veraksa et al., 2000), thus causing ectopic expression of CncB in an embryonic anterior

gradient. As shown in Fig. 2_64, expression of the ic-CRE is almost abolished in this ectopic CncB background, most prominently in embryos that receive a double dose of maternal driver input (mothers homozygous for the driver on the second chromosome). Transcription of the endogenous *hh* intercalary stripe is also partially suppressed (Fig. 2_65).

Fig. 2_64. NBT: ic-CRE_tgfp, FRED: *cnc* (i, ii, iii); *col* (iv)
i,ii wt background

iii, ic_CRE_tgfp/ UAS_CncB/ *pnos_GAL4/GCN4_bcd 3' UTR*

iv, ic_CRE_tgfp/ UAS_Col/ *pnos_GAL4/GCN4_bcd 3'UTR*

ic-CRE expression outcome is greatly reduced (arrow in iii) in ectopically expressed CncB background while it remains unaltered in ectopically expressed Collier using the same maternal driver.

Fig. 2_65. *hh* (NBT)_*cnc* (Fred)
i, ii wt background
iii, iv UAS_CncB/ *pnos_GAL4/GCN4_bcd 3' UTR*
Transcription of the endogenous *hh* in the intercalary segment is also partially suppressed when ectopic procephalic expression of CncB is driven early in development using the *pnos_GAL4/GCN4_bcd 3'UTR* maternal driver. Embryos in i, iii and ii, iv are of approximately the same stage.

56

Therefore, it is concluded that CncB function seems to direct a negative regulatory input in the intercalary-specific transcriptional control of *hh* expression *in vivo* (§3.7.1). This conclusion is consistent with the overlap in the expression patterns of Collier and CncB in the anterior most part of the mandibular segment where the ic-CRE is not activated (Fig. 2_67_c).

Overlap between *col* and *cnc* expression in the presumptive anterior mandibular region is apparent from blastoderm stage; at st. 5/6 *collier* exhibits a posterior border of parasegmental register of expression which is abutting the mandibular stripe of *hh* cells, while *cnc* is detected in the mandibular *hh* cells exceeding in addition two rows of cells anterior thereto (Fig. 2_66_d and Crozatier et al., 1999). At stage 10, the posterior parasegmental border of *col* expression has been shifted anteriorly and *col* and *cnc* overlap only in the anterior most part of the mandibular segment (Fig. 2_67_a, b). An anterior parasegmental border of *col* expression however is evident defined by the *wg* cells of the intercalary segment (Fig. 2_66 e, f). By stage, 10 *cnc* (CncB) is expressed throughout the mandibular compartment precisely abutting the *hh* positive cells of the posterior part of the intercalary segment and the *hh* cells of the mandibular segment (Fig. 2_68_a). During stage 10, a secondary site of *cnc* transcription is initiated by Collier activity in the *hh* expressing cells of the posterior part of the intercalary segment (Fig. 2_68_b; and Crozatier et al., 1999; Seecoomar et al., 2000). Notably from that stage on, detection of *hh* in epidermal cells of the intercalary ectoderm is gradually reduced until 'turning off' by the end of stage 11, in a way similar to the antennal and ocular *hh* procephalic ectodermal stripes (Fig. 2_56). Detection of the ic-CRE-driven transcripts is also greatly reduced (Fig. 2_55). Expression of *cnc* in the anterior most part of the mandibular segment is also under positive regulation by *collier* (Seecoomar et al., 2000); *collier* null mutants apart from lacking the late (stage 10-11) *cnc* expression in the posterior part of the intercalary segment (Crozatier et al., 1999), they also lack *cnc* expression in the anterior most cells of the mandibular compartment which are the progenitors of the hypopharyngeal lobes (Seecoomar et al., 2000). Later during development (stage 11) transcription of *collier* is restricted to cells of the posterior part of the intercalary segment while being under positive auto-regulation control (Crozatier et al., 1999) (Fig. 2_69), meaning that the overlap between *col* and *cnc* expression in the anterior part of the mandibular segment is gradually reduced. Co-expression of Collier and CncB in the anterior most part of the mandibular segment by stage 10 and later co-expression in the posterior part of the intercalary segment is also revealed by fluorescent immunostaining (Fig. 2_70).
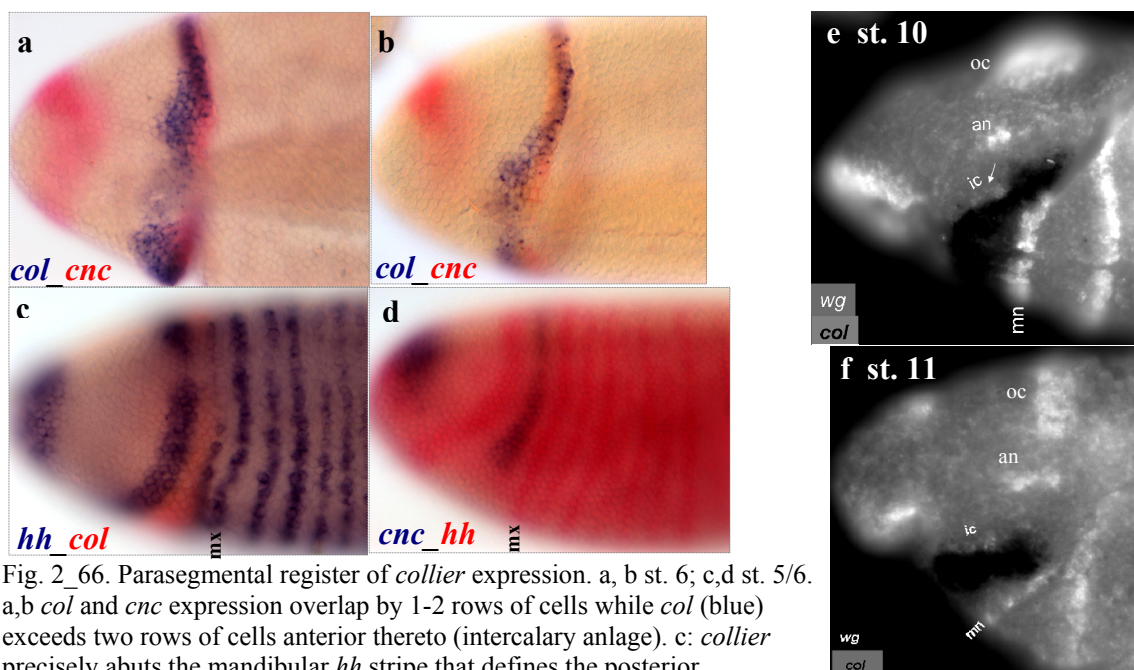
Fig. 2_66. Parasegmental register of *collier* expression. a, b st. 6; c,d st. 5/6. a,b *col* and *cnc* expression overlap by 1-2 rows of cells while *col* (blue) exceeds two rows of cells anterior thereto (intercalary anlage). c: *collier* precisely abuts the mandibular *hh* stripe that defines the posterior parasegmental border of expression. d: *cnc* overlaps the mandibular *hh* stripe and exceeds one row of cells anterior thereto. e (st.10) and f (st.11): anterior parasegmental border of *col* expression is defined by abutting the *wg* expressing cells of the anterior part of the intercalary segment.
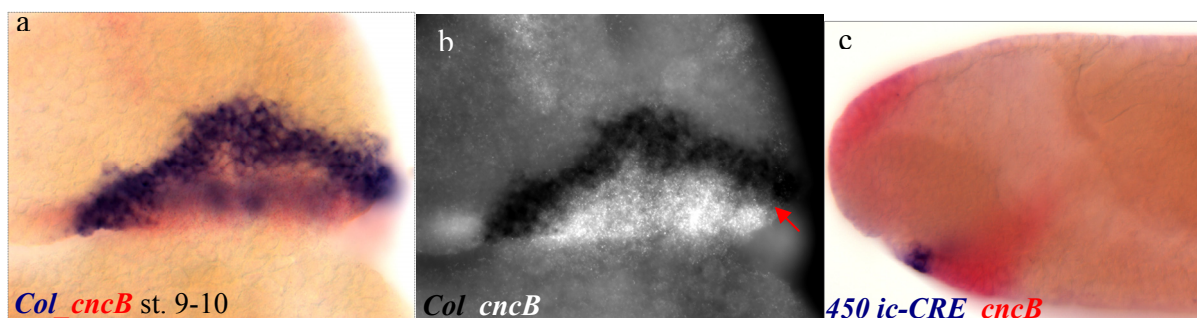
Fig. 2_67. a, b: Overlap in the expression domains of *col* and *cncB* in the anterior most part of the mandibular segment. Red arrow in b depicts overlap of *col* and *cnc* expression in the anterior most part of the mandibular segment during stages 9-10. This is the region where Collier fails to activate the ic-CRE (c); *cncB* (red) expression abuts the ic-CRE reporter expression (blue) in the *hh* expressing cells of the posterior part of the intercalary segment.
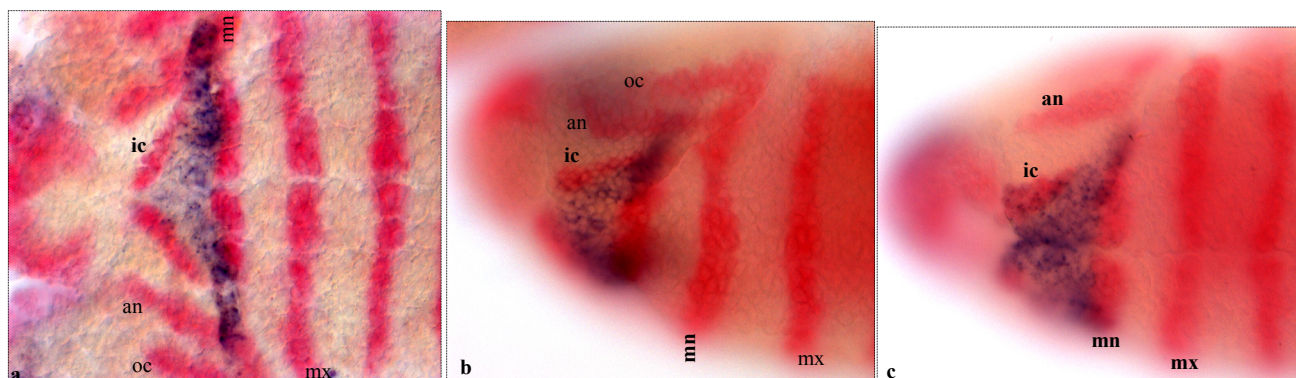
Fig. 2_68. *cnc* (NBT) *hh* (FastRed). *cnc* expression in the mandibular compartment abuts the intercalary stripe of *hh* expressing cells (a). Early at stage 10 (a, flat preparation; b, ventral-lateral view) a secondary site of *cnc* expression is initiated by *collier* activity in the *hh* expressing cells of the posterior part of the intercalary segment (see text for citation). Expression of *cnc* in the *hh* expressing cells of the posterior part of the intercalary segment becomes evident at late stage 10 (c).
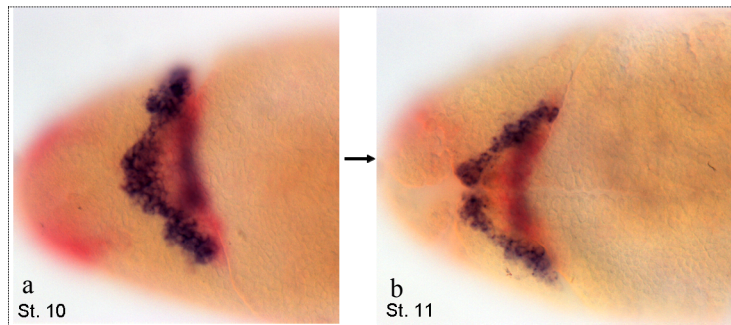
Fig. 2_69. *cnc* (FastRed), *col* (NBT). Gradual decrease in the overlap of expression initially evident in the anterior most part of the mandibular segment (a). During development towards st. 11 *col* expression is gradually getting restricted to the anterior part of the parasegment 0, i.e. to the posterior part of the intercalary segment, while being under positive autoregulation (see text for citation).
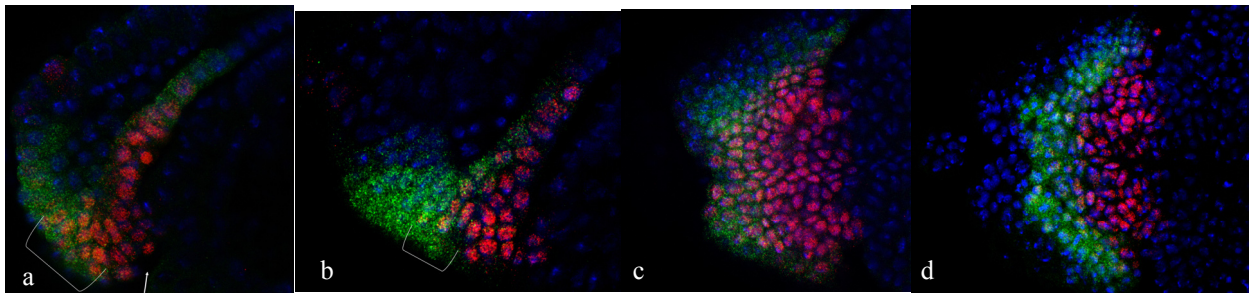
Fig. 2_70. Fluorescent immunostaining of Col and CncB – single focal planes. Anterior-left, dorsal-up (a, b). Arrow in (a) marks the cephalic furrow. c, d: anterior-left, ventral focal planes. Alexa 488 (green) Collier; Cy3 (red) CncB; Hoechst (blue) – nuclear staining. a (st. 8), b (st. 9/10), c st 10: overlap of expression in the anterior most part of the mandibular segment. d, st. 11 Col has been restricted to the posterior part of the intercalary segment. CncB exhibits nuclear localization while Collier is detected both in the nucleus and the cytoplasm.

Fluorescent immunostainings also revealed that only a small fraction of the expressed Collier protein is nuclear in contrast to CncB that seems exclusively targeted to the nucleus. *In silico* analysis of the primary protein sequences for prediction of nuclear localization signals (NLS) (http://cubic.bioc.columbia.edu/cgi/var/nair/resonline.pl) generates no results for Collier while CncB contains a NLS (RRRGKNKVAAQNCRKRK aa 622-638) within the bZIP domain (aa 617-680). Interestingly, Collier carries a perfect SUMOylation motif in the very N-terminus, predicted with the highest threshold value (Ren et al., 2009; http://sumosp.biocuckoo.org/prediction.php). Protein sequence TSLKEEP at position 44 matches the motif Ψ-K-X-E. Additional members of the COE factor family contain a SUMOylation motif at this conserved position (see sequence alignment of 'Additional File 2' from Daburon et al., 2008). Apart from antagonizing ubiquitin-mediated degradation and modifying transcriptional activation/repression potential of transcription factors, sumoylation has also been implicated in protein nucleo-cytoplasmic translocation (reviewed in Zhao, 2007). Alternatively, in the absence of a nuclear localization signal, Collier import in the nucleus may be realized by heterodimerization with a protein that carries a NLS, possibly a distinct HLH factor. That would increase probability that Collier enters into combinatorial control mechanisms, which nevertheless has already been implicated in muscle specification (Dubois et al., 2007) (§3.7.2).

Although in the ectopic CncB expression system a negative regulatory aspect involved in transcriptional control of procephalic *hh* expression became apparent (Fig. 2_65), still it was not

possible to detect posterior derepression in *cnc* loss-of-function background (§3.7.1). Expression of the ic-CRE was tested both in homozygotes of *cnc* mutation *K22* (loss-of-function; Veraksa et al., 2000) and in mutation *VL110* which lacks the entire *cnc* open reading frame and it was generated following imprecise P element excision (Mohler et al., 1995) (§5.3). No *cnc* transcripts are detected by *in situ* hybridization in *VL110* null mutants. In addition, in these *in situ* hybridization stainings abnormal persistence of *hh* transcripts in ectodermal cells of the intercalary lobes of late stage 11 null mutant embryos could be observed (Fig. 2_71; §3.7.1). Irregular persistence of the ic-CRE driven transcripts in the intercalary lobes of stage 11 *VL110* mutant embryos could also be detected (Fig. 2_72). These data serve as further evidence for a repressive role of CncB on *hh* expression in the posterior part of the intercalary segment.
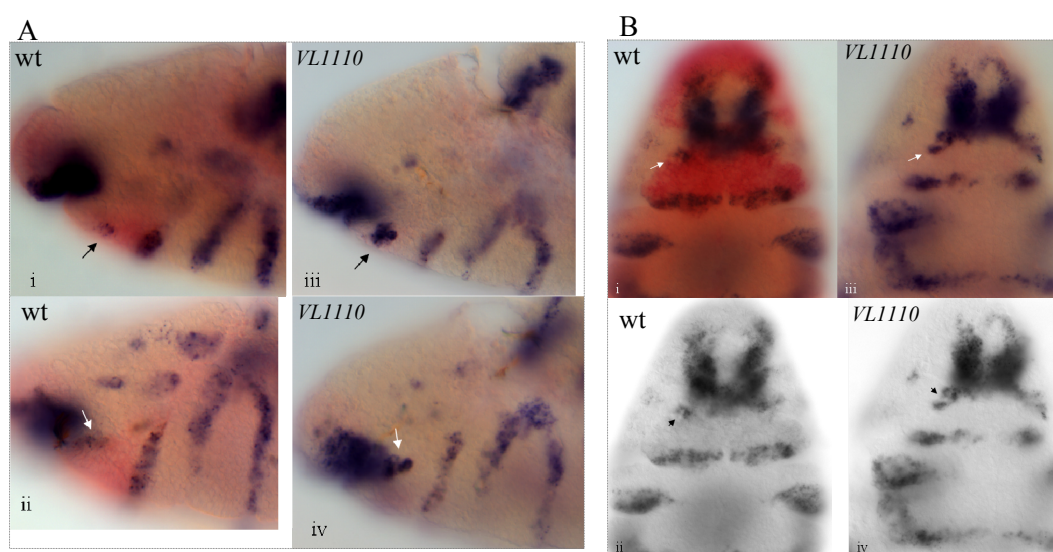
Fig. 2_71. *hh*-NBT, *cnc*-FastRed. In each panel: i,ii wt. iii, iv *VL110* nulls. Arrow marks intercalary *hh* expression. In panel A: i, ii and iii, iv are different focal planes of the same embryo respectively. Detection (arrow in iii, iv) of irregular persistence of *hh* transcripts in cells of the intercalary lobes at st. 11 in the *VL110* embryos.
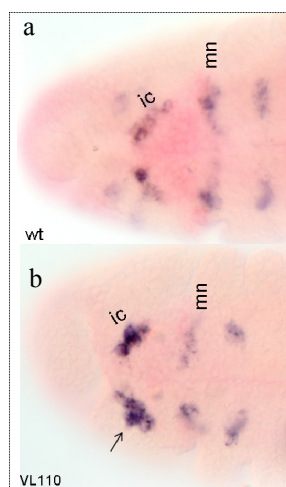
Fig. 2_72. *ic-CRE_tgfp,* NBT; *cnc,* FastRed. a, wt; b, *VL110.* Detection of irregular persistance of ic-CRE driven transcripts in the intercalary lobes of stage 11 *VL110* mutant embryos (arrow in b). The 335 bp minimum ic-CRE which shows partial trunk derepression at st. 11 was used in these experiments.

3. Discussion

3.1 Conclusions regarding *cis*-regulatory information governing expression of segment polarity genes in the anterior head region.

3.1.1 Distinct spatial cis-regulatory elements control procephalic expression of segment polarity genes

One first conclusion extracted from the outcome of functional dissection of segment polarity gene enhancers is that *cis*-regulatory information governing procephalic expression is distinct from *cis*-information controlling expression in the rest of the embryo. For example, anterior head-segment specific *cis*-elements are not included in a total of 8.1 kb enhancer sequence upstream of *wingless* transcriptional unit, which sufficiently drives expression in the trunk and the anterior most terminal embryonic region (clypeolabrum) (§2.1.1.). The intronic enhancer of *engrailed* is functional only for the gnathal and trunk stripes but not for the anterior head region (§2.1.2). An opposite example stands for the 6.43 kb upstream sequence of *hedgehog* which acts as an early enhancer only for the presumptive procephalic region but not for the rest of the trunk where expression is only mediated at postblastodermal stages by this element (Fig. 2_14 a-c). In addition, a 1 kb enhancer element of *hh* that drives expression specifically in the intercalary segment was isolated and further functionally analyzed (§2.2).

3.1.2 Dorso-ventral distinct cis-regulatory information underlying transcription of segment polarity genes within procephalic segmental units

8.4 kb sequence upstream of *wg,* spanning region [-16.2_-7.8 kb], contains *cis*-regulatory information that drives expression only in the ventral part of the ocular segment (Fig. 2_6 g,h). In more primitive insects the anterior procephalic expression domain of *wg* splits into expression subdomains, namely the median protocerebral neuroectoderm expression domain (mpn), the dorsal protocerebral (dpn) and the ventral protocerbral neuroectoderm domain (vpn) (Liu et al., 2006). It corresponds to a true segmental unit, i.e. the ocular segment (Schmidt-Ott and Technau, 1992) based on data of the phylogenetically conserved expression profile of *engrailed* (Schmidt-Ott and al., 1994a; Urbach and Technau 2003a). On the contrary, in *Drosophila* the anterior procephalic expression domain of *wingless* in the ocular segment remains intact constituting the 'head blob'. The ventral protocerebral neuroectodermal domain has been specifically lost in *Drosophila* and, as it is reported in Liu et al., 2006, the contiguous protocerebral neuroectoderm domain or head blob may be equivalent either to **a)** the median **or** the dorsal protocerebral neuroectoderm expression domain **b)** the primordial yet non-dissociated protocerebral ectoderm domain of the primitive insects. Data

indicating dorso-ventral distinct *cis*-regulatory information controlling expression of *wg* within the ocular segment would rather support the second possibility with the ventral-specific expression subdomain of the contiguous *wg* head blob most likely corresponding to the median protocerebral expression domain of the primitive insects, with respect to topological orientation.

In the case of *hedgehog* procephalic expression pattern, dorso-ventral differences in regulation reported as well elsewhere (Gallitano-Mendel and Finkelstein, 1999) can be in fact reflected on transcriptional level of control; although the 6.43 kb upstream enhancer sequence of *hh* sufficiently drives expression in all procephalic segments, -3.17 kb sequence drives expression only in the dorsal most part of the antennal segment (Fig. 2.15 c), indicating again that transcriptional control of segment polarity genes in the procephalic region involves dorso- and ventral-specific *cis*-regulatory elements functional within the very same segmental unit. In a developmental physiological context, this molecular mechanism may potentially be involved in ensuring that distinct cell-fates will arise from dorsal and ventral parts of the ocular and antennal segments. Nevertheless, and at least in the case of dorsal epidermal embryonic patterning, functionality of segment polarity genes has been implicated in stabilizing pre-existing cell-fates rather than being involved in *de novo* cell-type specification (Vincent et al., 2008).

3.1.3 Redundancy in cis-regulatory information controlling gnathal and trunk expression could not be observed for procephalic control of segment polarity gene expression

During functional dissection of *cis*-regulatory regions it became apparent that extended enhancer fragments share a high degree of redundancy; *cis*-elements controlling expression of segment polarity genes in the gnathal and trunk segments are dispersed along extended enhancer regions. For example, two non-overlapping fragments of the *wingless* upstream enhancer, namely regions [-16.2_-7.8 kb] and -4.8 kb upstream of *wg* transcriptional unit, both drive expression of the reporter in the trunk in overlapping but not totally coincident expression patterns (Fig. 2_6 and Lessing and Nusse, 1998). This result indicates that the two distant enhancer regions are functionally redundant to some extent. The same was true for the first intron of *engrailed* which drives expression in the trunk and gnathal stripes as well as the 8.8 kb upstream enhancer sequence does (§2.1.2). On the other hand, during the functional dissection analysis, no case became apparent that would imply redundancy in *cis*-regulatory information controlling procephalic expression of segment polarity genes. This outcome underlines once more that distinctive DNA-protein interactions within the context of transcriptional regulatory mechanisms are involved in the establishment of procephalic segment polarity gene expression.

Collectively, implication of distinct transcriptional mechanisms controlling procephalic segment-polarity gene expression is in agreement with the previously described implication of discrete gene networks and regulatory interactions in patterning of the anterior head region and formation of the procephalic segments (reviewed in Rogers and Kaufman, 1997; Gallitano-Mendel and Finkelstein, 1997).

3.1.4 Procephalic expression of segment polarity genes is also under temporal control

Furthermore, additional *cis*-regulatory elements are involved in temporal control of distinct spatial sub-patterns of the whole expression pattern of segment polarity genes throughout development. That could be also observed particularly for control of procephalic expression pattern during the functional enhancer dissection analysis.

Intercalary stripe of reporter expression driven by the *wg* upstream enhancer [-16.2_-7.8 kb] is present at stage 11 but not at stage 10 (Fig. 2_6). This delay in reporter expression coming-up may indicate that intercalary-specific *wg* expression is under the control of independently operating or co-operating *cis*-binding elements recognized by distinct transcription factors during different developmental time windows. If such 'early control' elements and 'late control' elements exist, then the enhancer region tested lacks the early *cis*-controlling elements. An alternative interpretation is that only one type of intercalary-specific *cis*-regulatory elements lies within the [-16.2_-7.8 kb] upstream enhancer of *wg*, but the early activation step requires synergistic function of *cis*-binding elements not included in this upstream enhancer sequence tested.

Such a molecular mechanism seems to actually underlie the temporal control of the early onset of intercalary specific-expression of *hh* (§2.5); Although enhancer fragment [-3914_-3465 bp] drives specific expression in the intercalary segment during stages 9-11, early onset at stage 8 is ensured by enhancer sequence [-4014_-3985 bp] which acts as an additional temporal control element.

3.1.5 Search of antennal- and ocular-specific cis-regulatory information in the hh upstream region

As presented in §2.6, in an effort to localize *cis*-regulatory elements that control expression of *hh* in the antennal and ocular segments, eight (~600 bp) partially overlapping subfragments spanning region [-6.43_-4 kb] were tested. However, none of them mediated specific expression of the reporter (the constructs were assayed by the *attB-attP* site specific integration system using the *attP* landing site '96E ' which was functional for the ic-CRE expression outcome). Failure during this enhancer dissection analysis to confine *cis*-regulatory information that would mediate segment-specific embryonic expression raised the question if this is due to disrupting individual *cis*-binding elements during the dissection analysis, or because a high degree of synergy between individual *cis*-

elements dispersed along different regions of the enhancer is required to determine anterior head segment-specific expression in the ocular and antennal segments. If this second possibility is indeed a prerequisite to be fulfilled, then a molecular mechanism mediating functional interactions among distant enhancer elements (and/or with core promoter elements) has to be considered. Could it be that the putative zeste site found within the ic-CRE is involved in a 'looping mechanism' serving such interactions? Or is it rather the overlapping 7 bp element recognized *in vivo* by (a) factor(s) mediating a similar effect? Notably, this 7 bp element is present once in the promoter and at a single distant position within the full-length upstream enhancer which leads to speculations whether this could be the positional point for a loop-formation. In the light of very recent publications, considering the occurrence of promoter elements in the upstream enhancer sequence (including the PCE-resembling 7 bp element which is found only once in the full-length upstream enhancer within the ic-CRE, §3.3) triggers the postulation that maybe the ic-CRE could also function as a distant centre of general transcription factor (GTF) recruitment, thus facilitating promoter transcription initiation (Szutorisz et al., 2005; Kim et al., 2006; Krysinska et al., 2007).

Moreover, regarding antennal-specific transcriptional regulation of *hh,* elements controlling expression in the most dorsal part of that segment are included in the [-3.17 kb] enhancer region (§2.1.4) which lacks however elements controlling expression in the rest of the segment, implying that segment-specific *cis*-regulatory information may indeed be dispersed along the full-length upstream enhancer.

In order to address the above assumptions, a series of reporter constructs have been prepared aiming to elucidate transcriptional mechanisms controlling *hh* expression in the antennal and ocular segments. That is: a) testing enhancer fragment [-6.43_-4 kb] in combination with the endogenous promoter region in order to confirm that the *cis*-elements required for antennal and ocular expression are indeed sufficiently included in that region. Then further dissection in a different way from that originally employed could be followed to isolate segment-specific *cis*-information in a more defined context (i.e. dissection generating 5' and 3' nested deletion series) b) Additionally three more reporter constructs bearing point mutations affecting independently the 7 bp enhancer element, the overlapping zeste site found within the ic-CRE and the functional Collier binding site have been generated (Fig. 3_1).
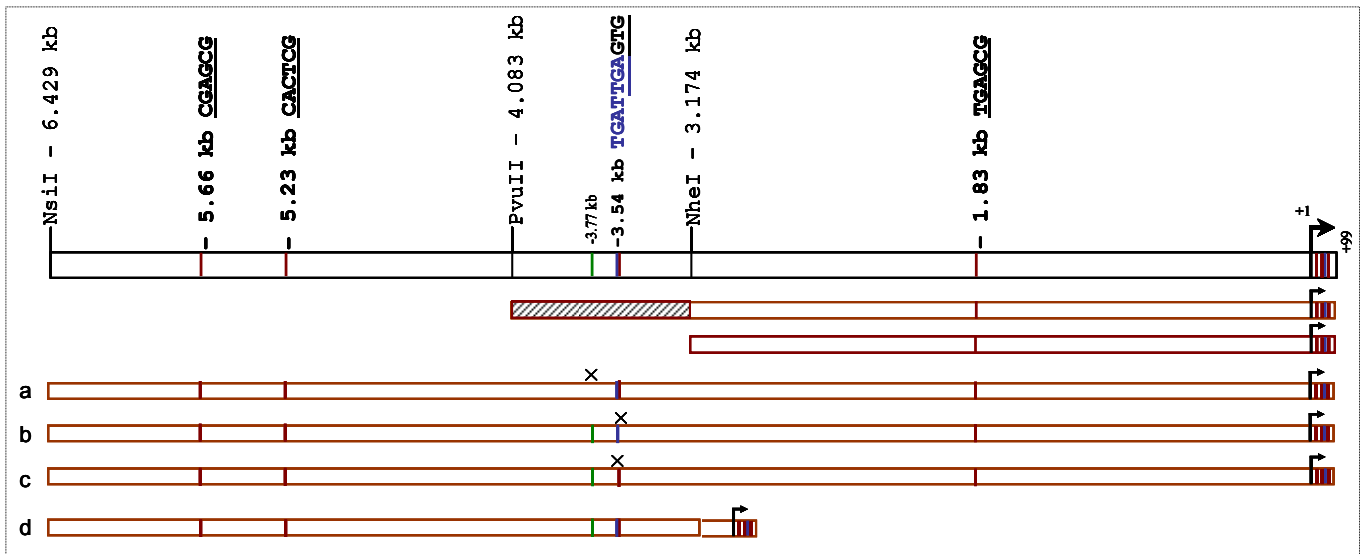
Fig. 3_1. Full-length upstream enhancer of *hh* (-6.43 kb) sufficiently drives head and trunk segmented expression pattern of the reporter (Fig. 2_14). Overlapping subfragments (a-h) of ~600 bp spanning *hh* upstream enhancer (-6.43_-4.08 kb) do not mediate specific expression pattern. The intercalary *cis*-regulatory element (ic-CRE) (-4.083_-3.174 kb) is marked. Green bar at -3.77 kb represents the functional identified Collier binding site. In total, four conserved putative zeste sites (CRCTCR) can be found in the full-length upstream enhancer. Constructs generated to be tested are d: wild type; a, b, c,: point mutant versions. Construct d aims to elucidate whether the *cis*-regulatory elements that control procephalic expression of *hh* in the antennal and ocular segments are sufficiently included within the region (-6.43_-3.17 kb) (tested in combination with the endogenous promoter region (-120_+99 bp)). Construct (a) disrupting the functional identified Collier binding site in the context of the full-length upstream enhancer is expected to affect expression of the reporter specifically in the intercalary segment. Construct (c) aims to elucidate a potential role of the conserved 7 bp element (blue bar) partially overlapping the zeste site of the ic-CRE, and found once in the full-length upstream enhancer sequence and once in the promoter. Construct (b) assayed allows to compare expression outcome of c with the mutant version that affects the overlapping zeste site (-3.54 kb) in the context of the -6.43 kb upstream enhancer.

3.2 Functional role of HMG-box factors in transcriptional regulation during early embryonic development in regard to the early onset of the ic-CRE expression

One mechanism involved in transcriptional control underlying patterning during early embryonic development is regulation of chromatin structure in order to facilitate assembly of multiple protein-DNA complexes and DNA-bound protein-protein interactions. Members of the High Mobility Group (HMG)-domain superfamily are involved in modulating chromatin structure by displaying a functional architectural role; although they may be either incapable of direct transactivation (in the case of several SOX subfamily members; Kamachi et al., 1995; Yuan et al., 1995) or they may contain a transactivation domain functional only in a specific context of protein-protein interactions (in the case of Lef-1; Giese et al., 1993), the property of the HMG proteins to induce strong DNA bending upon binding (Giese et al., 1997; Dragan et al., 2004) suggests that they can facilitate assembly and stabilization of transcription factor-DNA complexes.

The HMG domain constitutes a discrete DNA-binding structure of ~80 aa. The superfamily can be divided in two subgroups; members of the TCF/SOX/MATA group contain a single sequence-specific HMG domain which binds in the minor groove of DNA recognizing variants of the motif sequence `WWCAAAG` (Laudet et al., 1993). Members of the HMG/UBF subgroup have multiple HMG domains which bind DNA in a less sequence-specific manner (Grosschedl et al., 1994; Soullier et al., 1999). Both sequence-specific (SS) and non-sequence-specific (NSS) HMG-box proteins can modify chromatin structure by bending and unwinding DNA (Giese et al., 1997; Dragan et al. 2004), meaning that their functional role is basically architectural, i.e., to facilitate simultaneous binding of other sequence-specific transcription factors. An example stands for sequence-specific binding of the Sox2 HMG-box protein which participates in complexes with POU domain factors to facilitate transcription during early embryonic development (Dailey and Basilico, 2001). A consequence of their architectural role is that DNA-recognition and binding by the HMG-box factors leads to transcriptional activation only in a specific transcription factor-binding site context (Giesse and Grosshedi, 1993).

SOX proteins (Sry-related High-mobility-group Box) are related to the mammalian testis determining factor SRY, sharing at least 60% identities in their HMG-box DNA-binding domains (Laudet et al., 1993). Sox family members have been implicated in cell fate specification and differentiation, in processes such as male differentiation, neurogenesis and skeletogenesis (reviewed in Lefebvre et al., 2007). In *Drosophila* eight Sox genes have been identified and their expression patterns determined (Cremazy et al., 2001).

In §2.5.2, results supporting the role of HMG-related activity in temporal control of the early onset of the ic-CRE expression are presented. After 5' dissection of the enhancer sequence [-4085_-3799 bp] it came up that fragment [-4014_-3985 bp] is essential for ensuring early onset of reporter expression in the intercalary anlage at stage 8 (Fig 2_35_B). This 30 bp element consists of two highly conserved (12 sp.) blocks `GGAT`**`C`**`AAAaGG` and `GTTGA`**`C`**`AAAt` (Capitals represent conserved residues). Both conform to the general HMG-box binding consensus [`W`**`C`**`AAAS`] (NCBI Conserved Domain Database; 'cd01388 Sox-TCF_HMG-box'; Love et al., 1995; Werner et al., 1995) with one mismatch being in both cases a non-conserved nucleotide. Additionally, 3' juxtaposed to the second block one more putative HMG-box recognition site `TA`**`C`**`AAAC` lies in reverse complement orientation (-3977_-3983) matching the `W`**`C`**`AAAS` consensus. This sequence is filtered through 11 species phylogenetic conservation, with the *D. yakuba* species sequence being divergent (i.e. not in capitals in Fig. 2_37).

In particular the first conservation block `GG`<u>`AT`**`C`**`AAA`</u><u>`a`</u>`GG` scores the binding matrix of dTCF/Pangolin [`WT`**`C`**`AAAS`] (MatInspector; Lee and Frasch 2000). It also strongly resembles the consensus binding sequence of dTCF determined by PCR-based site selection [`GAT`**`C`**`AAAGG`] (van de Wetering et al., 1997), which matches the canonical binding site consensus of the mammalian Lef1/TCF [`WWT`**`C`**`AAAGG`] (van de Wetering et al., 1993). The predicted site displays only one mismatch to dTCF binding consensus which is the additional non-conserved `A`. Indeed, the wild type sequence is recognized *in vitro* by dTCF with very low affinity, but a mutated version of the probe removing the extra `A` leads to specific shift complex formation with *in vitro* expressed dTCF (Fig. 2_36). It is interesting that the one extra A in the sequence that distorts matching to dTCF consensus (`W`**`TC`**`AAA`**`S`**) and abolishes *in vitro* recognition and binding, is not phylogenetically filtered and disturbs the conservation block. Nevertheless, 11 *Drosophila* species carry an `A` at position 7 and it's only *D. grimshawi* that carries a `C` instead. *In vivo,* it is likely that the site is recognized by another member of the HMG family as it still strongly conforms to the <u>`WW`**`C`**`AAW`</u> HMG-box binding consensus (Lee and Frasch, 2000).

Although *in vitro* expressed dTCF does not efficiently recognize in EMSAs any of the wild-type HMG-box putative binding sequences (found within the DNA stretch that confers the early onset of ic-CRE expression), it cannot currently be excluded that *in vivo* the ubiquitously expressed dTCF contributes to regulation of the ic-CRE expression outcome through direct binding. It was not however possible to detect a delay in the onset of reporter expression in the intercalary segment, controlled by the 'γ1_mF3_hh R4 550 bp' enhancer fragment (Fig. 2_35_I) in a loss-of-function background; dTCF activity was down-regulated by RNAi (UAS-*pangolin* hairpin (VDRC) driven from the maternal driver {*pnos-GAL4/GCN4_bcd 3' UTR*}, §5.3). Regarding the resolved function

of dTCF/Pangolin as a *wingless* signaling effector (van de Wetering et al., 1997; Lee and Frasch, 2000), even if dTCF activity is involved in the early establishment of intercalary-specific expression of *hh in vivo,* then it is certainly independent of upstream *wingless* activity, since onset of *wingless* expression in the intercalary segment succeeds *hh*.

Furthermore, both highly conserved block sequences of the 30 bp fragment conferring early onset of the ic-CRE expression, plus the 3' juxtaposed site (found at reverse-complement orientation), they all strongly resemble consensus binding sequence of the Sox-subfamily of HMG-box proteins [WWCAAW] (reviewed in Lefebvre et al., 2007 and references therein; Churchill et al., 1995). Specificity of DNA sequence, which consists of a cluster of three putative HMG-box recognition sites, in combination with previously reported critical functions of HMG-box factors in transcriptional regulation during early embryonic development (for example, Dailey and Basilico, 2001), supports that *Drosophila* HMG-box activity, perhaps other than dTCF, may indeed be involved in the establishment of *hh* expression in the procephalic intercalary segment anlage, ensuring the early onset of expression at stage 8.

Interestingly, one of the previously characterized *Drosophila* Sox homologues (Cremazy et al., 2001), *Dichaete/Fish/Sox70D* displays an early embryonic expression pattern that would recruit it as a very good candidate for being directly involved in the early activation steps of the ic-CRE expression. At stages 5/6 it is expressed in a procephalic ventral-lateral region, corresponding to the presumptive intercalary anlage, overlapping the early anterior procephalic expression domain of *hh* (Nambu and Nambu, 1996; Russell et al., 1996). During stages 7-10 strong expression is detected in the procephalic neuroectoderm including the intercalary segment. Defects appearing in the tritocerebrum anlage (Sorriano and Russell, 2000) of homozygous mutants indicate that the early embryonic function of *Dichaete* is involved in the establishment of cell-fates within the intercalary segment.

Moreover, *Dichaete* null mutants exhibit severe segmentation defects, including organization of head structures (Nambu and Nambu, 1996), and Sox-HMG-box function of *Dichaete* has been implicated in segment-specific transcriptional regulation of pair-rule as well as independently segment polarity gene expression. For example, loss of *wg* expression in the maxillary and labial segments was reported. Although it was suggested that it is rather unlikely that *Dichaete* is required for the initial activation of segment polarity gene expression, at least not in the gnathal and trunk segments, still a potential involvement in temporal control of the ic-CRE early onset of expression remains to be clarified. Nevertheless, it was proposed that *Dichaete* may directly modulate regulatory function of transcriptional complexes required for segment polarity gene expression. In addition, Sox-domain proteins can physically associate with transcription factors from divergent

families (reviewed in Wilson and Koopman, 2002) and by providing their strong DNA-bending properties they facilitate the assembly and stabilization of transcription factors-DNA binding complexes, thus providing a crucial architectural role in the establishment and coordination of transcriptional regulatory interactions.

Finding of a cluster of three HMG-box sites in the DNA stretch conferring early onset of the ic-CRE expression, all conforming to the SOX DNA-binding sequence consensus, as well as early procephalic ectodermal expression pattern of *Dichaete* Sox-encoding gene overlapping the anterior procephalic broad expression domain of *hh* and reported function of *Dichaete* in the development of the intercalary segment and transcriptional regulation of segment polarity genes *wg* and *en,* would all in a concert constitute reasonable directions to examine if Sox function of the *Dichaete* candidate is involved in ensuring early onset of the ic-CRE expression. This would subsequently lead to examine whether the Sox function of *Dichaete* is also involved in early procephalic transcriptional regulation of *hh* expression.

3.3 Molecular basis of an enhancer-promoter specific interaction

As mentioned in §2.1.4, transcriptional outcome of the intercalary-specific *cis*-regulatory element depends on the presence of the endogenous promoter region (-120 bp_+99 bp) in the reporter construct. The ic-CRE is not functional in combination with a *hs43* TATA-box (`TATAAA`) minimal promoter (Thummel and Pirrotta, 1991). *In silico* analysis and phylogenetic footprint of the 220 bp TATA-less endogenous *hh* promoter sequence used in the functional reporter constructs generates the following predictions;

Fig. 3_2. *In silico* analysis of the *hh* promoter region. Capitals represent conserved nucleotides.

- An Initiator element (**Inr**); Consensus `TCAKTY` (Lim et al., 2004)

- A Downstream Promoter Element (**DPE**); `RGWYV(T)` (Butler and Kadonaga, 2001)

- Three putative **Zeste** recognition sites `CRCTCR`

- A Motif Ten Element (**MTE**) overlapping the third putative zeste site and the first two nucleotide residues of the DPE; Consensus `CSARCSSAACGS` (Lim et al., 2004)

- A putative GAGA factor recognition element; `GAGAG` (Orihara et al., 1999)

The presumptive function of the *in silico* identified elements is presented later in this section. Additionally, one more phylogenetically conserved 7 bp sequence [TCAATCA] is found at +54, 3' juxtaposed to the Inr element (Fig. 3_2). Interestingly, this highly conserved '7 bp element' is also present only once in the full-length upstream enhancer, falling within the ic-CRE sequence (position -3533_-3539 bp) where it partially overlaps on opposite strand a putative zeste site (§2.2.3). Sequence of the 7 bp element in the reverse complement orientation resembles an eight nucleotide sequence TGATTGAT which was identified as the PCE (Pbx consensus element). Although the PCE is present in many eukaryotic promoters its function and relevance in transcriptional regulation remains unknown (Knoepfler and Kamps, 1997). Interestingly here, it is also present at a distant position (-3539_-3532; TGATTGAg) falling within the ic-CRE.

In order to identify functional sites mediating interaction of the ic-CRE with the endogenous *hh* promoter, a series of constructs bearing point mutations have been generated. Because of several reports implicating Zeste to display in many cases a key function in mediating enhancer-promoter interactions (presented later in this paragraph), the focus was initially placed on the predicted conserved putative zeste sites which match consensus sequence CRCTCR, and they are present both in the promoter and the ic-CRE sequence (§2.2.3). As mentioned in §2.2.1, a 344 bp fragment ('F6_R5') 5' overlapping the 335 bp '*F5_R4*' fragment fails to drive reporter expression in the intercalary segment, although it contains the functional activating Collier binding site (§2.7). Interestingly, this fragment lacks the sixth conservation block of the minimum ic-CRE consisting of the zeste site and the overlapping 7 bp element (§2.2.3). A point mutation disrupting the zeste recognition site in the context of the 335 bp minimum ic-CRE abolishes reporter expression. However, the same point mutation also affects the 7 bp element, therefore it cannot be excluded that loss of reporter expression is not due to preventing Zeste binding, but rather because disrupting the function of the 7 bp element. Thus, in a new round of point mutagenesis screen in the context of the 450 bp 'γ1_mF5' ic-CRE sequence (§2.5.1) two additional point mutations have been constructed that independently affect the 7 bp element (a**tga**ttgagtgc>a**cag**ttgagtgc) and the putative zeste site (atga**ttgagtg**c>atga**ttgacat**c) (sequenced in the attB vector). Additionally, two point mutations have been generated in the context of the promoter that independently affect the 7 bp element, and the DPE (Listed in Table 5-1.4). The two mutant versions of the promoter can be combined either with the ic-CRE sequence or with the full-length upstream enhancer to study effects on the transcriptional outcome after integration of the reporter constructs at the same genomic position using the attP-attB site-specific recombination system. This will shed some light on functionality of individual core promoter elements of the *hh* promoter region. Also in combination with the results from the enhancer point mutagenesis screen affecting separately the 7 bp element and the zeste site found within the ic-CRE enhancer, conclusions can be easier extracted regarding the prerequisite in ic-CRE expression for a promoter-specific interaction .

During functional dissection of the *hh* upstream enhancer, it became evident that the 1 kb intercalary-specific *cis*-regulatory element [-4.08_-3.077 kb] as well as the overlapping subfragments that retain expression in the intercalary segment ('γ1 620 bp; F5_R4 335 bp', §2.2.1) were functional only when combined with the endogenous promoter region (-120 bp_+99 bp) but not with a *hs43* basal promoter. Since expression outcome of a distinct *cis*-regulatory module, namely the *wingless* clypeolabrum-specific distant enhancer element (Fig. 2_5) was feasible in combination with the *hs43* basal promoter, it was concluded that transcriptional outcome of the *hh* intercalary-specific *cis*-regulatory element depends on a functional specific interaction or 'communication' with the *hh* promoter, and that the observed specificity is not due to failure or inability of the *hs43* basal promoter to respond to potential activators and/or functional enhancers-input activity.

In order to understand the molecular basis of this enhancer-promoter specificity, focus was initially placed on the specific characteristics or discrete core elements of the *hh* endogenous promoter. One major difference between the *hs43* minimal promoter and the *hh* promoter region is that the *hs43* basal promoter has a functional TATA-box (`TATAAA`) (Thummel and Pirrotta, 1991) while the *hh* promoter is a TATA-less promoter and instead the presence of a Downstream Promoter Element (DPE) is one of its crucial core elements. Core promoters that contain either a TATA or a DPE motif typically depend on these elements for core promoter transcriptional activity (by assembly of the basal transcriptional machinery) and are therefore termed TATA- or DPE-dependent promoters (Butler and Kadonaga, 2001). Both the TATA-box element plus the DPE are recognition sites for binding by TFIID. Addition of a DPE at its normal downstream position in a TATA-dependent promoter may compensate *in vitro* for a mutation in the TATA motif. Occurrence of the DPE motif in *Drosophila* core promoters is as common as the TATA-box (Kutach and Kadonaga, 2000).

A study reported in Butler and Kadonaga (2001) revealed occurrence of transcriptional enhancers in *Drosophila* that are specific for promoters containing either DPE or TATA box elements. More specifically, the hypothesis tested was the following: is it possible that some transcriptional enhancers function specifically either with a TATA-box or with a DPE and what is the molecular basis of such specificity? In a genomic screen (using a double recombination transgenic system and enhancer trapping) some enhancers were identified as DPE-specific, some as TATA-specific and some as non-specific. Importantly, the DPE- or the TATA-specific enhancers appeared to activate transcription only from its cognate core promoter. A similar mechanism could also underlie promoter-specificity of the *hh* ic-CRE; this is not the case for the *wg* clypeolabrum-specific enhancer which also functions with a TATA-box *hs43* promoter. The *wg* endogenous promoter is also TATA-less but notably there is no detectable DPE sequence matching consensus `RGWYV(T)`. The only

detected core promoter element in the *wg* promoter sequence is the Initiator element (Inr) [TC<u>A</u>KTY], here typically encompassing the transcription start site (A +1).

Specificity of some enhancer elements for DPE-core promoters versus TATA-box promoters indicates that the core promoter functions not only to specify transcription initiation but also as a transcriptional regulatory element. Enhancer-promoter specificity extends beyond preference for either DPE or TATA-box core element; in other cases transcription from a TATA-less, Inr-containing core promoter was found to strictly depend on the Inr core element even if a TATA-box is added at the functional position of -30 (Garraway et al., 1996). The initiator element (Inr) TCAKTY typically encompasses the transcription start site (the A is typically designated the +1 of the core promoter) which functions either by itself or in conjugation with a TATA-box or with a DPE motif to direct accurate transcription initiation. Collectively, it is inevitable to consider the activity of all core promoter elements (TATA, Inr, BRE (Gershon et al., 2008), and DPE) in regulation of enhancer function. In other words, core promoters not only mediate initiation of transcription by the basal transcriptional machinery but also function as regulatory elements in determining transcriptional outcome of enhancers. They are indispensable active components of the regulatory mechanism of gene activity. In any case, requirement of an enhancer for its cognate core promoter constitutes one additional level of specificity that has to be fulfilled in regulation ensuring spatial- and temporal-specific expression outcome of gene activity.

Specificity between enhancers and promoters was also observed in other cases (Li and Noll, 1994; Merli et al., 1996) but it was not clarified whether the enhancer-promoter specific interaction effect on the transcriptional outcome is caused by proximal-promoter sequences containing recognition sites and bound by sequence-specific transcription factors (others than basal transcriptional machinery components) or whether it is rather due to functional presence of specific core-promoter elements. In the study of Butler and Kadonaga (2001) data presented made it clear that the enhancer-promoter specificities observed *in vivo* were dependent upon functional DPE versus TATA-box core element requirement.

The mechanistic basis or the molecular interactions underlying enhancer-core promoter specificity has just began to get resolved. Two examples from the most recent literature are mentioned below:

a) Transcription factors (sequence specific) that bind to the DPE-specific enhancers might recruit basal transcription factor NC2 (Negative Cofactor 2) which activates DPE-transcription and represses TATA-transcription (Willy et al., 2000). Still, the biochemical basis of NC2- core element-dependent DPE-activation or TATA-repression mechanisms has just become to get elucidated (Hsu et al., 2008), and it is yet unknown which factors recruit

NC2 and what is their binding specificity (or their sequence-specific recognition sites) within the DPE-dependent enhancer elements. Nevertheless, the proposed mechanism can at least explain the reason why DPE-specific transcriptional enhancers fail to activate transcription from TATA-box promoters.

b) At this very moment there is only one recently reported case of a sequence-specific transcription factor, which binds distant control enhancer elements *in vivo*, and for which the inherent transcriptional activation potential was proven to depend upon a specific core-promoter element; Caudal was shown to function as a DPE-specific transcriptional activator in *Drosophila* Schneider S2 transfection assays (Gershon et al., 2008) in correlation with the fact that identified downstream gene targets all contain TATA-less, DPE-dependent promoters.

Could it be that the observed specificity of the *hh* ic-CRE for the endogenous promoter region, versus the *hs43* basal promoter, actually lies on DPE-specific input activity for one of the crucial transcription factors controlling transcriptional outcome of the ic-CRE? In particular, could Collier function as a DPE-specific activator failing to trigger transcription from a TATA-dependent core-element promoter? A first experimental approach would be S2 transfection assays co-transfecting plasmids expressing the two Collier isoforms with reporter constructs bearing the identified functional Collier binding site (in tandem repeats for increased occupancy and hence efficiency) in combination with different core-element-dependent promoter sequence versions. Nevertheless, in order to determine if the ic-CRE specificity for the *hh* promoter region versus the TATA-box *hs43* promoter lies on a functional dependence upon the DPE core element, a mutated version of the promoter disrupting the DPE has been created to be tested (Table 5-1.4). However, it has been shown that loss of transcriptional activity upon mutation of a DPE (or a TATA-box) can be compensated by the addition of a discrete downstream core promoter element, namely the Motif Ten Element (MTE; Lim et al, 2004). The *hh* promoter actually already contains an MTE (Fig. 3_2) at a potentially functional downstream position, partially overlapping the first two nucleotides of the DPE sequence. In addition, it was reported that the MTE functions greatly synergistically with the DPE (as well as with the TATA-box). Therefore, in an effort to resolve the underlying molecular mechanism of the ic-CRE dependence upon the *hh* promoter, two additional mutated versions of the promoter sequence should be generated, i.e. disrupting independently the MTE and in the last mutated version both the MTE and the DPE. In conclusion, results from this promoter mutant analysis should indicate if functional specificity of the ic-CRE for the endogenous TATA-less promoter sequence, versus the TATA-box *hs43* basal promoter, actually depends upon the DPE, or on synergism also with the MTE core element, and it is not due to the presence in the proximal promoter sequence (within the -120_+99 bp region) of binding sites recognized by sequence specific transcription factors (such as the Zeste or the GAGA sites). Nevertheless, it cannot currently be

excluded that the latter contribute as well to the transcriptional outcome of the intercalary-specific enhancer element, as it is in the following discussed.

One additional feature of the *hh* promoter region used in the functional reporter constructs, absent from the *hs43* promoter, is a phylogenetically conserved GAGA element (GAGAG) found at position +89 (Fig. 3_2). GAGA elements are also present in the endogenous *en* promoter and they are thought to augment reporter expression pattern driven by the *en* intronic stripe enhancer (Orihara et al., 1999). *Drosophila* GAGA factors (GAGAG element binding) are encoded by the *Trithorax-like* gene (*Trl*) which has been shown to be required for expression of several genes of a broad spectrum of developmental roles (Farkas et al., 2000; Farkas et al., 1994). GAGA facilitates enhancer-promoter communication by bringing distal enhancer elements into close distance of proximal- and core- promoter elements, thereby looping out the intervening DNA.

In §3.2, role of HMG proteins as chromatin remodeling factors employing their strong DNA bending properties upon sequence specific or less specific DNA binding is pointed out. Although GAGA factors do not exhibit DNA bending properties, still their function is involved in establishment and/or maintenance of chromatin architecture; major reason why GAGA factors are needed for initiation of RNA polymerase II transcription is because they can relieve chromatin repression (Mahmoudi et al., 2002 and references therein). On the other hand GAGA function is also involved in heterochromatin-mediated silencing, since *Trl* mutations enhance position effect variegation (Farkas et al., 1994). Function of GAGA factors to mediate enhancer looping depends on their N-terminal POZ/BTB domain *(poxvirus and zinc figer/broad complex tramtrack bric-a-brac)* which mediates protein-protein interactions. The zinc-finger DNA binding domain of GAGA recognizes and specifically binds to the GAGAG element. Usually multiple GAGA elements are found in the vicinity of promoter sequences. Therefore, the POZ domain mediates multimerization into higher order GAGA oligomers that bind the multiple sites in a co-operative manner (Katsani et al., 1999). Function of GAGA factors depends exactly on this property of the formed GAGA oligomers with their ability to bind two DNA molecules simultaneously, thus facilitating enhancer looping and enhancer-promoter communication *in cis* or even *in trans* (Mahmoudi et al., 2002). Furthermore, presence of GAGA binding sites not only in the promoter proximity but also in distal control elements further facilitates long-range interactions mediated by GAGA. Indeed, in the case of the *hh cis*-regulatory region, a GAGA site found at +89 is not the only one as three additional sites can be found within the 1 kb ic-CRE element (-4025(rev), -3250(rev), -3126bp) (and some more dispersed along the 6.43 kb upstream enhancer). However, since the 450 bp ic-CRE sequence (-3914_-3465 bp) or the minimum 335 bp ic-CRE sequence do not contain any GAGA sites, and with only one GAGA site within the promoter (+89) it seems unlikely that in the case of the ic-CRE mediated expression enhancer-promoter specificity is due to GAGA mediated enhancer-promoter

communication. Nevertheless, it cannot be excluded that expression of the 1 kb ic-CRE is a bit stronger than the 450 bp ic-CRE outcome because of enhancer-promoter communication facilitated by GAGA function; the three sites found in the 1 kb ic-CRE sequence and the one site in the promoter, in the context of the reporter construct, may allow for the formation of a GAGA oligomer which might participate in higher order – expanding – chromatin pressure-relieving complexes *in vivo*.

Another striking characteristic of the used endogenous promoter not shared by the *hs43* basal promoter is the presence of zeste sites (`CRCTCR`). As shown in Fig. 3_2, three putative zeste sites are found within the *hh* promoter region used in the functional reporter constructs, one of them overlapping the MTE. In addition, four conserved putative zeste sites are found within the 6.43 kb upstream enhancer (Fig. 3_1); the third one falls within the ic-CRE while overlapping the '7 bp element'. Finding of zeste sites both in the promoter and the enhancer sequence of *hh* is consistent with the primary knowledge about Zeste being a sequence-specific factor that binds to enhancers and promoters of many developmental genes (Benson and Pirrotta, 1988; Pirrotta et al., 1988). Zeste is thought to regulate gene expression through modulating chromatin structure due to its role in transvection and position effect variegation (Judd, 1995). *In vitro* studies have shown that Zeste molecules, by cooperative binding to multiple sites, can form higher-order oligomers which provide a protein bridge among distant DNA regions (Benson and Pirrotta, 1988; Chen and Pirrotta, 1993). Thus, in a way similar to the GAGA-mediated function discussed above, enhancer and promoter-proximal bound Zeste protein molecules contact each other and loop out the intervening DNA. Zeste protein also supports interchromosomal enhancer-promoter interactions (transvection) which requires both DNA binding and self interaction (Bickel and Pirrotta, 1990). Since in the context of the 335 bp minimum enhancer fragment disruption of the putative zeste site abolishes reporter expression, and since zeste sites are also present in the promoter sequence, a first postulation was that a functional zeste-mediated enhancer-promoter interaction might underlie the displayed enhancer-promoter specificity. Still, this would not explain why the activating 335 bp enhancer fragment juxtaposed to the promoter sequence in the context of the reporter construct would require a zeste-mediated interaction since it seems that juxtaposing the two elements overcomes requirement for looping out intervening DNA. However, perhaps both enhancer-bound and proximal-promoter bound Zeste molecules are co-required to establish or enhance or stabilize a transcriptional synergy between the enhancer and the proximal promoter by facilitating simultaneous interactions between protein factors bound to these regulatory elements (i.e. the Collier activator and a X regulator bound at a promoter-proximal position) with the core promoter general transcriptional machinery (DPE-dependent). That would result in augmentation of the transcriptional outcome directed by the minimum ic-CRE. This idea was based on the outcome of a previous study proving that Zeste-mediated activation by a developmental enhancer does not result from cooperative Zeste binding to

distal and proximal-promoter sites, nor results in increase of proximal-promoter-bound Zeste (Laney and Biggin, 1997); although Zeste-mediated enhancer activation is independent of cooperative Zeste binding to distal and promoter-proximal elements, still both are required for an efficient transcriptional outcome of Ubx reporter constructs, indicating an underlying functioning synergistic molecular mechanism. The perspective that expression outcome of the minimum ic-CRE depends on a Zeste-mediated function was still prominent even after it turned out that expression is retained in a *zeste* null mutant background (z^a; data not shown); *zeste* gene is largely redundant and specific *Ubx* reporter constructs as well as the endogenous gene do not respond to *zeste* loss-of-function. It was concluded that this is due to functional presence of distinct *cis*-regulatory elements occupied by yet unknown factors that redundantly share the function of Zeste (Laney and Biggin, 1996).

Unfortunately, the above scenario implicating a Zeste-mediated mechanism underlying transcriptional outcome of the ic-CRE collapsed when it turned out that disrupting the enhancer zeste site in the context of the 450 bp ic-CRE sequence did not affect the intercalary expression pattern (actually it seems it increased in intensity (Fig. 2_29_c)) and additionally caused derepression of the reporter in the trunk which accumulates (increases) during germ band extension towards late stages of development (followed up until late stage 11, Fig. 2_29). At the same time it became necessary to reconsider a potential functional role for the overlapping 7 bp element which was also affected by the enhancer zeste-site disrupting point mutation (§2.2.3). Therefore, in order to elucidate the issue, two novel mutations in the context of the 450 bp ic-CRE sequence have been generated that independently affect the putative zeste recognition nucleotides and some of the '7 bp element' recognition nucleotides of the two overlapping sites (mentioned above – third paragraph of this section – Fig. 2_37; Table 5-1.5). Moreover, due to the recently observed derepression effect caused by the zeste mutation in the context of the 450 bp ic-CRE (Fig. 2_29) it became necessary to consider a role for *zeste* function in repression of the reporter expression in the trunk.

3.4 Repression as a regulatory mechanism in development in regard to the late derepression effect of some specific ic-CRE sequences

The derepression effect caused by the zeste mutation in the context of the 450 bp ic-CRE fragment is reminiscent of the late derepression pattern displayed by the 'γ1 620 bp' enhancer fragment (-4085_-3465 bp) (Fig. 2_18_g) in contrast to the well defined intercalary-specific pattern driven by the full-length 1 kb ic-CRE 'α' fragment (-4085_-3077) (§2.2.1) or the 450 bp 'γ1mF5_hhR4' ic-CRE. In addition, the minimum 335 bp ic-CRE also exhibits partial derepression effect at stage 11 (Fig. 2_19) but this is not as prominent as in the case of the 'γ1_620 bp' fragment or in the case of the zeste mutation in the context of the 450 bp ic-CRE sequence. Since in all cases trunk derepression gradually increases and becomes more prominent during germ band extension reaching

a maximum by the end of stage 11, it seems like a failure in the maintenance mechanism of a repressed (or silenced) state initially established in the cells of the ectodermal trunk stripes. In other words, the derepression effect resembles a kind of a loss in epigenetic 'cell memory' like in the case of Polycomb group (PcG) activity-mediated gene silencing. The fact that the mutation disrupting the zeste recognition site causes a similar gradually established derepression effect further enhances possibility that regulation of *hh* expression, at least in the case of the trunk pattern, involves PcG-mediated repression mechanisms.

A few recent studies have implicated a role of the otherwise positive transcriptional regulators, Zeste and GAGA in repression mediated by Polycomb (Hur et al., 2002; Mulholland et al., 2003; Mahmoudi et al., 2003; Dejardin and Cavalli, 2004). In the model proposed for *Ubx* transgenic reporter repression (Hur et al, 2002) both Zeste and GAGA factors display dual transcriptional activities (positive and negative) in a complex mechanism that chooses which kind of maintenance system (trxG-activation or PcG-repression) will be targeted to a specific promoter. Basic component that allows this mechanism to function is the intrinsic function of Zeste proteins to bind promoter-proximal sites regardless of whether the promoter is actively transcribed or not (Hur et al., 2002 citing Laney and Biggin, 1997). If the gene is targeted for activation (by sequence-specific transcription factors bound to distant/proximal functional spatial-control elements) early in the embryo then the proximal-promoter bound Zeste participates in the activating transcriptional initiation complexes formed at the promoter region. In the complexed situation, Zeste protein surfaces and that what would otherwise be exposed is now masked and not contacted by PcG proteins or PcG-recruiting factors. GAGA role in PcG-mediated repression is incorporated in this scenario in a similar way. This model is further supported by the fact that GAGA has been isolated in complex with PcG proteins from nuclear extracts (Horard et al., 2000) and also Zeste is part of a Polycomb-containing complex (Saurin et al., 2001). In addition, PcG proteins are frequently attached to promoter regions that contain proximal GAGA and Zeste recognition sites (this specific site composition is also the case for the endogenous *hh* promoter region), as it was shown in the case of the *Ubx* promoter (Orlando et al., 1998). Since PcG proteins do not exhibit sequence specificity maybe they are recruited via GAGA- and Zeste-mediated protein interactions.

In a slightly alternative model (Mulholland et al., 2003) Zeste function in PcG-mediated repression depends on the intrinsic biochemical property of Zeste which has functionally separated activation and repression transcriptional potential (Rosen et al., 1998). Authors suggest that when incorporated in a PcG complex, Zeste protein may attain a configuration that exposes only surfaces responsible for transcriptional repression. This is further supported by an experimental result *in vitro*; including Zeste in a Polycomb-containing complex ('PRC1') results in greater repression of transcription from

templates that contain no Zeste recognition sites. In this model GAGA role is incorporated only as a PcG-recruiting factor, which nevertheless results in a repression effect.

   In the case of the intercalary specific *cis*-regulatory element of *hh*, the question arises why in the context of the 335 bp minimum ic-CRE sequence (Fig. 3_3) disrupting the zeste site causes silencing of the reporter while in the context of the 450 bp ic-CRE element (Fig. 3_3) the same mutation causes late trunk derepression. First of all, it is noticeable that this derepression effect caused by the point mutation is reminiscent of the late de-repression pattern displayed by the wild type sequence 'γ1 620 bp' enhancer fragment (Fig. 3_3; Fig. 2_18_g) in contrast to the intercalary-specific pattern driven by the full-length 1 kb ic-CRE α fragment (Fig. 3_3; Fig. 2_16). A first conclusion that can be drawn from this observation is that the sequence spanning (-3465_-3077 bp; Fig. 3_3) may act as a PcG-responsive element (PRE). As mentioned above, PcG proteins do not display sequence specificity and are rather recruited by sequence-specific binding factors, for which very little is currently known. GAGA and Zeste proteins as recruiting factors were discussed above. Two GAGA sites (GAGAG) (-3250(rev);-3126) but no Zeste elements (CRCTCR) can be found in this fragment (Fig. 3_3). Other sequence specific recruiting factors are Pleiohomeotic (Pho) (Fritch et al., 1999), with footprinting core consensus GCCAK (Mahmoudi et al., 2003), and Pho-like (Brown et al., 2003). Pho and Pho-like recruit both PcG and TrxG. In addition, DSP1 (Dorsal Switch Protein 1; Dejardin and Cavalli, 2005) and Grainyhead (Blastyak et al, 2006) cooperate with Pho in recruiting PcG proteins and thus promoting silencing. Binding consensus of Grainyhead is [YGGTTT(T/M)] (Transfac M00110; Dynlacht et al., 1989). Interestingly, several sites matching the Pho binding sequence consensus (total seven) and three sites matching the Grainyhead consensus are found within the 1 kb ic-CRE sequence (Fig. 3_3), but none in the promoter.
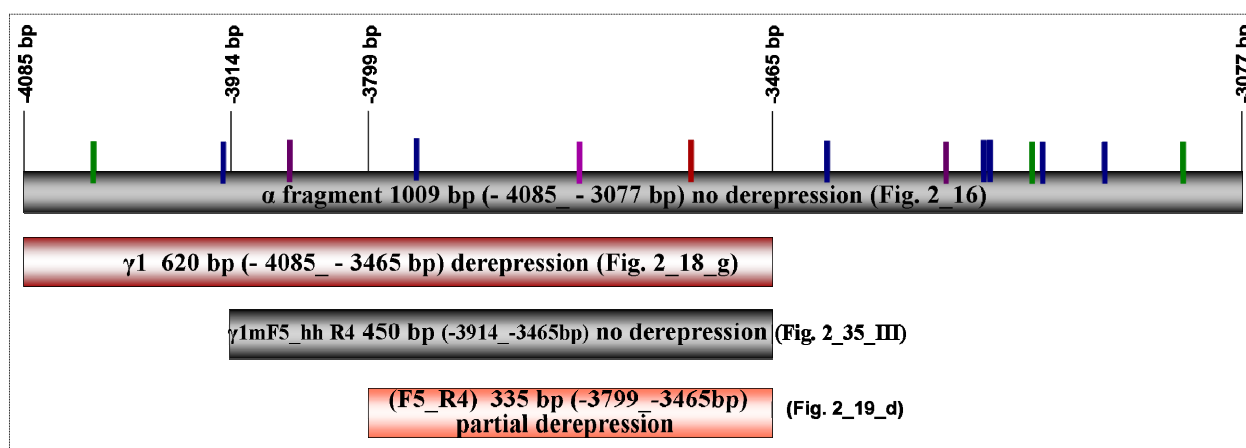

Fig. 3_3. Late (st. 11) trunk derepression effect of some ic-CRE subfragment sequences. Green bars: GAGA sites (GAGAG), Red bar: Zeste site (CRCTCR), Blue bars: Pho sites (GCCAK), Purple bars: Grainyhead sites (YGGTTT(T/M)) (the second lighter purple is CGGTTTg). Disrupting the Zeste site in the context of the 450 bp ic-CRE sequence causes late trunk derepression (Fig. 2_29).

The recruitment of PcG-mediated repression requires binding sites for early repressors (such as Hunchback and Kruppel) which actually start the repression process in the cells of the pregastrual embryos where they are expressed. The early repressor binding sites are also found within the PREs. Within the 1 kb ic-CRE sequence of *hh* such putative early repressor binding sites can be found for Hairy and Knirps. Notably, not only Hox genes but also other developmental control gene expression is subject to PREs-mediated regulation. Analysis of a PRE binding site-context has been performed in the case of *en* (Brown et al, 2005). In addition, PREs contain binding sites not only for early repressors but also activators (Tillib et al., 1999) (in the 1 kb ic-CRE sequence putative sites for Paired can be found). In the case of regulation of *Ubx* maintenance of expression it was suggested that multiple regulatory maintenance modules are essential in a coordinated mode of function for the proper maintenance of the embryonic expression patterns. Each of the modules contains both PcG-responsive elements (PREs; responsible for establishing and maintaining stable gene silencing) and trxG-responsive elements (TREs; in charge to counteract silencing and enable gene activation). Further dissection of such a regulatory maintenance module of *Ubx* indicated that independent protein complexes are formed at juxtaposed but separable DNA sequences of elements that respond to PcG-negative or trxG-positive regulatory maintaining activity. In the most recently published study employing genome-wide ChIP-on-Chip analysis (Schuettengruber et al., 2009) it was revealed that several DNA sequence-features discriminate between PcG- and Trx-N-terminal-fragment-bound chromosomal regions, indicating that the underlying sequence contains critical information to drive PREs and TREs towards silencing or activation. Furthermore, an isolated upstream *en* element that can act as a PRE and mediate PcG silencing is also activating the endogenous promoter by receiving signals from distant enhancers (Devido et al., 2008). It was concluded that *engrailed* PREs can mediate a looping mechanism and thus facilitate both positive and negative interactions between distant elements.

Perhaps, a similar complex mechanism underlies establishment and maintenance of *hh* trunk expression with maintenance regulatory modules occupying the full length upstream enhancer, and involving overlapping positive and negative regulatory elements. Finding of several dispersed along the enhancer binding sites for 'primary' recruiters (such as GAGA, Zeste, Pho, Pho-like, Dsp1) indicates that these DNA binding proteins may serve as platforms to assist PcG and trxG binding (Schuettengruber et al., 2009); PcG (negative) and trxG (positive) responsiveness input signals are then coordinately integrated into an 'orchestrated' promoter transcriptional outcome, perhaps by the involvement of a looping mechanism, which leads to the accurate *hh* expression pattern.

While the -4.08 kb region (Fig. 2_15) as well as the isolated 1 kb ic-CRE sequence (Fig. 2_16) is silent in the trunk, overlapping subfragments that derepress may not contain the DNA elements that would allow proper PcG responsiveness or may reflect isolated trxG-independently-functioning

responsive elements (Fig. 3_3). Separable but closely situated DNA sequences, specifically recognized by binding factors, still in distinct total DNA sequence context, may create a differential PcG- or trxG- recruiting environment, or even integrate competitive signals. Thus, although the 1 kb isolated 'α' element (Fig. 3_3) integrates into promoter transcriptional outcome a presumptively PcG-repressive response, the 'γ1 620 bp' subfragment (Fig. 3_3) represents an isolated trxG element that, in the absence of the 3' half ic-CRE [-3465_-3077 bp; Fig. 3_3] (which presumptively acts as a competitive PcG-responsive sequence) integrates into transcriptional outcome a signal that reflects positive maintaining regulation of the endogenous *hh* expression pattern. The idea that the 3' half ic-CRE sequence (-3465_-3077 bp) might function as an antagonizing PcG-responsive element (functional in the context of the full length enhancer and participating in coordination to specify the accurate maintaining expression pattern) is supported by the fact that trunk expression outcome driven by the 5' half ic-CRE sequence [γ1 620 bp subfragment] is much stronger than the endogenous *hh* expression pattern and even ectopic in cells posterior to the normal segment polarity stripes (Fig. 2_18_g). (This is also the case for the ectopic expression and late trunk derepression displayed by the zeste mutation in the context of the 450 bp ic-CRE (Fig. 2_29) which further suggests Zeste-dependent PcG-responsive function of the enhancer).

Moreover, in distinct sequence-contexts Zeste protein may be a component of differentially formed recruiting complexes and either required to maintain active chromatin states (Dejardin and Cavalli, 2004; Schuettengruber et al., 2009) or acting cooperatively with Pho and GAGA as a crucial determinant of PcG-mediated repression, as discussed above. Furthermore, when incorporating a potential role for GAGA factors (one site in the promoter and three sites within the 1 kb ic-CRE sequence), one should consider apart from its property to recruit PcG components to PHO-sites-containing PREs (Mahmoudi et al., 2003) also its positive chromatin remodelling – via nucleosome disruption – function connected to Trx recruitment to promoters (Schuettengruber et al., 2009 and references therein). In addition, some of the sequence-specific bound factors that serve as a platform for PcG- or trxG- protein recruitment may have an intrinsic activation or repression potential themselves, which additionally may be defined by secondary protein-protein interactions (which in turn differentially establish on distinct DNA sequence contexts). Therefore a potential antagonistic input effect has also to be integrated from the primarily DNA-bound transcription factors. That might be for example part of the reason why in the case of some ic-CRE sequences, derepression effect is exerted rather gradually during development; if transcription factors (bound to negative spatial control *cis*-regulatory elements) act as repressors in the trunk but serve as a platform for trxG-proteins recruitment then an antagonistic effect takes place and primary trunk-specific repressor activity is gradually overcome by trxG-mediated positive maintaining regulation. Since PcG and trxG maintaining mechanisms expand and establish over long chromatin regions, in such a

case a 'long-range' activating mechanism would have to antagonize a 'short-range' (resulting from an enhancer-promoter-interaction) repression mechanism.

A slightly different idea would be the following: The 1 kb ic-CRE contains a PRE component, or it exerts an intrinsic PcG-responsive activity, as it can be supported by the presence of multiple predicted Pho sites (seven in total), three Grainyhead sites, one Zeste site and three GAGA sites. This PRE activity is responsible to confer silencing of the 1 kb ic-CRE in the trunk dependent on Pho activity (Fujioka et al., 2008). However, on slightly different isolated (and partially overlapping) sub-fragment sequences of the 1 kb element distinct trunk-expression-controlling primary transcriptional states establish (due to distinguishable binding site context) that differentially regulate activity of a Pho-containing complex to mediate (trxG-) activation or (PcG-) silencing (Fujioka et al., 2008). However, in order to explain the observed late trunk derepression effect with such a scenario, a prerequisite would be that at the same time transgenic embryos exhibit the early trunk expression pattern; this feature does not regard the ectodermal segment-polarity trunk stripes although it can be observed in some dorsal epidermal ectopically reporter-expressing cells of early embryos (Fig. 2_29 b). Collectively and conclusively, perhaps different sets of cells in the trunk employ different mechanisms to establish and maintain epigenetic cell memory.

Finally, the detail that the derepression effect is observed only in the trunk segments and ectopically in cells of the intercalary segment but not in the anterior procephalic region is in good agreement i) with the fact that the most anterior segment expressing Hox genes is the intercalary and ii) with the possibility that homeodomain recognition sequences might serve as binding sites for factors that recruit or facilitate recruitment of trxG and PcG protein complex effectors.

3.5 Functional role of homeodomain proteins in development

In silico analysis identified several putative homeodomain recognition sites within the ic-CRE sequence (§2.2.3). However, none of the point mutations in the third or the fifth conservation block affected expression outcome of the ic-CRE. Moreover, each of the homeodomain sequences generated predictions for several putative recognition factors. This is due to the very similar DNA-binding properties shared by the homeodomain transcription factors; very fine structural and other means of determinants are employed *in vivo* to specify the precise DNA binding specificity (Noyes et al., 2008).

The putative homeodomain sequence within the fifth conservation block also resembles the binding site of Labial which is the homolog of the mammalian HoxA1 (Mann and Chan, 1996). *labial* belongs to the ANT-C Hox cluster and marks the intercalary segment which is the anterior most segment expressing a Hox gene. Hox genes encode a subclass of homeodomain transcription

factors being involved in cell-fate specification along the A/P axis (reviewed Pearson et al., 2005; Wiellette and McGinnis, 1999). Due to homeotic transformations in loss-of-function backgrounds, Hox factors act as regional selectors. Homeodomain proteins bind DNA with relatively low DNA binding specificity *in vitro* as well as *in vivo* (Carr and Biggin, 1999; Liang and Biggin, 1998). Nevertheless, individual Hox mutations have specific phenotypes; Hox target selection is enhanced by co-factors complex formation which determines DNA binding specificity (Ebner et al, 2005). Still, these enhanced target-selectivity complexes consist of factors which show a broad developmental spectrum of expression, so the question remains how cell group-specific transcriptional regulation is achieved. A recently reported mechanism involves direct competition exerted between a Hox protein complex (Exd/Hth/Abd) and a repressor (Senseless) for the same DNA binding sites on a cell-group-specific (sensor organ precursors) *cis*-regulatory element, with the transcriptional outcome dependent upon relative DNA binding occupancies (Li-Kroeger et al., 2008). Although Hox/co-factors complexes regulate gene expression by either activating and/or repressing transcription, with even the possibility of a single Hox input performing both functions in the same group of cells, transcriptional outcome is determined by collaboration with additional factors and such a mechanism underlies Hox-mediated cell-type specification (Kroeger et al., 2008 and references therein).

Not only Hox, but also in principle DNA binding specificity of homeodomain proteins is modified by forming homodimers, heterodimers or even higher-order complexes, with recognition properties being further specified by protein-protein interactions. In addition, not only intermolecular but also intramolecular interactions perform as DNA-binding specificity determinants, while single amino acid changes at DNA-contacting residues are sufficient to alter specificity (Noyes et al., 2008). Nevertheless, this does not mean that the defined specificity determinants prevent homeodomains from binding *in vivo* to non-target sites with minor affinity. In fact, ChIP experiments showed that homeodomain proteins do not bind *in vivo* only within specific enhancer elements of known targeted function but also occupy several sites scattered between enhancers and promoters as well as throughout long genomic regions, in agreement with their low DNA binding specificity *in vitro* (Carr and Biggin, 1999). In addition, they bind to sequences of genes not expected to be targets, collectively indicating that most developmental control genes of *Drosophila* embryogenesis are regulated to a lesser or greater extent by homeodomain factors (Liang and Biggin, 1998). These observations challenged the previously established belief that homeodomain proteins acts as gene-target specific activators and repressors; their widely observed binding occupancy, allowed by their low DNA binding specificities, may not be related to direct control or contribution to the transcriptional outcome of nearby promoters, but may instead be involved in a distinct process of great biological significance. As proposed by Dorsett (1999) widely distributed homeodomain proteins could help fulfill the function of enhancer-facilitators, participating in higher order protein-

complexes with the *Drosophila* factor Chip (*Chi*), to facilitate and promote intrachromosomal enhancer-promoter communication. This function of facilitators is slightly distinct from the looping formation mediated by GAGA and Zeste (§3.3), as it involves Chip-mediated multimeric- and cooperative homeodomain protein binding to multiple sites along the enhancer; this results in formation of a series of smaller loops that brings distant enhancer elements close to the promoter. Interestingly, Chip function is involved in regulation of several segmentation genes and embryos lacking maternal Chip show segmentation defects (Morcillo et al., 1997).

In the point mutagenesis screen (§2.2.3), affecting the putative homeodomain binding sites in the context of the ic-CRE sequence, expression of the reporter was not altered, despite their phylogenetic conservation implying a functional role. Perhaps, this is due to functional redundancy shared by individual homeodomain sites within the isolated enhancer fragment or it could also be that they are not involved in controlling the transcriptional outcome of the ic-CRE. Perhaps they have a functional regulatory role in the context of the endogenous *hh* gene locus, consistent with their scattered distribution along the full-length upstream enhancer sequence.

3.6 Transcription factor Collier mediates direct transcriptional control of segment polarity gene expression in the intercalary segment

After functional isolation of the intercalary-specific *cis*-regulatory element of *hh*, sequence of the ic-CRE was submitted to *in silico* analysis (§2.2.3). One of the predictions scored the binding matrix of the mammalian Helix-Loop-Helix COE transcription factor Olfactory-1 (Olf-1)/EBF (Early B-Cell factor) (Wang et al., 1993) which is the mammalian homolog of Dm_Collier (Crozatier et al., 1996). Olf-1/EBF and Collier share 86% identities in their DNA binding domains which would allow for a similar DNA binding specificity. Therefore, the predicted sequence was regarded as a putative Collier recognition site found within the ic-CRE. This prediction displayed quite an interest from a functional aspect, since Collier was the first factor to be introduced as a second-level regulator in the anterior head region. More specifically, the genetic interaction that was examined involves *collier* as an activator of segment polarity gene expression in the intercalary segment while being under the strict control of the head gap gene *buttonhead* (Crozatier et al., 1996).

During early embryonic development, at blastoderm stage (st. 5/6) *collier* is expressed in a domain extensively overlapping the mandibular segment anlage but slightly anteriorly displaced (Crozatier et al, 1996 and Fig. 2_66). During germ band elongation, at stage 10, *collier* is expressed in cells of the En intercalary spot (Crozatier et al., 1996), consequently it overlaps the *hh* expressing cells of the intercalary segment. At stage 10-11 it is clear that it precisely abuts the *wg* expressing cells of the intercalary segment (Fig. 2_66 e, f). At the same time, during germ band extension, it is also

expressed in cells directly posterior to the En intercalary spot which belong to the anterior most part of the mandibular segment. The cells of the anterior compartment of the mandibular segment are the hypopharyngeal progenitors (Seecoomar et al., 2000); Collier activity was reported to be required for the formation of the hypopharyngeal lobes in late germ band, while activating expression of *cap `n` collar* early in development in cells of the anterior compartment of the mandibular segment (in the hypopharyngeal progenitors cells) (Seecoomar et al., 2000). Collectively, *col* appears to be specifically expressed in cells of the posterior part of the intercalary segment and cells of the anterior part of the mandibular segment, which in whole comprises a region corresponding to the presumptive parasegment 0 (Jürgens et al., 1996; Martinez-Arias et al., 1985). Moreover, mandibular and intercalary-derived larval structures are affected in *collier* null mutants (Crozatier et al., 1996; Crozatier et al., 1999; Seecoomar et al., 2000). This result, in combination with the fact that *collier* activity is required for the establishment of segment polarity gene expression in the intercalary segment, led to the conclusion that *collier* has a parasegmental register of expression, while being a segment-specific patterning gene required for the establishment of the PS(-1)/PS0 parasegmental border and formation of the intercalary segment (Crozatier et al., 1999).

Identification of a functional Collier binding site within the intercalary-specific *cis*-regulatory element of *hh* further supports that Collier functions as a second-order regulator of embryonic head development in the procephalic region. Providing evidence that Collier directly recognizes and occupies the identified binding site, and that the established protein-DNA interaction is functional for the transcriptional outcome of the ic-CRE, leads to characterizing the segment polarity gene *hh* as a direct downstream activation target of Collier activity in the intercalary segment. In this direction, further analysis ensued *in vitro* and *in vivo* to prove that the predicted Collier recognition site is in fact a functional target-binding sequence *in vivo*, inevitably setting the ic-CRE under the direct transcriptional control by Collier. First of all, expression outcome of the ic-CRE, in concurrence with the endogenous intercalary-specific expression pattern of *hh,* depends on wild type Collier activity, since both are abolished in loss-of-function backgrounds (§2.7.1). Expression of the ic-CRE was tested in a *collier* null mutation (*col^1^*; Crozatier et al., 1999) and also in a generated transgenic line producing antisense RNA against the coding sequence of the Collier A isoform, thus leading to RNAi-mediated knocking down of the *collier* gene activity (Fig. 2_40). Second, the *in silico* predicted Collier recognition site was proven functional for transcriptional outcome of the ic-CRE, since insertion of a point mutation disrupting the site abolished expression of the reporter in the point mutagenesis screen (§2.2.3). Third, the binding site is enriched in chromatin immunoprecipitation samples with an anti-Col specific antibody (§2.7.3). Fourth, the site is specifically recognized in mobility shift assays with *in vitro* expressed Collier protein (§2.7.2). Specific complex formation is down-regulated in the presence of wild type competitor but not with a mutant version bearing the same mutation that abolished expression of the reporter *in vivo*.

Collier is one of the primarily independently cloned members of the discrete COE family of transcription factors (Dubois and Vincent, 2001). *COE* stands for Collier/ Olfactory-1 (rat)/ Early B cell factor (mouse). Additional members of the family include the mammalian EBF-2 and EBF-3, the *Xenopus* XCOE2 and XCOE3, the *Zebrafish* ZCOE2 and the *Caenorhabditis elegans* Unc-3. In general, COE proteins have various functions in different organisms. For example the mammalian COE homologues play a role in B-cell differentiation and also in neuronal differentiation (Hagman et al., 1993; Wang and Reed., 1993). All three COEs are expressed in olfactory neurons (precursor neuronal cells and mature olfactory neurons) as well as in the developing nervous system during embryogenesis. In *Xenopus* the *Xebf2/Xcoe2* homolog functions as an early embryonic regulator during primary neurogenesis (involved in the transcriptional cascade specifying primary neurons; Dubois et al., 1998), while the *Xebf3/Xcoe3* functions as a regulator of neuronal differentiation (Pozzoli et al., 2001). The *Zebrafish* homolog is also expressed during primary neurogenesis, at later stages in anterior brain neurons and it also marks olfactory placodes (Bally-Cuif et al., 1998). In *Drosophila* apart from the anterior head patterning and development of the intercalary segment (Crozatier et al., 1996,1999), Collier function has also been involved in imaginal-disk derived wing vein specification (Crozatier et al., 2002), in combinatorial specification of a single muscle lineage (along with a yet unknown factor; Crozatier and Vincent., 1999), and finally in specification and regulation of a single hemocyte type differentiation involved in immuno-response (lamellocytes; Crozatier et al., 2004). Later in embryogenesis, *collier* is also expressed in subsets of neurons of the CNS and PNS (Crozatier et al., 1996) and a critical combinatorial function of *collier* is involved in neuronal differentiation (Baumgardt et al., 2007). The nematode homolog *unk-3* is required in axonal guidance and/or neuronal differentiation (Prasad et al., 1998). Collectively, the COE family is involved in neurogenesis both in vertebrates and invertebrates.

The COE family members contain a Helix-Loop-Helix dimerization motif and a discrete zinc finger DNA binding domain while they lack the N-terminal basic region which mediates DNA interactions conducted by the basic-HLH (bHLH) transcription factors. *In vitro*, EBF factors (mouse, rat homologues) bind as homo- or heterodimers an imperfect palindromic DNA consensus sequence (binding matrix depicted in §2.2.3; Hagman et al., 1991; Travis et al., 1993; Hagman et al., 1995). Significantly, the COE protein family members share a high degree of identities in their DNA binding domains which reaches up to 86% between the mammalian and the *Drosophila* homologues. This characteristic allows for a similar DNA binding specificity and it was shown that both Collier and the *Xenopus* homologues recognize and bind cognate target sequences of the mammalian homolog EBF1 in electrophoretic mobility shift assays (Daburon et al., 2008; Pozzoli et al., 2001). For this thesis, DNA-binding specificity of Collier for a novel endogenous site, identified as functional within the context of the intercalary-specific *cis*-regulatory element of *hh,* has been

analyzed and is presented in §2.7.1-2.7.3. Further, recognition efficiency for additional putative sites predicted *in silico* within the full-length upstream enhancer of *hh* was assayed *in vitro* (§2.7.4).
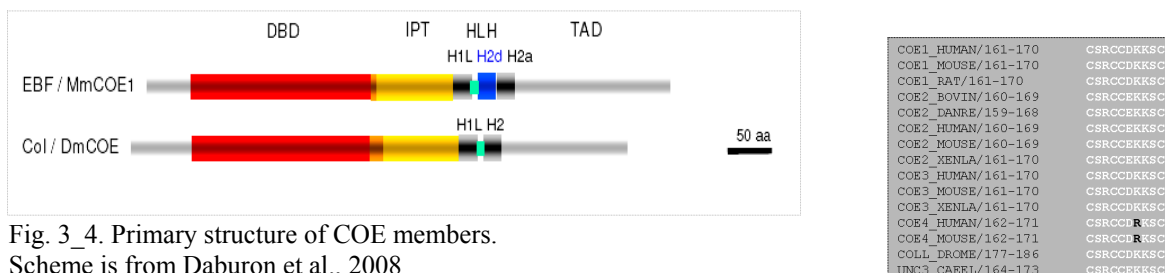
Fig. 3_4. Primary structure of COE members.
Scheme is from Daburon et al., 2008

Fig. 3_5. All members of the COE family exhibit a ten aminoacid residue signature-pattern which is located in the DNA binding domain and includes four cysteines that are part of a conserved zing finger motif. Pattern consensus is {C-S-R-C-C-[DE]-[KR]-S-C} is derived from the depicted alignment (Prosite Entry PS01345).

In more detail, the non-basic HLH COE proteins consist of i) an N-terminal DNA-binding domain (~210 aa) which shows high degree of primary sequence conservation (86% identity between the mammalian homolog and Dm_Collier), thus it is the signature of the COE family and it contains an atypical Zc coordination motif [**H**-X3-**C**-X2-**C**-X5-**C**] (Hagman et al., 1995) ii) a Helix-loop-Helix motif which mediates dimerization. All vertebrate homologues contain a duplication of the Helix2 (H2ancestral-H2duplicate) which – although is not a prerequisite for dimerization – is postulated to increase partnership flexibility and thus functional versatility of the vertebrate COE proteins (Daburon et al., 2008) (Fig. 3_4) iii) a C-terminal domain without significant primary sequence conservation which in the case of EBF was proven to function as a transcriptional activation domain (Hagman et al., 1995). In addition, an Ig-like/Plectin/Transcription Factor (IPT) domain between the DBD and the HLH domain is present but the function of this domain remains unknown. Recent experimental data presented from Daburon et al. (2008) suggest that the presence of the IPT domain interferes with the ability of the H2d-H2a domain to mediate homophilic interactions, and thus interferes with dimerization of a truncated EBF form which lacks the H1 helix. The high degree of sequence conservation of the COE IPT domain suggests that it is subject to very stringent structural and functional constraints and the relative positioning to the HLH domain may substantially affect COE dimerization (Daburon et al., 2008).

The mammalian homolog EBF/Olf-1 was initially isolated as a nuclear factor recognizing functionally important *cis*-regulatory sequences within the *mb-1* promoter which is specifically expressed in early B-lymphocytes (Hagman and Travis., 1991). EBF/Olf-1 was further characterized *in vitro* to bind as a homodimer variations of the perfect palindromic DNA sequence ATTCCCNNGGGAAT (Hagman et al., 1993; Travis et al., 1993). As mentioned above, allowed by the high degree of conservation in the DNA binding domain, the *Xenopus* COE homologues also shift oligos encompassing a binding site of EBF1 (Pozzoli et al., 2001). The EBF1 recognition site that

was used as a probe in this report is 5'-ACCCATGCTCTGG<u>TCCCCAAGGA</u>GCCTGTC-3' (Kudrycki et al., 1993). It was in addition shown that *Drosophila* Collier can bind in EMSAs as a homodimer or heterodimer with EBF either to a 125 bp DNA fragment containing the *mb-1* promoter sequence (from -250 to -115), which includes the EBF binding site AGACTCaaGGGAAT, or to an oligo probe which contains the palindromic site ATTCCCaaGGGAAT (Daburon et al., 2008). The shift generated by Collier with the mb-1 probe was assigned to be due to homodimer binding complex formation since it migrates with the same mobility as the previously characterized EBF homodimer complex. In addition, authors report no evidence for faster migrating complexes that would correspond to monomer binding. Finally, Collier was shown to efficiently heterodimerize with EBF and authors report that heterodimer complex formation prevailed over EBF homodimer shift complex, without however concluding for sure if that observation was due to favored protein heterodimerization (over EBF homodimer formation) or rather due to higher DNA binding affinity of the heterodimer situation on the mb-1 sequence. Nevertheless, and since the duplicated H2d helix of EBF is not required for dimerization of the COE proteins, it was further suggested that inclusion of the helix duplication which occurred early in the vertebrate lineage may result in higher resourcefulness in partnership, thus increasing functional potential of the vertebrate COE homologues.

As described in §2.7.2, DNA binding specificity of Collier was assayed for an endogenous *Drosophila* wild-type sequence which acts as a functional *cis*-regulatory element within the intercalary-specific *hh* enhancer fragment; in addition this sequence is enriched in anti-Col-specific chromatin IP samples (§2.7.3). Since an intercalary-specific expression pattern for ColA could not be concluded by double *in situ* hybridization using a conspecific and a ColB-specific RNA probe (§2.7.2), both isoforms were considered to be potentially involved in regulation of *hh* expression and both were assayed for DNA recognition and binding in EMSAs.

Testing the potential of the two Collier isoforms to recognize the endogenous functional binding site revealed that both proteins generate a major shift complex which according to previous studies (Daburon et al., 2008) should be due to homodimer binding. In the case of Collier B, but not Collier A, one faster migrating complex of minor intensity can be detected which is most probably due to the ability of Collier B to bind the site also as a monomer. Monomer binding specificity with little affinity has also been reported for the mammalian homolog EBF (Hagman et al., 1995). Since the two Collier isoforms differ only in their very C-terminus (29 aa for ColA versus 47 aa for ColB) with the C-terminal region of the Collier B isoform constituting a histidine-rich stretch, perhaps monomer binding-complex formation observed in the case of ColB is stabilized by additional DNA contacts mediated by the ColB-specific C-terminal tail.

In addition, from the titration experiments (Fig. 2_45) it is concluded that at least *in vitro* and on that specific DNA binding sequence, Collier A/Collier B heterodimer situation binds the DNA oligo with intermediate affinity than the homodimers, lesser than the Collier A homodimer and higher than Collier B homodimer. It also appears that in the presence of Collier A protein, Collier B molecules prefer to heterodimerize leading to diminishment of monomer complex formation (Fig. 2_45_b).

In order to investigate whether apart from affecting DNA binding affinity of the protein, the isoform-specific C-terminal regions also possess differential transcriptional activation (or even negative) properties, the potential of the factors was assayed by *in vitro* transcription run-off assays using crude embryonic nuclear extracts as a source of RNA polymerase II and basal transcriptional machinery components (§5.9). However, most probably due to carry-over of general transcription factors from the *in vitro* transcription and translation reactions (TNT; §5.7), first assays performed with TNT-expressed Collier proteins could not lead to safely reproducible results in correlation with intrinsic control experiments (data not shown).

A distinct experimental approach would be to check transactivation potential of the two Collier protein isoforms in S2 Schneider cells transfection assays using the functional binding site in repeats to drive expression of a luciferase reporter. With the same experimental setup and by using different core promoter elements in the transfected reporter constructs it could also be determined if Collier displays a specificity for a DPE core promoter element versus a TATA-box, which could underlie the reason why the ic-CRE expression outcome depends on the presence of the endogenous promoter region in the reporter construct whereas it is not functional with a *hs43* basal promoter (§3.3). In a similar study it was recently determined that a key developmental regulator, Caudal, is actually a DPE-specific activator in concurrence with the observation that identified downstream target genes contain DPE-dependent promoters (Juven-Gershon et al., 2008).

In regard with the analyzed DNA binding specificity of Collier, if the core sequence of the identified functional recognition site is used to perform a genome-wide scan analysis using the Genome Enhancer 2.0 (http://opengenomics.org/) allowing only for one subtle position (A<u>B</u>TCCCCAATGGC) still in agreement with the Olf-1 binding matrix (§2.7.2), then four predictions come up; a putative site predicted 360 bp upstream of *unplugged*; a site within the fourth intron of *zipper* (in both cases: ACTCCCCAATGGC); a site within the third intron of *zinc finger homeodomain 1* (*zfh1*) (AATTCCCCAATGGC); and a site within the second intron of CG42389 (ATTCCCCAATGGC). Notably, all four predictions are within non-coding DNA sequence with *zipper* and *zfh1* predictions displaying 8 species conservation. In addition, expression patterns of the first three mentioned genes coincide with subpatterns of *collier* expression and additionally they are

involved in developmental processes in which a functional role for Collier has also been implicated. Therefore, it is feasible that the three predictions could be direct downstream targets of Collier activity. More specifically, *unplugged* and *zipper* are both expressed in the CNS (Chiang et al., 1995; Zhao et al., 1988; Cote et al., 1987). *zipper* encodes a putative integral membrane protein required for normal axon patterning during *Drosophila* neurogenesis. However, overlap in expression with subsets of CNS neuroblasts that express *collier* has to be confirmed by *in situ* hybridization assay. Moreover, Zipper protein is most similar to the metazoan smooth myosins (Ketchum et al., 1990) and it is also involved in muscle cell differentiation (Bloor et al., 2001). Collier is also involved in combinatorial control of muscle cell lineage specification (Crozatier and Vincent., 1999). *zfh1* is expressed in CNS as well as in procephalic ventral ectoderm and procephalic mesoderm (Lai et al., 1991) where *collier* is also expressed (Seecoomar et al., 2000). Again, precise overlapping expression in sets of cells has to be confirmed by *in situ* hybridization assay but from a first glance in BDGP database it seems likely that *collier* and *zfh1* expression patterns overlap in the procephalic region. Additionally, *zfh1* function has been implicated in lymph gland development and immune response (Sellin et al., 2006), a process where *collier* is also involved (Crozatier et al., 2004).

3.7 Examine a means of restricting positive regulatory function of Collier to the posterior part of the intercalary segment

Up to now, a line of results has been presented and analyzed supporting a positive regulatory function of Collier directly involved in the activation of the segment polarity gene *hh* in the posterior compartment of the intercalary segment (§2.7, §3.6). From the embryonic *in situ* hybridizations and fluorescent immunostainings and in agreement with the reports in literature (Crozatier et al., 1999; Seecoomar et al., 2000), expression pattern of *collier* is not strictly restricted to the ectoderm of the posterior part of the intercalary segment but also exceeds into the anterior most part of the mandibular segment. Moreover, it is also expressed in the procephalic mesoderm.

In regard to the activating input exerted by Collier on the transcriptional outcome of the ic-CRE, the function of an underlying mechanism responsible for restricting ic-CRE-mediated expression only to the *hh* expressing cells of the posterior part of the intercalary segment has to be considered. This could be i) requirement of an additional synergistic activator involved in combinatorial positive regulation of the ic-CRE, presumptively of the endogenous intercalary-specific *hh* expression pattern, absent from the anterior part of the mandibular segment where Collier fails to directly activate target expression outcome ii) presence of a negative regulator in the anterior part of the mandibular segment inhibiting or competing with Collier positive function.

3.7.1 CncB as a negative regulator of the ic-CRE expression

It could be extracted from the literature and further confirmed by *in situ* hybridization and immunostaining that the expression pattern of the bZIP factor encoding gene *cap `n` collar* (Mohler et al., 1995) would qualify as an optional candidate to fulfill the above second mentioned possibility. *cnc* is required for specification of the mandibular segmental identity by suppressing the maxillary structures-promoting function of *Deformed* (McGinnis et al., 1998). From among the three Cnc isoforms produced, only CncB displays this functional role in head development (McGinnis et al., 1998; Veraksa et al., 2000).

At blastoderm stage *cnc* is expressed in a domain corresponding to the mandibular anlage (Fig. 2_66_d). At stage 10 it is expressed throughout the mandibular compartment precisely abutting the *hh* expressing cells of the intercalary segment (Fig. 2_68_a). Therefore, overlapping with *collier* expression is only in the anterior most cells of the mandibular segment (Fig. 2_67_a, b) where Collier fails to trigger the ic-CRE transcriptional outcome (Fig. 2_67_c). With onset at stage 10, a secondary site of *cnc* expression is initiated by Collier in the posterior cells of the intercalary segment (Fig. 2_68 and Crozatier et al., 1999). Consequent with the assumption that CncB could act as a negative regulator of Collier positive function, from late stage 10 on and by the end of stage 11, expression of the ic-CRE greatly diminishes in intensity to get restricted to 2-3 cells of the intercalary spot (Fig 2_55). Expression of the endogenous *hh* pattern in the intercalary segment is also decreased but this reduction is more gradual (Fig. 2_56). This difference in progress of expression pattern attenuation between the ic-CRE and the endogenous gene could be because expression of *hh* in the whole is also under the control of additional regulatory inputs, while a positive regulatory input exerted by Collier on the transcriptional outcome of solely 450 bp *cis*-regulatory element could be more easily competed by a negative regulatory input of another factor, presumptively by CncB.

Expression of *cnc* in the anterior most portion of the anterior compartment of the mandibular segment, which comprises the progenitors cells of the hypopharyngeal lobes, is also under positive regulatory control by Collier, since *col* null mutants lack *cnc* expression in the anterior most part of the mandibular segment from early gastrulation stage and the hypopharyngeal lobes subsequently fail to develop (Seecoomar et al., 2000).

Collectively, the extent of overlap in the very early expression domains of *col* and *cnc* in the blastoderm 'fate map' (Fig 2_66_a, b and Crozatier et al., 1999; Seecoomar et al., 2000) would allow for the following conclusion; the anterior row of cells expressing only *col* will give rise to the posterior part of the intercalary segment – *hh* expressing cells – since *collier* null mutants specifically lack intercalary *hh* expression (Fig. 2_39). The middle 1-2 cells-wide row of the early

blastodermal overlapping expression domain will give rise to the hypopharyngeal lobes, which fail to form in *collier* null mutants; expression of *cnc* in *col* null mutants is also defected in the anterior most cells of the mandibular segment which are the progenitor cells of the hypopharyngeal lobes. Finally, the posterior row of cells expressing only *cnc* will give rise to the mandibular lobes which nevertheless normally form and normally express *cnc* in *col* nulls. Later in development, with onset during stage 10, a secondary site of *cnc* expression is initiated by *collier* activity in the *hh* cells of the posterior part of the intercalary segment. Notably as mentioned above, ectodermal expression of *hh* in the posterior part of the intercalary segment from that point on seems to gradually diminish.

Considering *cnc* function in a negative regulatory aspect opposing positive regulatory input activity of *collier,* and taken all together, it would not look unlikely that *collier* ensures attenuation of its own activating function, at a later maintenance phase of targets expression, by early enough turning on an antagonizing negative gene function. As described in §2.8.1-2.8.3, it seems likely that the negative regulatory effect of CncB on the the ic-CRE expression outcome is indirect and most probably resulting from competing the positive regulatory input of Collier at the protein level, involving a physical protein-protein interaction. It is less likely that CncB is directly involved in transcriptional repression of the ic-CRE expression outcome. The following reasons support this conclusion: First, the ic-CRE-mediated expression pattern does not posteriorly expand, meaning it exhibits no posterior derepression in *cnc* null mutants (*VL110*; Mohler et al., 1995). Although this result could also postulate the presence of distinct bZIP factors that redundantly share a negative repressive function of CncB in the anterior most part of the mandibular segment, still it stands insubstantial to suggest that CncB inhibits ic-CRE expression in the anterior most part of the mandibular segment. Nevertheless, it seems rather safe to propose that CncB could potentially function to prevent or to soothe possibility that the ic-CRE is expressed in the anterior most part of the mandibular segment by competing with Collier positive regulatory input activity. The last can be supported by two experimental outcomes: a) expression outcome of the ic-CRE is greatly reduced, almost abolished in *cnc* gain-of-function background (Fig 2_64) by using a UAS/GAL4 system to ectopically drive procephalic CncB expression under the control of a double dose of maternal {*pnos_GAL4/GCN4_ bcd 3' utr* } driver input; transcription of the endogenous *hh* intercalary stripe is also suppressed (Fig. 2_65) b) conversely, transcripts of *hh* and ic-CRE – reporter irregularly persist in the epidermal cells of the intercalary lobes of stage 11 *cnc* null mutant embryos (Fig. 2_71, 2_72).

Second, the above 'competition' scenario, involving function of CncB in antagonizing the positive regulatory input activity of Collier, could be further supported i) *in vitro* in mobility shift assays; including increasing amounts of CncB factor in the binding reaction attenuates complex formation generated by Collier with its cognate functional target – oligo sequence – found within the ic-CRE

(Fig. 2_61). CncB-mediated inhibition of Collier complex formation on this very specific DNA target sequence, which nevertheless lacks functional recognition sites for CncB, is exerted in a dominant negative way as it could be concluded by the EMSA competition assays (Fig. 2_62). Providing CncB factor to Collier-DNA-binding complex, that has reached the point of binding equilibrium, leads to progressive 'stripping-off' of Collier molecules from the DNA-protein complex, consequently causing attenuation of shift intensity. The described negative effect of CncB protein on Collier-DNA binding complex formation indicates that a physical interaction between CncB and Collier antagonizes Collier binding to its DNA target sequence. This physical interaction could be further supported by *in vivo* co-immunoprecipitation assays from crude embryonic nuclear extracts (§2.8.3). In addition, the antagonizing effect on binding of the HLH-factor Collier to its cognate site exerted by a competitive bZIP-factor CncB physical protein interaction is reminiscent of a previously described direct physical association between the factors c-Jun and MyoD in differentiating myoblasts (Bengal et al., Cell, 1992); in fact this interaction was biochemically proven to be mediated by the leucine zipper domain of c-Jun and the HLH domain of MyoD and it results in functional antagonism at the transcriptional level; in transfection assays the presence of c-Jun inhibits transactivation by MyoD of a reporter linked to MyoD DNA-binding sites and vice-versa. Furthermore, this interaction appeared to be primary sequence-specific since the bZIP protein Fos did not interact with MyoD, although Fos and c-Jun form heterodimers via the leucine zippers. In a similar way, described in §2.8.2, *Drosophila* bZIP factor C/EBP, which is ubiquitously expressed during embryogenesis (Rorth and Montell., 1992; Montell et al., 1992) does not efficiently interact with Collier and thus does not interfere with binding to its cognate site, although CncB and C/EBP interact, most probably by forming bZIP heterodimers which fail to shift an oligo that does not contain a recognition site for the CncB partner (Fig. 2_59, 2_60).

In regard to the previously studied DNA-binding specificity of CncB, it was shown that it binds in heterodimers, in a highly cooperative manner with Maf-S bZIP-protein a consensus target sequence defined by PCR-based site selection (Veraksa et al., 2000). Multimers of the heterodimer recognition site result in transcriptional activation of a linked reporter *in vivo* in response to wild type CncB activity (Veraksa et al., 2000). Strong intrinsic transactivation potential of CncB was further proven in transfection assays (Veraksa et al., 2000). Mammalian bZIP homologues of the CNC-class have also been shown to possess transcriptional activation properties, and only those members that contain the BTB domain (such as Bach1 and Bach2) function primarily as transcriptional repressors (Veraksa et al., 2000 and references therein). Conclusively, function of CncB at a transcriptional level *in vivo* is more likely to result in activation of targets, requiring Maf-S as a cofactor. Nevertheless, the half site of the heterodimer consensus binding sequence which is contacted by CncB (KTCAT) matches the binding site of the *C. elegans* CNC homolog Skn-1 (Blackwell et al., 1994) which binds DNA as a monomer. However, the putative recognition site found within the ic-

CRE juxtaposed to the Collier cognate site was not recognized *in vitro* by CncB or CncA (§2.8.2). Collectively, it seems unlikely that CncB is directly involved in transcriptional control of the ic-CRE expression outcome.

In spite of the intrinsic transcriptional activation properties of Cnc homologues, CncB acts to suppress both the expression and the homeotic selector (maxillary structures-promoting) function of *Deformed* in the mandibular segment (Mohler et al., 1995; McGinnis et al., 1998). Specifically, although CncB is not capable of repressing *Dfd* expression before stage 11, after that stage CncB represses the maintenance phase of *Dfd* transcription in the mandibular cells, most probably by interfering with positive regulatory function of Deformed protein within the *Dfd* autoactivation circuit, which is established during stages 9 and 10 (McGinnis et al., 1998; Veraksa et al., 2000; Zeng et al., 1994). Moreover, overexpression of CncB partially represses Dfd-responsive transcriptional target elements (McGinnis et al., 1998). Interestingly, Veraksa et al. (2000) report an observation of an interaction between CncB and Dfd proteins in GST pull-downs, with an as yet unclear biological relevance. In regard to the experimental outcomes of this thesis, perhaps the negative regulation of *Dfd* expression and function exerted by CncB results from CncB interfering with Dfd binding to its functional-target *cis*-regulatory elements *in vivo*, as a consequence of a direct physical interaction at protein level.

3.7.2 Collier in combinatorial control of the ic-CRE expression

As described earlier in this section, combinatorial control of the ic-CRE involving apart from Collier an additional factor required to direct a positive synergistic input could be a distinct means of restricting expression outcome to the posterior part of the intercalary segment or it could also contribute to the restriction mechanism (§2.8.2, §2.8.3, §3.7.1). Experimental results that would line up to this conclusion are in the following discussed. First of all, the HLH domain of Collier *de facto* creates maximum prospect for cooperative partnership with other members of the extensive HLH family of transcription factors. Interestingly, primary sequence of Collier protein does not have a nuclear localization signal (NLS). That is most probably the reason why localization of Collier as revealed by fluorescent immunostaining is not strikingly nuclear, as it is for example in the case of the NLS-carrying CncB protein, but it is mainly detected in the cytoplasm (Fig. 2_70). Notably, Collier exhibits a perfect SUMOylation signal in the N-terminus which is highly conserved across the COE proteins. Apart from modification of transcriptional activation/repression potential of transcription factors, sumoylation has also been implicated in nucleo-cytoplasmic translocation of several proteins, although its function has only recently started to get resolved (reviewed in Zhao, 2007). Another means of Collier import to the nucleus would be heterodimerization with a protein that carries an NLS, possibly a distinct HLH factor.

Second, it was not possible to detect ectopic expression of the ic-CRE in a Collier gain-of-function background (*UAS-kn.M*; Mohler et al., 2000 (Bloomington #9554)) driven by {*pnos_GAL4/GCN4_3' bcd utr*} maternal input (Fig. 2_64 iv). Although this result could imply response of the ic-CRE to transcriptional repressors outside the posterior part of the intercalary segment, it could also implicate Collier involvement in positive combinatorial control of the ic-CRE expression outcome. Further, disrupting the Collier DNA-binding site was not the only mutation abolishing reporter expression in the intercalary segment. Interestingly, point mutations in a putative fork head recognition site found within the fourth conservation block of the ic-CRE also abolished reporter expression (§2.2.3). This fork head recognition sequence is a very good match of a site protected in DNaseI footprinting assay by Slp1, found within the proximal promoter sequence of *ftz* (Yu et al., 1999). Although the site was not efficiently recognized by Slp1 in EMSAs, still expression of the ic-CRE driven reporter is greatly reduced in response to *slp1* loss-of-function via RNAi (Fig. 2_30). Positive response of the ic-CRE to Slp1 activity or a direct involvement of Slp1 in transcriptional control of the ic-CRE expression outcome has to be further elucidated. Furthermore, intercalary-expression of the endogenous *hh* in *slp1* loss-of-function (null) background remains to be resolved, since in previous studies only effects on *en* and *wg* segment polarity gene expression were examined (Cadigan et al., 1994).

In addition, and as discussed in §3.6 the precise transactivation potential of the two Collier isoforms and their heterodimer form remains to be resolved by *in vitro* transcription assays and/or transfection assays. Taken all together, at the moment it cannot be excluded that apart from Collier, additional factors might as well contribute a positive input to the expression outcome of the ic-CRE, thus subsequently being involved in establishment and regulation of *hh* expression in the intercalary segment. Whether such factors act directly at transcriptional level, by recognizing and occupying binding sites on the ic-CRE sequence, consequently triggering activation of the promoter and contributing to the positive transcriptional outcome of the ic-CRE, or whether they rather act as co-factors or co-regulators enhancing positive regulatory function of the Collier activator remains to be further elucidated.

At the same time, a parallel functional mechanism that restricts expression of the ic-CRE to the posterior part of the intercalary segment potently involves a negative regulatory function exerted by CncB in the anterior most part of the mandibular segment, which is most probably indirectly exerted by competing with Collier binding to its cognate DNA regulatory element. High affinity recognition of the DNA-binding site by Collier allows for transactivation effect in the cells of the posterior part of the intercalary segment even though a very small fraction of the expressed Collier protein is nuclear localized. Conversely, nuclear accumulation of CncB in the anterior-most cells of the mandibular segment results in sequestering nuclear Collier from DNA binding. *collier* sets off

attenuation of its own positive function by initiating *cncB* expression in the cells of the posterior part of the intercalary segment during stage 10, thus triggering the DNA-binding competing mechanism at the onset of a 'maintenance' phase that leads to continuing soothing of target activation by the end of stage 11.

Summary

Head segmentation in insects, as reflected by the model system *Drosophila melanogaster*, is less well understood in comparison to the trunk. While the gnathal segments (mandibular, maxillary, labial) are patterned in a way similar to the trunk, distinct molecular mechanisms underlie the metamerization of the anterior head region into the procephalic (ocular, antennal, intercalary) segment anlagen (reviewed in Rogers and Kaufman, 1997). In that context, actually a unique mode of establishment for each of the procephalic segments has been suggested based on anterior head segment-specific cross-regulatory interactions among segment polarity genes (Gallitano-Mendel and Finkelstein, 1997). So far it remains unclear how the establishment of the procephalic segment polarity gene expression is controlled, although a functional input from the head gap-like genes has been elucidated and a second level of regulation in patterning of the anterior head region has been implicated (Crozatier et al., 1999).

In this thesis, a bottom-up approach was employed to identify key components of the head segmentation gene network by the detection and functional dissection of *cis*-regulatory regions of the segment polarity genes *wingless*, *hedgehog* (*hh*) and *engrailed* in *Drosophila* and in a comparative approach also in the short-germ band red flour beetle *Tribolium castaneum*. During this study an intercalary-specific *cis*-regulatory element (ic-CRE) of Dm_*hh* was isolated, which further supports the theory of a unique establishment of each procephalic head segment. *In silico* analysis on the ic-CRE sequence identified a recognition site for the HLH-COE factor Collier. This binding site is essential for the function of the ic-CRE that also depends on *collier* activity. DNA-binding specificity of Collier to this endogenous sequence was further supported *in vitro*. Therefore, Collier directly controls expression of the segment polarity gene *hh*. Since *collier* itself is activated by the head gap-like gene *buttonhead* (Crozatier et al., 1999), Collier functions as a second order regulator in patterning of the intercalary segment.

Furthermore, the ic-CRE also integrates temporal control; nested dissection led to the isolation of HMG-box binding sites required to ensure the early onset of expression. Regarding the means of restricting positive Collier activity to the posterior part of the intercalary segment, the effect of a physical interaction between Collier and the bZIP factor CncB was examined. The *in vitro* results indicate that CncB antagonizes DNA-binding of Collier to its target sequence, which is also reflected *in vivo* by a negative regulatory aspect of CncB misexpression on the ic-CRE-mediated expression. Moreover, functionality of the ic-CRE is exerted *in vivo* only in combination with the endogenous *hh* promoter, but not with a TATA-box basal promoter. This result indicates that the intercalary-specific transcriptional control of *hh* involves an enhancer-promoter specific interaction. However, the molecular basis of this mechanism remains to be further elucidated.

5. Materials and Methods

Standard methods of Molecular Biology and *Drosophila* Biochemistry and Cell Biology employed during this thesis were essentially as described in i) Sambrook and Russel, Molecular Cloning, A Laboratory Manual, Third Edition, CSHL Press ii) *Drosophila* Protocols, edited by Sullivan, Ashburner, Hawley, 2000, CSHL Press. Experimental procedures modified from protocols of those two sources or not included therein are presented in this section. Preparation of solutions and chemicals was also as described in those two references.

5.1 Procedures for functional dissection of *cis*-regulatory regions

5.1.1 Cloning vectors

turboGFP reporter [tgfp_SV40 952 bp sequence] was excised with AgeI (site T4 blunted)_AflII from pTGFP_PRL (EVROGEN) and inserted into PmlI (blunt)_AflII of pslaf1180af vector (Horn and Wimmer, 2000) to create pslaf_tgfp_af.2 (psltgfp.2) (map1). Internal NotI and XbaI after *tgfp* termination codon (TAA) of the original pTGFP_PRL were primarily destroyed (with restriction, T4 blunting and religation) so that psltgfp.2 contains a single NotI site to be used in the polylinker. Promoter *hs43* 156 bp sequence was excised from *pCasper_hs43_lacZ* (Thummel and Pirotta, 1991) with XhoI_PstI(T4 blunt) and subcloned in XhoI_SpeI(T4 blunt) of psltgfp.2. Promoter sequences were cloned with primers XhoI_Forward/ SpeI_Reverse. *cis*-regulatory sequences to be tested were amplified with primers bearing restriction sites found within the polylinker of pslaf1180af (upstream of XhoI), T/A cloned in PCRII vector (Invitrogen), sequenced and subcloned in the psltgsp.2 vector. DNA sequences were isolated with PCR on genomic DNA template isolated with genomic DNA extraction (0-12 h embryos).

5.1.2 Transgenesis

For *piggyBac*-transgenesis, cassettes consisting of [*cis*-regulatory region_promoter_tGFP-SV40] were excised with AscI from psltgfp.2 and subcloned in pBac[3xP3_EGFPaf] (Horn and Wimmer, 2000). *piggyBac* constructs were coinjected at 500ng/µl with helper plasmid providing transposase activity (*phspBac*) at 300ng/µl. The strain injected was *Drosophila w*⁻ and procedure as desribed in Sullivan et al., 2000. Briefly, 0-20 min embryo collections were dechorionated for two minutes in 50% chlorix (if any chorion remained it was manually removed), embryos were extensively washed with salt free water and aligned on cover-slips (on a thin stripe of tape glue in heptane to immobilize them). After short desiccation embryos were covered with a thin layer of Voltalef oil (10 S) and injected using an Eppendorf Femtojet Microinjector (Eppendorf AG, Hamburg).

For site specific transgenesis using the *attB-attP φC31* – mediated integration system (Bischof et al., 2007), the full-length *attB* sequence 314 bp was excised from pTA-attB (Calos MP, Stanford University, personal communication) with EcoRI; restriction generated ends were blunted with T4 DNA polymerase and subcloned in *pBac*[3xP3_EGFPaf] linearized with BglII (T4 blunted) generating vector *pBac_attB* (map 2). Orientation was checked by colony PCR and confirmed with subsequent sequencing. Reporter cassettes were subcloned in the AscI of the pBac_attB vector (map 2). Line injected was the one bearing the *attP*-landing site at position 96E of the 3rd chromosome (Ac. Num EF362408). This is a combined line carrying on the X chromosome a codon-optimized *φC31* integrase driven under the control of *vasa* promoter (Bischof et al., 2007). Marker for the *attP* landing site insertion was *3xP3_RFP*, while for the *vasa*-integrase insertion was *3xP3_eGFP*. Since the *attB* construct also carries the EGFP marker, male offspring individuals of the [injected x w-] crossing are specifically selected for recombination events (integrase deficient). A combined line bearing the *attP* landing site on the 2nd chromosome (position 51D) was also tested but this line was not further used due to position effect causing ectopic mesodermal reporter expression.
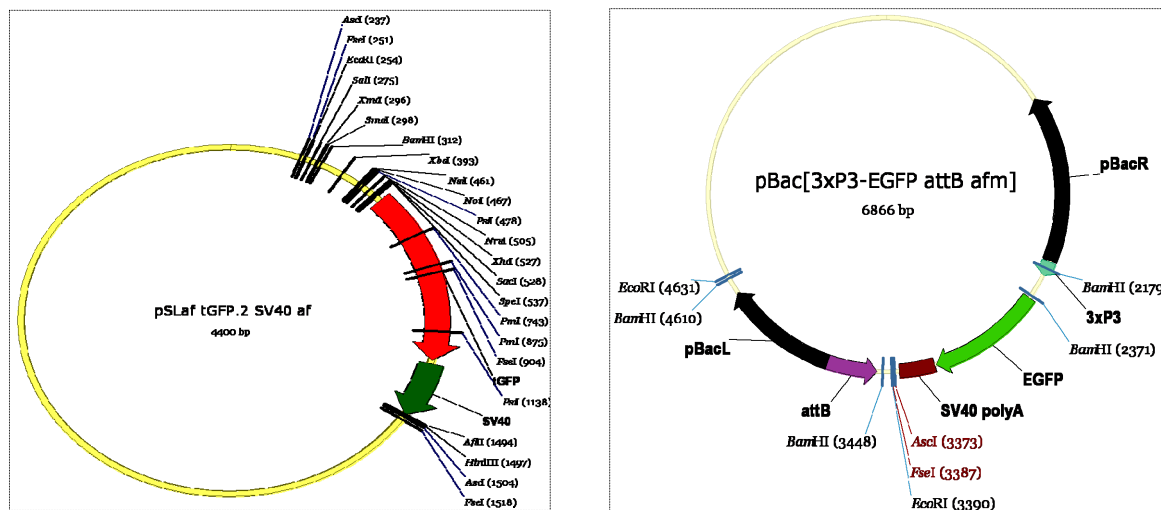
Fig. 5_1. Vector maps.

5.1.3 Construction of deletion mutants

Deletion mutant constructs lacking highly conserved sequence blocks in the context of the minimum 335 bp ic-CRE (§2.2.2) were prepared in two-step cloning; 3' part of the ic-CRE was amplified using primers NruI_Forward (mutF1-mutF6) (Table 5-1) and XhoI_*hhR4* (Reverse) and subcloned in [psltgfp.2_hh_promoter (Xh_SpeI)]. 5' part of the ic-CRE was amplified with primers BglIIF5 (Forward) and NruI_Reverse (mutR1-mutR5) (Table 5-1.3.) and subcloned in the vector containing the respective 3'part. Schematically the procedure is depicted in Fig. 5_2.
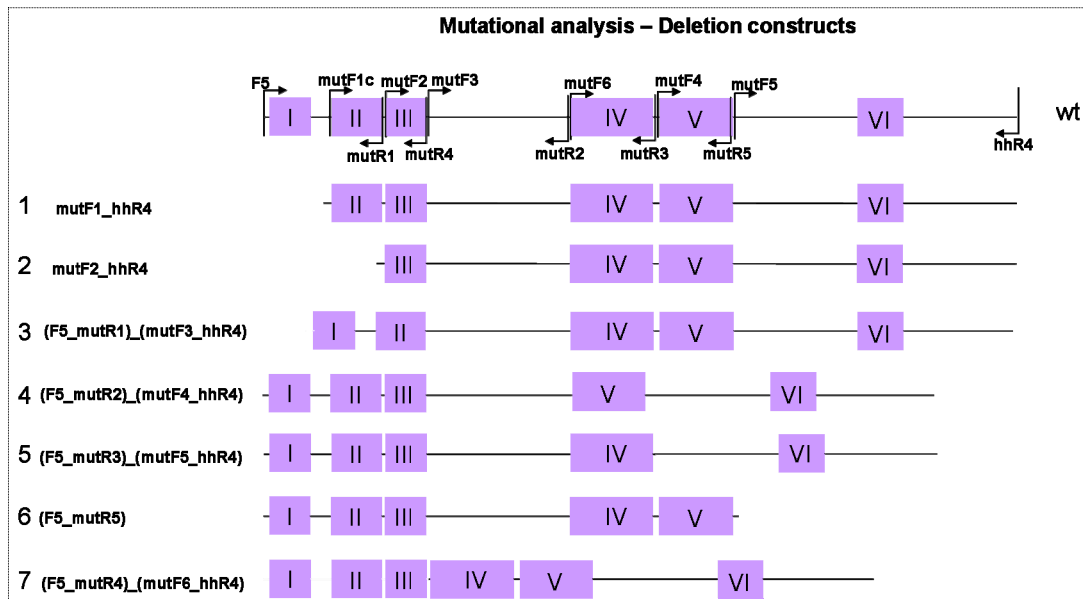
Fig. 5_2. Deletion mutant constructs.

5.1.4 Construction of point mutations

Constructs bearing point mutations in highly conserved blocks (§2.2.3) in the context of the 335 bp minimum ic-CRE or in the context of the 450 bp ic-CRE sequence were generated by overlapping PCR site-directed point mutagenesis. Overlapping forward and reverse primers were designed bearing point mutations affecting targeted nucleotide positions. Fragments [BglIIF5_'point_mut'Rev] or [BglIIγ1mF5_'point_mut'Rev] and ['point_mut'Forw_hhR4] were amplified in separate proofreading (Pfu polymerase) PCR reactions using minimum amount of plasmid or genomic template (~10ng). Product bands were excised from agarose gel, DNA was purified (Band excision protocol, QIAGEN) and eluted in 20 µl of TE pH 8.5. 2-3 µl of each elution of the corresponding 5' and 3' fragments were combined in a final 50 µl proofreading PCR reaction; after an initial round of denaturation and annealing (so that the overlapping 3' and 5' sequences of the fragments hybridize and function as primers for the next elongation round) full-length ic-CRE sequence was amplified with external primers BglIIF5_hhR4 (335 bp minimum ic-CRE) or with BglIIγ1mF5_hhR4 (450 bp ic-CRE). Procedure is schematically depicted in Figure 5_3.
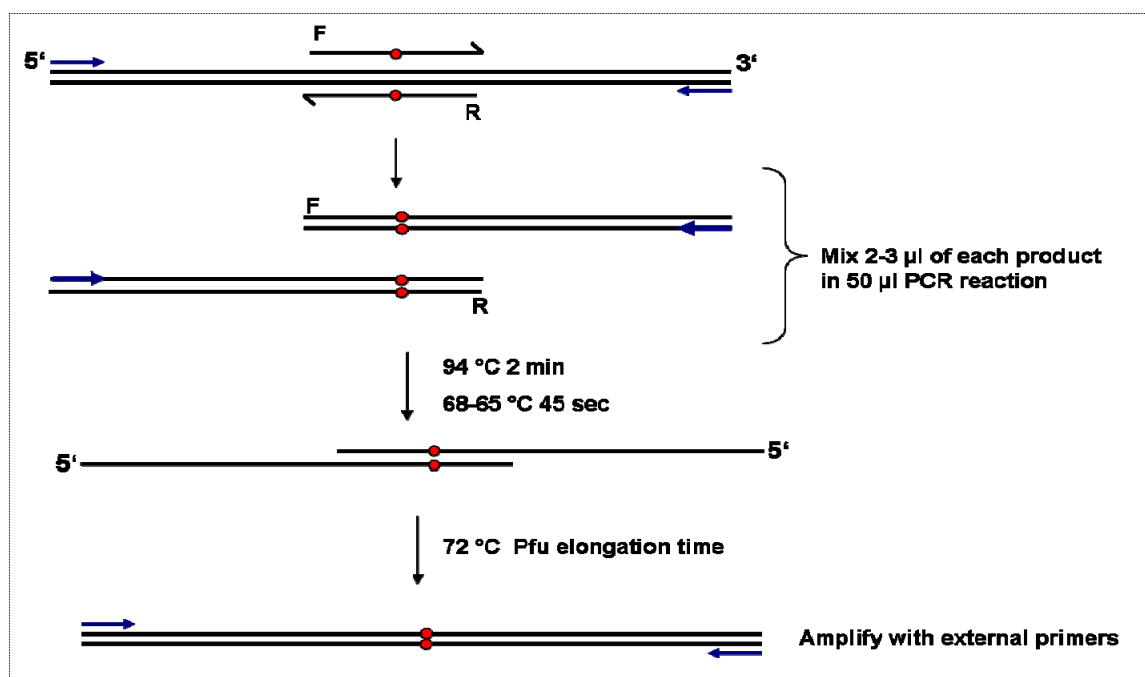
Fig. 5_3. Overlapping PCR site-directed point mutagenesis procedure.

5.1.5 Long Range PCR amplification

Amplification of large genomic DNA sequences (8-13 kb) from *Drosophila melanogaster* and *Tribolium castaneum* (on isolated genomic DNA template) was performed using the High Fidelity PCR Enzyme mix of Fermentas which is a blend of Taq DNA polymerase and a thermostable DNA polymerase with proofreading activity. Parameters considered for Long Range PCR reactions were the following;

Primer Design: Primers were 30-35 nucleotides long with Tm 68-72 °C for maximum specificity, GC% content 50-60%, maximum Tm difference between Forward and Reverse primers 2-4 °C. G-C rich stretches in the 3' primer were avoided as this may increase non-specific priming. Amplicon sequence was checked for internal priming sites. Primers were used in 0.2 µM concentration in the PCR reaction.

Template DNA: ~80-120 ng of *Drosophila* or *Tribolium* genomic DNA was used successfully as a template in long-range PCR 50 µl reactions. Template was first checked for integrity on agarose gel. During the DNA extraction procedure, sequential phenol – phenol-Sevag and CHCl3 extraction steps, as well as several 70% EtOH washing steps (with good hand-shaking – no vortex) after EtOH precipitation were performed to minimize contaminants and reduce salt presence. After preparation, DNA was completely air-dried and dissolved in TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA). For Long-Range PCRs repeated thawing-freezing was avoided. Instead, aliquots of genomic DNA template were kept. 0.5-1% DMSO was included in the PCR reaction to help amplification of GC-rich regions (DMSO functions as a 'template melting' reagent).

Cycling Conditions: A maximum of 2 minutes initial denaturation step and 20-25 seconds at 94 °C within the cycles were performed to minimize gradual loss of enzyme activity. Annealing temperatures were 2-5 °C lower than the lowest primer Tm. A few first cycles (5-7) were at an annealing temperature near the primers Tm and then the rest at [Tm-4°C] for increased specificity. Elongation was at 68°C, 1 min per kb (final elongation step 5-7 min). Number of cycles was 35-37.

Typical Long Range 50 µl PCR Reaction (For amplification 8-12 kb)

1x Buffer

2-2.5 mM MgCl2

0.2 mM dNTP

0.2 µM primers

100 ng genomic DNA

2.5 units enzyme blend (0.5 µl Fermentas)

1 % DMSO

Reactions were set on ice, enzyme was added last, gently mixed, spinned-down. Reactions were placed directly from ice to denaturation temperature.

5.1.6 Determination of transcription start sites

Determination of transcription start sites of segment polarity genes in *Drosophila* and *Tribolium*, for subsequent cloning of promoter regions in the reporter constructs, was performed by 5' RACE PCR. mRNA from 0-12 h embryos was isolated using the MicroPoly(A)Purist Kit (Ambion) or the protocol and the oligo(dT)-coated magnetic beads from Roche. cDNA RACE pools (5') were synthesized using the SMART PCR RACE cDNA Synthesis Kit (ClonTech).

5.2 Cloning of cDNA sequences

Cloning of cDNA sequences encoding open reading frames of genes of interest was performed for RNA probe generation or for subsequent protein (transcription factor) expression. mRNA isolation was as described in §5.1.6 and cDNA pool was synthesized using the SMART PCR cDNA Synthesis Kit (ClonTech). In the case of protein expression, sequences encoding the complete open reading frame of factors of interest were isolated by proofreading PCR and subcloned in an orientation-directed way in the *pTNT* vector (Promega) (§5.7).

Table 5–1. Oligos (*Operon*) used as primers in cloning procedures

5-1.1 *Primers used to clone open reading frames of transcription factors (cDNA pool template) that were expressed in vitro (TNT) for in vitro DNA-protein interaction assays (EMSA, in vitro transcription assay)*

Name	Sequence	Len	Tm	Purpose
Col_RA_F	GCTTCAAGCTGCGTTCCGAAGAG	23	66	Cloning of Collier A isoform 557 aa
Col_RA_R	GGGCCTGCCAATAGCCTCATTTAGTAG	27	67.6	
Col_RB_F	ATGGAGTGGGGCCGGAAGCTGTA	23	68	Cloning of Collier B isoform 575 aa
Col_RB_R	GGGAGTCCGGGAAATGCTTAAACG	24	66	
ColBspecF	TGTCCGCCGTGTCGTCGACGTG	22	70	Cloning of 465 bp fragment spliced out from ColA ORF, used as ColB specific probe
ColBspecR	CGTTAATCGTGTGTTGTCTGTGGTGTTTTG	30	66	
CncBF_Xh	CTCGAGACGAGATGTTGCAATCAGCGGTTC	30	70	CncB ORF 805aa
CncBR_RI	GAATTCATTCCTTGCGGTGCTGCTGTTGC	29	69	
CncAF	CTCGAGCTTCTGTCGCATGGTTGACAACAG	30	70	CncA ORF 533 aa
CncAR	GAATTCTCATTCCTTGCGGTGCTGCTGTTG	30	68.7	
Xh_LabF	ctcgagTTCCGGCTGGACAATATGATGGAC	30	70	Labial ORF 629aa
RI_LabR	gaattcATCGAAGAAAGCCCTTCAACTTTGC	31	66	
Zeste-F	CTCGAGAAACAAACTCAAATGTCGGCGCAG	30	68.7	Zeste ORF 575 aa
Zeste_R	GAATTCGTCGCTTATGTCCTCCAGACCTCATG	32	70	
TCF_F	CTCGAGTCAAAATCATGCCTCATACACACAG	31	67	TCF ORF (PanA) 751 aa
TCF_R	GGTACCATGGCGATCAGTTATGAAACG	27	66	
Slp1_F	CTCGAGATATCCGAATCAGAAAAATGGTGA	30	64.6	Slp1 ORF 322 aa
Slp1_R	GGTACCTGCCAAGATTTAGCTGGATGGAAAC	31	68.6	
DmC/EBP_F	CTCGAGCGCAACTGAAAATGCTTAACATG	29	66	
DmC/EBP_R	GAATTCTGGCTTTAGCTGGAAGCGGCTAC	29	69	

5-1.2	*Primers used during cloning of segment polarity genes cis-regulatory regions*			
Dm_hhprom_F	CAACGCGGAATGAACTCGAGGCGATAG	27	69	*hh* promoter [-120_+99] cloned XhoI_SpeI in the pslaftgfp.2
Dm_hhprom_R	AACTAGTTAGCTCTCGGTTCGGACAACCGTTG	32	67	
Dm_wgpromF	CTCGAGCAGGAGTCAGGGTATAGCTCCAC	29	65	*wg* promoter [-159_+121]
Dm_wgprom_R	ACTAGTTTCGATAGAATACACTCGGCTCGCTCTAG	35	65	
Dm_enprom_F	CCTCGAGCTCGCCCTCTCGCTCCCGCAC	28	73	*en* promoter [-104_+55]
Dm_enprom_R	TTACTAGTCTGTTGTTCCGACACGCACGTCTG	32	68	
Dm_hhupF1	CGCTGAACTAGATTCGAACAAATTCATCAGCTCTG	35	68	Long Range PCR Dm_*hh* full-length upstream [-8842_+265]
Dm_hhupR1	GCACTTCACTTTTGGCACACAGACACGCT	29	69	
Dm_wgupF1	CTCGACGGCAAACAGAGAAGGCGAGGAGTGACT	33	73	Long Range PCR Dm_*wg* full-length upstream [-8094_+193 (tsB)/_+2122 (tsA)] 10.216 kb RI excised from PCRII vector, subcloned in the reporter cassette
Dm_wgupR1	AGTCACTCCTCGCCTTCTCTGTTTGCCGTCGAG	33	73	
Dm_wgupF2	CGCTGCTCCAGATCATCAGCGTTGTACCAG	30	70	Cloning of [-16,212_-7,813 kb] upstream of *wg*
Dm_wgupR2	GAATCGGAATCGGGTTGGCTCGACCTCAC	29	71.4	
Dm_enupF	CAGTTTCATAGGAACTGCTGCGCTCAGGTGTTCA	34	71.4	Dm_*en* upstream [-8811_+165]
Dm_enupR	TGGGTCACTTGACACTGAGCCACTGATTCTTCTGA	35	71.4	
Tc_wgupF	CATTCGGCATTCATAGAGGGCATTGTTTTGACGCT	35	69.4	Cloning 10,188 kb of Tc_wg region (-8.088 kb upstream o tsA)
Tc_wgupR	ACCCAAGGCCGCTACGGTGCGGTAGCA	27	73.7	
Tc_enintronF	TTGGGTCTACTGCACGCGGTATTCGGA	27	69	
Tc_enintronR	CTTGTCCTCACCGTTGGGTTTGGACGGT	28	70.5	
Tc_enpromF	ATGCATGCTCAGAGAGAGACAGGCAAATGC	30	69	
Tc_enpromF	GCGGCCGCACCACACTCAGCCA	22	72	
Tc_hhpromF	AATCGGGCCGGCGTCAAGTAATCGTCCTG	29	71.6	

Tc_hhpromF	ACTAGTCTCCGAACGCACGCACCATCAGCACTC	33	70	
5-1.3	*Dissection of the 1 kb ic-CRE and construction of deletion mutations*			
ahhF	TCGCGAGCTGATAGCACAATGGACCCAC	28	68	a fragment 1009 bp [-4085_-3077] cloned NruI_XhoI in the pslaf_hhprom_tgfp.2_af
ahhR	CTCGAGTATCTAAAAGCCAATTTCGATTGTGAC	33	64	
hhR4	CTCGAGATCGAAACAGCGACGGTTCTCATATC	32	68.4	
F5	AGATCTCCATTTGCCTAATTTCTATTTCGGCAG	33	65.5	F5_R4 335 bp
γ1BglII_F	AGATCTAGCTGATAGCACAATGGACCCACTGCA	33	69.6	γ1_hhR4 fragment 620 bp
NruIicmutF1	TCGCGAGCAATTCCCCAATGGCATTTC	28	68	
NruIicmutF2	TCGCGAGATTTATGTGAATTAAACAGCCTGAAC	33	66	
NruImutR1	TTCGCGATCTAAGTGAAATGCCATTG	26	63	
NruImutF3	TCGCGAGCCTGAACACAGCCACTTC	25	67	
NruImutR2	TCGCGAGGATTGGGATTGGGTCG	24	67.3	
NruImutF4	TCGCGATCAAACCGTTAGTGGCATTTGTC	29	68	
NruImutR3	TCGCGAGGTCTCTAATGGATGTTGACATTTAC	32	66	
NruImutF5	TCGCGAGGCAGGCCGAAAGCAAC	23	69	
NruImutR4	TCGCGATCAGGCTGTTTAATTCACATAAATC	31	65	
NruImutF6	TCGCGAATCCCTGGTAGCCGTAAATGTCAAC	31	69	
XhoImutR5	TCGCGACTGCCTCATAGCGTGCGAC	65	69	
5-1.4	*Overlapping primers used in point mutagenesis of the hh promoter (external primers: DmhhpromF_DmhhpromR)*			
zesteprom_mutF	CAGGGAAATGTCATTTGTCAATCACAGTGCCTG	33	68.4	zeste mutation in the promoter
zesteprom_mutR	CAGGCACTGTGATTGACAAATGACATTTCCCTG	33	68.4	
DPEmutRev	ACTAGTTAGCTCTCGGTTCGGACCCTCGTTG	31	70	DPE mutation in the promoter – in combination with DmhhpromF

7bp_prom_mut F	CTCGGAAATGTCATTTGTCAACTGCAGTGCGAG	33	69.5	'7bp element' mutation in the promoter
7bp_prom_mut R	CTCGCACTGCAGTTGACAAATGACATTTCCGAG	33	69.5	

5-1.5 *Overlapping primers used in point mutagenesis screen (external: F5_hhR4 or γ1mF5_hhR4)*

homeomutF	CTTAGATCCATGTGAACCGAACAGCCTGAACAC	33	69.6	point mutation in the third block
homeomuR	GTGTTCAGGCTGTTCGGTTCACATGGATCTAAG	33	69.6	
7bpmutF	CACAACTCGCAGCACAGTTGAGTGCCTGATTC	32	71	to be injected
7bpmutR	GAATCAGGCACTCAACTGTGCTGCGAGTTG	30	70	
zestemut**new**F	CACAACTCGCAGCATGATTGACATCCTGATTC	32	68	to be injected
zestemut**new**R	GAATCAGGATGTCAATCATGCTGCGAGTTGTG	32	68	
zestemutF	CACAACTCGCAGCATGATTCATTGCCTGATTC	32	68.5	expression abolished in the 335 bp, derepressed in the 450 bp
zestemutR	GAATCAGGCAATGAATCATGCTGCGAGTTGTG	32	68.5	
slpsitemutF	GTAGCCGTAAATGTCTTCATCCATTAGAGACCTC	34	68	4th conservation block expression abolished
slpsitemutR	GAGGTCTCTAATGGATGAAGACATTTACGGCTAC	34	68	
labsitemutF	GTAAATGTCAACATCCAGGAGAGACCTCAAACC	33	68.4	5th conservation block expression unaffected
labsitemutR	GGTTTGAGGTCTCTCCTGGATGTTGACATTTAC	33	68.4	
colsitemutF	CTATTTCGGCAGCAATGAACCAATGGCATTTC	32	67	2nd block, expression abolished
colsitemutR	GAAATGCCATTGGTTCATTGCTGCCGAAATAG	32	67	
c/ebp_colsite_mut2F	CTATTTCGGCAAGAATTCCCCAATGGCATTTC	32	67	2nd block, expression abolished
c/ebp_colsite_mut2R	GAAATGCCATTGGGGAATTCTTGCCGAAATAG	32	67	

5-1.6 *Forward primers used in combination with hhR4 for 5' dissection of the γ1 620 bp construct (-4085_-3077 bp)*

γ1mutF1	AGATCTGCACAATGATAACTATGCCGCTTACGAC	34	68	-4063

γ1mutF2	CTTTTTGACTCTCGTGCCGGTCCG	28	69	-4034
γ1mutF3	AGATCTCCGGATCAAAAGGAGCCCGGTTG	29	70	-4014
γ1mutF4	AGATCTGTTTGTATGCACGCACACATCGAGAC	32	68.5	-3985
γ1mutF5	AGATCTCTGTGAATGTTTATGTTACAAGCCCCCTG	35	68	-3914
γ1mutF6	AGATCTGCCCATTTTGCTCCTGAACTGAAACC	32	68.5	-3888

5-1.7	*Primers used in 5' RACE PCRs*			
Tc_wg5RC	CCGACTGACCGTGAATCCGAACCCA	25	69.5	
Tc_hh5RC	CGGAATCCCACAACTTTGCCACCCATC	27	69	
Tc_hh5RCnest	CGAGCGATCTCGATCGCTGGTGGTGA	26	70	
Dm_hh5Race	TTGGAGCTGGAACTGGAACTGGAACTG	27	64	

5-1.8 *Primers used for constructs a-h (Fig. 2_38) spanning upstream enhancer sequence of hh. (NruI)Forward _(XhoI) or (SalI)Reverse. Fragments ~600 bp cloned in NruI_XhoI of psltgfp2_hhprom (XhoI_SpeI).*

ocF1	TTCGCGATTTTTCAGCTCTTTTGCATTGC	29	65	a [-6430_-5822 bp]
ocR2	CTCGAGCGATACTCTATATAAGTGCCATGCAC	32	64	
ocF2	TTCGCGACTGACTGACTGACTGACTGCCTG	31	70	b [-6140-5546 bp]
anR1	CTCGAGCTTAAGGGGTCTTAGAAAATTGCAC	31	64	
anF1	TCGCGAGCAACCCACTAAAGTGTGCATG	28	68	c [-5863_-5264 bp]
anR2	CTCGAGAGCCACATGCGTCCACTTGGAG	28	70	
anF2	TCGCGAGTGCAATTTTCTAAGACCCCTTAAGC	32	68.5	d [-5571_-4972 bp]
anR3	GTCGACTTAAGCTCTGCGGCTGCTGTTG	28	67.5	
anF3	TCGCGACACTCCAAGTGGACGCATG	26	69	e [-5287_-4719 bp]
anR4	TGTCGACGAACATCGAGCATCGAGCAC	27	68.6	
anF4	TTCGCGAGGATGCCGATGCAGCAAC	25	69.5	f [-5013_-4425 bp]

anR5	CTCGAGCATCATATATTGCAGCGATCAAAC	30	64.5	
anF5	TTCGCGATGCTCGATGTTCGTCGAG	25	68	g [-4738_-4141 bp]
anR6	TCGCGAGGTGAGTAAGCAGTGCCCGTTG	28	70	
anF6	TTCGCGAGTTTGATCGCTGCAATATATGATG	31	67.5	h [-4449_-3989 bp]
hhRevSal	GTCGACGTCAACCGGGCTCCTTTTGATC	28	70	

5.3 Transgenic RNAi and *Drosophila* strains

Coding sequence of Collier A was cloned in antisense orientation into the EcoRI site of the polylinker of the pSLafHSaf (Schmid and Wimmer (2000), Göttingen). Cassette consisting of {*heat-shock promoter_ColA_antisense_heat-shock 3' UTR*} was excised with AscI and subcloned in the pBac_attB vector (map 2; Fig. 5_1). Construct was injected in the attP-96E combined line (§5.1.2) to generate transgenic line HS.ColA_antisense #M3. For heat-shock inducible embryonic RNAi, transgenic embryos were collected for 1,5 h @ rt, let to develop for 2,5 h, heatshocked for 30-45 min @ 37 °C, and then let to develop at rt (22 °C). For *in situ* hybridizations embryo collections 6-8 hours after heatshock were dechorionated for 5-6 minutes instead of the three minutes of the standard protocol (§5.10)

UAS-hairpins for *cnc* (Transformant ID 37674), *slp1* (Transformant ID 15749) and *pan* (Transformant ID 3014) were ordered from VDRC stock centre (Vienna) and crossed to virgins homozygous on the second chromosome for the maternal driver {*pnos*-GAL4/GCN4-*bcd* 3' UTR} (Janody et al., 2000) (insertion 'P21B') and for the ic-CRE (attB-96E insertion) homozygous on the third chromosome.

Drosophila mutant alleles that were used as genetic background to examine expression outcome of the ic-CRE were the following; *Col[l]* (Crozatier et al., 1999); *cnc*[K22] (Veraksa et al., 2000); *VL110* (Mohler et al., 1995); *lab[14]* (Diederich et al., 1989); *z[a]* (Laney and Biggin, 1996). Ectopic expression of *collier* and *cncB* was achieved using the lines *UAS-kn* (Mohler et al., 2000); *UAS-CncB* (Veraksa et al., 2000) driven under the control of the maternal driver {*pnos*-GAL4/GCN4-*bcd* 3' UTR} (Janody et al., 2000) (insertion 'P21B').
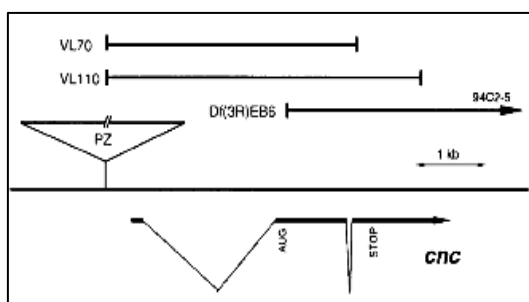

Fig. 5_4. Scheme is from Mohler et al., 1995. The mutation *VL110* lacks the entire *cnc* open reading frame and was generated following imprecise P element excision (Mohler et al., 1995).

5.4 Chromatin immunoprecipitation

Chromatin immunoprecipitation procedure was after a protocol published by Sandmann et al., 2006 [Nature Protocols 1(6): 2839].

Repetitive embryo collections (~2-9,5 hours ael) from large scale (w-) fly-populations (12-15 cages) were dechorionated for 2.5 min in 50% chlorix, washed, dried and resuspended in 10 ml PBT/ 1.5 g embryos. Agar plates were changed over a period of two days before the first used collection. Embryos in PBT solution were passed through Nitex membrane and dried by blotting on paper towels. Then the embryos were transferred with a spatula into a 50 ml Falcon tube containing 30 ml heptane, and 10 ml crosslinking solution containing 1.8% Formaldehyde v/v (MB Grade) was added. During the mild 15 min (shaking) fixation step protein-chromatin and protein-protein cross-linking takes place. Fixation time was essentially kept constant for all independent collections. Embryos were pelleted at 500 g for 1 min, supernananant was removed and 30 ml stop solution (125 mM glycine) was added (shaking). After repelleting (500 g, 1min) they were washed twice with 50 ml PBT. Then they were dried (solution was passed through Nitex membrane and the remaining on the membrane embryos were blotted on paper towels), the batch was weighed and quick-frozen in liquid nitrogen. ChIP was started with ~ 150-200 mg of embryos per single IP. Batches were combined to make the required total weight, and resuspended in 1:10 volumes of cold PBT supplemented with protease inhibitors and 1 mM PMSF. Embryos were transferred in PBT (10 ml/ g embryos) in a pre-cooled Dounce homogenizer kept on ice and 20 strokes with the Loose pestle were applied. The homogenate was centrifuged (400 g, 4 °C, 1 min). The supernatant was re-centrifuged (1100 g, 4 °C, 10 min). The cell pellet was resuspended in 1:10 volumes of cold cell lysis buffer (supplemented with protease inhibitors and 1 mM PMSF). Then it was transferred in a Dounce homogenizer, 20 strokes with the Tight pestle were applied. Nuclei were pelleted at 2000 g 4 °C 4 min and quick-frozen in liquid nitrogen. For every single IP nuclei aliquots corresponding to ~200 mg of initial embryo weight were combined on ice by resuspending in a maximum volume of 1.8 ml cold nuclear lysis buffer. After splitting in 6x300 µl (maximum volume for efficient sonication) samples were sonicated using a pre-cooled Bioruptor sonicator (Diagenode) (samples are immersed in ice-water-bath). A range of different parameters were tested.

Note: In Experiment A sonication using parameters [20 min; high settings] resulted in chromatin fragments 200-500 bp size, while [10 min; high settings] did not cause sufficient fragmentation of chromatin. The latter batch of chromatin was not used in the immunoprecipitation. In Experiment B sonication for [20 min; medium settings] resulted in fragments around 500 bp while [17 min; high settings] resulted in fragmentation efficiency 1- 2 kb (Fig 5_5).
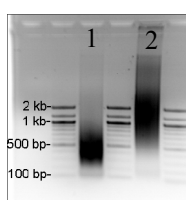

Fig. 5_5. In Exp. B sonication for [20 min; medium settings] resulted in fragments around 500 bp (lane 1) while [17 min; high settings] resulted in fragmentation efficiency 1- 2 kb (lane 2). The two samples were combined and used in the IP. All sonication steps were performed with *15 sec on/15 sec off*. To assess shearing efficiency of sonication 2.5-3.5 µg of each sample were loaded on a 1% agarose gel.

After chromatin shearing using the Bioruptor, samples was centrifuged at 20000g, 4 °C, for 10 min. Supernatants were pooled, 50 µl were removed to assess quality, and the rest quick-frozen (in ~400 µl aliquots). To assess fragmentation size on the 50 µl removed aliquot, 50 µl TE and 1.5 µl RNase A (10 mg/ml) were added and incubated for 45 min @37 °C. Then 2.5 µl SDS 20% and 2.5 µl Proteinase K (20 mg/ml) were added. Incubation was at 65-68 °C shaking for 6 h (this step reverses cross-links and extracts DNA from proteins), followed by Phenol-Sevag extraction and EtOH precipitation. DNA was resuspended in 30-40 µl TE. 2.5-3.5 µg were loaded on 1% agarose gel (Fig 5_5). Before starting the ChIP procedure, the Protein G-sepharose beads were washed three times in RIPA buffer (140 mM NaCl, 1% Triton X-100, 0.1 %SDS, 0.1% sodium deoxycholate, 10 mM Tris-HCL) (rotating in 1 ml 10 min 4 °C, 2000g 2 min 4 °C) and 50 % slurry in RIPA was prepared. An aliquot of the sonicated chromatin (~400 µl, 60-80 µg) was thawed and volume was adjusted to 500 µl with cold TE. The following solutions were sequentially added (with gentle mix after each step to gradually equilibrate chromatin in the RIPA buffer conditions): 100 µl 10% Triton-X, 100 µl 1% sodium deoxycholate, 100 µl 1% SDS, 100 µl 1.4 M NaCl, 10 µl 100 mM PMSF. 50 µl of 50% beads slurry was added to each sample, incubated for 1 h rotating at 4 °C (preclearing step). Beads were pelleted at 2000 g for 2 min @4°C, chromatin supernatant was transferred to siliconized eppendorf tube while avoiding any bead carryover. A volume of sample corresponding to 1% input was removed and retained at 4 °C. Antibody (1-10 µg) was added (in the presence of 0.1% BSA) and incubated rotating o/n at 4 °C. The beads were also overnight blocked in 0.1% BSA, 0.05% sheared salmon sperm DNA in 1 ml RIPA. Next day, beads were pelleted and the 50% slurry in RIPA was restored. 100 µl of slurry was added to each immunoprecipitation sample; incubation was for 3 h rotating at 4 °C. The antigen-antibody complexes were pelletted (1000 g, 2 min), washed twice with 1 ml RIPA, four times with RIPA500 (RIPA containing 500 mM NaCl), once with LiCl, twice with TE (all solutions cold). After last washing step bead complexes were resuspended in 100 µl TE. RNase treatment, reversing of the crosslinks and SDS-Proteinase K treatment were performed as described above, followed by Phenol, Phenol-Sevag and ChCl3 extractions. EtOH precipitation was with 1/10V NaAc pH 5.2 3M, 2.5-3V EtOH ABS, 250 mg/ml glycogen (carrier). Precipitation was at -80 °C 30 min, -20 °C 2 h or o/n. DNA was precipitated by centrifugation at full rpm 45 min 4 °C, washed 3 times (with good shaking to resuspend DNA pellet) in 70% EtOH, air-dried (or under vacuum), and resuspended in 35 µl TE. DNA prepared the above way is of appropriate quality to be used as template in quantitative PCR reactions (§5.5). All samples were processed the same way, in the same volumes. For the mock IP, the mouse_anti-BP102 was used in the same amount and same dilution volume as for the mouse _anti-Col (specific IP sample). The anti-BP102 recognizes *Drosophila* CNS axons (reveals connectives and commisures in immunostainings; Fujita et al., 1982).

5.5 Quantitative Real-Time PCR

Enrichment of chromatin immunoprecipitated samples in specific DNA target sequences was assessed by quantitative real-time PCR (qPCR). QPCR is the most robust method to analyze chromatin IP samples; it does not quantify the amount of PCR product at the end of the PCR reaction as the conventional PCR densitometry, but instead the initial amount of template DNA is calculated from the kinetics of the PCR reaction (Ct extraction and a calibration line are used; Fig. 5_6). SYBGreen fluorescent DNA-binding dye included in the reaction binds all newly synthesized double-stranded DNA complexes and fluoresces. Fluorescence accumulates as cycling of PCR continues and is measured at the end of each PCR cycle. Cycler and Software used were from BIORAD (Opticon Monitor Software – Chromo4 system). The fluorescence threshold was first defined by plotting fluorescence signal vs. cycle number and then the threshold line on the graph was set at a point were signal exceeds background noise and begins to increase exponentially. Because the detected and measured fluorescence signal reflects the amount of product, samples that initially have more template molecule copies (i.e units) will need less number of cycles (Ct) to reach the defined threshold. Thus the Ct value is inversely proportional to the log of the initial template units. This is due to the exponential nature of PCR amplification; doubling of fluorescence from one cycle to the next is directly proportional to the doubling of amplicons. Therefore samples generated by sequential 1:4 dilutions should have a difference of 2 in their Ct values. (For samples A, B if Concentration_UnitsB=Concentration_UnitsA/4 then CtB-CtA=2). This allows for the generation of a calibration line or standard curve from known initial sample dilutions against which Ct values of unknown samples are plotted. A very good reference describing analytically parameters to consider for ChIP followed by qPCR is from Haring et al., 2007.
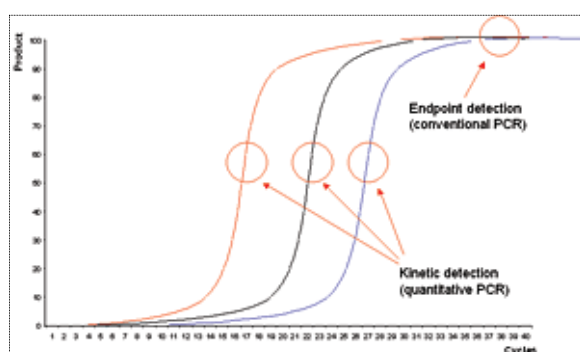

Fig. 5_6. Difference in DNA amount quantification between conventional and qPCR. Picture is from http://www.eppendorfna.com/

Design of primers for qPCR was done with Program Primer3 (http://frodo.wi.mit.edu/primer3/), setting the following parameters: Amplicon size 75-150 bp, Minimum primer Tm 59 °C, Optimal 60 °C, Maximum 62 °C and maximum Tm difference within the primer set 0.5 °C. For qPCR two parameters are of extreme importance: i) amplicon sequence to be amplified must have a maximum size of 150 bp to ensure that each amplification round in the quantitative PCR reaction reaches 100% amplification efficiency b) different positive (for target sequence amplification) and negative-control primer sets should be tested for a series of input dilutions and appropriately designed to produce the

same Ct value for a wide range of input sample template concentrations. If positive and negative primers should differ in their Ct values then it is crucial, for overcoming artifacts, that ChIP enrichment ratios (units_positive/units_negative) are normalized over the factor [Ct(pos)/Ct(neg)].The best two forward and two reverse primers (which were the output of the Primer3 Program on a given DNA sequence) were ordered and all 4 combinations were tested (on genomic DNA sample template or sonicated purified chromatin sample) for a broad series of dilutions starting from ~10 ng (1:4, 1:16, 1:64, 1:256, 1:1024). For assessing the ChIP samples, the best combination(s) of primers to be used should yield a single pick in the melting curve analysis (fluorescence plotted against temperature). This corresponds to a single denaturation event meaning that the primer set specifically amplifies a single target amplicon. In addition, for comparison purposes and extraction of the relative ratios, it was confirmed that different primer sets (for the target sequence and negative control sequence) displayed the same respective Ct values throughout the wide range of input dilutions. (If the Ct values of the different primer sets should differ and no alternative in the primer design exists then calculated sample units should be normalized against a factor regarding the difference between the Ct values. For example, if Ct values differ for positive and negative control sequences then Relative Ratio of Enrichment=Unitspos/Unitsneg x Ctpos/Ctneg.) Enrichment of the ChIP sample for target-sequence template molecules was estimated as a ratio over a negative control region. For this thesis, as a negative control, an amplicon within the second exon of *caudal* was picked since no Collier binding activity is expected in an extent of 2 kb around that region, which is approximately the maximum fragmentation size resulting from the sonication process (§5.4).

Reactions were performed in 15 µl final volume in qPCR – 64 well-plates suitable for the Biorad machine. Reactions consist of 7.5 µl 2xSYBR Green I Master Mix (BioRad), 0.9 µl primer mix (0.5 µmol) each, and 1 µl template. A master mix was first prepared, 14 µl were pipetted in each well and the 1 µl of template was added last. Reactions were pipetted in duplicates or triplets to extract mean values (and standard deviation). Program used was 10 min 95 °C (Hot Start) and 40 cycles of [15 sec 95 °C denaturation and 1 min 60 °C annealing/extension] (End hold @4 °C).

Table 5–2.

Primers used in quantitative real time PCRs after the ChIP experiments				
cad_2ndexF	GAGCTGGAGAAGGAGTACTGCAC	23	62	negative control primer set- amplicon 111 bp within second exon of *caudal*
cad_2ndexR	CGGTTCTGGAACCAGATCTTAAC	23	63	
col_ic_F	ATCCATCCATTTGCCTAATTTCT	23	60	Targeting the Collier site within the ic-CRE, product size 89 bp (-3804_-3716 bp)
col_ic_R	CTGTGTTCAGGCTGTTTAATTCAC	24	60	

5.6 Coimmunoprecipitation of proteins from crude embryonic extracts

Coimmunoprecipitation of protein factors from crude embryonic extracts was performed based on an immunoprecipitation protocol from Dr. Andreas Wodarz (Methods in Molecular Biology, *Drosophila*, Methods and Protocols, Edited by Dahmann C., 2008, Humana Press; Chapter 21, p. 335-345) and on a nuclear extracts preparation protocol (edited by Sullivan et al., *Drosophila* Protocols, Chapter 31, p. 553-557).

Embryos from large scale collections (~3-5 g) 1-10 h ael were dechorionated in 50% chlorix for 2 min, washed thoroughly, dried (blotted on paper towels), weighed and transferred in a Dounce homogeniser. 3 ml of Buffer I (15mM Hepes pH 7.9, 10 mM KCl, 350 mM sucrose, 1.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM EGTA, 1 mM DTT, 1 mM PMSF) was added per gram of embryos. All steps were carried on ice, in the presence of protein inhibitors, preferably in a cold room. After homogenization all steps were carried out as quickly as possible to minimize leakage of protein out of the nuclei. 10 strokes with the Loose pestle and 15 strokes with the Tight pestle were applied. Homogenate was filtered through a funnel lined with Miracloth (to remove debris and vitelline membranes) in 15 ml Falcon tubes. Nuclei were pelleted at 10000g @4 °C for 15 min (Sorvall SS-34). The supernatant was quick-frozen (corresponds to the cytoplasmic extract, but it usually also contains leaked-out nuclear proteins). The nuclei pellet was resuspended in Buffer I (1 ml/embryos), briefly shaked for washing nuclei and repelletted at 10000g for 10 min. Supernatant was decanted, the tube was inversed to quickly blot on a paper, nuclei pellet was weighed and resuspended in nuclear extraction buffer (15 mM Hepes pH 7.9, 100 mM 100 KCl, 50 mM NaCl, 0.1 mM EDTA, 0.1 mM EGTA, 1.5 mM $MgCl_2$, 1 mM DTT, 1 mM PMSF) 1 ml /g nuclei. Proteins were extracted while gently shaking on ice for 45 min, followed by ultracentrifugation for 1h 100,000 g (30,000 rpm AH 650 rotor). The thin lipid layer was aspirated, and the yellow-opaque liquid layer was transferred in a siliconized eppendorf. Protein concentration was assessed using a Nanodrop ND-1000 Spectrophotometer with protein settings '1 Abs/mg'. 0.8–1,5 mg of crude nuclear extract was used for each immunoprecipitation using 25-35 μg antibody anti-Col in a volume dilution 1:50 to 1:100. Same amount and same dilution of anti-BP102 was used in the mock immunoprecipitation. The final volume of the immunoprecipitation reaction was adjusted with IP buffer (25 mM Hepes pH 7.9, 10 mM Tris pH 8.0, 1 mM DTT, 1 mM PMSF, 100-150 mM NaCl) (maximum volume 1.2 ml). Immunoprecipitation was overnight, while gently rotating at 4 °C in the presence of protein inhibitors. Next day pre-equilibrated and pre-blocked (0.1% BSA in IP buffer) Protein G-sepharose or agarose beads were added to the immunoprecipitation reactions; 30-50 μl of 50% slurry were added per ml of IP sample and incubation was at 4°C rotating for 3 h. Then the beads carrying the antibody-antigen complexes were spinned-down, supernatant was aspirated and beads were washed five times (10 min each) with ice-cold IP buffer. After last washing step, 20 μl of 2x SDS sample buffer (100 mM Tris, pH 6.8, 4% SDS, 200 mM β-mercaptoethanol, 20% glycerol,

BB trace) was added to the beads and samples were heated to 95 °C for 5 min, briefly cooled on ice and spinned at maximum speed for 15 seconds. Supernatants, avoiding beads-carrying over, were subjected to SDS (8%) gel electrophoresis followed by semi-dry electroblotting (100 V, maximum mA, maximum Watts) for 1,5 h. Membrane used was from GE Healthcare (nitrocellulose Hybond ECL). Blot was briefly washed in TBS-Tw (TBS, 0.2% Tween), blocked in 3% nonfat dry milk plus 2% BSA in TBS-Tw, and primary antibody was applied o/n @4°C (a-Cnc 1:5000, a-Col 1:3000), followed by TBS-T washings (4x), 2nd antibody application (a-mouse- or a-rb-HRP at 1:10000, 1 h rt in TBS-T, 3% BSA) and HRP detection using a Lumiglo substrate (Cell Signaling Technology, Inc).

5.7 *In vitro* protein expression

After cloning of the open reading frames – encoding cDNA sequences of transcription factors of interest in the pTNT vector (§5.2), recombinant plasmid DNA was EtOH-precipitated (1/10 V NaAc 3M pH 5.3) followed by three 70% EtOH washing steps. 8.5 µg of plasmid DNA was used as a template for *in vitro* transcription and translation using the TNT wheat germ cell SP6 system from Promega. Reaction was at 25 °C for 2 h, then put on ice and subsequently stored at -20 °C. Mock reactions were performed by using as template the same amount of pTNT plasmid DNA. To check efficiency of the *in vitro* protein expression 1 µl of the reaction was loaded on 8% SDS gel followed by immunoblotting using the appropriate antibody. 0.5-4.5 µl of the TNT expressed protein factor was used in DNA-protein interaction assays (§5.8, §5.9).

5.8 Electrophoretic Mobility Shift Assay

5.8.1 Preparation of labeled DNA fragment

DNA oligos for upper and lower strand spanning the recognition site to be tested were ordered from Operon as HPLC purified. Sequences of the complementary oligos partially overlap leaving 3' overhangs of 3-8 nucleotides in the double stranded form to be filled in by the 5'-3' Klenow enzyme activity.

Oligo stocks were dissolved to 100 pmoles/ µl (100 µM stock concentration) in ddsH2O (mQ). Then 5 µg of each oligo strand were mixed in 100 µl TE; 150 mM NaCl. Solution was mixed by vortexing, spinned-down and placed at 94 °C for one minute. Heat-block was turned off and left to reach room temperature (~8 hours) to allow for perfect annealing of the complementary strands. Concentration of the double-stranded oligo form is 100 ng/µl.

200-300 ng of the ds-oligo form were used in the Klenow labeling reaction. Klenow enzyme used was from Biolabs (3'-5' exo⁻) (New England BioLabs, NEB). Isotope (α-^{32}P dCTP) was ordered from Hartman Analytic, Braunschweig. For Klenow labeling reaction of ds-oligos to be used in EMSAs the isotope used was of specific activity 800Ci/mmol.

Vfinal 20 µl Klenow labeling reaction

2µl NEB2 buffer (NEB)

2 µl d[AGT]P 2.5 mM each => final c= 0.25 mM

200-300 ng of ds-oligo with 3' overhangs

1 µl Klenow (3'-5' exo$^-$) 5 units (NEB)

1 µl α-^{32}P dCTP 10 µCi

Reaction was set on ice, enzyme was added last and incubation was at room temperature (22 °C) for 45 min. At the end of the incubation 30 µl of TE was added and probe was passed through G25 column (GE Healthcare) to eliminate presence of unincorporated nucleotides. 1 µl of the probe was brought juxtaposed to a Geiger-Müller detector (Type EP15, Mini 900 Series, Thermo Scientific) (Perspective Instruments Ltd, UK) to assess counts per second (cps)/µl. By using fresh isotope (less than one half-time) and 200-300 ng of template the probe generated was 500-1000 cps/µl. A dilution in TE was prepared to adjust counts to 150 cps/µl and from this 1 µl was used in the EMSA reactions. In the final dilution the molecular concentration of the probe was also calculated (that was used to assess the amount of cold oligo competitors at the designated molar excess). Probes generated the above way were used at a concentration of ~0.05-0.2 pmoles in the EMSA reaction. 150-200 cps enables signal detection using a Typhoon 9400 scanner (Amersham Biosciences; Image Quant 5.0 Software) after three hours of exposure at room temperature on a Phosphoimager Screen (Amersham Biosciences) or for autoradiography on X-ray film after over-night exposure at -80 °C.

5.8.2 EMSA – Binding reaction

EMSA reactions were set on ice in a final volume of 20µl including 1x Binding Buffer (10mM Hepes pH 7.9, 60mM KCl, 8.4 % glycerol, 1mM EDTA, 2.5mM MgCl2, 1mM DTT, 0.2µM ZnAc). polydIdC was added to a concentration 50-100 ng/µl and BSA (optional, MB grade, helps stabilize complex formation, helpful in super-shift reactions) to a final concentration 50-200 ng/µl. Protein factor was added (0.5-5 µl of TNT expressed protein or crude nuclear extract ~20 µg) and equilibrated in the binding buffer in the presence of dIdC for 10-15 min @ rt. Then the probe was added (150-200 cps) and binding reactions were incubated at room temperature for 30 min. For super-shift reactions antibody (0.5-1 µl anti-Col) and protein factor were preincubated for 20 min at rt and then probe was added. Binding reactions with crude nuclear extract were performed on ice for 45 minutes to minimize nuclease activity that degrades probe, in the presence of 100-200 ng polydIdC. MgCl2 was also omitted (included in nuclear extraction buffer). [Embryonic crude nuclear extracts were prepared essentially as described in Sullivan et al, 2000 *Drosophila* Protocols, CSHL Press, Chapter 31, p. 553-557.]

In the meanwhile the native gel was prerun at 4 °C at 120 V-10 mA to equilibrate in buffer conditions and minimize presence of Aps and Temed. Pre-ran buffer was exchanged with cooled-

fresh buffer (0.25xTBE pH 8.3, 0.1% glycerol) before loading the samples. Native gel composition was 6% acrylamide 59:1, 0.25xTBE pH 8.3, 2.5% glycerol (7.5% APS, 0.04% Temed). Reactions were run for 1,5-2 h at 160V, 12 mA. (Bromophenol blue buffer consisting of 10mM Tris pH 8.0, 1 mM EDTA ,10% glycerol, loaded at a separated lane or added in the free-probe reaction reaches at 2/3 gel distance). Then the gel was dried on 3MM Whatman paper under vacuum (80°C) and exposed.

Oligos used for mobility shift assays; after annealing either the 3' overhangs (underlined) were filled in by Klenow emzymic activity to generate labeled probe or used as competitors in 50-150x molar excess.			
Name	Sequence	Len	Purpose
TCFsite_HMG	GGTCCGGATCAAAAGGAGCCCGGTTGA	27	§2.5.2
TCFsite_HMG_Rev	AACATTTGTCAACCGGGCTCCTTTTGAT	28	
TCFsite_optimal	GGTCCGGATCAAAGGAGCCCGGTTGA	26	
TCFsite_optimal_Rev	AACATTTGTCAACCGGGCTCCTTTGAT	27	
TCFsite_compet	GGTCCGGATCAAAGGAGCCC		
TCFsite_competR	GGGCTCCTTTTGATCCGGACC		
HMG_compet	GCCCGGTTGACAAATGTT		
HMG_competR	AACATTTGTCAACCGGGC		
Colsite_F	TTTCGGCAGCAATTCCCCAATGGCATTT	28	§2.7.2
Colsite_R	TAAGTGAAATGCCATTGGGGAATTGCTG	28	
Colsite_optimal	TTTCGGCAGCAATTCCCCAAGGGCTTTT		
Colsite_optimalRev	TAAGTGAAAAGCCCTTGGGGAATTGCTG		
colsite_mut1	TTCGGCAAGGATTCCCCAATGGCATTTCAC		
colsite_mut1R	GTGAAATGCCATTGGGGAATCCTTGCCGAA		
colsite_mut2	TTCGGCAGCAATGAACCAATGGCATTTCAC		
colsite_mut2R	GTGAAATGCCATTGGTTCATTGCTGCCGAA		

oligo3compet	ATCCCAATCCCTGGTAGCCGTAAAT	25	-3674_-3650; Fig. 2_46
oligo3competR	ATTTACGGCTACCAGGGATTGGGAT	25	
oligo3mutcompet	ATCCCAACGGCTTTTAGCCGTAAAT		
oligo3mutcompetR	ATTTACGGCTAAAAGCCGTTGGGAT		
col_cnc_site	TCGGCAGCAATTCCCCAATGGCATTT	26	§2.8.1
col_cnc_siteRev	ATCTAAGTGAAATGCCATTGGGGAATTG	28	
bZIPsite_F	CCATTTGCCTAATTTCTATTTCGG		positive target site for C/EBP binding shift
bZIPsite_R	GGAATTGCTGCCGAAATAGAAATTAGG		
oligo4_fkhsite_F	GTAAATGTCAACATCCATTAGAGA		forkhead site in the 4th block matching Slp1 binding site
oligo4_fkhsiteR	GAGGTCTCTAATGGATGTT		

5.9 Run-off *in vitro* transcription assay

In vitro transcription assays were preformed based on a Promega protocol (#TB123) on linear DNA template consisting of the wild-type enhancer fragment sequence (which contains binding sites for the transcription factors assayed), the endogenous promoter (-120_+100 bp) and the first 145 nucleotides of the tgfp reporter. This linear DNA fragment was generated as a PCR product from primes *F5* (Table 5-1.3) and *tgfp_Rev* (5'GTGCTCTTCATCTTGTTGGTCATG 3') on plasmid DNA of the respective constructs. Run-off transcripts (245 nucleotides long) were run on urea-denaturing gels.

Preparation of DNA template: PCR product was excised from the agarose gel, DNA was extracted (Qiagen protocol) and further purified with EtOH precipitation (1/10 V NaAc 3M pH 5.2, 3 V EtOH ABS) and 3x 70% EtOH washing steps. After drying it was dissolved in 20 µl TE.

Reactions were set in a final volume of 12.5 µl including 50 ng of DNA template and 2-4.5 µl of crude nuclear extract (Soluble Nuclear Fraction, SNF) as a source of RNA polymerase II and basal transcriptional machinery components (Sullivan et al. (2000), *Drosophila* Protocols, CSHP, Chapter 31). Reactions also include 2.5 mM MgCl2, x µl Transcription buffer (20mM Hepes pH 7.9, 100mM NaCl, 0.2mM EDTA, 0.5mM DTT, 20% glycerol), 1µl of 12,5x rNTP (5mM rATP, 5mM rGTP, 5mM rCTP, 0.3mM rUTP) and 0.4 µl α-^{32}P rUTP of 1:500 stock dilution (800Ci/mmol; 10µCi/µl) (corresponds to 150 cps of fresh isotope). DNA template was preincubated with TNT-expressed protein factor at room temperature for 20 minutes. Then nuclear extract, rNTP mix and 150 cps α-^{32}P rUTP were added and *in vitro* transcription reactions were incubated at 30 °C for 30 minutes. An

equal volume of stop (loading) buffer was added (98% deionized formamide, 10 mM EDTA, bromophenol blue and xylene cyanol trace) and reactions were terminated by placing at 90 °C for 10 min (this also denatures secondary RNA forms). Reactions were placed shortly on ice and immediately loaded on urea denaturing gel. The gel was 6%, 7M Urea, 0.5xTBE. Reactions were run for 30-45 min 180 °C until BB dye reached the half of gel distance. Gel was dried under vacuum, 80 °C and exposed to PhosphoImager screen (Amersham).

5.10 Whole mount embryo *in situ* hybridization

Probe generation

To generated Dig- or Fluo-labeled RNA probes for *in situ* hybridization cDNA sequences of genes of interest were cloned in the PCRII vector (Invitrogen) and antisense-RNA was generated from T7 or Sp6 promoter (depending on insert orientation) using the respective RNA polymerase enzyme (from Roche, Fermentas or Ambion). Sequences of ~450 bp (minimum) to 1.2 kb were successfully used as probes. The template plasmid was first linearized at an appropriate restriction site 5' of the insert, excised from agarose gel, DNA was extracted and further purified by EtOH precipitation. 1-2 μg of linearized template was used in the *in vitro* transcription reaction (20-30 μl) according to the protocol of the enzyme's company in the presence of 10% DIG-labeling or Fluo-labeling rNTP mix (Roche). After 2 h incubation @37 °C, RNA was precipitated either with LiCl precipitation or EtOH-NaAc precipitation. Labeled RNA probe was resuspended in 50-80 μl ddsH20 and subsequently stored in the presence of 50% deionized formamide (for RNase inactivation and greater RNA stability during long period storage) at -80 °C.

Whole mount embryo fixation

0-10.30 hour collections were dechorionated for 3 min in 50 % chlorix, extensively washed with tab-water and mQH20 and fixed for 20 minutes, shaking 200-250 rpm rt, in fixation solution [2ml heptane, 1.5ml PEM (0.1 M PIPES, 1mM MgCl2, 1 mM EGTA, pH 6.9 (KOH)) and 187.5 μl formaldehyde). Lower phase was removed with a pasteur pipette, and an equal volume of MeOH was added followed by vigorous shaking and/or vortexing to achieve devitellinization. Settled embryos were repeatedly washed with MeOH (in maximum volume) followed by an overnight washing at 4 °C.

Fixed embryos can be stored in MeOH for long periods before *in situ* hybridization.

Double *in situ* hybridization using two probes (DIG- and Fluo-) was based on a protocol from Gleich NR. and Patel N. SDS detergent-treatment of whole mount embryos allows for the RNA probes to penetrate membranes, while at the same time embryos retain most of their original morphology in comparison to Proteinase K treatment-based protocols.

First, MeOH-stored embryos (embryos volume ~ 30-50 µl in 1.5 ml eppendorf tube) were gradually rehydrated in PBS (137 mM NaCl, 2.68 mM KCl, 10.14 mM Na2HPO4, 1.76 mM K2HPO4, pH 7.2) (by 10 min sequential washings 3:1, 1:1, 1:4) followed by a 20 min (maximum duration) secondary fixation in PBS, 10% formaldehyde. Then embryos were washed several times (3-5) in PBT exchanges (PBS, 0.1% Tween-20), followed by a 30 min SDS treatment while rotating at room temperature (1% SDS, 0.5% Tween-20, 50 mM Tris-HCl pH 7.5, 1 mM EDTA, 150 mM NaCl). Next, PBT washes (3-5 times from 10 min each) were performed. Embryos were brought gradually to hybridization solution conditions (SDS-Hybe: 50% Formamide, 5xSSC pH 4.5, 0.1% Tween-20, 0.3 % SDS, 50 µg/ml heparin, 100 µg/ml sonicated salmon sperm DNA) by an intermediate washing step in PBT:SDS-Hybe 1:1 (10 min rt). Embryos were washed once for 10 min in SDS-Hybe followed by prehybridation treatment in 100-200 µl prewarmed SDS-Hybe at 65 °C for 1-2 hours. In the meanwhile, RNA probes were prepared in SDS-Hybe solution (0.55 µl from 1:10 Dig-labeled probe dilution and 0.66 µl from 1:10 Fluo-labeled probe dilution in 50 µl hybridization volume were used). At the end of the prehybridation step, solution was aspirated from the embryos, probes were denatured for 1 min at 94 °C and shortly placed on ice before adding to the embryos. Hybridization was overnight at 65 or 68 °C.

Next day, probes were removed and embryos were washed several times in SDS-Hybe at hybridization temperature (on a heat-block shaker). Then through sequential warm washes in SDS-Hybe:PBT solution (4:1, 1:1, 1:4) they were gradually brought back to PBT condition (pH 7.2), followed by a few PBT washes at room-temprature. Embryos were blocked in BBT (PBT, 0.1%BSA) and anti-DIG antibody conjugated to alkaline phosphatase (a-DIG-AP) (1:2000 in BBT) was added (incubation 1 h rt gentle rotating). After several (4x20min) washes in BBT/PBT solution embryos were equilibrated directly in staining AP-buffer (100 mM pH 9.5, 100 mM NaCl, 50 mM MgCl2, 0.1% Tween-20). NBT/BCIP substrate (Sigma) was used for staining which was followed under the binocular scope and stopped in PBT buffer. Several washes afterwards help remove backround. In the continue, for double *in situ* staining, AP activity of the first antibody was inactivated in (50% formamide, 0.1 %Tween-20, 0.3%SDS, 5xSSC) for 15 min @65 °C followed by several PBT washes (5X20 min) at rt. After a blocking washing step in BBT, the second antibody (a-Fluo-AP) was added (1:2000) and incubated as mentioned above. For FastRed (Sigma) staining embryos were equilibrated in AP buffer pH 8.2 (5mM MgCl2, 100 mM NaCl, 0.1 mM Tween-20, 35 mM Tris pH 9.5, 65 mM Tris pH 9.5). Stained embryos were kept in PBT at 4 °C.

5.11 Fluoerescent Immunostaining (FIS)

For fluorescent immunostainings embryos were subjected to a milder fixation than for *in situ* hybridization. After dechorionization embryos were transferred in 2 ml heptane and an equal volume of 3.7 % formaldehyde in PEM (0.1 M PIPES, 1mM MgCl2, 1 mM EGTA, pH 6.9 (KOH)) was added (1.9 ml PEM+100 µl formaldehyde 37%). Thus, final formaldehyde concentration is 1.85%

(while prior *in situ* staining embryos are fixed in 2% formaldehyde followed by an additional secondary fixation).

Fixation was for 20 minutes at room temperature with shaking. Lower phase was removed with a pasteur pipette and an equal volume (2 ml) of MeOH was added. Embryos were devitellinized with vigorous shaking and/or vortexing and the settled ones were then washed with several MeOH exchanges. Embryos were additionally washed in MeOH over-night (rotating at 4°C).

On the first day of FIS embryos were gradually rehydrated from MeOH in PBX (PBS, 0.1%Triton-X) (4:1, 1:1, 1:4 and only PBX wash three times). Then they were blocked in BBX (PBX, 0.1% BSA) supplemented with 20% Western blocking reagent (Roche) for 1 h rotating rt. Primary antibodies were added in BBX; mouse_a-Col (monoclonal) was applied at 1:50 dilution, rb_a-Cnc at 1:200-1:300 and rb_a-Lab at 1:150. After overnight incubation with gentle shaking at 4°C, embryos were washed several times with BBX at rt and secondary antibodies (Alexa 488 a-mouse, Cy3 a-rabbit) were added at 1:300. Incubation was 1-2 h, rotating, rt. After several washes in PBX embryos were stored at 4 °C in PBX or PBT and mounted in glycerol for laser-scanning microscopy. During a PBX wash after 2nd antibody incubation, DAPI or Hoechst DNA-binding dye was added at a dilution 1:1000 for nuclear staining.

5.12 Microscopy

Embryos stained after *in situ* hybridization were mounted in glycerol (~90%) and documented with a Zeiss Axioplan 2 microscope (20x or 40x planes) using the ImageProPlus software (Version 6.2; MediaCybernetics). Pictures of fluorescent immunostained embryos (single planes or merged stacks) were taken with a Zeiss LSM 510 as previously described (Smith, Current Protoc Microbiol, 2006; Mavrakis et al., 2008, Curr Protoc Cell Biol).

APPENDIX

Stages of embryonic *Drosophila* development by Volker Hartenstein (1993).

The Stages of Embryogenesis
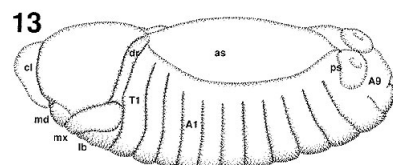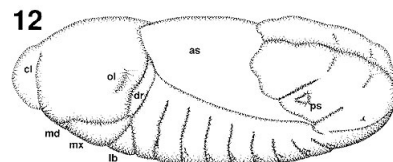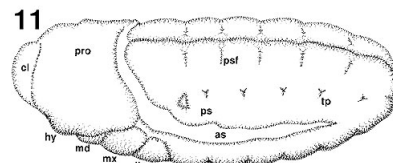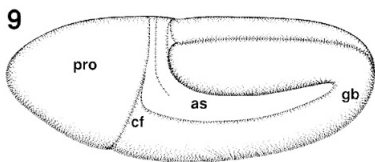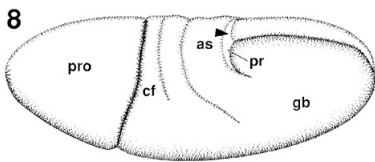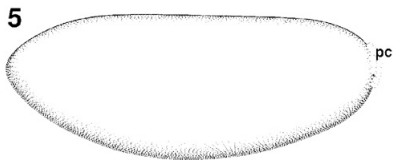
Stage 5: Embryogenesis starts with cleavage (stages 1- 4) during which the nucleus of the fertilized egg performs 13 rapid divisions. Most of the resulting nuclei become arranged in a single layer beneath the egg surface, and cell membranes are formed around the nuclei. This leads to the cellular blastoderm (stage 5), a homogeneous cellular sheet surrounding the central yolk. The germ-line cells, also called pole cells (*pc*), stand out as a cluster of 34-37 round cells at the posterior embryonic pole.

Stage 8: During gastrulation (stages 6-8), cells within the polar caps and the mid-ventral part of the blastoderm invaginate. The three germ Layers are generated by this movement. Most of the cells that remain at the surface represent the ectoderm; the invaginating cells form the endoderm (anterior and posterior midgut rudiments) and the mesoderm. A narrow mid-dorsal partition of the blastoderm gives rise to the amnioserosa (as), a thin membrane that covers the germ band dorsally. Coincident with gastrulation is the beginning of germ-band elongation, a movement that pushes the posterior tip of the germ band upward and then toward anterior. Several transitory furrows are formed during germ-band elongation. The deepest one of these furrows (cephalic furrow, *cf*) completely surrounds the embryo. Anterior to the cephalic furrow lies the procephalon (*pro*); posterior to the cephalic furrow lies the segmented germ band (*gb*). At the posterior tip of the germ band lies the amnioproctodeal invagination (*pr*), a pocket formed by the invaginating posterior midgut rudiment (endoderm) and the surrounding ectoderm that becomes the hindgut.

Stage 9: Following gastrulation, the germ band elongates further (stages 9-11). The ectoderm begins to split up into numerous different organ primordia (foregut and hindgut, CNS, epidermis). The cephalic furrow is still present; the other transient furrows have all but disappeared.

Stage 11: This stage is characterized by the parasegmental furrows (*psf*) that subdivide the germ band into metameric units. The first three segments (*md*, mandible; *mx*, maxilla; *lb*, labium) appear as conspicuous protuberances, the gnathal buds. The clypeolabrum (*cl*) and hypopharyngeal lobe (*hy*) are protuberances of the procephalon (*pro*) that mark the rudimentary first (labral) and third (intercalary) head segment. Tracheal pits (*tp*) appear in segments T2 to A8. The dorsal part of A8 gives rise to an additional invagination from which the posterior spiracle (*ps*) develops. The salivary gland invaginates from the ventral labium.

Stage 12: This stage begins when the germ band starts to retract and ends when germ-band retraction is complete. The external landmarks of the ectoderm (i.e., metameric furrows, gnathal buds, tracheal pits) that had appeared in stage 11 remain basically very similar throughout stages 12 and 13. A prominent feature of the head of a stage-12 embryo is the invaginating optic lobe (ol) and the dorsal ridge (dr), a protuberance formed by the dorsal part of the gnathal segments. During stage 12, important morphogenetic events take place in the endoderm and mesoderm.

Stage 13: The end of germ-band retraction marks the stage in embryogenesis at which cells in most organ primordia begin to differentiate, as testified by the expression of specific structural and biochemical markers. The main morphogenetic events that shape the embryonic surface after germ-band retraction are dorsal closure and head involution. During dorsal closure, the epidermal primordium stretches in the transversal axis, thereby gradually closing the gap that had so far persisted in the dorsal germ band. Dorsal closure begins anteriorly and posteriorly and progresses to mid levels of the germ band. By the end of stage 13, only A9 and the gnathal segments (i.e., the dorsal ridge) are closed. Head involution, which begins with stage 14, is a complex movement during which most head structures (i.e., the procephalon and the gnathal segments) vanish from the surface into the interior of the embryo. (*as*) Amnioserosa; (*A1, A9*) first and ninth abdominal segment, respectively; (*T1*) first thoracic segment.

From Campos-Ortega and Hartenstein (1997).

Overview of the Stages of Development
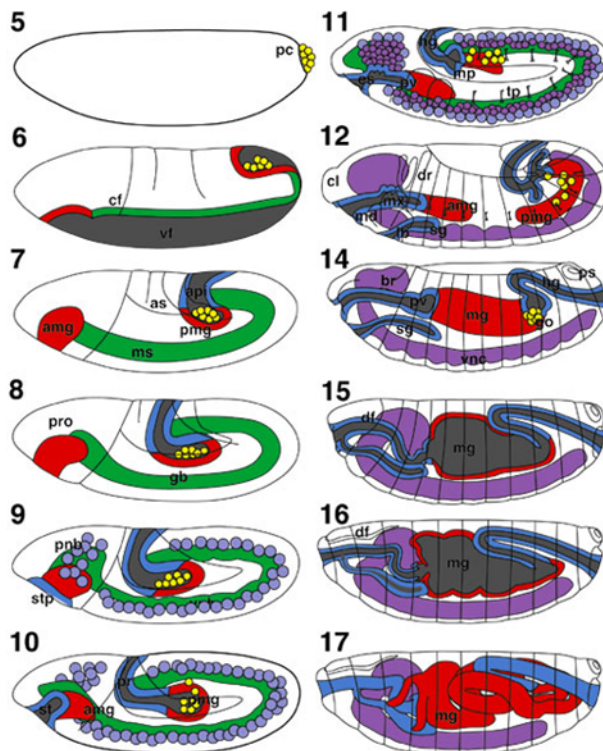
All embryos are in lateral view (anterior to the left). Endoderm, midgut; mesoderm; central nervous system; foregut, hindgut and pole cells in yellow.

(*amg*) (Anterior midgut rudiment; (*br*) brain; (*cf*) cephalic furrow; (*cl*) clypeolabrum; (*df*) dorsal fold; (*dr*) dorsal ridge; (*es*) esophagus; (*gb*) germ band; (*go*) gonads; (*hg*) hindgut; (*lb*) labial bud; (*md*) mandibular bud; (*mg*) midgut; (*mg*) Malpighian tubules; (mx) maxillary bud; (*pc*) pole cells; (*pmg*) posterior midgut rudiment; (*pnb*) procephalic neuroblasts; (*pro*) procephalon; (*ps*) posterior spiracle; (*po*) proventriculus; (*sg*) salivary gland; (*stp*) stomodeal plate; (*st*) stomodeum; (*tp*) tracheal pits; (*vf*) ventral furrow; (*vnb*) ventral neuroblasts; (*vnc*) ventral nerve

Stage 5: Blastoderm with the pole cells (pc) at the posterior pole

Stage 6: onset of gastrulation; anlagen that will invaginate during gastrulation: anterior midgut primordium anlage (am), mesoderm (ms), posterior midgut primordium (pm/pr). The cephalic furrow separates the procephalon from the prospective metameric germ band. Pole cells are included in the pm/pr. as, amnioserosa. anterior (atr) and posterior (ptr) transverse furrows are visible.

Stage 7: Beginning of germ band elongation. atr and ptr transverse furrows are approaching each other; the amnioproctodeal invagination (api) has deepened.

Stage 8: During stage 8 germ band elongation proceeds further.

Stage 9: Germ band elongation enters slow phase. The stomodeal plate becomes evident (stp).

Stage 10: Stomodeal invagination (st) takes place.

Stage 11: Epidermal segmentation becomes evident.

References

Abzhanov A, Kaufman TC. Homeotic genes and the arthropod head: expression patterns of the labial, proboscipedia, and Deformed genes in crustaceans and insects. Proc Natl Acad Sci USA. 1999; 96(18):10224-9.

Affolter M, Slattery M, Mann RS. A lexicon for homeodomain-DNA recognition. Cell. 2008;133(7):1133-5.

Akam M. The molecular basis for metameric pattern in the Drosophila embryo. Development. 1987;101(1):1-22. Review.

Bajusz I, Sipos L, Györgypál Z, Carrington EA, Jones RS, Gausz J, Gyurkovics H. The Trithorax-mimic allele of Enhancer of zeste renders active domains of target genes accessible to polycomb-group-dependent silencing in Drosophila melanogaster. Genetics. 2001;159(3):1135-50

Bally-Cuif L, Dubois L, Vincent A. Molecular cloning of Zcoe2, the zebrafish homolog of Xenopus Xcoe2 and mouse EBF-2, and its expression during primary neurogenesis. Mech Dev. 1998;77(1):85-90.

Baumgardt M, Miguel-Aliaga I, Karlsson D, Ekman H, Thor S. Specification of neuronal identities by feedforward combinatorial coding. PLoS Biol. 2007 5(2):e37.

Bejerano G, Siepel AC, Kent WJ, Haussler D. Computational screening of conserved genomic DNA in search of functional noncoding elements. Nat Methods. 2005 Jul; 2(7):535-45.

Bengal E, Ransone L, Scharfmann R, Dwarki VJ, Tapscott SJ, Weintraub H, Verma IM. Functional antagonism between c-Jun and MyoD proteins: a direct physical association. Cell. 1992;68(3):507-19.

Benson,M. and Pirrotta,V. The Drosophila zeste protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation. EMBO J. 1988;7(12):3907-15.

Berger et al., 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell. 2008;133(7):1266-76.

Bickel S, Pirrotta V. Self-association of the Drosophila zeste protein is responsible for transvection effects. EMBO J. 1990;9(9):2959-67.

Biggin MD, Bickel S, Benson M, Pirrotta V, Tjian R. Zeste encodes a sequence-specific transcription factor that activates the Ultrabithorax promoter in vitro. Cell. 1988;53(5):713-22.

Biggin MD, McGinnis W. Regulation of segmentation and segmental identity by Drosophila homeoproteins: the role of DNA binding in functional activity and specificity. Development. 1997;124(22):4425-33. Review.

Bischof J, Maeda RK, Hediger M, Karch F, Basler K. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci USA. 2007;104(9):3312-7.

Blackwell TK, Bowerman B, Priess JR, Weintraub H. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. Science. 1994;266(5185):621-8.

Blastyák A, Mishra RK, Karch F, Gyurkovics H. Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. Mol Cell Biol. 2006;26(4):1434-44.

Bloor JW, Kiehart DP. zipper Nonmuscle myosin-II functions downstream of PS2 integrin in Drosophila myogenesis and is necessary for myofibril formation. Dev Biol. 2001;239(2):215-28.

Brock HW, van Lohuizen M. The Polycomb group--no longer an exclusive club? Curr Opin Genet Dev. 2001;11(2):175-81. Review

Brown JL, Fritsch C, Mueller J, Kassis JA. The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development. 2003; 130(2):285-94.

Brown JL, Grau DJ, DeVido SK, Kassis JA. An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. Nucleic Acids Res. 2005;33(16):5181-9.

Bucher and Wimmer. Beetle a-head, *B.I.F. Futura* 20 (2005), pp. 164–169.

Bürglin TR. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. 1997;25(21):4173-80.

Butler JE, Kadonaga JT. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev. 2001;15(19):2515-9.

Cadigan KM, Grossniklaus U, Gehring WJ. Functional redundancy: the respective roles of the two sloppy paired genes in Drosophila segmentation. Proc Natl Acad Sci U S A. 1994; 91(14):6324-8.

Cadigan KM, Grossniklaus U, Gehring WJ. Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments. Genes Dev. 1994;8(8):899-913.

Campos-Ortega J.A. and Hartenstein V. The embryonic development of Drosophila melenogaster. Second Edition, Springer, 1997.

Carr A, Biggin MD. A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. EMBO J. 1999;18(6):1598-608.

Casas-Tinto S, Gomez-Velazquez M, Granadino B, Fernandez-Funez P. FoxK mediates TGF-beta signalling during midgut differentiation in flies. J Cell Biol. 2008; 183(6):1049-60.

Chan SK, Mann RS. A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. Proc Natl Acad Sci U S A. 1996;93(11):5223-8.

Chan SK, Pöpperl H, Krumlauf R, Mann RS. An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. EMBO J. 1996;15(10):2476-87.

Chang CP, Shen WF, Rozenfeld S, Lawrence HJ, Largman C, Cleary ML. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. Genes Dev. 1995;9(6):663-74.

Chen JD, Pirrotta V. Multimerization of the Drosophila zeste protein is required for efficient DNA binding. EMBO J. 1993;12(5):2075-83.

Chen JD, Pirrotta V. Stepwise assembly of hyperaggregated forms of Drosophila zeste mutant protein suppresses white gene expression in vivo. EMBO J. 1993;12(5):2061-73.

Chiang C, Young KE, Beachy PA. Control of Drosophila tracheal branching by the novel homeodomain gene unplugged, a regulatory target for genes of the bithorax complex. Development. 1995;121(11):3901-12

Churchill ME, Jones DN, Glaser T, Hefner H, Searles MA, Travers AA. HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. EMBO J. 1995; 14(6):1264-75.

Cohen S, Jürgens G. Drosophila headlines. Trends Genet. 1991;7(8):267-72.

Cohen SM, Jürgens G. Mediation of Drosophila head development by gap-like segmentation genes. Nature. 1990;346(6283):482-5.

Côté S, Preiss A, Haller J, Schuh R, Kienlin A, Seifert E, Jäckle H. The gooseberry-zipper region of Drosophila: five genes encode different spatially restricted transcripts in the embryo. EMBO J. 1987;6(9):2793-2801

Crémazy F, Berta P, Girard F. Genome-wide analysis of Sox genes in Drosophila melanogaster. Mech Dev. 2001;109(2):371-5.

Crozatier M, Glise B, Vincent A. Connecting Hh, Dpp and EGF signalling in patterning of the Drosophila wing; the pivotal role of collier/knot in the AP organiser. Development. 2002; 129(18):4261-9.

Crozatier M, Ubeda JM, Vincent A, Meister M. Cellular immune response to parasitization in Drosophila requires the EBF orthologue collier. PLoS Biol. 2004;2(8):E196.

Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A. Collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. Curr Biol. 1996;6(6):707-18.

Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A. Head versus trunk patterning in the Drosophila embryo; collier requirement for formation of the intercalary segment. Development. 1999;126(19):4385-94.

Crozatier M, Vincent A. Control of multidendritic neuron differentiation in Drosophila: the role of Collier. Dev Biol. 2008;315(1):232-42.

Crozatier M, Vincent A. Requirement for the Drosophila COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to notch signalling. Development. 1999;126(7):1495-504.

Daburon V, Mella S, Plouhinec JL, Mazan S, Crozatier M, Vincent A. The metazoan history of the COE transcription factors. Selection of a variant HLH motif by mandatory inclusion of a duplicated exon in vertebrates. BMC Evol Biol. 2008;8:131.

Dailey L, Basilico C. Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. J Cell Physiol. 2001;186(3):315-28.

Déjardin J, Cavalli G. Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. EMBO J. 2004;23(4):857-68. Epub 2004 Feb 12.

Déjardin J, Rappailles A, Cuvier O, Grimaud C, Decoville M, Locker D, Cavalli G. Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature. 2005;434(7032):533-8.

DeVido SK, Kwon D, Brown JL, Kassis JA. The role of Polycomb-group response elements in regulation of engrailed transcription in Drosophila. Development. 2008;135(4):669-76.

Diederich RJ, Merrill VK, Pultz MA, Kaufman TC. Isolation, structure, and expression of labial, a homeotic gene of the Antennapedia Complex involved in Drosophila head development. Genes Dev. 1989;3(3):399-414.

DiNardo S, Heemskerk J, Dougan S, O'Farrell PH. The making of a maggot: patterning the Drosophila embryonic epidermis. Curr Opin Genet Dev. 1994; 4(4):529-34.

Dorsett D. Cohesin, gene expression and development: lessons from Drosophila. Chromosome Res. 2009;17(2):185-200. Review.

Dorsett D. Distant liaisons: long-range enhancer-promoter interactions in Drosophila. Curr Opin Genet Dev. 1999;9(5):505-14. Review.

Dragan AI, Read CM, Makeyeva EN, Milgotina EI, Churchill ME, Crane-Robinson C, Privalov PL. DNA binding and bending by HMG boxes: energetic determinants of specificity. J Mol Biol. 2004; 343(2):371-93.

Dubois L, Bally-Cuif L, Crozatier M, Moreau J, Paquereau L, Vincent A. XCoe2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in Xenopus. Curr Biol. 1998;8(4):199-209.

Dubois L, Enriquez J, Daburon V, Crozet F, Lebreton G, Crozatier M, Vincent A. Collier transcription in a single Drosophila muscle lineage: the combinatorial control of muscle identity. Development. 2007 Dec;134(24):4347-55.

Dubois L, Vincent A. The COE--Collier/Olf1/EBF--transcription factors: structural conservation and diversity of developmental functions. Mech Dev. 2001;108(1-2):3-12.

Dynlacht BD, Attardi LD, Admon A, Freeman M, Tjian R. Functional analysis of NTF-1, a developmentally regulated Drosophila transcription factor that binds neuronal cis elements. Genes Dev. 1989; 3(11):1677-88.

Ebner A, Cabernard C, Affolter M, Merabet S. Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. Development. 2005;132(7):1591-600.

Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F. The Trithorax-like gene encodes the Drosophila GAGA factor. Nature. 1994;371(6500):806-8.

Farkas G, Leibovitch BA, Elgin SC. Chromatin organization and transcriptional control of gene expression in Drosophila. Gene. 2000 Aug 8;253(2):117-36. Review.

Finkelstein R, Perrimon N. The molecular genetics of head development in Drosophila melanogaster. Development. 1991 Aug;112(4):899-912. Review.

Finkelstein R, Perrimon N. The orthodenticle gene is regulated by bicoid and torso and specifies Drosophila head development. Nature. 1990;346(6283):485-8.

Fritsch C, Brown JL, Kassis JA, Müller J. The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. Development. 1999;126(17):3905-13.

Fujioka M, Yusibova GL, Zhou J, Jaynes JB. The DNA-binding Polycomb-group protein Pleiohomeotic maintains both active and repressed transcriptional states through a single site. Development.2008;135(24):4131-9.

Fujita SC, Zipursky SL, Benzer S, Ferrus A, Shotwell SL: Monoclonal antibodies against the Drosophila nervous system. Proc Natl Acad Sci U S A 1982, 79(24):7929-7933.

Gallitano-Mendel A, Finkelstein R. Ectopic orthodenticle expression alters segment polarity gene expression but not head segment identity in the Drosophila embryo. Dev Biol. 1998;199(1):125-37.

Gallitano-Mendel A, Finkelstein R. Novel segment polarity gene interactions during embryonic head development in Drosophila. Dev Biol. 1997;192(2):599-613.

Garraway IP, Semple K, Smale ST. Transcription of the lymphocyte-specific terminal deoxynucleotidyltransferase gene requires a specific core promoter structure. Proc Natl Acad Sci USA. 1996;93(9):4336-41.

Giese K, Grosschedl R. LEF-1 contains an activation domain that stimulates transcription only in a specific context of factor-binding sites. EMBO J. 1993; 12(12):4667-76.

Giese K, Pagel J, Grosschedl R. Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors. Proc Natl Acad Sci U S A. 1994; 91(8):3368-72.

Giese K, Pagel J, Grosschedl R. Functional analysis of DNA bending and unwinding by the high mobility group domain of LEF-1. Proc Natl Acad Sci U S A. 1997;94(24):12845-50.

Green NC, Rambaldi I, Teakles J, Featherstone MS.A conserved C-terminal domain in PBX increases DNA binding by the PBX homeodomain and is not a primary site of contact for the YPWM motif of HOXA1. J Biol Chem. 1998;273(21):13273-9.

Gregory PA, Mackenzie PI. The homeodomain Pbx2-Prep1 complex modulates hepatocyte nuclear factor 1alpha-mediated activation of the UDP-glucuronosyltransferase 2B17 gene. Mol Pharmacol. 2002;62(1):154-61.

Grosschedl R, Giese K, Pagel J. HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet. 1994;10(3):94-100. Review.

Grossniklaus U, Cadigan KM, Gehring WJ. Three maternal coordinate systems cooperate in the patterning of the Drosophila head. Development. 1994 Nov;120(11):3155-71.

Häcker U, Kaufmann E, Hartmann C, Jürgens G, Knöchel W, Jäckle H. The Drosophila fork head domain protein crocodile is required for the establishment of head structures. EMBO J. 1995; 14(21):5306-17.

Hagman J, Belanger C, Travis A, Turck CW, Grosschedl R. Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. Genes Dev. 1993; 7(5):760-73.

Hagman J, Grosschedl R. Regulation of gene expression at early stages of B-cell differentiation. Curr Opin Immunol. 1994;6(2):222-30. Review.

Hagman J, Gutch MJ, Lin H, Grosschedl R. EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. EMBO J. 1995;14(12):2907-16.

Hagman J, Travis A, Grosschedl R. A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. EMBO J. 1991;10(11):3409-17.

Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M. Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. Plant Methods. 2007; 3:11.

Hartenstein V. Atlas of Drosophila Development. Cold Spring Harbor Laboratory Press, 1993.

Hoch M, Jäckle H. Krüppel acts as a developmental switch gene that mediates Notch signalling-dependent tip cell differentiation in the excretory organs of Drosophila. EMBO J. 1998; 17(19):5766-75.

Horard B, Tatout C, Poux S, Pirrotta V. Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. Mol Cell Biol. 2000; 20(9):3187-97

Horn C, Wimmer EA. A versatile vector set for animal transgenesis. Dev Genes Evol. 2000;210(12):630-7.

Hsia CC, McGinnis W. Evolution of transcription factor function. Curr Opin Genet Dev. 2003;13(2):199-206. Review.

Hsieh-Li HM, Witte DP, Szucsik JC, Weinstein M, Li H, Potter SS. Gsh-2, a murine homeobox gene expressed in the developing brain. Mech Dev. 1995;50(2-3):177-86.

Hsu JY, Juven-Gershon T, Marr MT 2nd, Wright KJ, Tjian R, Kadonaga JT. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. Genes Dev. 2008;22(17):2353-8.

Hultmark D, Klemenz R, Gehring WJ. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell. 1986; 44(3):429-38.

Hur MW, Laney JD, Jeon SH, Ali J, Biggin MD. Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. Development. 2002;129(6):1339-43.

Ingham PW, Martinez Arias A. Boundaries and fields in early embryos. Cell. 1992;68(2):221-35

Jabet C, Gitti R, Summers MF, Wolberger C. NMR studies of the pbx1 TALE homeodomain protein free in solution and bound to DNA: proposal for a mechanism of HoxB1-Pbx1-DNA complex assembly. J Mol Biol. 1999;291(3):521-30.

Judd BH. Mutations of zeste that mediate transvection are recessive enhancers of position-effect variegation in Drosophila melanogaster. Genetics. 1995;141(1):245-53.

Juven-Gershon T, Hsu JY, Kadonaga JT. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. Genes Dev. 2008;22(20):2823-30.

Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter - the gateway to transcription. Curr Opin Cell Biol. 2008;20(3):253-9. Review.

Kamachi Y, Sockanathan S, Liu Q, Breitman M, Lovell-Badge R, Kondoh H. Involvement of SOX proteins in lens-specific activation of crystallin genes. EMBO J. 1995; 14(14):3510-9.

Kassis JA. Spatial and temporal control elements of the Drosophila engrailed gene. Genes Dev. 1990;4(3):433-43.

Katsani KR, Hajibagheri MA, Verrijzer CP. Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. EMBO J. 1999;18(3):698-708.

Kaufmann E, Müller D, Knöchel W. DNA recognition site analysis of Xenopus winged helix proteins. J Mol Biol. 1995;248(2):239-54.

Ketchum AS, Stewart CT, Stewart M, Kiehart DP. Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript: conserved sequences in the myosin tail and differential splicing in the 5' untranslated sequence. Proc Natl Acad Sci U S A. 1990;87(16):6316-20.

Kim S, Yamazaki M, Zella LA, Shevde NK, Pike JW. Activation of receptor activator of NF-kappaB ligand gene expression by 1,25-dihydroxyvitamin D3 is mediated through multiple long-range enhancers. Mol Cell Biol. 2006 Sep;26(17):6469-86.

Knoepfler PS, Calvo KR, Chen H, Antonarakis SE, Kamps MP. Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. Proc Natl Acad Sci U S A. 1997;94(26):14553-8.

Krysinska H, Hoogenkamp M, Ingram R, Wilson N, Tagoh H, Laslo P, Singh H, Bonifer C. A two-step, PU.1-dependent mechanism for developmentally regulated chromatin remodeling and transcription of the c-fms gene. Mol Cell Biol. 2007;27(3):878-87.

Kudrycki K, Stein-Izsak C, Behn C, Grillo M, Akeson R, Margolis FL. Olf-1-binding site: characterization of an olfactory neuron-specific promoter motif. Mol Cell Biol. 1993;13(5):3002-14.

Kutach AK, Kadonaga JT. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol. 2000;20(13):4754-64.

Kwong C, Adryan B, Bell I, Meadows L, Russell S, Manak JR, White R. Stability and dynamics of polycomb target sites in Drosophila development. PLoS Biol. 2009;7(1)

Lai ZC, Fortini ME, Rubin GM. The embryonic expression patterns of zfh-1 and zfh-2, two Drosophila genes encoding novel zinc-finger homeodomain proteins. Mech Dev. 1991; 34(2-3):123-34.

Laney JD, Biggin MD. Redundant control of Ultrabithorax by zeste involves functional levels of zeste protein binding at the Ultrabithorax promoter. Development. 1996;122(7):2303-11.

Laney JD, Biggin MD. zeste, a nonessential gene, potently activates Ultrabithorax transcription in the Drosophila embryo. Genes Dev. 1992;6(8):1531-41.

Laney JD, Biggin MD. Zeste-mediated activation by an enhancer is independent of cooperative DNA binding in vivo. Proc Natl Acad Sci U S A. 1997;94(8):3602-4.

Laudet V, Stehelin D, Clevers H. Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res. 1993; 21(10):2493-501.

Lee HH, Frasch M. Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. Development. 2000;127(24):5497-508.

Lee JJ, von Kessler DP, Parks S, Beachy PA. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. Cell. 1992;71(1):33-50.

Lefebvre V, Dumitriu B, Penzo-Méndez A, Han Y, Pallavi B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol. 2007; 39(12):2195-214. Review.

Lessing D, Nusse R. Expression of wingless in the Drosophila embryo: a conserved cis-acting element lacking conserved Ci-binding sites is required for patched-mediated repression. Development. 1998;125(8):1469-76.

Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell. 2007;128(4):707-19.Review.

Li X, Murre C, McGinnis W. Activity regulation of a Hox protein and a role for the homeodomain in inhibiting transcriptional activation. EMBO J. 1999;18(1):198-211.

Li X, Noll M. Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo. EMBO J. 1994 Jan;13(2):400-6.

Li X, Veraksa A, McGinnis W. A sequence motif distinct from Hox binding sites controls the specificity of a Hox response element. Development. 1999;126(24):5581-9.

Liang Z, Biggin MD. Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila. Development. 1998; 125(22):4471-82.

Li-Kroeger D, Witt LM, Grimes HL, Cook TA, Gebelein B. Hox and senseless antagonism functions as a molecular switch to regulate EGF secretion in the Drosophila PNS. Dev Cell. 2008;15(2):298-308.

Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev. 2004;18(13):1606-17.

Liu Z, Yang X, Dong Y, Friedrich M. Tracking down the "head blob": comparative analysis of wingless expression in the developing insect procephalon reveals progressive reduction of embryonic visual system patterning in higher insects. Arthropod Struct Dev. 2006; 35(4):341-56.

Lohmann I, McGinnis W. Hox Genes: it's all a matter of context. Curr Biol. 2002; 12(15):R514-6.

Love JJ, Li X, Case DA, Giese K, Grosschedl R, Wright PE. Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature. 1995;376(6543):791-5.

Lu Q, Wright DD, Kamps MP. Fusion with E2A converts the Pbx1 homeodomain protein into a constitutive transcriptional activator in human leukemias carrying the t(1;19) translocation. Mol Cell Biol. 1994;14(6):3938-48.

Macías A, Morata G. Functional hierarchy and phenotypic suppression among Drosophila homeotic genes: the labial and empty spiracles genes. EMBO J. 1996 Jan 15;15(2):334-43.

Mahmoudi T, Katsani KR, Verrijzer CP. GAGA can mediate enhancer function in trans by linking two separate DNA molecules. EMBO J. 2002;21(7):1775-81.

Mahmoudi T, Zuijderduijn LM, Mohd-Sarip A, Verrijzer CP. GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. Nucleic Acids Res. 2003;31(14):4147-56.

Mann RS, Chan SK. Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet. 1996;12(7):258-62. Review

Mannervik M. Target genes of homeodomain proteins. Bioessays. 1999; 21(4):267-70. Review.

Martinez Arias, A. (1993) Development and patterning of the larval epidermis of Drosophila. In Bate, M. and Martinez Arias, A. (eds), The Development of Drosophila Melanogaster. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 517–608.

Mavrakis M, Rikhy R, Lilly M, Lippincott-Schwartz J. i Fluorescence maging techniques for studying Drosophila embryo development. Curr Protoc Cell Biol. 2008;Chapter 4:Unit 4.18.

McGinnis N, Ragnhildstveit E, Veraksa A, McGinnis W. A cap 'n' collar protein isoform contains a selective Hox repressor function. Development. 1998;125(22):4553-64.

Merli C, Bergstrom DE, Cygan JA, Blackman RK. Promoter specificity mediates the independent regulation of neighboring genes. Genes Dev. 1996 May;10(10):1260-70.

Merrill VK, Diederich RJ, Turner FR, Kaufman TC. A genetic and developmental analysis of mutations in labial, a gene necessary for proper head formation in Drosophila melanogaster. Dev Biol. 1989;135(2):376-91.

Miele A, Dekker J. Long-range chromosomal interactions and gene regulation. Mol Biosyst. 2008; 4(11):1046-57

Mohd-Sarip A, Venturini F, Chalkley GE, Verrijzer CP. Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. Mol Cell Biol. 2002 Nov;22(21):7473-83.

Mohler J, Mahaffey JW, Deutsch E, Vani K. Control of Drosophila head segment identity by the bZIP homeotic gene cnc. Development. 1995; 121(1):237-47.

Mohler J. Spatial regulation of segment polarity gene expression in the anterior terminal region of the Drosophila blastoderm embryo. Mech Dev. 1995; 50(2-3):151-61. Review.

Montell DJ, Rorth P, Spradling AC. slow border cells, a locus required for a developmentally regulated cell migration during oogenesis, encodes Drosophila C/EBP. Cell. 1992;71(1):51-62.

Morcillo P, Rosen C, Baylies MK, Dorsett D. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. Genes Dev. 1997;11(20):2729-40.

Mulholland NM, King IF, Kingston RE. Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. Genes Dev. 2003 Nov 15;17(22):2741-6

Nagao T, Endo K, Kawauchi H, Walldorf U, Furukubo-Tokunaga K. Patterning defects in the primary axonal scaffolds caused by the mutations of the extradenticle and homothorax genes in the embryonic Drosophila brain. Dev Genes Evol. 2000;210(6):289-99.

Nambu PA, Nambu JR. The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. Development. 1996; 122(11):3467-75.

Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell. 2008;133(7):1277-89.

Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila. Nature. 1980; 287(5785):795-801.

Orihara M, Hosono C, Kojima T, Saigo K. Identification of engrailed promoter elements essential for interactions with a stripe enhancer in Drosophila embryos. Genes Cells. 1999; 4(4):205-18.

Orlando V, Jane EP, Chinwalla V, Harte PJ, Paro R. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. EMBO J. 1998;17(17):5141-50.

Pearson JC, Lemons D, McGinnis W. Modulating Hox gene functions during animal body patterning. Nat Rev Genet. 2005;6(12):893-904.

Pederson JA, LaFollette JW, Gross C, Veraksa A, McGinnis W, Mahaffey JW. Regulation by homeoproteins: a comparison of deformed-responsive elements. Genetics. 2000;156(2):677-86.

Perrimon N. Hedgehog and beyond. Cell. 80(4):517-20.

Peterson RS, Lim L, Ye H, Zhou H, Overdier DG, Costa RH. The winged helix transcriptional activator HFH-8 is expressed in the mesoderm of the primitive streak stage of mouse embryos and its cellular derivatives. Mech Dev. 1997;69(1-2):53-69.

Pinsonneault J, Florence B, Vaessin H, McGinnis W. A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. EMBO J. 1997;16(8):2032-42.

Pirrotta V, Bickel S, Mariani C. Developmental expression of the Drosophila zeste gene and localization of zeste protein on polytene chromosomes. Genes Dev. 1988; 2(12B):1839-50.

Pirrotta V. Vectors for P-mediated transformation in Drosophila. Biotechnology. 1988; 10:437-56. Review.

Popichenko D, Sellin J, Bartkuhn M, Paululat A. Hand is a direct target of the forkhead transcription factor Biniou during Drosophila visceral mesoderm differentiation. BMC Dev Biol. 2007; 7:49.

Pozzoli O, Bosetti A, Croci L, Consalez GG, Vetter ML. Xebf3 is a regulator of neuronal differentiation during primary neurogenesis in Xenopus. Dev Biol. 2001;233(2):495-512.

Prasad BC, Ye B, Zackhary R, Schrader K, Seydoux G, Reed RR. unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. Development. 1998;125(8):1561-8.

Ren J, Gao X, Jin C, Zhu M, Wang X, Shaw A, Wen L, Yao X, Xue Y. Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. Proteomics. 2009;9(12):3409-3412.

Ringrose L, Paro R. Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development. 2007;134(2):223-32. Review.

Rogers BT, Kaufman TC. Structure of the insect head as revealed by the EN protein pattern in developing embryos. Development. 1996; 122(11):3419-32.

Rogers BT, Kaufman TC. Structure of the insect head in ontogeny and phylogeny: a view from Drosophila. Int Rev Cytol. 1997;174:1-84. Review.

Rørth P, Montell DJ. Drosophila C/EBP: a tissue-specific DNA-binding protein required for embryonic development. Genes Dev. 1992;6(12A):2299-311.

Russell SR, Sanchez-Soriano N, Wright CR, Ashburner M. The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. Development. 1996; 122(11):3669-76.

Sambrook and Russell, Molecular Cloning, A Laboratory Manual, Third Edition (2001), Cold Spring Harbor Laboratory Press.

Sánchez-Soriano N, Russell S. Regulatory mutations of the Drosophila Sox gene Dichaete reveal new functions in embryonic brain and hindgut development. Dev Biol. 2000;220(2):307-21.

Sandmann T, Jakobsen JS, Furlong EE. ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos. Nat Protoc. 2006;1(6):2839-55.

Sanson B. Generating patterns from fields of cells; Examples from Drosophila segmentation. EMBO reports 2, 12, 1083–1088 (2001).

Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. Nature. 2001 Aug 9;412(6847):655-60.

Schmidt-Ott U, González-Gaitán M, Jäckle H, Technau GM. Number, identity, and sequence of the Drosophila head segments as revealed by neural elements and their deletion patterns in mutants. Proc Natl Acad Sci U S A. 1994;91(18):8363-7.

Schmidt-Ott U, Technau GM. Expression of en and wg in the embryonic head and brain of Drosophila indicates a refolded band of seven segment remnants. Development. 1992;116(1):111-25.

Schmidt-Ott U. Different ways to make a head. Bioessays. 2001 Jan;23(1):8-11

Schöck F, Purnell BA, Wimmer EA, Jäckle H. Common and diverged functions of the Drosophila gene pair D-Sp1 and buttonhead. Mech Dev. 1999;89(1-2):125-32.

Schöck F, Reischl J, Wimmer E, Taubert H, Purnell BA, Jäckle H. Phenotypic suppression of empty spiracles is prevented by buttonhead. Nature. 2000; 405(6784):351-4.

Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, van Lohuizen M, Tanay A, Cavalli G. Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol. 2009 13;7(1)

Schwartz YB, Pirrotta V. Polycomb complexes and epigenetic states. Curr Opin Cell Biol. 2008;20(3):266-73. Review.

Seecoomar M, Agarwal S, Vani K, Yang G, Mohler J. knot is required for the hypopharyngeal lobe and its derivatives in the Drosophila embryo. Mech Dev. 2000; 91(1-2):209-15.

Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature. 2008 Jan 31;451(7178):535-40.

Segal E, Widom J. From DNA sequence to transcriptional behaviour: a quantitative approach. Nat Rev Genet. 2009;10(7):443-56.

Sellin, J., Albrecht, S., Koelsch, V., Paululat, A. (2006). Dynamics of heart differentiation, visualized utilizing heart enhancer elements of the Drosophila melanogaster bHLH transcription factor Hand. Gene Expression Patterns 6(4): 360--375.

Smith CL. Basic confocal microscopy. Curr Protoc Cell Biol. 2001 May;Chapter 4:Unit 4.5.

Soriano NS, Russell S. The Drosophila SOX-domain protein Dichaete is required for the development of the central nervous system midline. Development. 1998;125(20):3989-96.

Soullier S, Jay P, Poulat F, Vanacker JM, Berta P, Laudet V. Diversification pattern of the HMG and SOX family members during evolution. J Mol Evol. 1999; 48(5):517-27.

St Johnston D, Nüsslein-Volhard C. The origin of pattern and polarity in the Drosophila embryo. Cell. 68(2):201-19.

Sullivan W, Ashburner M, Scott Hawley R., Drosophila Protocols, 2000, Cold Spring Harbor Laboratory Press.

Sylvain Poux, Béatrice Horard, Christian J. A. Sigrist‡ and Vincenzo Pirrotta. The Drosophila Trithorax protein is a coactivator required to prevent re-establishment of Polycomb silencing. Development 129, 2483-2493 (2002).

Szutorisz H, Dillon N, Tora L. The role of enhancers as centres for general transcription factor recruitment. Trends Biochem Sci. 2005;30(11):593-9. Review.

Tillib S, Petruk S, Sedkov Y, Kuzin A, Fujioka M, Goto T, Mazo A.Trithorax- and Polycomb-group response elements within an Ultrabithorax transcription maintenance unit consist of closely situated but separable sequences. Mol Cell Biol. 1999;19(7):5189-202.

Tour E, Hittinger CT, McGinnis W. Evolutionarily conserved domains required for activation and repression functions of the Drosophila Hox protein Ultrabithorax. Development. 2005;132(23):5271-81.

Travis A, Hagman J, Grosschedl R. Heterogeneously initiated transcription from the pre-B- and B-cell-specific mb-1 promoter: analysis of the requirement for upstream factor-binding sites and initiation site sequences. Mol Cell Biol. 1991;11(11):5756-66.

Travis A, Hagman J, Hwang L, Grosschedl R. Purification of early-B-cell factor and characterization of its DNA-binding specificity. Mol Cell Biol. 1993; 13(6):3392-400.

Urbach R, Technau GM. Early steps in building the insect brain: neuroblast formation and segmental patterning in the developing brain of different insect species. Arthropod Struct Dev. 2003 Aug;32(1):103-23.

Urbach R, Technau GM. Molecular markers for identified neuroblasts in the developing brain of Drosophila. Development. 2003 Aug; 130(16):3621-37.

Urbach R, Technau GM. Segment polarity and DV patterning gene expression reveals segmental organization of the Drosophila brain. Development. 2003;130(16):3607-20.

Urbach, R., Technau, G.M. (2004). Neuroblast formation and patterning during early brain development in Drosophila. BioEssays 26(7): 739-751

Vactor DV, Sink H, Fambrough D, Tsoo R, Goodman CS: Genes that control neuromuscular specificity in Drosophila. Cell 1993, 73(6):1137-1153.

van de Wetering M, Cavallo R, Dooijes D, van Beest M, van Es J, Loureiro J, Ypma A, Hursh D, Jones T, Bejsovec A, Peifer M, Mortin M, Clevers H. Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. Cell. 1997;88(6):789-99.

van de Wetering M, Oosterwegel M, Dooijes D, Clevers H. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J. 1991; 10(1):123-32.

van de Wetering M, Oosterwegel M, van Norren K, Clevers H. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. EMBO J. 1993; 12(10):3847-54.

Vardhanabhuti S, Wang J, Hannenhalli S. Position and distance specificity are important determinants of cis-regulatory motifs in addition to evolutionary conservation. Nucleic Acids Res. 2007; 35(10):3203-13.

Veraksa A, McGinnis N, Li X, Mohler J, McGinnis W. Cap 'n' collar B cooperates with a small Maf subunit to specify pharyngeal development and suppress deformed homeotic function in the Drosophila head. Development. 2000;127(18):4023-37.

Vincent A, Blankenship JT, Wieschaus E. Integration of the head and trunk segmentation systems controls cephalic furrow formation in Drosophila. Development. 1997;124(19):3747-54.

Vincent S, Perrimon N, Axelrod JD. Hedgehog and Wingless stabilize but do not induce cell fate during Drosophila dorsal embryonic epidermal patterning. Development. 2008;135(16):2767-75.

Wang MM, Reed RR. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature. 1993;364(6433):121-6.

Wang MM, Tsai RY, Schrader KA, Reed RR. Genes encoding components of the olfactory signal transduction cascade contain a DNA binding site that may direct neuronal expression. Mol Cell Biol. 1993;13(9):5805-13.

Werner MH, Huth JR, Gronenborn AM, Clore GM. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell. 1995; 81(5):705-14.

Wiellette EL, McGinnis W Hox genes differentially regulate Serrate to generate segment-specific structures. Development. 1999;126(9):1985-95.

Willy PJ, Kobayashi R, Kadonaga JT. A basal transcription factor that activates or represses transcription. Science. 2000;290(5493):982-5.

Wilson M, Koopman P. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr Opin Genet Dev. 2002; 12(4):441-6. Review.

Wimmer EA, Carleton A, Harjes P, Turner T, Desplan C. Bicoid-independent formation of thoracic segments in Drosophila. Science. 2000;287(5462):2476-9.

Wimmer EA, Cohen SM, Jäckle H, Desplan C. buttonhead does not contribute to a combinatorial code proposed for Drosophila head development. Development. 1997;124(8):1509-17.

Wodarz A. Extraction and immunoblotting of proteins from embryos. Methods Mol Biol. 2008;420:335-45.

Yu Y, Yussa M, Song J, Hirsch J, Pick L. A double interaction screen identifies positive and negative ftz gene regulators and ftz-interacting proteins. Mech Dev. 1999; 83(1-2):95-105.

Yuan H, Corbi N, Basilico C, Dailey L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev. 1995; 9(21):2635-45.

Zeng C, Pinsonneault J, Gellon G, McGinnis N, McGinnis W. Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. EMBO J. 1994; 13(10):2362-77.

Zhao DB, Côté S, Jähnig F, Haller J, Jäckle H. Zipper encodes a putative integral membrane protein required for normal axon patterning during Drosophila neurogenesis. EMBO J. 1988;7(4):1115-9.

Zhao J. Sumoylation regulates diverse biological processes. Cell Mol Life Sci. 2007;64(23):3017-33.

Zhao, G., Boekhoff-Falk, G., Wilson, B.A., Skeath, J.B. (2007). Linking pattern formation to cell-type specification: Dichaete and Ind directly repress achaete gene expression in the Drosophila CNS. Proc. Natl. Acad. Sci. U.S.A. 104(10): 3847-3852.