OMB and ORG-1:
Homologous *Drosophila* T-box proteins
with functional specificity

The figures on the cover page show the consequences of ectopic *org-1* (left) and *omb* (right) expression on appendage development in *Drosophila*.

Left. Habitus of a young *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] female showing antenna to leg transformations, stunted legs, and vestigial wings (25x magnification).
Right. Habitus of a pharate adult *dpp.blk1*-Gal4; UAS-*omb* fly with an ectopic pair of wings (data taken from Grimm and Pflugfelder, 1996; Grimm, 1997).

eingereicht am: ....

Mitglieder der Promotionskommision:
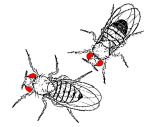
Vorsitzender: ....

Gutachter: Prof. Dr. Gert O. Pflugfelder
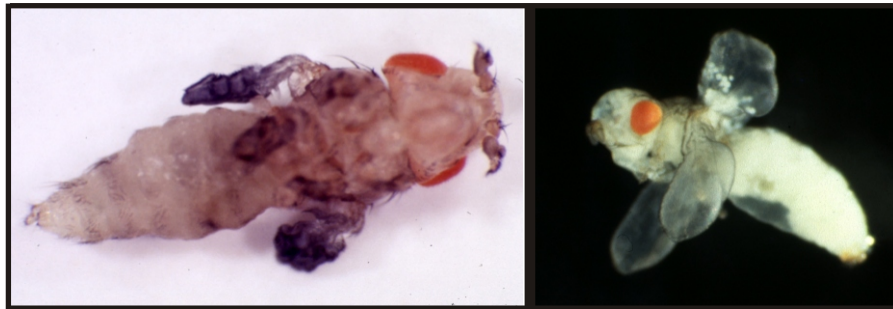
Gutachter: Prof. Dr. Georg Krohne

Tag des Promotionskolloquiums: ....

Doktorurkunde ausgehändigt am: ....

OMB and ORG-1:
Homologous *Drosophila* T-box proteins
with functional specificity

1. Introduction

Members of the T-box gene family encode transcription factors that play key roles during embryonic development and organogenesis of invertebrates and vertebrates. The defining feature of T-box proteins is an about 200 aa large, homologous DNA binding motif, the T domain. Phylogenetic analysis indicates an ancient origin of the T domain in the evolution of the animal kingdom. T-box genes are expressed in dynamic and highly specific patterns during the formation and differentiation of many tissues and organs. Their importance for proper development is highlighted by the dramatic phenotypes of T-box mutant animals. Five T-box genes are associated with clinical syndromes in humans.

Most importantly, ulnar-mammary syndrome and DiGeorge syndrome are caused by haploinsufficiency of *TBX3* and *TBX1*, the putative human orthologs of the *Drosophila* T-box genes *omb* and *org-1* under our investigation. The marked dosage-sensitivity of T-box factors appears to be a consequence of cooperative DNA binding and synergistic effects on target gene regulation. Homeodomain proteins and signaling molecules have been identified both upstream and downstream of T-box proteins indicating that T-box transcription factors are crucial components of developmental programmes closely interconnected with other master regulators of animal development.

The discovery of the T-box

Our knowledge on this gene family emerged from the molecular characterization of two renowned, initially unrelated mutants, mouse Bra*chyury* (*Bra*) (Dobrovolskaïa-Zavadskaïa, 1927; Gluecksohn-Schoenheimer, 1938) and *Drosophila optomotor-blind* (*omb*) (Heisenberg, 1972; Heisenberg and Götz, 1975).

Bra*chyury* (<greek: "*brakhus*", short, "*oura*", tail>), or *Tail* (*T*), has been identified as a semidominant mutation with a short-tail phenotype in $T^{+/-}$ heterozygous mice. Homozygous $T^{-/-}$ embryos lack the notochord (the precursor structure of the spine) and the entire posterior region as a consequence of insufficient mesoderm formation. The failure of the allantois, a mesoderm-derived extraembryonal organ, to connect with the maternal circulation precedes embryonic death at about the

10th day of gestation (Gluecksohn-Schoenheimer, 1944; Herrmann *et al.*, 1990).

In the original *Drosophila omb* mutant *H31*, the absence of a subset of giant neurons in the mutant brain correlates with a defective optomotor-turning response. Although visually competent, *omb*^*H31* flies show impaired reactions to moving stimuli and, thus, are partially motion-blind. Subsequently isolated *omb* null alleles were all late pupal lethal and resulted in pharate adults with severely reduced optic lobes and rudimentary wings (Heisenberg *et al.*, 1978; Bausenwein *et al.*, 1986; Pflugfelder *et al.*, 1992a).

The genes underlying the *Bra* and *omb* mutations were both identified in laborious positional cloning approaches (Herrmann *et al.*, 1990; Pflugfelder *et al.*, 1990, 1992a). Pflugfelder and co-workers soon recognized a high sequence similarity between the central region in OMB and the N-terminal half of Bra and demonstrated that the conserved domain confers general DNA binding affinity to OMB. In addition, a comparable sequence organization in Bra and OMB, such as the distribution of SPXX motifs or charged residues with regard to the homologous domains, was suggestive of a molecular function common to both proteins, possible in transcriptional regulation (Pflugfelder *et al.*, 1992b). Reports on a cell-autonomous function of T, its nuclear localization, and a predicted helical secondary structure of the conserved domain were consistent with this hypothesis (Rashbass *et al.*, 1991, Schulte-Merker *et al.*, 1992; Cunliffe and Smith, 1994; Kispert and Herrmann, 1994). Bra was subsequently shown to be a sequence-specific DNA binding protein that acts as a transcription factor (Kispert and Herrmann, 1993; Kispert *et al.*, 1995a).

As *omb* and *T* appear unlikely to represent functional homologs (inferred from the different protein architecture and different mutant phenotypes) and as, in general, distinct DNA binding motifs are shared by multiple members of larger protein families, Bollag and colleagues set up a PCR screen for the amplification of additional T domains from the mouse genome. Several related T domains were identified and established a new family of transcription factors, the so-called T-box proteins, which was estimated to comprise up to 20 members in the mouse (Bollag *et al.*, 1994).

The evolution of the T-box gene family

During the past decade, numerous T-box genes were identified and cloned in functional studies, homology screens or genomic sequencing approaches. Genome projects found 22 T-box genes in *C. elegans*, 8 T-box genes in *Drosophila*, and at least 18 members in humans, but no T-box sequences in yeast, prokaryotes or plants (Ruvkun and Hobert, 1998; Papaioannou, 2001). Phylogenetic analysis indicates an ancient origin of the T-box gene family at the outset of the metazoan evolution (Agulnik *et al.*, 1996; Papaioannou, 2001). The existence of *Bra1*, a probable *Brachyury* ortholog in the radial-symmetrical polyp *Hydra vulgaris*, a cnidarian, further supports our current view that an ancestral T-box gene must have arisen very early in the metazoan evolution (Technau and Bode, 1999). The topology of the phylogenetic tree subdivides the T-box genes into 5 subfamilies: The *T* subfamily, including *Bra* and its *Drosophila* ortholog *T related gene* (*Trg*), also known as *brachyenteron* (*byn*), the *Tbr-1* subfamily with mouse *T-Brain-1* (*Tbr-1*) and closely related genes, the *Tbx6* subfamily which, among others, contains mouse *T-box6* (*Tbx6*) and three highly similar and linked *Drosophila* genes, the *Tbx2* subfamily containing vertebrate *Tbx2-Tbx5* and *Drosophila omb*, and finally the *Tbx1* subfamily with mammalian *Tbx1*, *Drosophila org-1* and *H15*, and additional close relatives (Figure 1; Papaioannou, 2001). The presence of invertebrate and vertebrate members within individual T-box subgroups demonstrates their existence prior to the separation of the protostomia and deuterostomia lineages about 600 mio years ago and indicates an early expansion of the ancient T-box progenitor gene in the evolution of the animal kingdom.

Within the T-box family, the evolution of the *Tbx2* subfamily has been most intensively studied. This subgroup comprises 4 vertebrate members, *Tbx2-Tbx5* and a single T-box gene in *Drosophila*, *omb*. Phylogenetic analysis shows that *Tbx2* and *Tbx3* as well as *Tbx4* and *Tbx5* are congnate pairs of paralogous genes. These 4 genes were found to form two linked gene pairs in the vertebrates (Agulnik *et al.*, 1996). However, not the most closely related genes are paired with each other, but *Tbx2* with *Tbx4* and *Tbx3* with *Tbx5* (the T domains of Tbx2 and Tbx3 have 95% identity, Tbx4 and Tbx5 94.4% identity, whereas Tbx2 or Tbx3 with Tbx4 or Tbx5 have only 65-67.5% identity). This observation led Agulnik and co-workers to propose a model for the evolution of the *Tbx2* sub-

family, in which an initial tandem duplication of a single ancestral gene by unequal crossing-over formed a two-gene cluster that later duplicated and dispersed to different chromosomal locations (Agulnik *et al.*, 1996; Papaioannou, 2001). The recognition of extended regions of paralogy around the *Tbx* gene clusters supports this hypothesis. Accordingly, the duplication of the *Tbx2/3* and *Tbx4/5* gene cluster occurred *en masse* prior to the divergence between bony fish and tetrapods around 400 million years ago (Ruvinsky and Silver, 1997).

Based on evolutionary distance, the tandem duplication of the primordial *omb/Tbx2/3/4/5* gene was estimated to have occurred in a common ancestor of arthropods and vertebrates (Agulnik *et al.*, 1996). If this scenario holds true, however, the duplicated gene must have been lost along the *Drosophila* lineage subsequent to the separation from the vertebrates. Conversely, the primary tandem duplication occurred in the vertebrate lineage after the divergence from the arthropods. The ancestral gene would then have evolved into the present *omb* in the *Drosophila* lineage, while in the vertebrate line, *Tbx2/3* remained structurally and functionally conserved, thereby relieving the novel *Tbx4/5* gene from selective pressure, so that it could rapidly evolve to acquire new functions (Figure 2). In this view, *omb* represents the putative *Drosophila* ortholog of the vertebrate *Tbx2* and *Tbx3* genes. The maintenance of the *Tbx* gene clusters over a long evolutionary distance implies a selective advantage of this genomic arrangement. It is conceivable that *cis* regulatory elements exist that work on both cluster members.

Expression and function of T-box genes

T-box genes are characteristically expressed in dynamic and specific patterns during embryogenesis and organogenesis. Each T-box gene has a unique expression profile, although overlapping expression domains exist especially among close relatives, suggesting partially redundant functions for T-box genes. A clear preponderance of T-box genes is expressed in mesodermal tissues (Papaioannou, 2001; Smith, 2001), indicating that T domain transcription factors are of special importance for the induction and differentiation of mesoderm (Smith, 2001).

An additional feature of T-box genes is a marked conservation of expression patterns among paralogs or orthologs from different species.
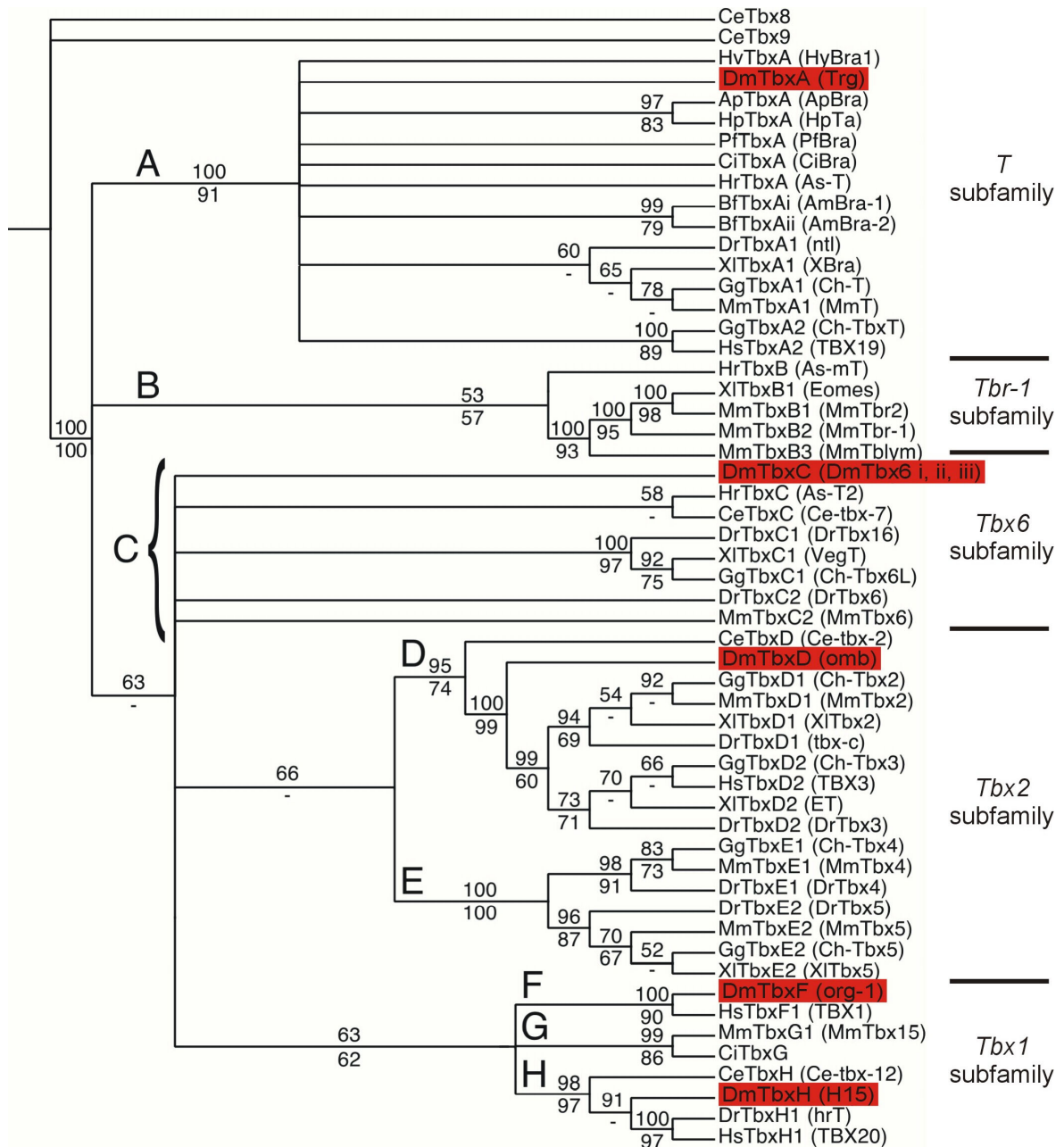
Figure 1. Phylogenetic tree of the T-box gene family.
The 5 subfamilies of the T-box gene family are indicated. Drosophila T-box genes are highlighted in red.
This phylogenetic analysis and the tree construction were made by Kevin J Peterson and Albert Erives, Pasadena, CA, USA.

The paralogous *Tbx2-5* genes, for instance, display strong similarities in their overall expression patterns (Chapman *et al.*, 1996). The spatiotemporal expression is especially similar between the cognate gene pairs *Tbx2/ Tbx3* and *Tbx4/ Tbx5*. In the mouse embryo, *Tbx2* and *Tbx3* are both expressed in the epithelium of the inner ear, the dorsal region of the retina, in the CNS, in the developing limb buds, and in the body wall. Likewise, the expression patterns of the paralogs *Tbx4* and *Tbx5* strikingly resemble each other in many areas. Tran-scripts of both genes are detectable in the allantois, the developing heart, lung mesenchyme, the body wall of the thorax. However, some notable differences exist in their expressions. *Tbx5*, but not *Tbx4*, is transcribed in the optic vessicle, and, most interestingly, *Tbx5* is exclusively expressed throughout the forelimb bud, whereas *Tbx4* mRNA is predominantly found in the hindlimb bud (Chapman *et al.*, 1996; Gibson-Brown *et al.*, 1996).
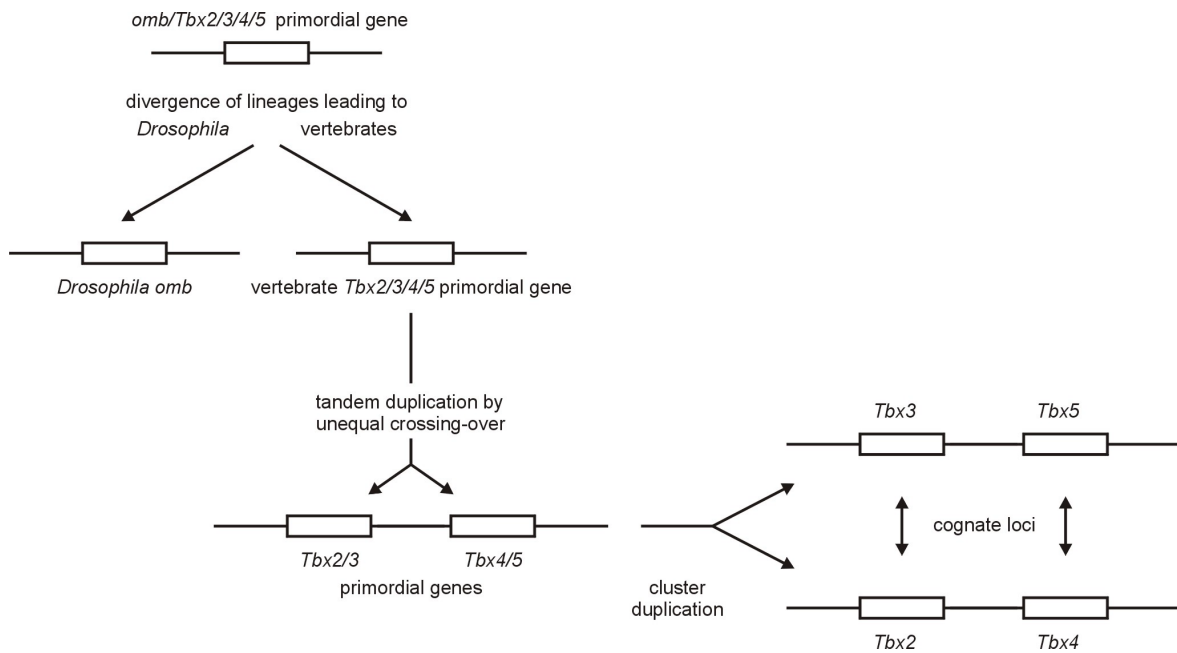
Figure 2. A model for the evolution of the *Tbx2* subfamily.
The model and this figure are slightly modified after Agulnik *et al*., 1996. Details are described in the text.

Tbx4 and *Tbx5*: the control of limb type identity

The complementary expressions of *Tbx5* and *Tbx4* throughout the developing forelimbs and hindlimbs, respectively, suggested that these genes might be involved in the specification of limb-type identity (Gibson-Brown *et al*., 1996). Their roles in the differential specification of fore- versus hindlimb identity has been studied in the chick model, where ectopic limbs can easily be induced by exogeneous FGF in the flank of the embryo (Gibson-Brown *et al*., 1998a; Logan *et al*., 1998; Isaac *et al*., 1998; Ohuchi *et al*., 1998). The identity of the ectopic limb thereby depends on the position along the rostro-caudal axis, at which a FGF-source is provided. FGF induces ectopic wings in vicinity of the endogenous wing field and promotes growth of ectopic legs when supplied caudally. Mosaic limbs with both wing and leg structures develop from FGF-soaked beads that are implanted in the middle of the flank. The expression of *Tbx5* and *Tbx4* thereby strictly correlated with wing and leg identity, respectively. Grafting experiments, in which wing and leg bud tissue are reciprocally transplanted, revealed that transplants retain the identity of their donor tissue as well as the expression of the appropriate limb-specific T-box gene (Gibson-Brown *et al*., 1998a;

Logan *et al*., 1998; Isaac *et al*., 1998; Ohuchi *et al*., 1998).
Misexpression of *Tbx5* in the presumptive hindlimb region causes a partial transformation of the leg into wing, resulting in wing/ leg mosaic limbs. Conversely, ectopic *Tbx4* in the developing wing promotes the growth of leg-like structures (Rodriguez-Esteban *et al*., 1999; Takeuchi *et al*., 1999; Logan and Tabin, 1999). *Tbx4* expression is activated by Pitx1, a paired-type homeodomain transcription factor, and is repressed by Tbx5 (Logan and Tabin, 1999). Thus, *Tbx4* and *Tbx5* have antagonistic functions for the selection of distinct limb identities with *Tbx5* responsible for wing identity and *Tbx4* responsible for leg identity.
Recent work by Saito and colleagues could demonstrate that specification and determination of the limb-type identities precede the onset of *Tbx4* and *Tbx5* transcription. Therefore, expression of *Tbx4* and *Tbx5* does not specify or determine limb identity, but mediates the differentiation of the distinct limb types (Saito *et al*., 2002).
The roles of *Tbx4* and *Tbx5* in limb development appear to be conserved among all vertebrates, since both genes are comparably expressed in the limb buds of mouse, chick, newt, and zebrafish (Gibson-Brown *et al*., 1996; Simon *et al*., 1997; Gibson-Brown *et al*., 1998a; Logan *et al*., 1998; Isaac *et al*., 1998; Ohuchi *et al*., 1998; Tamura *et al*., 1999).

Tbx2-5 expression in mouse, chick and *Xenopus* is similar beyond the developing limbs too, although some species-specific temporal or spacial variations exist (Chapman *et al.*, 1996; Gibson-Brown *et al.*, 1998b; Takabatake *et al.*, 2000). Even *omb* in *Drosophila* shares common areas of expression with its putative vertebrate orthologs, such as the developing eyes, wings and legs (Grimm and Pflugfelder, 1996; Grimm, 1997; Brook and Cohen, 1996; Chao *et al.*, in preparation).

Brachyury: posterior mesoderm formation and notochord differentiation

The Bra*chyury* subfamily provides a further example for conservation between T-box orthologs of different species. *Bra* was first described as a haploinsufficient mutant with shortened tails, while homozygous *Bra* embryos die during gestation and lack the notochord and all somites posterior to somite 7 (Dobrovolskaïa-Zavadskaïa, 1927; Gluecksohn-Schoenheimer, 1938). *Bra* transcripts become first detectable in the primitive streak, a morphological stripe that extends in anteroposterior direction along the two-layered early mouse embryo. Cells from the inner layer, the epiblast, migrate in the primitive streak region between the epiblast and the upper layer, the visceral endoderm, to become mesoderm. *Bra* expression is also evident in the nascent mesoderm surrounding the primitive streak and the node, from where axial mesoderm and, subsequently, the notochord derive. It fades in differentiating mesodermal cells, except in the notocord and the tail bud, where *Bra* expression remains high through late gestation (Wilkinson *et al.*, 1990; Wilson *et al.*, 1993; Kispert and Herrmann, 1994). Thus, Bra*chyury* expression is seen in all tissues affected in the *Bra* mutant.

Bra*chyury* orthologs have been cloned in several other vertebrate species including *Xenopus*, zebrafish, and chicken, and their embryonic expression patterns were found to be very similar to that of mouse *Bra* (Smith *et al.*, 1991; Schulte-Merker *et al.*, 1992; Kispert *et al.*, 1995b). The function of *Bra* appears to be conserved as well. The zebrafish mutant *no tail* (*ntl*) resembles mouse *Bra* embryos, as it does not form enough posterior mesoderm and lacks the notochord and the caudal region. *ntl* is caused by mutations in the zebrafish Bra*chyury* ortholog *zfT* (Schulte-Merker *et al.*, 1994). Experiments in *Xenopus* embryos demonstrated that *Xenopus Bra* (*XBra*) is both necessary and sufficient for mesoderm formation (Cunliffe and Smith, 1992). Interference with *XBra* function produces *Xenopus* embryos with an absent notochord and posterior region (Conlon *et al.*, 1996).

The importance of T-box genes in developmental processes can be best evaluated from the consequences of loss-of-function situations on normal development. A number of T-box gene mutations could be identified in genetic studies of species ranging from *C. elegans* to humans. Without any exeption, all described T-box mutants revealed profound phenotypes implicating that members of the T-box family play crucial roles in the regulation of embryonic development. T-box gene function has been found to be particularly required in cell fate assignment and cell differentiation, morphogenic movements, inductive tissue interactions, and organogenesis.

In *C. elegans mab9* mutants, the cell fate transformation of two blast cells leads to defects in hindgut and male-tail development (Woollard and Hodgkin, 2000). *Drosophila brachyenteron* encodes an essential function with a similar role in gut formation (Kispert *et al.*, 1994; Singer *et al.*, 1996).

More T-box genes in formation and differentiation of mesoderm

spadetail (*spt*) is a second known zebrafish T-box mutant besides *ntl* (Kimmel *et al.*, 1989; Griffin *et al.*, 1998). It results from a mutated *tbx16* gene, and manifests, like *ntl*, mesoderm deficiencies. *spt* embryos lack trunk somites, but are relatively normal in notochord and tail development (Griffin *et al.*, 1998). Thus, *spt* and *ntl* have complementory areas of function in mesoderm formation, with *spt* predominantly regulating trunk mesoderm and *ntl* mainly controlling notochord and tail mesoderm, although *spt* and *ntl* are both expressed in trunk and tail mesoderm progenitors (Griffin *et al.*, 1998). *spt* is the zebrafish ortholog of *VegT/Xombi/Antipodean/Brat* that in *Xenopus* functions in mesoderm formation as well (Zhang and King, 1996; Lustig *et al.*, 1996; Stennard *et al.*, 1996; Horb and Thomsen, 1997).

Eomesodermin (*Eomes*) and *T-box 6* (*Tbx6*) are additional T-box genes with essential functions in mesoderm formation and specification.
Eomesodermin (<greek: "*eos*", dawn>) was first cloned in *Xenopus* and named according to its early key function in gastrulation and mesoderm differentiation (Ryan *et al.*, 1996). *Eomes* is among the first genes that are transcribed in panmesoderm in response to signals from vegetal cells. Its expression occurs in a ventral-to-dorsal gradient of increasing Eomes concentration, which defines the differential activation of a spectrum of mesodermal genes mediating mesodermal differentiation. Overexpression of *Eomes* dorsalizes the ventral or lat-

eral mesoderm; *Eomes* misexpression within animal caps can induce mesodermal structures such as notochord or muscle, while inhibition of *Eomes* function halts the development of *Xenopus* embryos at the onset of the gastrulation (Ryan *et al.*, 1996). Hence, *Xenopus Eomes* is both necessary and sufficient for mesoderm formation and the determination of mesodermal cell fate. In mouse *Eomes*$^{-/-}$ embryos, prospective mesodermal cells in the pregastrulation epiblast fail to migrate into the primitive streak. As a consequence, *Eomes*$^{-/-}$ individuals arrest in the blastocyst stage as unorganized embryos prior to the formation of the mesodermal germ layer. This suggests that *Eomes* has a conserved function required for morphogenetic movements underlying gastrulation (Russ *et al.*, 2000). Deficient morphogenic cell movement has also been implicated to cause the failure of mesoderm formation in the *Bra/ntl* and *spadetail* mutants (Wilson *et al.*, 1995; Griffin *et al.*, 1998; Conlon and Smith, 1999; Smith, 2001).

Lack of another murine T-box gene, *Tbx6*, results in the differentiation of posterior paraxial tissue into ectopic neural tubes instead of somites (somites are segmental units of mesoderm occuring in pairs along the notochord) (Chapman and Papaioannou, 1998). In contrast to the mutants described above, however, impaired cellular movements do not account for the mesodermal defect in *Tbx6*$^{-/-}$ embryos, since the prospective paraxial cells ingress into the primitive streak and properly migrate laterally during gastrulation. This observation indicates that the differentiation of posterior mesoderm requires *Tbx6* independently of morphogenic cell migration. Neuralization of presumptive mesoderm has also been recognized by the functional inhibition of *XBra* or *Eomes* in *Xenopus* embryos (Rao, 1994; Ryan *et al.*, 1996). Hence, the assignment of mesodermal versus neural cell fate appears to be a function common to several T-box factors.

A total of six T-box genes mouse mutants have been described so far: *Bra*, *Tbx1*, *Tbx5*, *Tbx6*, *Eomes*, and *Tbr-1*. Mutant analysis revealed that they all encode essential functions in areas, where these genes are normally expressed. Different from most vertebrate T-box genes, *Tbr-1* is predominantly expressed in postmitotic cells of the CNS, where *Tbr-1* transcripts are mainly restricted to the cerebral cortex. Embryonic *Tbr-1* expression is also seen in cells of the cerebellum, the skin, and the epithelium of the tongue (Bulfone *et al.*, 1995). *Tbr-1*$^{-/-}$ mice develop smaller brains, have small olfactory bulbs and lack olfactory tracts that connect the olfactory bulb with the primary olfactory cortex. *Tbr-1* mutant mice die postnatally because of a failure of nursing. Appearently, they cannot smell and rec-

ognize their mothers (Bulfone *et al.*, 1998). Mutations in *Tbx1* and *Tbx5* will be discussed below.

Two T-box genes have been associated with terminal cell differentiation, *Tbet/TBX21* and *Tpit/TBX19* (Szabo *et al.*, 2000; Lamolet *et al.*, 2001). *Tbet* expression has been found to direct naive T helper cells into the differentiation pathway of the Th1 cell lineage by activating the Th1 marker *INFγ* and repressing the opposing Th2 differentiation programme (Szabo *et al.*, 2000). *Tbet* is the mouse ortholog of human *TBX21*, a member of the *Tbr-1* subfamily.
Tpit was identified as a transcription factor required for the activation of the *pro-opiomelanocortin* (*POMC*) gene (Lamolet *et al.*, 2001). *Tpit* expression itself is highly restricted to two *POMC* expressing cell types of the pituitary gland, the ACTH-producing corticothrophs and the α-MSH producing melanotrophs. Ectopic expression of *Tpit* in the rostral tip of the early pituitary was sufficient to initiate POMC cell differentiation *in vivo*. Furthermore, mutations in the human *TPIT/TBX19* gene were discovered in patients with deficiency of pituitary adrenocorticotropic hormone (ACTH) and secondary adrenal insufficiency (Lamolet *et al.*, 2001). *TPIT/TBX19* belongs to the *T* subfamily.

TBX mutations and human syndromes

The great medical relevance of T-box genes is evident from the association of four further family members with human developmental syndromes. Mutations in human *TBX3* and *TBX5* are responsible for ulnar-mammary syndrome and Holt-Oram syndrome, respectively (Bamshad *et al.*, 1997; Li *et al.*, 1997; Basson *et al.*, 1997).
Ulnar-mammary syndrome (UMS) is a rare pleiotropic disorder affecting limb, apocrine gland, tooth, hair, and genital development (Schinzel, 1987; Bamshad *et al.*, 1997). UMS is caused by haploinsufficiency of *TBX3* and follows an autosomal dominant inheritance. The expressivity of the various mutant foci is highly variable, even among UMS patients carrying identical mutant alleles. The characteristic feature of UMS is a variable malformation of posterior structures of upper limb that derive from the ulnar ray. The limb phenotype is frequently associated with aplasia or hypoplasia of the breast, a lack of axillary hair, and, less commonly, with ectopic or missing canines and genital hypoplasia, indicating that TBX3 participates in inductive processes of ectoderm and mesoderm. The pleiotropic defects in UMS are concordant with

sites of *Tbx3* expression in mouse and chick embryos that include mammary bud, jaw mesenchyme and genital papilla (Chapman *et al.*, 1996; Gibson-Brown *et al.*, 1998b). During limb development, *Tbx3* mRNA is first abundantly detected in the posterior mesenchyme of both limb buds. At later stages, *Tbx3* transcripts are also seen at the anterior margins, albeit less extended distally than at the posterior margin (Gibson-Brown *et al.*, 1996; Gibson-Brown *et al.*, 1998a).

Holt-Oram syndrome (HOS) belongs to a group of developmental disorders called heart-hand syndromes that are characterized by upper limb malformations and heart defects. HOS is a rare haploinsufficiency disorder caused by mutations in human *TBX5*. It occurs with an incidence of 1/100.000 live births. As seen in UMS, HOS expressivity is highly variable, even among affected family members segregating the identical mutation. The skeletal abnormalities of the forelimb, for instance, can range from clinodactyly (<greek: "klinein", to slope, "daktulos", finger>) to severe reduction deformities (phocomelia). Heart abnormalities commonly include atrial and/ or ventricular septal defects. Some HOS patients additionally suffer from an absent muscle pectoralis major or from ocular defects, consistent with the expression of vertebrate *Tbx5* in the devolping heart, forelimb bud, body wall and optic vessicle (Li *et al.,* 1997; Chapman *et al.,* 1996; Gibson-Brown *et al.,* 1998b).

Mutations in *TBX22* have recently been found to be responsible for X-linked cleft palate with ankyloglossia (CPX; <greek: "*ankulosis*", stiffening of the joints, "*glossa*", tongue>) (Braybrook *et al.*, 2001). Haploinsufficiency of *TBX22* with variable expressivity and penetrance underlies CPX, as *TBX22* mutations manifest in all hemizyous males, while both affected and unaffected carrier heterozygous females are observed (Braybrook *et al.*, 2001).

DiGeorge syndrome (DGS) and Velocardiofacial syndrome (VCFS) belong to a number of dominantly inherited disorders all associated with deletions or translocations involving human chromosome 22q11 (Scambler, 2000). More than 80 distinct birth defects or malformations have been associated with 22q11 deletions, occuring in many combinations and a with a wide range of severity. Clinical features of the various described 22q11 deletion syndromes largely overlap, suggesting that the different diagnoses may result from variable expressivity of a common genetic defect. The spectrum of DGS/VCFS phenotypes includes defects in the outflow tract of the heart, branchial arch arter-

ies defect, aplasia/hypoplasia of thymus and parathyroid gland, craniofacial dysmorphism, and neuropsychicatric problems (Scambler, 2000).

DGS/VCFS patients typically have deletions of about 3 Mb. The overlap of such deletions defines an approximately 750 kb large DGS chromosomal region. However, since DGS/VCFS patients with atypical deletions have also been described, the gene(s) underlying this haploinsufficiency syndrome remained elusive, although the complete DNA sequence of the DGS region was determined at an early stage of the human genome sequencing project (Kirsch *et al.*, 2000). Subsequent work on a mouse model of DGS turned out to be crucial for a molecular understanding of DGS/VCFS.

A targeted 1 Mb deletion, *Df(16)1*, of the DGS chromosomal region in the mouse genome resulted in haploinsufficient mice with cardiovascular defects similar to those of DGS patients (Lindsay *et al.*, 1999). Two research groups subsequently used sets of nested deletions and bacterial or P1 artificial chromosome (BAC or PAC) transgenic mice to map the responsible gene within the deleted interval (Lindsay *et al.*, 2001; Merscher *et al.*, 2001). Among a few candidate genes, *Tbx1*, the putative vertebrate ortholog of *org-1*, appeared most promising concerning its expression pattern and its homology to haploinsufficient genes. Human *TBX1* has previously been mapped into the DGS chromosomal region (Chieffo *et al.*, 1997; Porsch *et al.*, 1998; this work). Indeed, both research teams and a third laboratory simultaneously showed that heterozygous *Tbx1* $^{+/-}$ mice develop aortic arch abnormalities which mimic one of the major phenotypes of the human syndrome (Lindsay *et al.*, 2001; Merscher *et al.*, 2001; Jerome and Papaioannou, 2001).

Homozygous *Tbx1* $^{-/-}$ mutant mice fail to inflate their lungs and suffocate as neonates. When investigated earlier during embryogenesis, *Tbx1* $^{-/-}$ individuals show a broad spectrum of phenotypes commonly associated with DGS/VCFS, however, with stronger expressivity: aortic arch and cardiac outflow tract defects, cleft palate, abnormal middle-ear ossicles and mis-shaped or absent external ears, weak cartilages of the neck, aplasia of the thymus and parathyroid gland. These defects could be traced back to the abnormal development of pharyngeal arches and pouches, head mesenchyme, and otic vesicles, areas, where *Tbx1* is normally expressed (Jerome and Papaioannou, 2001; Chapman *et al.*, 1996). Thus, *TBX1* appears to be the key gene in the etiology of DGS/VCFS. However, it seems likely that additional linked genes contribute to the 22q11 deletion syndrome, because a significant number of patients with clinical suspicion of DGS/VCFS but without detectable

deletions did not reveal mutations within the coding region of *TBX1*, and because DGS/VCFS patients with deletions outside of the *TBX1* locus were described (Lindsay *et al.*, 2001; Jerome and Papaio-annou, 2001).

The finding that *TBX2*, a transcriptional repressor, is capable of downregulating the tumor suppressor gene *Cdnk2a (p19ARF)* and that *TBX2* is amplified in a subset of human breast cancers further underlines the importance of T-box genes in development and disease (Jacobs *et al.*, 2000).

Dosage-sensitivity of T-box factors

As described above, haploinsufficiency of *Bra*, *TBX1*, *TBX3*, *TBX5*, and *TBX22* produce dominant phenotypes implying that functional levels of T-box genes are critical for normal development. Acute dosage-sensitivity of T-box genes is evident not only in situations with reduced gene dose, but also in cases, in which the level of a given T-box gene is elevated. BAC transgenic mice containing four human genes including *TBX1* have cardiac and conotruncal defects, thymus hypoplasia, and ear defects similar to those of *Tbx1* [+/-] mice and/ or DGS/VCFS patients (Merscher *et al.*, 2001; Funke *et al.*, 2001). Humans with a chromosomal duplication of 12q24, a region involving *TBX3* and *TBX5*, have congenital anomalies with HOS features (Melnyk *et al.*, 1981; Vaughan and Basson, 2001; Hatcher and Basson, 2001). Furthermore, overexpression of *Tbx5* in the heart of mouse and chick embryos produced animals with heart defects similar to those of HOS (Liberatore *et al.*, 2000; Hatcher *et al.*, 2001); overexpression of *Tbx5* and *Tbx4* in their endogenous domains during chick limb development, the forelimb and hindlimb bud, respectively, leads to truncated limbs (Rodriguez-Esteban *et al.*, 1999). Thus, both too little and too much T-box protein appears to be deleterious for proper development and may cause similar phenotypes.

It is important to note that syndromes underlying *TBX* haploinsufficiency are manifested only in some, but not all tissues in which a given T-box gene is expressed. Furthermore, the phenotypic severity frequently differs among distinct symptoms even within affected individuals (Li *et al.*, 1997; Basson *et al.*, 1997; Bamshad *et al.*, 1999; Scambler, 2000). These observations imply a tight dosage-sensitivity of T-box factors in a tissue-specific manner.

Two plausible mechanisms may account for this phenomenon, functional redundancy and concen-

tration-dependent target gene regulation, and arguments for both exist. In UMS, haploinsufficiency of *TBX3* leads to upper limb defects that are restricted to the posterior and distal region. *Tbx3* expression in mouse and chick embryos, however, is seen at both posterior and anterior margin of the developing limb buds, with the posterior *Tbx3* expression being assymmetrically extended distally. *Tbx2* is similarly expressed at both limb bud margins, exept that the distal posterior *Tbx2* expression is absent. It is therefore conceivable that reduced level of normal *TBX3* may be compensated by a closely related factor in areas of overlapping expressions. According to that, functional redundancy with *TBX2* might suppress a phenotypic expression of *TBX3* haploinsufficiency in common regions of expression.

In *Xenopus* embryos, *Eomes* is expressed with increasing strength along the ventral to dorsal axis and regulates target genes in a concentration-dependent way (Ryan *et al.*, 1996). Similarly, different concentrations of *XBra* produce different mesodermal subtypes: low concentrations induce the formation of ventral mesoderm, whereas higher concentrations cause dorsal mesoderm (Cunliffe and Smith, 1992; O'Reilly *et al.*, 1995).

T protein is required in increasing quantities along the rostrocaudal axis during posterior axis formation (Stott *et al.*, 1993; Schulte-Merker *et al.*, 1992, 1994). Taken together, these data suggest that the regulation of downstream target genes in different tissues may require different concentrations of a T-box protein. In this view, it is conceivable, that a haploinsufficient organism still produces adequate T-box protein to maintain some functions, but insufficient for others. Cooperative binding to multiple promoter elements and synergistic transcriptional regulation provide us a molecular basis to understand, how concentration-dependent target gene regulation may be established by T-box transcription factors (Bruneau *et al.*, 2001; Kusch *et al.*, 2002).

T-box proteins: DNA binding and transcriptional regulation

DNA binding characteristics

A function of the T domain in DNA binding has originally been identified in OMB, in which the central region possesses general DNA binding affinity and shows homology to the N-terminal half of mouse Brachyury (Pflugfelder *et al.*, 1992b). The N-terminal 229 aa of Brachyury/T were subsequently shown to be necessary and sufficient for

sequence-specific DNA binding (Kispert and Herrmann, 1993).

The DNA sequence preferentially bound by the T protein *in vitro* was determined in a PCR based binding site selection experiment. Sequences that were isolated from a pool of random oligonucleotides defined a 20 bp nearly palindromic TG/CACACCT * AGGTGTGAAATT consensus sequence with an invariant AGGTG core sequence (Kispert and Herrmann, 1993). Full-length T protein also binds a perfect consensus palindrome, the T site, but not to a single half site, of which at least two copies are required to allow binding *in vitro* to occur (Kispert and Herrmann, 1993; Kispert *et al.*, 1995a). Conflicting data exist for Bra binding the T site either as a monomer or dimer (Kispert and Herrmann, 1993; Papapetrou *et al.*, 1997, Grimm and Pflugfelder, in prep.).

The X-ray crystallographic structure of the XBra T domain in complex with its target DNA revealed a new mode of sequence-specific protein-DNA interaction. The C-terminal helix of the T domain is deeply embedded into the minor groove of the palindromic T site and contacts specific bases in the minor groove without bending the DNA. Interactions with the major groove take place as well. The X-ray structure showed that the isolated T domain forms a dimer upon DNA recognition, although the protein is a monomer in solution (Müller and Herrmann, 1997).

A cyclic *in vitro* binding site selection has also been carried out with OMB (Grimm and Pflugfelder, in prep.). The compilation of selected sequences identified the consensus sequence AGGTGTGA, which corresponds to a half site of the palindromic Bra target sequence. A second, generally imperfect, half site was frequently co-selected by OMB. Both half sites preferentially formed everted palindromes with a central 4 bp spacer or occurred in variably spaced tandem repeats. The palindromic arrangement of half sites as in the T site was not obtained. OMB is capable of binding the T site, as do all other T-box proteins hitherto tested including T, TBX1, TBX2, TBX3, TBX5, TBX6 and Tbr-1 (Kispert and Herrmann, 1993; Papapetrou *et al.*, 1997; Grimm and Pflugfelder, in prep.; Sinha *et al.*, 2000; Carreira *et al.*, 1998; Carlson *et al.*, 2001; Bruneau *et al.*, 2001; Ghosh *et al.*, 2001; Papapetrou *et al.*, 1999; Hsueh *et al.*, 2000). However, T-box proteins revealed differences in the recognition of a single half site *in vitro*, with OMB, TBX2, Tbx5 and the Bra T domain being able to bind to, while full-length Bra and TBX1 can not (Kispert and Herrmann, 1993; Grimm and Pflugfelder, in prep.; Sinha *et al.*, 2000; Bruneau *et al.*, 2001; Ghosh *et al.*, 2001). These data suggest that most, if not all

T-box proteins have very similar *in vitro* DNA sequence specificities but differ in half site recognition, in the preference for certain arrangements of half sites, and in dimerization characteristics (Kispert and Herrmann, 1993; Grimm and Pflugfelder, in prep.; Sinha *et al.*, 2000).

T-box target genes: Evidence for cooperative DNA binding and synergism

How are "perfect" *in vitro* binding sequences related to genuine T-box factor binding elements (TBEs) *in vivo*? An increasing number of downstream targets of T-box proteins has been identified and their promoter sequences have been molecularly analyzed. Data of these studies are summarized in Table 1. It is evident from a survey of known target gene promoters that natural TBEs resemble the *in vitro* selected binding sites in the way, that they usually contain multiple, heterogeneously arranged T half sites. Thereby, the endogenous half sites predominantly occur in variably spaced direct repeats and, less frequently, in imperfect palindromic arrangements.

Two types of half sites can be distinguished according to sequence, binding affinity, and effect on reporter gene activation. High affinity binding sites (strong sites, type A) are very similar to the half-site consensus sequence, are recognized by T-box protein *in vitro*, and, at least in several copies, can direct reporter gene expression. Low affinity binding sites (weak sites, type B), on the other hand, have substantial deviations from the consensus, are not bound *in vitro*, and are insufficient for reporter gene activation (Casey *et al.*, 1998; Kusch *et al.*, 2002). Weak type B binding sites, however, may strongly synergize with type A sites and are required for efficient reporter gene transactivation. For example, the *Xenopus eFGF* promoter contains a perfect half site located 936 nucleotides upstream of the transcription start site and a second, related TBE located 123 nucleotides downstream of the transcription start site. Although only the distal half site, but not the proximal site, can be bound by XBra *in vitro*, both sites are required for the *eFGF* promoter induction by XBra and the deletion of either site results in an equal decrease of reporter gene expression (Casey *et al.*, 1998). Likewise, several Byn binding regions with a total of 15 half sites were determined in DNase I protection experiments within the 5' regulatory region of the hindgut specific promoter of *orthopedia* (*otp*), a downstream target of Byn in *Drosophila*. The seven most distal Byn sites all deviate from the consensus and thus appear to be type B sites. A distal fragment of the *otp* promoter containing the seven half sites proved to be insufficient for Byn to induce luciferase

expression. However, in conjunction with a central fragment containing several type A sites, the reporter gene expression was markedly increased beyond the level that has been obtained with the central promoter fragment alone (Kusch *et al.*, 2002). Moreover, transcriptional activation assays with synthetic promoter constructs revealed that type B sites exhibit clear synergistic effects with type A sites, when combined in antiparallel orientation suggesting that Byn molecules cooperate in DNA binding to activate target gene transcription. Cooperative DNA binding has also been described for Tbx5 at the *ANF* promoter (Bruneau *et al.*, 2001).

The finding that (some) half sites are bound by T-box proteins implys that the half site, rather than a palindrome, represents the functional DNA binding unit of the T domain. In this view, it is intriguing, why most of the identified T-box target genes contain at least two TBEs in their promoter regions (see Table 1). Only two promoters have hitherto been identified with a single T-box binding site each: the *pro-opiomelanocortin* (*POMC*) promoter regulated by Tpit and the *natriuretic peptide precursor type A* (*Nppa*) promoter activated by Tbx5 (Lamolet *et al.*, 2001; Hiroi *et al.*, 2001). Interestingly, in both promoters, the single half site is juxtaposed to a binding site for the homeodomain (HD) transcription factors Pitx1 or Nkx2-5, respectively, and Tpit and Pitx1 were found to cooperatively activate the *POMC* promoter, as do Tbx5 and Nkx2-5 the *Nppa* promoter (Lamolet *et al.*, 2001; Hiroi *et al.*, 2001). Tpit and Pitx1 and Tbx5 and Nkx2-5 bind to their contiguous target sites in the *POMC* and *Nppa* promoters as heterodimers in tandem, forming ternary protein-protein-DNA complexes (Lamolet *et al.*, 2001; Hiroi *et al.*, 2001). Similarly, Tbx5 and Nkx2-5 also bind to neighboring target sites in the *atrial natriuretic factor* (*ANF*) promoter and show synergistic transactivation (Bruneau *et al.*, 2001). Pull-down assays demonstrated that Tbx5 physically interacts with Nkx2-5 (Hiroi *et al.*, 2001; Bruneau *et al.*, 2001). The HD of Nkx2-5 is thereby necessary and sufficient for this interaction, while the N-terminal 90 aa including 28 aa of the T domain are required in Tbx5 (Hiroi *et al.*, 2001; Bruneau *et al.*, 2001). Furthermore, the HDs of Nkx2-5 and Pitx1 proved to be both necessary and sufficient for synergistic transactivation with their T-box protein partners, although additional parts of Nkx2-5 and Pitx1 outside the HD are required for full synergism.
Since the HD is a conserved DNA binding motif with little effect on transcriptional regulation *per se* (Kornberg, 1993), the seen synergism presumably results from cooperative DNA binding. Support for

this idea comes from the observation that Tbx5 does not synergize with Nkx2-5 on a Nkx2-5 specific promoter without Tbx5 binding sites (Hiroi *et al.*, 2001), suggesting that interactions between HD and T-box factors result in synergistic effects only, when both proteins are directly bound on DNA. Conceivably, the contacts with HD proteins thereby enhance the affinity of bound T-box factors to their DNA target sites. It has been shown previously that the strength of T protein-DNA complexes is strongly increased by the addition of T antibodies (Kispert and Herrmann, 1993).
Byn proteins also appear to make physical contacts *inter se* when bound to various Byn sites in the *otp* promoter (Kusch *et al.*, 2002). The Byn-Byn interaction occurs via a central region of the T domain (Kusch *et al.*, 2002). This region contains a stretch of aa that forms most of the dimer interface of two XBra T domains when bound to the T site (Müller and Herrmann, 1997).
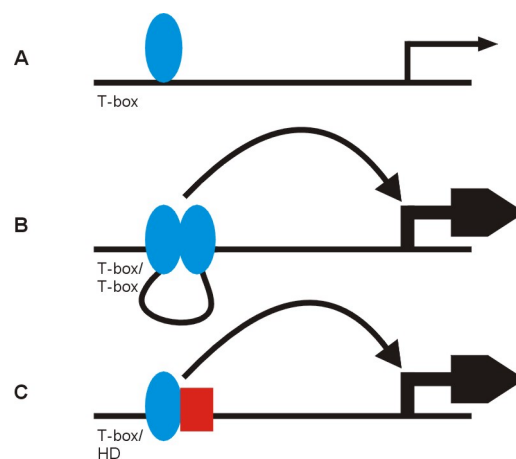
Figure 3. A model for the cooperative DNA binding and synergism between T-box proteins and/ or HD transcription factors.
A. A promoter with a single T half site is only weakly bound by a T-box factor (blue) and has basal activtity. B. Two (or multiple) appropriately spaced and oriented T half sites are cooperatively bound by T-box proteins that interact with each other. The protein-protein interactions enhance individual T-box protein-DNA contacts and lead to synergistic effects on the target promoter (activation or repression). C. Cooperative binding and synergism may also result from contacts between a T-box protein and a HD factor (red) bound in tandem to contiguous T-box and HD binding sites.

Taken together, these data suggest that the binding of T-box proteins to natural target sites is generally accompanied by protein-protein interactions which stabilize individual T-box protein-DNA contacts. Cooperative DNA binding of T-box proteins is

thereby established on most target promoters via multiple half sites *inter se*, but can also occur in combination with HD transcription factors on juxtaposed HD and T-box binding sites. All 13 analyzed T-box target genes followed this promoter architecture suggesting that DNA binding cooperativity is obligatory on T-box target gene regulation and is responsible for synergism (Figure 3). This model may account for how T-box transcription factors achieve tissue-specific target gene expression in a dose-dependent way.

T-box proteins: Transcriptional activators and repressors

T-box proteins are specific transcription factors that regulate the expression of downstream target genes. Both transcriptional activators and repressors were observed among T-box factors, with a clear preponderance of transactivators. A few molecular studies exist in which regulatory domains of T-box transcription factors have been characterized. Analysis of T deletion proteins or Gal4-T fusion proteins in in reporter gene assays revealed a complex domain architecture within the mouse T protein. Two pairs of transactivation and repression domains are alternately located within the C-terminal half of T protein that, in overall, acts as a transcriptional activator (Kispert *et al.*, 1995a). Most other T-box proteins function as transcriptional activators as well (see Table 1). However, a dominant repression domain was found in *Xenopus* ET, its human ortholog TBX3 and in the closely related TBX2, making these T-box proteins to behave as transcriptional repressors (He *et al.*, 1999; Carlson *et al.*, 2001; Sinha *et al.*, 2000; Paxton *et al.*, 2002). In TBX2 and TBX3, as in T, several effector domains exist outside of the T domains that regulate transcription. Consistently, the known TBX2 target genes, *TRP-1* and *Cdkn2a(p19^{ARF})*, are both downregulated by TBX2 (Carreira *et al.*, 1998; Jacobs *et al.*, 2000).

Interestingly, cofactor-mediated transactivation has recently been described, too (Hsueh *et al.*, 2000). A yeast two hybrid screen identified Tbr-1 as specific binding partner for CASK/LIN-2, a membrane-associated guanylate kinase. Binding to Tbr-1 locates CASK/LIN-2 to the nucleus, where CASK/LIN-2 acts as a coactivator of Tbr-1 to induce transcription of *Tbr-1* target genes, including *reelin* (Hsueh *et al.*, 2000). The interaction with different cofactors may enable a T-box protein to either activate or repress transcription of target genes. A dual mode of transcriptional regulation was indicated for the T-box factors T-bet and Tbx2 (Szabo *et al.*, 2000; Chen *et al.*, 2001).

Nuclear localization of T-box proteins

All T-box proteins for which their subcellular localization has been investigated proved to be exclusively located within the cell nucleus consistent with their function as transcriptional regulators (Schulte-Merker *et al.*, 1992; Kispert and Herrmann, 1994; Grimm, 1997; Hsueh *et al.*, 2000; Carlson *et al.*, 2001). A few attempts were made map nuclear localization signals (NLS) within T-box proteins. The T protein appears to contain several complex NLS between residues 137 and 320 (Kispert *et al.*, 1995a). This region comprises the C-terminal part of the T domain and a part of the regulatory region of the T protein. The Tbx3 NLS consists of a cluster of basic aa at the C-terminus of the Tbx3 T domain (RREKRK, aa 292-297) (Carlson *et al.*, 2001). Site and sequence of this motif are fully conserved in TBX2 suggesting that this stretch of basic aa may direct nuclear localization in TBX2 as well (Table 2). Interestingly, some sequence similarity to the Tbx3 NLS can also be found in some other T-box proteins at the corresponding position, albeit functional significance needs to be experimentally tested (Table 2).

hTBX3	292	**RRE**K**RK**	297
hTBX2	282	**RRE**K**RK**	287
Dmel OMB	518	**KR**EK**NCYR**	525
mT	221	**K**ERND**HK**	227

Table 2. TBX3 nuclear localization signal and similar sequences at the C-terminus of other T domains.
The TBX3 NLS and similar sequences at the corresponding position of TBX2, OMB, and T protein are aligned. Numbers indicate the position of the shown aa in the protein. Amino acids with basic side chains are shown in bold.

Upstream and downstream of T-box factors

In *Xenopus* embryos, mesoderm formation is initiated through an inductive interaction of vegetal cells with overlying equatorial cells (reviewed by Harland and Gerhart, 1997). It is well known that vegetal cells provide a source of secreted signaling molecules which diffuse into the animal hemisphere of the embryo to induce mesoderm in a concentration-dependent manner. Morphogens of the fibroblast growth factor (FGF) and the transforming growth factor β (TGFβ) families were found

to be especially potent mesoderm inducers. Of these, TGFβ-like factors such as activins are signals for mesoderm of a dorsal character, whereas basic FGF induces ventral mesoderm.

Given the important roles of several T-box factors in mesoderm formation, it is not surprising that T-box genes are under control of those signaling factors. Expression of *XBra*, for example, can ectopically be induced in prospective ectodermal cells by activinA (Smith *et al.*, 1991). The response to activin is thereby tightly concentration-dependent: only moderate activin concentrations induce *XBra*, while high and low concentrations of activin do not (Smith, 2001; and references therein). Induction of *Bra/ntl* by activin has also been observed in zebrafish and chick and, thus, appears to be conserved among vertebrates (Schulte-Merker *et al.*, 1992; Kispert *et al.*, 1995b). *Eomes* expression can also be stimulated in cultured animal caps by the addition of activinA (Ryan *et al.*, 1996).

FGF signaling is crucial during gastrulation for the development of trunk and tail mesoderm, where it activates and/ or maintains the expression of *XBra/ntl* and *spt* (Isaacs *et al.*, 1994; Schulte-Merker and Smith, 1995; Griffin *et al.*, 1998). Moreover, *Bra* is a direct target of Wnt3a, a secreted signaling protein of the Wnt family, during paraxial mesoderm formation (Yamaguchi *et al.*, 1999). Intriguingly, the expression of *Drosophila omb* in wing development is also controlled by morphogens of the TGFβ and Wnt families, Decapentaplegic (Dpp) and Wingless (Wg), respectively, suggesting that T-box genes are regulated by upstream signals conserved between vertebrates and invertebrates (Grimm and Pflugfelder, 1996; Hofmeyer, 2001).

Aside from signaling proteins, HD transcription factors were shown to control the expression of T-box genes. Pitx1, a paired-type HD transcription factor, activates *Tbx4* during hindlimb development (Logan and Tabin, 1999). An other example gives *Goosecoid* which is activated by high concentrations of activin and can suppress *XBra* in the vegetal hemisphere of the *Xenopus* embryo (Smith, 2001; and references therein).

During the past years, several downstream targets of T-box factors could be identified (Table 1). A study of known T-box target genes reveals that many of these encode hormones, transcription factors or signaling molecules. Since such proteins are well known to function concentration-dependently, the observed dosage-sensitivity of T-box transcription factors appears to be at least in part a consequence of dosage-dependent target genes.

It is intriguing to see that both upstream regulators and downstream targets of T-box proteins include HD transcription factors and signaling molecules of the FGF, TGFβ or Wnt families (Figure 4). The close interconnections with those key regulators make T-box transcription factors to crucial components of developmental programmes governing the development of complex organisms.
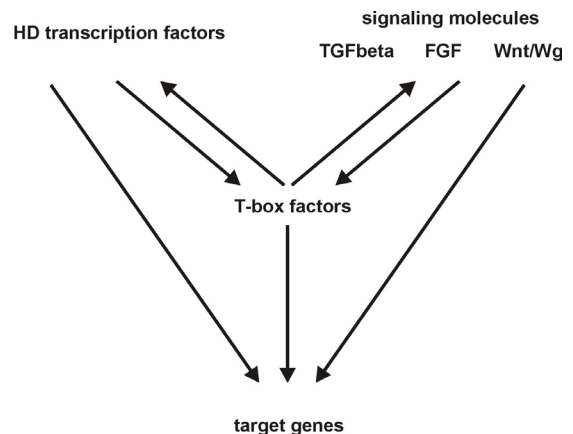
Figure 4. Relationships between T-box factors, HD factors and signaling molecules in animal development.

This work now reports on the functional analysis of the *Drosophila* T-box gene *org-1* and includes genetic experiments to isolate *org-1* mutants as well as ectopic *org-1* expression studies. A second main project is presented in which we investigated the molecular determinants of the functional specificity in OMB and ORG-1 using chimeric transgenes.

T-box factor	target gene	gene product	transcriptional regulation	promoter element	promoter element/DNA binding sequence	reference
***in vitro* binding site selections**						
mouse T				1 p →←	T(G/C)ACACCTAGGTGTGAAATT	Kispert and Hermann, 1993
Drosophila OMB				frequently 2 h s; ←→ or →→	AGGTGTGA	Grimm and Pflugfelder, in prep.
human TBX5				1 or 2 h s; → or →→→, ←→, →←	(A/G)GGTGT(C/G/T)(A/G)	Ghosh et al., 2001
identified target genes						
Xbra	*Xenopus eFGF*	signaling molecule	activation	2 h s → distal, -936 / → proximal, +123	TTTCACCCT / AACCACACCT	Casey et al., 1998
Xbra, VegT	*Xenopus Bix4*	HD transcription factor	activation	2 h s → distal, -85 / → proximal, -66	CTTCACACCT / ATTCACACGT	Tada et al., 1998
VegT	*Xenopus Xnr1*	secreted signaling molecule, TGFβ-like	activation	2 h s → TBX1 / ↓ TBX2 / or: → -12 / ↓ +64	CATAGGTGTGAA……… / AGCTCACTCCTA / AGGTGTGAAG / GCATTACACC	Hyde and Old, 2000 / Kofron et al., 1999
Ci-Bra	*Ciona intestinalis* (ascidia) *Ci-trop*	structural protein	activation	1 p, 1 tr, 1 h s →← Ci-Bra proximal / →→→ Ci-Bra distal / → Ci-Bra #3	TTTATCACTATCCTGTGAAG / TCTCGCACCCGGCACCTCTT / CGTCACACCT	Di Gregorio and Levine, 1999
AsT	*Halocynthia roretzi* (ascidia) *AsT*	T-box transcription factor	activation	1 p →←	TTTGTTACCTAGGTGTGGAAA	Takahashi et al., 1999
Tbx2	mouse *TRP-1*	melanogenic hormone	repression	2 h s ← MSEu, -237 / ← MSEi, +1	GTGTGA / GTGTGA	Carreira et al., 1998
TBX2	human *Cdkn2a(p19^{ARF})*	CDK inhibitor, tumor suppressor	repression	within −19 to +54 of Cdkn2a promoter	unknown	Jacobs et al., 2000

identified target genes (continued)

T-box factor	target gene	gene product	transcriptional regulation	promoter element/DNA binding sequence	reference
Tbx5	mouse *Nppa*	secreted protein	activation; synergistic transactivation with Nkx2-5	1 h s next to a Nkx2-5 site. → -252 to -237 — TCACACCT [TGAAGTG]	Hiroi *et al.*, 2001
Tbx5	mouse *Cx40*	structural protein	activation; synergistic transactivation with Nkx2-5	5 h s: ← CX1 -67; ← -160; ← CX2 -500; ← -770; ← -1000 — GTGGGA, GTGAGA, GTGACA, GTGTAA, GTGACA	Bruneau *et al.*, 2001
Tbx5	mouse *ANF*	secreted protein	activation; synergistic transactivation with Nkx2-5	3 h s. 1 h s flanked by Nkx2-5 sites, 1 h s next to Nkx2-5 site. ← TBE1 -90; → TBE2 -252; ← TBE3, -485 — GCAAGTGA ACAGAATG, TCACACC.. [TGAAGTG], GGTGTGA	Bruneau *et al.*, 2001
Tpit	mouse *POMC*	hormone	activation; synergistic transactivation with Pitx1	1 h s next to Pitx1 site. → CE3 — TCACACCA...... [TAAGCC]	Lamolet *et al.*, 2001
Byn	*Drosophila orthopedia*	HD transcription factor	activation	15 h s, 4 in tr, 2 in nested inverted repeat — AAAACCGCACAA, CTGAATGCACAT, GCGATTTCTCCC, CAATTTGCACAA, GTATTAACACTA, TTAAATTCACGC, ATAAACACACTT, GTTGTAACACAT, ATTTTTTCACCT, GAAATCGCACTT, ATAGCGCCACTT, AAAGTCGCACTT, GATATCGCACCT, AAATTAGCACTT, AATTTTACACAT	Kusch *et al.*, 2002

potential target genes

T-box factor	target gene	gene product	transcriptional regulation		promoter element/DNA binding sequence		reference
Xbra	*Xenopus Xwnt11*	signaling molecule	activation	unknown			Tada and Smith, 2000
Tbr-1	mouse *reelin*	extracellular matrix protein	coactivator CASK/LIN-2 mediated activation	2 h s		unpublished	Hsueh *et al.*, 2000
T-bet	mouse *IL-2*	cytokine	repression	1 p ←→	-220	<u>AAACTGCCACCTAAGTGTGGGCTA</u>	Szabo *et al.*, 2000
T-bet	mouse *IFNγ*	cytokine	activation	3 p →← →← →←	-2291 -1948 +4655	GACAGCTCACACTGGTGTGGAGCA CTTCTGTCACCTGAGTGTCTGGGA TGCTATGCACAACAGTGAGAATCA	Szabo *et al.*, 2000

Table 1. T-box protein binding sites and T-box target genes.

In vitro selected binding sites and similar sequences in the promoters of T-box target genes are listed. Underlined sequences indicate binding sites for which DNA recognition has been demonstrated *in vitro*. Arrows indicate the orientation of half sites according to the T protein *in vitro* binding site. Used abbreviations are: h s, half site; tr, tandem repeat; p, palindrome.

2. Material and Methods

2.1 *Drosophila* stocks and rearing conditions

stock	genotype	GOP stock #	reference
dpp-Gal4 K54	w; Cy/Sp; K54/TM6, Tb Hu	530	Staehling-Hampton *et al.*, 1994
E132-Gal4	w? P[w+ Gal4]E132/Y	502	Halder *et al.*, 1995
30A-Gal4	w; pGawB/CyO	567	Brand and Perrimon, 1993
GMR-Gal4	II	786	Bloomington stock #1104
omb^{P3}-Gal4	y w *omb^{P3}*/FM7	55	
UAS-omb	w^{1118}; P[w$^+$ UAS:omb] 4-15 (II)	255	Grimm, 1997
hs-Gal4 (III)	w; hs-Gal4(89-2-1)	796	Bloomington
C31	X	184	Strauss and Trinath, 1996
C31 x attX	C(1)DX, y w f/C31	185	Strauss and Trinath, 1996
Δ2-3 TM3 Sb/Dr	yw; Δ2-3 TM3, Sb/Dr	404	
Δ2-3/TM3, Ser	w; P[Δ2-3], Sb/TM3 Ser	33	
w, C31	w C31		this work
EP 3668	EP/TM6, Tb EP insertion in *vmd2*	168	Bloomington

Table 3. *Drosophila* fly stocks.

Numerous fly stocks were characterized or generated in the course of this work. These lines will be described and listed in subsequent chapters. A complete list of my fly stocks is provided in the appendix section and can also been found on the accompanying CD-ROM [inventory/fly stocks]. A list of Gert Pflugfelder's fly stocks is saved there, too.

Fly stocks were raised at 18°C or 25°C (maintenance of stocks or propagation, respectively) on standard *Drosophila* medium containing cornmeal, agar, molasses, yeast, and Nipagin.

2.2 *Drosophila* germline transformation

Transgenic *Drosophila* lines were generated by transforming modified P elements into the germline of *Drosophila* embryos (Santamaria, 1986; Spradling, 1986). 12 µg of pUAST constructs and 4 µg of pUChsπΔ2-3 helper vector (Rio and Rubin, 1985) were co-precipitated and resuspended in 25 µl injection buffer (5mM KCl, 0,1 mM Na-phosphate buffer pH 6.8). Resuspended DNA was microinjected into w^{1118} embryos. Injected flies were mated to w^{1118} flies and transformants could be identified by the presence of the *white$^+$* marker of the pUAST vector. The transgenes were then chromosomally mapped by segregation analysis and, if homozygotically viable, made homozygous. The procedure has thoroughly been described previously (Hofmeyer, 1996; Heindel, 1998).

2.3 Determination of the relative expression strength of individual UAS-transgenic lines

The UAS-transgenic lines are crossed to *hsp70*-Gal4 flies and female transheterozygotes for the *hsp70*-Gal4 and the UAS-transgenes are selected among their offspring. Groups of such 12-36 h old flies are then exposed to a single 45 min heat shock at 37°C (flies were transferred into empty food vials containing a moistened piece of paper and subsequently put into the 37°C room) that induces ubiquitous Gal-4 expression in the adult fly. At distinct time points after the heat shock, some 10 flies are decapitated and heads are homoge-

nized in 10µl/ head SDS PAGE loading buffer using glass tissue grinders [Kontes]. The homogenate is incubated 5 min at 95°C, centrifuged, and stored in aliquots à 10 µl at -20°C until the samples are separated on a conventional SDS-PAGE and blotted. Western blots were then simultaneously incubated with anti-HA (mab 12CA5, 1:1000) [Roche] or anti-MYC (mab 1-9E10.2, 1:75) [American Type Culture Collection] (Grimm, 1997) and anti-SAP47 (nc46/1, 1:1000). The ECL kit [Amersham] was used for signal detection according to the supplier's manual.

2.4 EMS mutagenesis

The mutagen ethyl methanesulfonate (EMS) was administered to adult males by feeding them on a sucrose solution containing EMS. All steps that included the handling with EMS were carried out by Gert Pflugfelder in a fume hood and according to Grigliatti (1986), with minor modifications as follows:

About 3 days old males were starved and desiccated in empty vials at room temperature for 1-3 h and then transferred as batches of approximately 50 flies into clean vials containing two mashed paper towels moistened with 7 ml buffer (100 mM Tris-HCl, pH 7.5, 10% sucrose, 25mM EMS). The buffer in a control vial did not contain EMS. Flies are allowed to feed on the EMS-sucrose solution for about 24 h. Then, they were returned into empty vials, where they could excrete residual EMS for several hours. Subsequently, the EMS-treated males were allowed to recover on ordinary food vials and crossed to virgin females. Four times, the males were separated from females and mated to new virgins each day. Fertilized females were transferred to new food vials every other day.

Remaining EMS buffer or contaminated material was inactivated in denaturing solution (4 g of NaOH and 0.5 ml of thioglycolic acid in 100 ml of H_2O) after use in a fume hood.

2.5 Scanning electron microscopy

Flies were anaesthesized by CO_2, selected, and killed with chloroform. Flies were then fixed in 6.25% glutar aldehyde, 100 mM Na-phosphate buffer pH 7.3 at 4°C ON. Subsequently, the flies were dehydrated in a series of aceton/ Sörensen-phosphate buffer pH 7.4 (obtained from Claudia Gehrig, Würzburg) with increasing concentration of aceton.
Dehydrated objects are kept in pure aceton, until dried at the critical point. The preparation is then

sputtered with gold and investigated at a scanning electron miscroscope [Zeiss DSM 962]. This work was made possible by Prof. Krohne, Würzburg, and was guided by members of his laboratory.

2.6 Preparation of adult *Drosophila* appendages and abdominal cuticle

Body appendages were carefully removed using a pair of fine tweezers and embedded in Euparal [Chroma].
Abdominal cuticle preparations were performed with help from Christian Leipold, Würzburg, who made a longitudinal cut along the ventral abdomen using a pair of fine scissors. The cuticle was flattened and pinned with tiny needles. The preparation was then incubated for several hours in 10% KOH at 50°C, washed with PBS, dehydrated in an EtOH series, and embedded in Euparal [Chroma].

2.7 Molecular biology

Material and methods routinely used in molecular biology were previously described elsewhere (Porsch, 1997; Roth, 1998; Sambrook *et al.*, 1989; Ausubel *et al.*, 1994).

2.8 Oligonucleotides

All oligonucleotides that I have ordered in the course of my work are listed in the appendix section. A map of primers within the corg-1M2 sequence is provided as well. Both can also been found on the accompanying CD-ROM [inventory/oligonucleotides].

2.9 DNA sequencing

DNA sequencing was performed using the ABI PRISM[TM] BigDye[TM] Terminator Cycle Sequencing Ready Reaction Kit [PE Applied Biosystems]. Sequencing reactions were carried out in a Hybaid thermal cycler and routinely contained:

terminator reaction mix	2.0 µl
oligonucleotide [2 µM]	4.5 µl
template DNA	
plasmid DNA	300 ng
or	
PCR product	10-100 ng
sterile, bidestilled H_2O	*ad* 10 µl

PCR products were gel-purified [gel purification kit, QIAGEN] prior to sequencing.

Sequencing reactions were set up on wet ice and overlaid with 40 µl mineral oil. Cycle sequencing comprised 25 cycles of 96°C for 15 sec, 50°C for 1 sec, and 60°C for 4 min.

BAC clones were sequenced with 6.0 µl terminator reaction mix, 1 µl oliogonucleotide solution [20 µM], and 0.5-2 µg BAC DNA in a final reaction volume of 15 µl. The modified sequencing programme includes an initial denaturation step 96°C for 4 min, followed by 100 cycles of 96°C 10 sec, 50°C for 10 sec, and 60°C for 4 min.

Extension products were purified by ethanol/ sodium acetate precipitation as described in the ABI protocol. Electrophoresis of purified products was carried out by Ellen Fecher, Würzburg, on an ABI PRISM™ 310 Genetic Analyzer.

2.10 *org-1* 5' RACE

reverse transcription

1 µg *Drosophila* poly A RNA from embryonic stage E4 (a gift from Gert Pflugfelder) was incubated with 3-20 pmoles corg1-5'end primer at 70°C for 10 min and subsequently chilled on wet ice. Reverse transcription was then performed in 1 mM MgCl$_2$, 400 µM dNTPs, 10 mM DTT, and started after a 2 min pre-incubation at 42°C by adding 200 U Superscript II reverse transcriptase (RT) [Gibco BRL] to the reaction mix. The reaction was stopped by heat inactivation (15 min at 70°C), and the RNA template was removed by RNase H digestion (1U RNase H [Gibco BRL], 20 min at 55°C). Synthesized cDNA was purified using the PCR purification kit [QIAGEN]. *org-1* 5' RACE was performed following two alternative methods, self-ligation and oligo C tailing.

oligo C tailing

Synthesized cDNA was oligo C tailed by terminal deoxynucleotidyl transferase (TdT) [TaKaRa] at 37°C for 1.5 h in a reaction containing 1x TdT MgCl$_2$ buffer (10x buffer contains 1 M sodium cacodylate pH 7.2, 20 mM MnCl$_2$, 1 mM DTT; sterile filtrated). cDNA was first denatured at 94°C for 3 min, before dCTP (200 µM final conc.), TdT MgCl$_2$ buffer, and 13 U TdT were added. Unincorporated nucleotides and enzyme were removed by column purification (PCR purification kit [QIAGEN]), before second strand synthesis was performed using

Klenow [Gibco BRL] and anchor primer annealing to the oligo C tail. Products were subsequently PCR amplified using anchor primer und nested *org-1* primers corg1-5'endN1Sal and corg1-5'endN2Sal. Amplicates were gel-purified (gel extraction kit [QIAGEN]) and cloned into pGEM-T [Promega].

self-ligation

The self-ligation method required the use of a 5' phosphorylated primer for the reverse transcription. Therefore, corg1-5'end primer was initially phosporylated by T4 polynucleotide kinase [MBI; 10 U used] in a reaction containing appropriate 1x reaction buffer [MBI] and 2 µM ATP. Synthesized cDNA was self-ligated by T4 RNA ligase in order to provide a circular template DNA for inverse PCR amplification. The self-ligation reaction contained 25% PEG 6000, 0.02% BSA, 1x T4 RNA ligase buffer, and 40 U T4 RNA ligase [TaKaRa], and was performed at 14°C ON. Column-purified circular *org-1* cDNA served as template for inverse PCR amplification using two nested primer pairs, org-1back1/ org-1forward1 and org-1back2/ org-1forward2. Amplicates were gel-purified and cloned into pGEM-T [Promega] or pBATSK (Grimm, 1997) vectors. Positive clones were identified by colony hybridization with a probe derived from the 5' end of corg-1M2.

2.11 Single fly PCR

Selected flies are individually placed in a 0.5 ml safelock Eppendorf tube and kept on wet ice. The fly is thouroughly mashed with a 200 µl plastic pipet tip containing 50 µl squishing buffer (SB) without expelling any buffer (some liquid escapes from the tip). Then the remaining SB is released and the crude homogenate is incubated at 37°C for 1 h. Subsequently, the proteinase K is inactivated by heating to 95°C for 5 min. A 5 min spin down removes fly parts from the solution.

The DNA within these preparations is stable for months, when stored at 4°C.

1-10 µl of the 50 µl preparations are used as template DNA in PCR (2-6 µl gave maximal yield). Conventional PCRs are set up on wet ice and, prior to the addition of *Taq* DNA polymerase, are incubated at 95°C for 8 min in order to properly denature the complex genomic template DNA.

Squishing buffer (SB) contains:

Tris-HCl pH 8.2   10 mM
EDTA   1 mM
NaCl   25 mM
sterile, bidest. H_2O

200 µg/ ml proteinase K, diluted fresh from a frozen aliquot each day.

2.12 Amplification of P element flanking genomic sequences by inverse PCR

The *in vitro* amplification of P element ends neighboring genomic sequences by inverse PCR (iPCR) (Ochman *et al*., 1988; Sentry and Kaiser, 1994; Spradling *et al*.,1995; Dalby *et al*., 1995) was performed using a slightly modified protocol of the Berkeley *Drosophila* Genome Project (BDGP) [http://www.fruitfly.org/about/methods/inverse.pcr.html].

Genomic DNA of some 50 flies of a P element line to be investigated is conventionally isolated and re-suspended in 2 µl TE per fly. Resuspended ge-nomic DNA of about 8 flies is digested with *Cfo* I or *Sau3A* I restriction endonuclease in a 20 µl reac-tion. Upon heat inactivation of the enzyme, the re-striction digest is self-ligated at 4°C ON in a 100 µl reaction containing 2 units T4 DNA ligase [Gib-coBRL]. These reaction conditions favor the in-trachromosomal circularization of restriction frag-ments (Collins and Weissman, 1984; Ochman *et al*., 1988). The ligation reaction is directly EtOH precipitated without prior phenol-chloroform extrac-tion and resuspended in 100 µl bidestilled (bidest.) H_2O to yield the iPCR template DNA solution. The iPCR contains:

dNTPs (2 mM)   5.0 µl
10x Taq PCR buffer (includes 15 mM $MgCl_2$)
 Eppendorf]   5.0 µl
forward primer (20 µM)   2.5 µl
reverse primer (20 µM)   2.5 µl
Taq DNA polymerase (5 U/ µl)
[Eppendorf]   0.2 µl
Cfo I or *Sau3A* I digested, self-ligated genomic DNA (100 µl resuspension in bidest. H_2O)
                                            10.0 µl
sterile, aliquotted bidest. H_2O 24.8 µl
final reaction volume   50.0 µl

For the amplification of P{lacW} element 5' and 3' ends, the forward and reverse primers are Plac1 and Plac4 or pry1 and pry2, repectively [http://www.fruitfly.org/about/methods/inverse.pcr.html]. iPCR amplification started with a 3 min hot start at 94°C, followed by 30 cycles consisting of

94°C for 30 sec (strand separation), 60°C for 30 sec (primer annealing), and 72°C for 2 min (primer extension). A final extension step at 72°C for 10 min completed the iPCR.

Obtained iPCR products were gel-purified [Qiagen] and sequenced using dye terminator technology [ABI Prism BigDye cycle sequencing, Perkin-Elmer].

2.13 Expression and purification of recombinant ORG-1 protein and raising ORG-1 antisera

These procedures are described in detail in chapter 3.3.

3. Molecular analysis of *org-1*, *TBX1*, and *vmd2*

3.1 Molecular characterization of *org-1*

3.1.1 Cloning and sequencing of a full-length *org-1* cDNA

Raimond Miassod, Marseille, isolated an original 2,8 kb *org-1* cDNA clone, corg-1M1, from a pNB40 embryonic cDNA library (Brown and Kafatos, 1988). Sequence analysis demonstrated that corg-1M1 is incomplete, since a long ORF extends beyond the 5' end of the cDNA (Porsch, 1997). PCR based screens of the pNB40 library to obtain the missing 5' end only gave amplifies corresponding to the 2,8 kb or smaller cDNAs. Therefore, 500.000 phages of a size-selected embryonic λgt 11 library were screened by Raimond Miassod. Two of the eight positive clones contained larger inserts of which the largest was subcloned *EcoR* I into pKS (by Gert Pflugfelder). The resulting clone, named corg-1M2, was completely sequenced on both strands. Individual sequences, the final corg-1M2 sequence, and sequence alignments can be found on the accompanying CD-ROM [DNAseq/*org-1* molecular analysis/corg-1M2]. The corg-1M2 is 3168 bp in size and encodes the full-length *org-1* ORF of 708 aa (Figure 5). corg-1M2 thereby extends the original *org-1* cDNA by 314 bp on the 5' site but has an identical 3' end. However, the two *org-1* cDNAs reveal numerous polymorphic sites that affect the peptide sequences, too (Figures 6 and 7).

3.1.2 *org-1* 5' RACE

Northern blot analysis indicated that *org-1* is expressed as a single transcript of about 3800 nt throughout all developmental stages, most abundantly during mid-embryogenesis (Porsch, 1997). Since our longest *org-1* cDNA, corg-1M2, contains only 3168 bp and ends with a poly A run, this *org-1* cDNA appears to be 5' incomplete lacking the first some 300-400 bp, if one assumes an average poly A tail of 200-250 residues. As P elements have a marked tendency to integrate into the genome at 5' regions of genes (Spradling *et al.,* 1995), I was advised to determine the *org-1* promoter region prior to an *org-1* P element insertion mutagenesis. Therefore, 5' rapid amplification of cDNA ends

(RACE) technology was applied in order to clone the missing 5' end of the *org-1* transcript. *Drosophila* poly A RNA from embryonic stage E4 was reverse transcribed using the *org-1* specific primer corg1-5'end. Template RNA was subsequently removed by RNase H digestion. We then employed two alternative methods for cDNA amplification: self-ligation in combination with inverse PCR and oligo C tailing with conventional PCR.

For the self-ligation (SL) approach, the corg1-5'end primer was 5' phosphorylated prior to its use for reverse transcription allowing the circularization of the synthesized cDNA by T4 RNA ligase. Circular cDNAs were subsequently amplified by inverse PCR with two sets of nested primer pairs (Figure 8, self-ligation).

Alternatively, an oligo C tail was added to the complementary DNA catalyzed by terminal deoxynucleotidyl transferase (TdT), so that modified cDNAs could by amplified with an anchor primer and a nested *org-1* primer (Figure 8, oligo C tailing).

PCR amplificates were obtained for both strategies. Products obtained with the SL method were separated on agarose gels, blotted, and hybridized with a probe from the 5' end of corg-1M2. Most amplificates gave strong hybridization signals which, however, predominantly appeared as a smear on audioradiographs, indicating that the majority of products is of single-stranded DNA. Some distinct signal bands of 400-700 bp (~250 bp known *org-1* sequence expected) could be obtained, too. Likewise, several TdT reactions also revealed visible amplificates of 400-800 bp (~340 bp known *org-1* sequence expected). A number of PCR products from the SL and TdT approaches was cloned and sequenced (Figure 8). All determined DNA sequences can be found on the accompanying CD-ROM [DNAseq/*org-1* molecular analysis/RACE]. The four SL clones investigated did not expand the known *org-1* transcript further, however, they all revealed sequences of the first *org-1* intron 5' to the exon 2 sequence. Surprisingly, all clones extend in 3' direction beyond the putative corg1-5'end annealing site at position bp 393 of corg-1M2. They heterogeneously end at bp 404 (clones SLII 10 and SLII 12), at bp 415 (SLII cl 26) or at bp 787 (SLII cl 14), suggesting that misannealing of corg1-5'end primed *org-1* cDNA synthesis from various positions along the *org-1* transcript, although a search of corg-1 M2 for sequences complementary to corg1-5'end discovered no significant matches besides at bp 393. Importantly, the *org-1* cDNA sequence of SLII cl 14 starts within exon 3, but does not contain intron 2 sequence, indicating that this cDNA (and presumably the other amplificates which contain intron 1 sequence as well) derived from a partially spliced *org-1* RNA rather than from

```
   1  ccaccagcgcttgacggaac  ggatgttgcacagtgagtga  gtgagtgagtgaatgagtga  gtgattgtgtgtgggtgcaa  taatcaaatttgacaactgc   100
 101  gcgaataacgttcgagatca  aaacaaattgtgccggaata  aatggaataaatgggccaag  aactcgctgcgatttccttt  cttccactgtgagcatcaca   200
 201  tatccagccatcatatatat  acatatatatatatattcca  acgatcacgctcgccatgac  gcacctgatgggccccactg  agtgcgccggcgccatgatg   300
   1                                                                                  M   T   H  L  M  G  P  T  E   C  A  G  A  M  M   15

 301  accaccacatccatgcagtt  cctggacaccagcctaacgg  actacaactgctatggcaac  gactactggacatcgccgta  catgaccggcggactcagtc   400
  16  T   T   T  S  M  Q  F   L  D  T  S  L  T  D   Y  N  C  Y  G  N   D  Y  W  T  S  P  Y   M  T  G  G  L  S  P    49

 401  ccatgaagcagatcgaagcc  tgcatccaaacggctggcaa  ggatcgcagctcgtacaagc  cgctggagcagatcgatgcc  aaattggcggacatcgagac   500
  50    M  K  Q  I  E  A   C  I  Q  T  A  G  K   D  R  S  S  Y  K  P   L  E  Q  I  D  A   K  L  A  D  I  E  T    82

 501  gcacagtacaggcagcactg  gcaccgcgaacagcaacagc  agcaccagcagcatctcgaa  tcccagttgcccggatcagt  cgtcgtcgtcgtcatcgtcg   600
  83  H   S  T  G  S  T  G   T  A  N  S  N  S   S  T  S  S  I  S  N   P  S  C  P  D  Q  S   S  S  S  S  S  S    115

 601  tccgtatcgctgccaaccga  ttatgccggcgtacacagtg  aagcctcgatggcaccaaca  gccggcggcacggcagtcac  aacgacatcagctggcggag   700
 116  S   V  S  L  P  T  D   Y  A  G  V  H  S  E   A  S  M  A  P  T   A  G  G  T  A  V  T   T  T  S  A  G  G  V   149

 701  tcagtgcatccaccgcgtcc  aaaaagttcaagggacagca  caaaaaagacaacaacagtg  cggagaacggtacagtgaag  cccaatagccataatatcag   800
 150    S  A  S  T  A  S   K  K  F  K  G  Q  H   K  K  D  N  N  S  A   E  N  G  T  V  K   P  N  S  H  N  I  S   182

 801  caaaggtgaatcggagccag  tgcatccatcgctggcccag  gccattgtggtgctggagac  gaaggcgctgtgggatcagt  tccatgcccagggcaccgaa   900
 183  K   G  E  S  E  P  V   H  P  S  L  A  Q   A  <u>I  V  V  L  E  T   K  A  L  W  D  Q  F   H  A  Q  G  T  E   215</u>

 901  atgatcatcaccaagacggg  ccgacgcatgtttcccacgt  ttcaggtgaggatcggtggt  ttggatccacatgccaccta  catttgcatgatggactttg  1000
 216  <u>M  I  I  T  K  T  G   R  R  M  F  P  T  F   Q  V  R  I  G  G   L  D  P  H  A  T  Y   I  C  M  M  D  F  V   249</u>

1001  tgcccatggatgacaaacgc  tatcgctacgccttchacaa  ctcctgctgggtggtggctg  gcaaggcggatcccatttcc  ccgcccaggattcatgtgca  1100
 250  <u> P  M  D  D  K  R   Y  R  Y  A  F  H  N   S  C  W  V  V  A  G   K  A  D  P  I  S   P  P  R  I  H  V  H   282</u>

1101  tcccgactcgccagccgtcg  gctccaattggatgaagcag  atcgtgtcctttgacaaatt  gaagctcaccaataaccagc  tggacgaaaatggacatatc  1200
 283  <u> P  D  S  P  A  V  G   S  N  W  M  K  Q   I  V  S  F  D  K  L   K  L  T  N  N  Q  L   D  E  N  G  H  I   315</u>

1201  attctgaactccatgcatcg  ctaccagccgcgtttccatc  tggtttatctgccaccgaag  aacgcctccttggatgagaa  cgagcactccagccactttc  1300
 316  <u>I  L  N  S  M  H  R   Y  Q  P  R  F  H  L   V  Y  L  P  P  K   N  A  S  L  D  E  N   E  H  S  S  H  F  R   349</u>

1301  gcactttcatctttccggaa  acgagctttacggccgtaac  tgcctaccagaatcagcggg  tgacacagctgaagatctcc  agcaatccattcgccaaagg  1400
 350    <u>T  F  I  F  P  E   T  S  F  T  A  V  T   A  Y  Q  N  Q  R  V   T  Q  L  K  I  S   S  N  P  F  A  K  G   382</u>

1401  ctttcgggatgatggcacca  acgatgtaaccactggcggt  ggcagcagcatgtcctccat  gagtcacgaaagtcaggcgc  gcatgaagcagcaacagcag  1500
 383  <u> F  R  D  D  G  T  N</u>  D  V  T  T  G  G   G  S  S  M  S  S  M   S  H  E  S  Q  A  R   M  K  Q  Q  Q  Q   415

1501  caacagcagcagcagcagca  gcagcaactgcagcagcaac  agcaacagcagcagcaactc  aaggagcgaacggcagcaac  cagcaactttggcctaagtt  1600
 416  Q   Q  Q  Q  Q  Q  Q   Q  Q  L  Q  Q  Q  Q   Q  Q  Q  Q  L   K  E  R  T  A  A  T   S  N  F  G  L  S  C   449

1601  gcagcgaactggccattgag  caacagcagcagcagcaaca  gcaacagggagttctgcagc  taccggccacgccctccagc  agctccacctccggcaattc  1700
 450    S  E  L  A  I  E   Q  Q  Q  Q  Q  Q  Q   Q  Q  G  V  L  Q  L   P  A  T  P  S  S   S  S  T  S  G  N  S   482

1701  acccgacttgctgggctacc  agatggagcagcaactgcaa  cagcaacaccaacagcagca  gcaacagcaacaccagtccc  agcagcaacatctccaccag  1800
 483  P   D  L  L  G  Y  Q   M  E  Q  Q  L  Q   Q  Q  H  Q  Q  Q  Q   Q  Q  Q  H  Q  S  Q   Q  Q  H  L  H  Q    515

1801  caacaccaggctaaccagca  acaatcgctgctccaacaga  gccagaatcacacgcaatat  ggcagctatcatcatgcata  ccaggcacaggtgcagtcgc  1900
 516  Q   H  Q  A  N  Q  Q   Q  S  L  L  Q  Q  S   Q  N  H  T  Q  Y   G  S  Y  H  H  A  Y   Q  A  Q  V  Q  S  H   549

1901  atcccctaacgccgcactcc  agcagctccgcatccccgcc  agcaactgctgcgccgggag  caagtgcagcaacagcagca  gtagcagcagcagcagcagc  2000
 550    P  L  T  P  H  S   S  S  S  A  S  P  P   A  T  A  A  P  G  A   S  A  A  T  A  A   V  A  A  A  A  A  A   582

2001  agcagcagccgtagcagggg  gaggagcaggagcaggcgga  gcaacgtcagccacacaagt  gatgagtgcggccaatatct  attcgagcattggacaaccg  2100
 583  A   A  A  V  A  G  G   G  A  G  A  G  G   A  T  S  A  T  Q  V   M  S  A  A  N  I  Y   S  S  I  G  Q  P   615

2101  tatgcccaggagcagagcaa  ctttggtgcgatctaccatc  ataatgctgctgccgcagcg  gccgctcactatcatcatgg  tcatgcccacggccacgccc  2200
 616  Y   A  Q  E  Q  S  N   F  G  A  I  Y  H  H   N  A  A  A  A  A   A  A  H  Y  H  H  G   H  A  H  G  H  A  H   649
```

```
2201  atagccatgcgcacgcccac gcccatggaccatatgccag cgcctacgacaagctgaagg tgtcgcgtcatgcggcagct gccgcctatggcatgggcgc  2300
 650    S  H  A  H  A  H    A  H  G  P  Y  A  S    A  Y  D  K  L  K  V    S  R  H  A  A  A    A  A  Y  G  M  G  A   682

2301  cacctatccaagtttttacg gttcggctgcacatcaccag atgatgcgaccgaatagcta catagatctggtgccgcgct aagggagcagcaacttggaa  2400
 683    T  Y  P  S  F  Y  G    S  A  A  H  H  Q    M  M  R  P  N  S  Y    I  D  L  V  P  R  *                       709

2401  gagaaggatttcggatttcg gatttcggatactctatgga attaactgcacttacacttg cctgtaaaaatgattgtaaa atccaaacttagactacgtc  2500
2501  atctatagccaaagctatac atatacatatatgtgtaaat ctcatgccaaagattcgttc taaaatcaagaatctatttc caaagtttagaaaggaagcc  2600
2601  tttaattttcgcccattaaa aaatgttttaacaaaacaaa aacataactaagcttaagcc aaaactataataacaggaat tatttttttagcaagcttaat  2700
2701  ttttaagcattcaaattcat tctttcgcgaaacatttgga atttggagcgatttgattct tgattttagaatcaatttca agtattagcagccagaaaac  2800
2801  caaaaataaatgcaacaagt attacaagtatttctacata caaaaattaccattaaaagt taaaatattttttttttttct agcttaggacgtaaatttta  2900
2901  ttgatttgtgtgaaactgaa aacgcataaaacatttcggt gtaaactgtagtgtaatttt aatatacatattattattat tattttttttttttgcttaa  3000
3001  cactctaggtttttttttct atgtaaatacaagtacatat gtatgtcgctatatatat atatatatatatatttaa gaactgcaacagtttcaagc  3100
3101  aataaaaacaaagaaaattt taaaccgaaactctagcaaa cagaagcataaattaaccaa aaaaaaaa                                      3168
```

Figure 5. Nucleotide sequence of corg-1M2 and the predicted ORG1 amino acid sequence.
The conserved T domain is underlined.
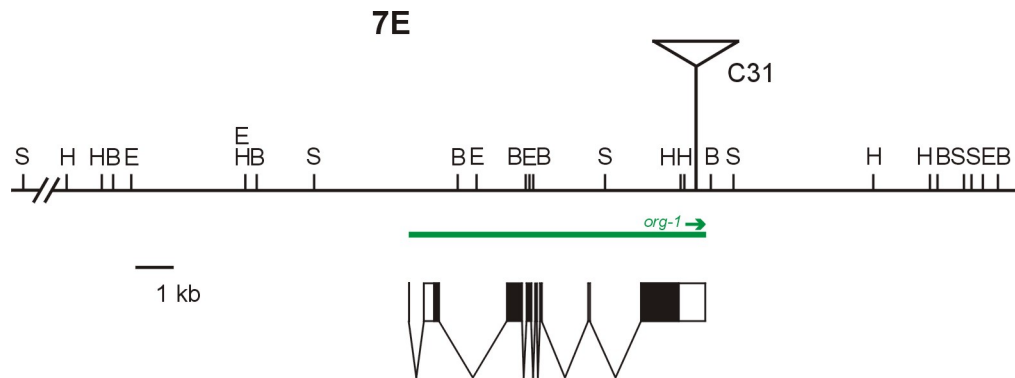
Figure 9. *org-1* exon-intron structure.
The architecture of the *org-1* gene is shown with exons indicated by boxes and introns by thin lines. Filled boxes represent the coding region. The genomic locus is shown above with restriction sites for *BamH* I (B), *EcoR* I (E), *Hind* III (H), and *Sal* I (S). The *C31* I element insertion site is indicated. A 1 kb scale bar is given. Below the exon-intron structure, sizes of the *org-1* exons and introns are listed.

a genomic DNA contamination. Additional six clones were characterized from TdT experiments. They all have uniform 3' ends at bp 322, the annealing site of the nested corg1-5'end N2 Sal primer. All clones contained intron 1 sequences in variable length again. Clones rxn3-5 clR7, rxn3-18 R15, and rxn3-18 R16 contained the complete intron 1 of 422 bp followed by exon 1 and unknown sequences of 19 bp, 12 bp, and 49 bp, respectively. Analysis of the new sequences revealed that these preceed the known exon 1 in the genomic sequence. These short 5' extensions suggest that the full-length *org-1* transcript is indeed only moderately longer than corg-1M2. However, a search for predicted promoter elements [http://www.fruitfly.org/seq_tools/promoter.html] within the *org-1* genomic sequence upstream of exon 1 remained unsuccessful.

Taken together, the results of the RACE analysis expand the *org-1* transcript further in 5' direction by 49 bp. Moreover, all analyzed RACE clones contained *org-1* intron 1 sequences indicating that the *org-1* transcripts predominantly include the first intron in the used RNA preparation. We then questioned, whether the *org-1* transcript on Northern blots would also contain the first intron and, therefore, hybridized RNA blots with an intron 1 specific *org-1* probe. The intronic probe, however, did not recognize the *org-1* transcript on Northern blots.

3.1.3 Exon-intron structure of *org-1*

The exon-intron structure for *org-1* was determined by sequencing genomic DNA fragments of the *org-1* locus and sequence comparison with the *org-1* cDNA sequence. This analysis identified 8 exons separated by 7 introns for *org-1* (Figure 9). *omb* contains the same number of exons and introns, however, has by far larger intron sizes that account for its unusually long 75 kb large transcription unit (Pflugfelder and Heisenberg, 1995). With an about 8 kb large primary transcript, the *org-1* gene is of moderate size. Several intron positions within the T domain encoding region are conserved between *org-1* and related T-box genes (Porsch *et al.,* 1998; Wattler *et al.,* 1998).

3.2 Molecular characterization of the *Drosophila* mutant *C31*

The *Drosophila* line *C31* was isolated by Roland Strauss in a behavioral screen for flies with defects in walking (Strauss 1995). *C31* mutant flies show a decaying locomotor activity, walk slower than wild type flies, do not show fast phototaxis, and are inable to fly, probably due to an abnormal held-out wing posture. Furthermore, three of the four neuropilar structures that make up the central brain of *Drosophila* are affected: the ellipsoid body is ventrally opened and kidney-shaped, the fan-shaped body has a dorsal cleft, and the noduli are disordered, whereas the protocerebral bridge of the central complex appears unaltered in *C31* brains. All decribed *C31* phenes are uncovered by the deficiency *Df(1)RA2*, but not by the partially overlapping deficiencies *Df(1)KA14* and *Df(1)GE202*. These results and data from two recombination experiments with regard to the brain defect and the walking impairment of *C31* (Strauss, 1995) map the affected locus between chromosome bands 7E3-4 and 7F1-2 on the X chromosome. The same chromosomal interval was determined for *org-1* (Porsch *et al.,* 1998) making *C31* a candidate for an *org-1* mutant. The molecular analysis indeed revealed a restriction fragment length polymorphism (RFLP) between *C31* and several wild type strains at the *org-1* locus. Cloning and sequencing of this polymorphism revealed that the 3' end of a retrotransposable I element is inserted within the last *org-1* exon (Figure 9) (Porsch, 1997). The I element insertion interrupts the corg-1M2 sequence at position bp 2971. No I element sequences were found at this position in five wild type strains tested including Berlin^{Tue} which was used in the EMS mutagenesis from which *C31* derived (Porsch, 1997).

I elements (or I factors) are LINEs (long interspersed nuclear elements) frequently found in the *Drosophila* genome. Complete I factors contain two long ORFs. ORF1 has similarity to the nucleic acid binding domain of the retroviral Gag polypeptide, whereas the larger ORF2 encodes a putative RNase H with homology to reverse transcriptases (Jensen *et al.,* 1994 and references therein). I elements are devoid of long terminal repeats but have A-rich 3' ends that commonly follow to polyadenylation signals. The 5' end of many I factors is heterogeneously truncated. All these characteristics appear to result from retrotransposition, the mechanism with which these factors propagate in the host genome via an intermediate RNA product and its reverse transcription.

Molecular analysis revealed that the insertion in *org-1* of the *C31* mutant contains all features common to I elements. This factor is incomplete at its 5' end and only contains 1265 bp of the 3' end of full-length I elements (Figure 10). Furthermore, the *org-1/ C31* I element has the conserved 3' sequence CTATCATAA followed by four repeats of TAAA, and is flanked by a duplication of the insertion target site TATACATAT (Figure 10).

```
corg-1M1.seq      ----------------------------------------------------------------------------------
corg-1M2.seq    1 CCACCAGCGCTTGACGGAACGGATGTTGCACAGTGAGTGAGTGAGTGAGTGAATGAGTGAGTGATTGTGTGTGGGTGCAA

corg-1M1.seq      ----------------------------------------------------------------------------------
corg-1M2.seq   81 TAATCAAATTTGACAACTGCGCGAATAACGTTCGAGATCAAAACAAATTGTGCCGGAATAAATGGAATAAATGGGCCAAG

corg-1M1.seq      ----------------------------------------------------------------------------------
corg-1M2.seq  161 AACTCGCTGCGATTTCCTTTCTTCCACTGTGAGCATCACATATCCAGCCATCATATATATACATATATATATATATTCCA

corg-1M1.seq    1 -------------------------------------------------------------------------TGCAGTT
corg-1M2.seq  241 ACGATCACGCTCGCCATGACGCACCTGATGGGCCCCACTGAGTGCGCCGGCGCCATGATGACCACCACATCCATGCAGTT

corg-1M1.seq    8 CCTGGACACCAGCCTAACGGACTACAACTGCTATGGCAACGACTACTGGACATCGCCGTACATGACCGGCGGACTCAGTC
corg-1M2.seq  321 CCTGGACACCAGCCTAACGGACTACAACTGCTATGGCAACGACTACTGGACATCGCCGTACATGACCGGCGGACTCAGTC

corg-1M1.seq   88 CCATGAAGCAGATCGAAGCCTGCATCCAAACGGCTGGCAAGGATCGCAGCTCGTACAAGCCGCTGGAGCAGATCGATGCC
corg-1M2.seq  401 CCATGAAGCAGATCGAAGCCTGCATCCAAACGGCTGGCAAGGATCGCAGCTCGTACAAGCCGCTGGAGCAGATCGATGCC

corg-1M1.seq  168 AAATTGGCGGACATCGAGACGCACAGTACAGGCAGCACTGGCACCGCGAACAGCAACAGCAGCACCAGCAGCATCTCGAA
corg-1M2.seq  481 AAATTGGCGGACATCGAGACGCACAGTACAGGCAGCACTGGCACCGCGAACAGCAACAGCAGCACCAGCAGCATCTCGAA

corg-1M1.seq  248 TCCCAGTTGCCCGGATCAGTCGTCGTCGTCGTCATCGTCGTCCGTATCGCTGCCAACCGATTATGCCGGCGTACACAGTG
corg-1M2.seq  561 TCCCAGTTGCCCGGATCAGTCGTCGTCGTCGTCATCGTCGTCCGTATCGCTGCCAACCGATTATGCCGGCGTACACAGTG

corg-1M1.seq  328 AAGCCTCGATGGCACCAACAGCCGGCGGCACGGCAGTCACAACGACATCAGCTGGCGGAGTCAGTGCATCCACCGCGTCC
corg-1M2.seq  641 AAGCCTCGATGGCACCAACAGCCGGCGGCACGGCAGTCACAACGACATCAGCTGGCGGAGTCAGTGCATCCACCGCGTCC

corg-1M1.seq  408 AAAAAGTTCAAGGGACAGCACAAAAAAGACAACAACAGTGCGGAGAACGGTACAGTGAAGCCCAATAGCCATAATATCAG
corg-1M2.seq  721 AAAAAGTTCAAGGGACAGCACAAAAAAGACAACAACAGTGCGGAGAACGGTACAGTGAAGCCCAATAGCCATAATATCAG

corg-1M1.seq  488 CAAAGGTGAATCGGAGCCAGTGCATCCATCGCTGGCCCAGGCCATTGTGGTGCTGGAGACGAAGGCGCTGTGGGATCAGT
corg-1M2.seq  801 CAAAGGTGAATCGGAGCCAGTGCATCCATCGCTGGCCCAGGCCATTGTGGTGCTGGAGACGAAGGCGCTGTGGGATCAGT

corg-1M1.seq  568 TCCATGCCCAGGGCACCGAAATGATCATCACCAAGACGGGGCCGACGCATGTTTCCCACGTTTCAGGTGAGGATCGGTGGT
corg-1M2.seq  881 TCCATGCCCAGGGCACCGAAATGATCATCACCAAGACGGGGCCGACGCATGTTTCCCACGTTTCAGGTGAGGATCGGTGGT

corg-1M1.seq  648 TTGGATCCACATGCCACCTACATTTGCATGATGGACTTTGTGCCCATGGATGACAAACGCTATCGCTACGCCTTTCACAA
corg-1M2.seq  961 TTGGATCCACATGCCACCTACATTTGCATGATGGACTTTGTGCCCATGGATGACAAACGCTATCGCTACGCCTTTCACAA

corg-1M1.seq  728 CTCCTGCTGGGTGGTGGCTGGCAAGGCGGATCCCATTTCCCCGCCCAGGATTCATGTGCATCCCGACTCGCCAGCCGTCG
corg-1M2.seq 1041 CTCCTGCTGGGTGGTGGCTGGCAAGGCGGATCCCATTTCCCCGCCCAGGATTCATGTGCATCCCGACTCGCCAGCCGTCG

corg-1M1.seq  808 GCTCCAATTGGATGAAGCAGATCGTGTCCTTTGACAAAATTGAAGCTCACCAATAACCAGCTGGACGAAAATGGACATATC
corg-1M2.seq 1121 GCTCCAATTGGATGAAGCAGATCGTGTCCTTTGACAAAATTGAAGCTCACCAATAACCAGCTGGACGAAAATGGACATATC

corg-1M1.seq  888 ATTCTGAACTCCATGCATCGCTACCAGCCGCGTTTCCATCTGGTTTATCTGCCACCGAAGAACGCCTCCTTGGATGAGAA
corg-1M2.seq 1201 ATTCTGAACTCCATGCATCGCTACCAGCCGCGTTTCCATCTGGTTTATCTGCCACCGAAGAACGCCTCCTTGGATGAGAA

corg-1M1.seq  968 CGAGCACTCCAGCCACTTTCGCACTTTCATCTTTCCGGAGACGAGCTTTACGGCCGTAACTGCCTACCAGAATCAGCGGG
corg-1M2.seq 1281 CGAGCACTCCAGCCACTTTCGCACTTTCATCTTTCCGGAAACGAGCTTTACGGCCGTAACTGCCTACCAGAATCAGCGGG

corg-1M1.seq 1048 TGACACAGCTGAAGATCTCCAGCAATCCATTCGCCAAAGGCTTTCGGGATGATGGCACCAACGATGTAACCACTGGCGGT
corg-1M2.seq 1361 TGACACAGCTGAAGATCTCCAGCAATCCATTCGCCAAAGGCTTTCGGGATGATGGCACCAACGATGTAACCACTGGCGGT

corg-1M1.seq 1128 GGCAGCAGCATGTCCTCCATGAGTCACGAAAGTCAGGCGCGCATGAAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCA
corg-1M2.seq 1441 GGCAGCAGCATGTCCTCCATGAGTCACGAAAGTCAGGCGCGCATGAAGCAGCAACAGCAGCAACAGCAGCAGCAGCAGCA

corg-1M1.seq 1208 GCAGCAGCAACTGCAGCAGCAACAGCAACAGCAGCAGCAACTCAAGGAGCGAACGGCAGCAACCAGCAACTTTGGCCTGA
corg-1M2.seq 1521 GCAGCA---ACTGCAGCAGCAACAGCAACAGCAGCAGCAACTCAAGGAGCGAACGGCAGCAACCAGCAACTTTGGCCTAA

corg-1M1.seq 1288 GTTGCAGCGAACTGGCCATTGAGCAACAGCAGCAGCAGCACAGCAACAGGGAGTTCTGCAGCTACCGGCCACGCCCTCC
corg-1M2.seq 1598 GTTGCAGCGAACTGGCCATTGAGCAACAGCAGCAGCAGCAACAGCAACAGGGAGTTCTGCAGCTACCGGCCACGCCCTCC

corg-1M1.seq 1368 AGCAGCTCCACCTCCGGCAATTCACCCGACTTGCTGGGCTACCAGATGGAGCAGCAACTGCAACAGCAACACCAACAGCA
corg-1M2.seq 1678 AGCAGCTCCACCTCCGGCAATTCACCCGACTTGCTGGGCTACCAGATGGAGCAGCAACTGCAACAGCAACACCAACAGCA

corg-1M1.seq 1448 GCAGCAACAGCAACACCAGTCCCAGCAGCAACATCTCCACCAGCAACACCAGGCTAACCAGCAACAATCGCTGCTCCAAC
corg-1M2.seq 1758 GCAGCAACAGCAACACCAGTCCCAGCAGCAACATCTCCACCAGCAACACCAGGCTAACCAGCAACAATCGCTGCTCCAAC

corg-1M1.seq 1528 AGAGCCAGAATCACACGCAATATGGCAGCTATCATCATGCGTACCAGGCACAGGTGCAGTCGCATCCCCTAACGCCGCAC
corg-1M2.seq 1838 AGAGCCAGAATCACACGCAATATGGCAGCTATCATCATGCATACCAGGCACAGGTGCAGTCGCATCCCCTAACGCCGCAC

corg-1M1.seq 1608 TCCAGCAGCTCCGCATCCCCGCCAGCAACTGCTGCGCCGGCCGCAAGTGCAGCAACAGCAGCAGTAGCAGCAGCAGCAGC
corg-1M2.seq 1918 TCCAGCAGCTCCGCATCCCCGCCAGCAACTGCTGCGCCGGGACGCAAGTGCAGCAACAGCAGCAGTAGCAGCAGCAGCAGC

corg-1M1.seq 1688 AGCAGCAGCAGCCGTAGCAGGGGGAGGAGCAGGAGCAGCAGGAGCAGGAGGAGCAACGTCAGCCACACAAGTGATGAGTG
corg-1M2.seq 1998 AGCAGCAGCAGCCGTAGCAGGGGGAGGAGCAGGA---------GCAGCGGGAGCAACGTCAGCCACACAAGTGATGAGTG

corg-1M1.seq 1768 CGGCCAATATCTACTCGAGCATTGGACAACCTATGCCCAGGAGCAGAGCAACTTCGGTGCGATCTACCATCATAATGCT
corg-1M2.seq 2069 CGGCCAATATCTATTCGAGCATTGGACAACGTATGCCCAGGAGCAGAGCAACTTTGGTGCGATCTACCATCATAATGCT

corg-1M1.seq 1848 GCTGCCGCAGCGGCCGCTCACTATCATCATGGTCATGCCCACGGCCACGCCCATAGCCATGCGCACGCCCACGCCCATGG
corg-1M2.seq 2149 GCTGCCGCAGCGGCCGCTCACTATCATCATGGTCATGCCCACGGCCACGCCCATAGCCATGCGCACGCCCACGCCCATGG

corg-1M1.seq 1928 ACCATATGCCAGCGCCTACGACAAGCTGAAGGTGTCGCGTCATGCGGCAGCTGCCGCCTATGGCATGGGCGCCACCTATC
corg-1M2.seq 2229 ACCATATGCCAGCGCCTACGACAAGCTGAAGGTGTCGCGTCATGCGGCAGCTGCCGCCTATGGCATGGGCGCCACCTATC

corg-1M1.seq 2008 CAAGTTTTTTACGGTTCGGCTGCACATCACCAGATGATGCGACCGAATAGCTACATAGATCTGGTGCCGCGCTAAGGGAGC
corg-1M2.seq 2309 CAAGTTTTTTACGGTTCGGCTGCACATCACCAGATGATGCGACCGAATAGCTACATAGATCTGGTGCCGCGCTAAGGGAGC
```

```
corg-1M1.seq 2088 ACCAACTGGGAAGAGAAGGATTTCGGATTTCGGAT-------ACTCTAAGCAATTAACTGTA----------------A
corg-1M2.seq 2389 AGCAACTTGGAAGAGAAGGATTTCGGATTTCGGATTTCGGATACTCTATGGAATTAACTGCACTTACACTTGCCTGTAAA

corg-1M1.seq 2144 AATTATTGTAAAATCCAAACAATGACGTAGTCTATCATCCATAGCCAAAGCTATACATATA----TATGTGTAAATCTCA
corg-1M2.seq 2469 AATGATTGTAAAATCCAAACTTAGAGTACGTC----ATCTATAGCCAAAGCTATACATATACATATATGTGTAAATCTCA

corg-1M1.seq 2220 TGCCAAAGATTCGTTCTAAAATCAAGAATCTATTTCCAAAGTTTAGAAAGGAAGCCTTTAATTTTCGCCCATTAAAAAAT
corg-1M2.seq 2545 TGCCAAAGATTCGTTCTAAAATCAAGAATCTATTTCCAAAGTTTAGAAAGGAAGCCTTTAATTTTCGCCCATTAAAAAAT

corg-1M1.seq 2300 GTTTTAACAAAACAAAAACATAACTAAGCTTAAGCCAAAACTATAATAACAGGAATTATTTTTTAGCAAGCTTAATTTTT
corg-1M2.seq 2625 GTTTTAACAAAACAAAAACATAACTAAGCTTAAGCCAAAACTATAATAACAGGAATTATTTTTTAGCAAGCTTAATTTTT

corg-1M1.seq 2380 AAGCATTCAAATTCATTCTTTCGCGAAACATTTGGAATTTGGAGCGATTTGATTCTTGATTTTAGAATCAATTTCAAGTA
corg-1M2.seq 2705 AAGCATTCAAATTCATTCTTTCGCGAAACATTTGGAATTTGGAGCGATTTGATTCTTGATTTTAGAATCAATTTCAAGTA

corg-1M1.seq 2460 TTAGCAGCCAGAAAACCAAAAATAAATGCAACAAGTATTACAAGTATTTCTACATACAAAAATTACCATTAAAAGTTAAA
corg-1M2.seq 2785 TTAGCAGCCAGAAAACCAAAAATAAATGCAACAAGTATTACAAGTATTTCTACATACAAAAATTACCATTAAAAGTTAAA

corg-1M1.seq 2540 ATATTTTTTTTTT-CTAGCTTAGGACGTAAATTTTATTGATTTGTGTGAAACTGAAAACGCATAAAACATTTCGATGTAA
corg-1M2.seq 2865 ATATTTTTTTTTTTCTAGCTTAGGACGTAAATTTTATTGATTTGTGTGAAACTGAAAACGCATAAAACATTTCGCTGTAA

corg-1M1.seq 2619 ACTGTAGTGTAATTTTAATATACATATTATTATT----TTTTTTTTTTTTTGCTTAACACTCTAGGTTTTTTTTTCTATGT
corg-1M2.seq 2945 ACTGTAGTGTAATTTTAATATACATATTATTATTATTATTTTTTTTTTTTTTGCTTAACACTCTAGGTTTTTTTTTCTATGT

corg-1M1.seq 2695 AAATACAAGTACATATGTATGTCGCTATATATATATATATATATAT------TTAAGAACTGCAACAGTTTCAAGCAATA
corg-1M2.seq 3025 AAATACAAGTACATATGTATGTCGCTATATATATATATATATATATATATATTTAAGAACTGCAACAGTTTCAAGCAATA

corg-1M1.seq 2769 AAAACAAAGAAAATTTTAAACCGAAACTCTAGCAAACAGAAGCATAAATTAACC----------
corg-1M2.seq 3105 AAAACAAAGAAAATTTTAAACCGAAACTCTAGCAAACAGAAGCATAAATTAACCAAAAAAAAAA
```

Figure 6. Alignment of corg-1M1 and corg-1M2 sequences.

Polymorphic sites within the *org-1* cDNAs are highlighted.

```
ORG-1M1   1  --------------------QFLDTSLTDYNCYGNDYWTSPYMTGGLSPMKQIEACIQTA
ORG-1M2   1  MTHLMGPTECAGAMMTTTSMQFLDTSLTDYNCYGNDYWTSPYMTGGLSPMKQIEACIQTA

ORG-1M1  41  GKDRSSYKPLEQIDAKLADIETHSTGSTGTANSNSSTSSISNPSCPDQSSSSSSSSSVSLP
ORG-1M2  61  GKDRSSYKPLEQIDAKLADIETHSTGSTGTANSNSSTSSISNPSCPDQSSSSSSSSSVSLP

ORG-1M1 101  TDYAGVHSEASMAPTAGGTAVTTTSAGGVSASTASKKFKGQHKKDNNSAENGTVKPNSHN
ORG-1M2 121  TDYAGVHSEASMAPTAGGTAVTTTSAGGVSASTASKKFKGQHKKDNNSAENGTVKPNSHN

ORG-1M1 161  ISKGESEPVHPSLAQAIVVLETKALWDQFHAQGTEMIITKTGRRMFPTFQVRIGGLDPHA
ORG-1M2 181  ISKGESEPVHPSLAQAIVVLETKALWDQFHAQGTEMIITKTGRRMFPTFQVRIGGLDPHA

ORG-1M1 221  TYICMMDFVPMDDKRYRYAFHNSCWVVAGKADPISPPRIHVHPDSPAVGSNWMKQIVSFD
ORG-1M2 241  TYICMMDFVPMDDKRYRYAFHNSCWVVAGKADPISPPRIHVHPDSPAVGSNWMKQIVSFD

ORG-1M1 281  KLKLTNNQLDENGHIILNSMHRYQPRFHLVYLPPKNASLDENEHSSHFRTFIFPETSFTA
ORG-1M2 301  KLKLTNNQLDENGHIILNSMHRYQPRFHLVYLPPKNASLDENEHSSHFRTFIFPETSFTA

ORG-1M1 341  VTAYQNQRVTQLKISSNPFAKGFRDDGTNDVTTGGGSSMSSMSHESQARMKQQQQQQQQQQ
ORG-1M2 361  VTAYQNQRVTQLKISSNPFAKGFRDDGTNDVTTGGGSSMSSMSHESQARMKQQQQQQQQQQ

ORG-1M1 401  QQQQQLQQQQQQQQQLKERTAATSNFGLSCSELAIEQQQQQQQQQQGVLQLPATPSSSSTS
ORG-1M2 421  QQQQ-LQQQQQQQQQLKERTAATSNFGLSCSELAIEQQQQQQQQQQGVLQLPATPSSSSTS

ORG-1M1 461  GNSPDLLGYQMEQQLQQQHQQQQQQQHQSQQQHLHQQHQANQQQSLLQQSQNHTQYGSYH
ORG-1M2 480  GNSPDLLGYQMEQQLQQQHQQQQQQQHQSQQQHLHQQHQANQQQSLLQQSQNHTQYGSYH

ORG-1M1 521  HAYQAQVQSHPLTPHSSSSASPPATAAPGASAATAAVAAAAAAAAAVAGGGAGAGAGGA
ORG-1M2 540  HAYQAQVQSHPLTPHSSSSASPPATAAPGASAATAAVAAAAAAAAAVAGGGAGA---GGA

ORG-1M1 581  TSATQVMSAANIYSSIGQPYAQEQSNFGAIYHHNAAAAAAAHYHHGHAHGHAHSHAHAHA
ORG-1M2 597  TSATQVMSAANIYSSIGQPYAQEQSNFGAIYHHNAAAAAAAHYHHGHAHGHAHSHAHAHA

ORG-1M1 641  HGPYASAYDKLKVSRHAAAAAYGMGATYPSFYGSAAHHQMMRPNSYIDLVPR
ORG-1M2 657  HGPYASAYDKLKVSRHAAAAAYGMGATYPSFYGSAAHHQMMRPNSYIDLVPR
```

Figure 7. Sequence alignment of ORG-1M1 and ORG-1M2.

Predicted ORG-1 peptide sequences from corg-1M1 and corg-1M2 are aligned and polymorphisms are highlighted.
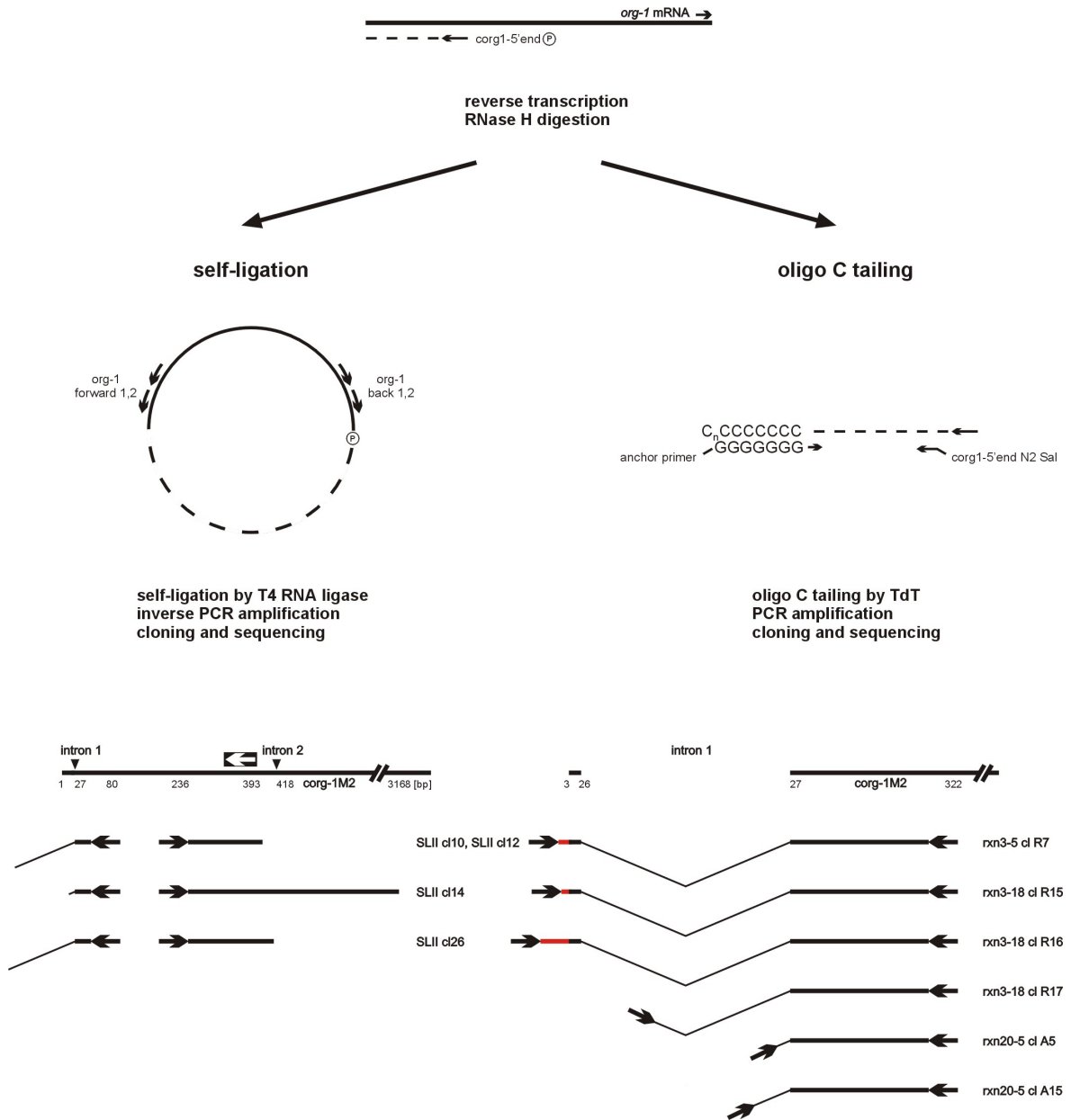
Figure 8. *org-1* 5' RACE.
org-1 5' RACE was performed using self-ligation (left) or oligo C tailing (right) of the synthesized *org-1* cDNA. Analyzed amplificates are shown below schematic drawings of the 5' end of corg-1M2. Arrows indicate primers, thick black lines corg-1M2 sequence, thin lines *org-1* intronic sequences. Red lines represent 5' sequence extensions of the *org-1* transcript. The white arrow above the corg-1M2 sequence marks the expected annealing site for corg1-5'end which was used to prime reverse transcription.
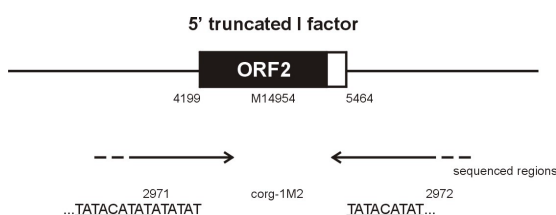
Figure 10. *C31* I element insertion within *org-1*.
The I element in *C31* within the *org-1* gene is 5' truncated containing bp 4199-5464 of a complete I factor (Genbank accession number: M14954; Fawcett *et al.,* 1986). The sequenced region of the *C31* I element is indicated below. The duplicated I element target site is underlined. The I element interrupts the corg-1M2 sequence at position bp 2971.

The *org-1* I element insertion in the *C31* mutant and its absence in 5 wild type strains strongly suggested that this polymorphism is specific for *C31* and that it might be causative for the *C31* syndrome.

3.3 Generation of ORG-1 antisera

Recombinant ORG-1$_{17-708}$ was expressed and purified as His-tag fusion protein in *E. coli* cells using the pET Expression System 15b [Novagen].
Therefore, a 692 aa long ORF lacking the first 16 aa of full-length ORG-1 was *Pfu* PCR amplified from the *org-1* cDNA corg-1M1 and cloned into pET 15b via *Xho* I in frame with a preceding His-tag (Porsch, 1997). The resulting clone pETcorg1 was transformed into *E. coli* BL (DE3). Expression of His-ORG-1$_{17-708}$ was induced by the addition of IPTG to the culture of transformed bacteria (Porsch, 1997). Bacterial cells were harvested, sonicated and centrifuged. ORG-1 protein was located within inclusion bodies which were finally resuspended in 1x binding buffer containing 6 M urea. His-ORG-1$_{17-708}$ was subsequently purified by Ni^{2+} column chromatography under denaturing conditions without complications essentially as described in the pET System Manual [Novagen]. Bound protein was eluted from the Ni^{2+} column with a 100 mM –1 M gradient of imidazol/ 6 M urea. Most of the recombinant His-ORG-1$_{17-708}$ eluted at 200-300 mM imidazol and was collected in fractions 4-6. These fractions à 5 ml contained homogeneously purified, denatured His-ORG-1$_{17-708}$ in the following concentrations:

fraction 4:	1,12 mg/ml
fraction 5:	400 µg/ml
fraction 6:	100 µg/ml

Six mice and a rabbit were immunized against purified, denatured His-ORG-1$_{17-708}$ with subcutaneous injections of 120 µl (mice) or 500 µl (rabbit) containing 1x adjuvance antibody-multiplier (ABM-S, ABM-ZK or ABM-N) [Linaris] and different amounts of recombinant ORG-1:

mouse 1	2 µg ORG-1
mouse 2	4 µg ORG-1
mouse 3	6 µg ORG-1
mouse 4	8 µg ORG-1
mouse 5	10 µg ORG-1
mouse 6	10 µg ORG-1
rabbit	100 µg ORG-1

The animals were boosted in intervals of three weeks 7-8 times prior to the final bleeding, except mouse 2 which died earlier. The blood was allowed to clot at room temperature and subsequently centrifuged. Sera were aliquotted and stored at -20°C. Six aliquots à 10 µl per mouse serum and about 30 aliquots à 500 µl of the rabbit antiserum are stored at -20°C in Matze's freezer, box "antisera". ORG-1 antisera were immunoreactive on Western blots with recombinant ORG-1 and *Drosophila* protein extracts. Dot blot analysis indicated that the mouse ORG-1 antisera recognize recombinant ORG-1 in 100 times higher dilutions than the rabbit antiserum. ORG-1 antisera have not been used yet for immunohistochemistry to determine the expression pattern of ORG-1 in *Drosophila*.

3.4 Chromosomal mapping of human *TBX1*

A BLAST search of the Genbank/ EMBL database with the *org-1* cDNA sequence revealed among numerous T-box genes also a closely related human EST clone, C22_821 (Genbank accession number: H55663) (Porsch, 1997). Clone C22_821 contains 172 bp encoding part of the T domain of human *TBX1* and derived by exon amplification from a flow-sorted human chromosome 22 cosmid library (Trofatter *et al.,* 1995; Porsch, 1997). Since *TBX1* is the putative human ortholog of *org-1*, we wanted to learn more about this gene, in particular, if *TBX1* like the homologous genes *TBX3* or *TBX5* might be associated with human syndromes. As a first step towards this goal, we intended to confirm and to refine the localization of *TBX1* on chromosome 22. Therefore, a pair of oligonucleotides was designed that was able to discriminate between human *TBX1* and the very similar mouse ortholog. *TBX1* was unambiguously mapped to chromosome 22 in a PCR analysis of a panel of human x ham-

ster hybrid cell lines. The chromosomal sublocalization was performed with hamster x human and mouse x human hybrid cell lines containing chromosome 22 translocation products, placing human *TBX1* in 22q11 (Figure 11; Porsch *et al.,* 1998). These experiments were carried out in the laboratory of Bernhard Weber, Würzburg.
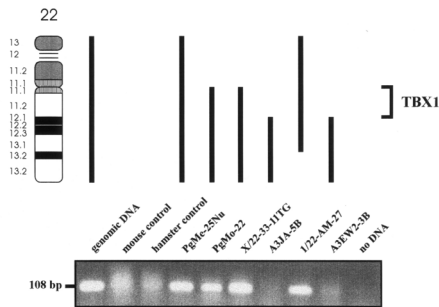
Figure 11. Localization of the human *TBX1* gene using a chromosome 22 hybrid panel.
The human DNA content retained in each hybrid and the probable breakpoints of chromosome 22 are indicated by vertical bars. The absence or presence of the 108 bp PCR product suggests a localization of *TBX1* in chromosomal region 22q11 (this figure was done by Bernhard Weber; taken from Porsch *et al.,* 1998).

Most interestingly, deletions involving 22q11 are associated with more than 80 different birth defects or malformations occurring in many combinations and with variable expressivity (Scambler, 2000). These symptomes are linked with several diagnostic syndromes including DiGeorge syndrome (DGS), velocardiofacial syndrome (VCFS), conotruncal anomaly face, Cayler syndrome and Opitz GBBB syndrome (Scambler, 2000; Emanuel *et al.,* 1998). Clinical features of these dominant syndromes largely overlap, suggesting that they are variable outcomes of the same underlying genetic defect. These syndromes are collectively referred to as 22q11 deletion syndrome (22q11 DS).

Most 22q11 DS patients have an interstitial deletion of about 3 Mb. The overlap of such deletions defines an approximately 750 kb large DGS chromosomal region (Scambler, 2000). Our analysis mapped *TBX1* to the center of the DGS chromosomal region.

We concluded from the chromosomal location of *TBX1* and the homology to the haploinsufficient genes *T*, *TBX3*, and *TBX5* that *TBX1* might be a candidate gene for DGS/ VCFS. To investigate a possible role of *TBX1* in 22q11 DS, we set up preparations for the cloning of human *TBX1* and for

a mutation analysis of DGS patients without cytologically visible deletions. In the course of this work, however, Chieffo *et al*. (1997) reported a detailed molecular study on human *TBX1*. We, therefore, stopped our own investigation.

Recently, three groups indepently showed that *Tbx1* mouse mutants diplay developmental anomalies that model the symptomes of DGS/ VCFS patients, indicating that *TBX1* in humans is a key gene in the etiology of DGS/ VCFS (Jerome and Papaioannou, 2001; Lindsay *et al.,* 2001; Merscher *et al.,* 2001).

3.5 The *Drosophila vmd2* gene

Vitelliform macular dystrophy (VMD2), also known as Best's disease, is an autosomal dominant disorder with a juvenile onset of macular degeneration that causes progressive loss of visual acuity in affected patients (Best, 1905; Marquardt *et al.,* 1998). Genetic linkage analysis placed the VMD2 disease locus within a 980 kb interval on chromosome 11q12-q13.1, the Best's disease critical region. This region was cloned and systematically analyzed for transcripts and, subsequently, for the presence of mutations in VMD2 patients (Marquardt *et al.,* 1998 and references therein). Indeed, one of the characterized genes, initially termed *TU15b*, is exclusively expressed in the retina pigment epithelium and was mutated in all VMD2 patients tested (Marquardt *et al.,* 1998). Therefore, the *TU15b* gene was renamed *VMD2*. *VMD2* encodes a predicted protein of 585 aa with significant sequence similarity to several putative proteins from *C. elegans*, *Drosophila* and mouse, indicating that VMD2 is a member of a conserved protein family. Since molecular work on VMD2 proteins has not been perfomed yet, nothing is hitherto known about the molecular function of VMD2 or its homologs (Marquardt *et al.,* 1998).

This prompted Bernhard Weber, Würzburg, to initiate a functional analysis of the *Drosophila* homolog of *VMD2* in cooperation with our laboratory. Bernhard Weber identified two *Drosophila* EST clones, LD 22528 and GH 28445, encoding the *Drosphila vmd2* gene. We ordered both clones and completely sequenced the larger clone LD 22528 on both strands. LD 22528 contains a *vmd2* cDNA of 2862 bp encoding a predicted polypeptide of 721 aa (Figure 12).

BLASTP searches of the SwissProt protein database with the *Drosophila* VMD2 peptide sequence found human VMD2 as the most closely related peptide sequence.

```
   1  gagcgcggacgtgagcatgt atttctgtttgagtgtgtgt gagtgttagtgtttgtgtaa gaagttcggcggcaacgaaa acgtaaaatagtgaagcata  100
 101  aaggcacaaagtgaagaaat actcgcacataaaccgatgt tagtgtgtttgtctaagccc ttctacctctttttttgcta cctgccaatttgttaacttt  200
 201  attgttgctaccgcttgcgt gccgtgaatcaaagtaacaa caaccgccacaacaacaaca tgcacaaat... aaatgtgaaga gtggaactttcattttcgac  300
 301  aaacaacaatgtgtgagacg cgataagacagtcggaaaga agaacagcaatctcagctat aaaagagcactataaacaaa ctaaattggaggtggaatta  400
 401  aaataggagagaacaatgac aattacgtacacaggtgaag tggccacttgtcgcggcttt ggctgttttctcaaattgct gctcagatggcgaggaagca  500
   1                      M  T  I  T  Y  T  G  E  V   A  T  C  R  G  F   G  C  F  L  K  L  L   L  R  W  R  G  S  I   29

 501  tttacaaactggtttggcta gatcttctggccttcttgac catttactatgcgatcaaca tggtgtatcgctttggcctc aaccccgcacaaaagaaac  600
  30     Y  K  L  V  W  L   D  L  L  A  F  L  T   I  Y  Y  A  I  N  M   V  Y  R  F  G  L   N  P  A  Q  K  E  T   62

 601  ctttgaggccattgttcagt actgtgatagttacagagaa ctcatacccctgtccttcgt gcttggtttctatgtatcga ttgtgatgacccgttggtgg  700
  63     F  E  A  I  V  Q  Y   C  D  S  Y  R  E   L  I  P  L  S  F  V   L  G  F  Y  V  S  I   V  M  T  R  W  W   95

 701  aatcagtacacctccattcc ctggccagatcccatcgccg tgtttgtcagctcgaatgtc catggccaggatgagcgagg acgcatgatgaggcgaacaa  800
  96     N  Q  Y  T  S  I  P   W  P  D  P  I  A  V   F  V  S  S  N  V   H  G  Q  D  E  R  G   R  M  M  R  R  T  I  129

 801  taatgcgatatgtgtgcctt tgcctgacgatggtcctggc gaatgtttcgccgagggtga agaagcgtttccccggccta aataatctggtggaagcggg  900
 130     M  R  Y  V  C  L   C  L  T  M  V  L  A   N  V  S  P  R  V  K   K  R  F  P  G  L   N  N  L  V  E  A  G  162

 901  tctgctaaatgacaatgaaa agaccatcatcgagaccatg aacaaggcctttcccagacc ttcgaagcactggctgccca tcgtttgggctgccagtatt 1000
 163     L  L  N  D  N  E  K   T  I  I  E  T  M   N  K  A  F  P  R  P   S  K  H  W  L  P  I   V  W  A  A  S  I  195

1001  ataaccagggccagaaagga aggtcgcattcgtgatgatt ttgctgtgaagaccatcatc gatgagctaaataagtttcg tggtcagtgtggactcctca 1100
 196     I  T  R  A  R  K  E   G  R  I  R  D  D  F   A  V  K  T  I  I   D  E  L  N  K  F  R   G  Q  C  G  L  L  I  229

1101  tcagctacgataccattagt gtacctctggtgtacaccca agtggtgaccctggcggtgt attcgtacttccttacctgc tgcatgggtcaacaatggac 1200
 230     S  Y  D  T  I  S   V  P  L  V  Y  T  Q   V  V  T  L  A  V  Y   S  Y  F  L  T  C   C  M  G  Q  Q  W  T  262

1201  cgatggcaaggtggtgggca ataccacatacctgaacaag gtggatctatactttcctgt atttacaacgctgcagttct tcttctacatgggttggctc 1300
 263     D  G  K  V  V  G  N   T  T  Y  L  N  K   V  D  L  Y  F  P  V   F  T  T  L  Q  F  F   F  Y  M  G  W  L  295

1301  aaggtggccgagtcgctgat aaatccatttggcgaagacg atgatgattttgaggtcaac tggatggtggatcgcaatct tcaggtgtcctatctgatcg 1400
 296     K  V  A  E  S  L  I   N  P  F  G  E  D  D   D  D  F  E  V  N   W  M  V  D  R  N  L   Q  V  S  Y  L  I  V  329

1401  tcgacgagatgcaccatgac catccggagctgttaaagga tcagtactgggacgaggtgt tccccaacgagctgccctac acaatagctgccgaacgatt 1500
 330     D  E  M  H  H  D   H  P  E  L  L  K  D   Q  Y  W  D  E  V  F   P  N  E  L  P  Y   T  I  A  A  E  R  F  362

1501  ccgggagaatcatccagagc cgtccactgccaagatcgag gtgcccaagaatgcggccat gccatcgacaatgtcgtccg ttcgcatcgatgaaatggcc 1600
 363     R  E  N  H  P  E  P   S  T  A  K  I  E   V  P  K  N  A  A  M   P  S  T  M  S  S  V   R  I  D  E  M  A  395

1601  gatgatgccagtggcattca cttctcagctggaaatggca aaatgcgcctggattcctcg ccctcgctggtgagcgtttc gggaactctatcccgggtga 1700
 396     D  D  A  S  G  I  H   F  S  A  G  N  G  K   M  R  L  D  S  S   P  S  L  V  S  V  S   G  T  L  S  R  V  N  429

1701  atacggtggcctcggccctc aaacgtttcctgagccgcga cgatagcaggccgggatcgg caacgcccagtcaggatcag ccctacaaattcccggccag 1800
 430     T  V  A  S  A  L   K  R  F  L  S  R  D   D  S  R  P  G  S  A   T  P  S  Q  D  Q   P  Y  K  F  P  A  S  462

1801  cgccagctcggcgagtttat cgggtgccgtggtaggatcg gctacatcggccggaaaacc agctggcagtcttaggatta cgcagcaagtgatcgaggag 1900
 463     A  S  S  A  S  L  S   G  A  V  V  G  S   A  T  S  A  G  K  P   A  G  S  L  R  I  T   Q  Q  V  I  E  E  495

1901  gtggacgaacaggcgaccat aacatccatgagagccaatg atccacgtcccaatgtcatg gacatattcgcacaaacctc gtcgggagctggaacctctg 2000
 496     V  D  E  Q  A  T  I   T  S  M  R  A  N  D   P  R  P  N  V  M   D  I  F  A  Q  T  S   S  G  A  G  T  S  G  529

2001  gaccgctgcagccaccaccg gcccactcggaaccggtgga catcccgtcacgtccgccct catacaatcgggcccaatcc cagtacgaacccaacctatt 2100
 530     P  L  Q  P  P  P   A  H  S  E  P  V  D   I  P  S  R  P  P  S   Y  N  R  A  Q  S   Q  Y  E  P  N  L  F  562

2101  tccacctggcggagtggatg cactgctcagtacttcagct cctgcgggcggaagtcccct gctcctgtctaatgcagcca ctgcacccagttcgccagtg 2200
 563     P  P  G  G  V  D  A   L  L  S  T  S  A   P  A  G  G  S  P  L   L  L  S  N  A  A  T   A  P  S  S  P  V  595

2201  ggcgagagctccaagtccct atacgatccacaaaagggcg ccagccgagagacagtggag agcatggacctgaggtcctc cacggatctactcggcgatg 2300
 596     G  E  S  S  K  S  L   Y  D  P  Q  K  G  A   S  R  E  T  V  E   S  M  D  L  R  S  S   T  D  L  L  G  D  A  629

2301  cggcagtgcagcccgaagac gagggcgatgacttcgataa gctgaaggcggaacgcgaga aggagaaactgatgcgacag caaaagaatctggccagaac 2400
```

```
630    A  V  Q  P  E  D   E  G  D  D  F  D  K   L  K  A  E  R  E  K   E  K  L  M  R  Q   Q  K  N  L  A  R  T   662

2401 tattagcaccgctccgggaa tggaagccacggctgtgccg atggtgccaatggtcccagt gaacgtggcagtgcaacagg cacagctgcaaccagttgca 2500

663    I  S  T  A  P  G  M   E  A  T  A  V  P   M  V  P  M  V  P  V   N  V  A  V  Q  Q  A   Q  L  Q  P  V  A   695

2501 tccagtgccgatcttctggc cggcggagatcagttctcca attcgacgatgaaatcggag gacgccatcaacggcagttg aaggacacctagtattttgt 2600

696    S  S  A  D  L  L  A   G  G  D  Q  F  S  N   S  T  M  K  S  E   D  A  I  N  G  S  *                        722

2601 ttcgtacacttacctagttt aagtctagtgcattatttag ttccctaagctgataagcta aattacctatactatattct agacgtatatgccgctgaca 2700

2701 tatactggacatacgcaaat taacggacagtttaagaatg ctcataatgtctaaaacgag ctcgagtgatgatggaccta attaacgcagttaataacac 2800

2801 aaatactaatatctaagtag aaagtttttgaattttacaa atataagttattttgtaaat tg                                          2862
```

Figure 12. Nucleotide sequence of *Drosophila vmd2* clone LD 22528 and the predicted *Drosophila* VMD2 amino acid sequence.

The other EST clone, GH 28445 was end-sequenced only. GH 28445 represents a 5' truncated *vmd2* cDNA (the sequence starts within the fourth *vmd2* exon), but expands the LD 22528 *vmd2* cDNA sequence further 3' by 240 bp (Figure 13). Individual sequences, the final *Drosophila vmd2* sequence, and sequence alignments can be found on the accompanying CD-ROM [DNAseq/*vmd2*]

The *vmd2* cDNA was used for a BLASTN search to identify genomic *Drosophila* sequences. Clones BACR28B01 and BACR39F04 both mapped to 85F, both contain the *vmd2* locus, thus placing the gene to this cytological position on the right arm of the III chromosome. The genomic sequence AC019521 includes the *vmd2* gene, too. The exon-intron architecture of *Drosophila vmd2* was determined by an alignment of the cDNA sequence to the genomic sequence AC019521 (Figure 13). The *Drosophila vmd2* gene is disrupted (at least) by 6 introns into 7 exons. Their sizes are given in Figure 13. We identified four additional *vmd2* EST clones in database searches, clones LD 04433, GH 18342, LP05915, LP07975 (Figure 13). The existence of two EST clones each from embryonic (LD), larval/ early pupal (LP) or adult head (GH) cDNA libraries suggests that *vmd2* is widely expressed during *Drosophila* development.

To eventually find P element insertions associated with *vmd2*, we also screened all available *Drosophila* sequences with the genomic sequence of AC019521 and the BAC clones. Luckily, we indeed identified P element insertion line EP (3) 3668 that carries a modified P element just 102 bp upstream of the known *vmd2* transcript. The inserted P element is of the EP (enhancer-promoter) type that contains at the 3' end a *hsp70* minimal promoter preceded by an enhancer with 14 Gal4 binding sites, allowing the directed expression of neighboring endogenous genes in any temporal or spacial pattern when combined with appropriate Gal4 driver lines (Rørth, 1996). An alignment of the P

element flanking sequence with the AC019521 sequence revealed that the EP element is inserted in 5' → 3' orientation upstream of the *vmd2* transcription unit. Thus, this P element insertion should be of great use for *vmd2* misexpression experiments. Line EP (3) 3668 was subsequently ordered from the Bloomington *Drosophila* stock center. We observed that this line is semi-lethal. We, however, have not determined yet whether the semi-lethality is caused by the P element insertion nor have we done a phenotypic characterization of the mutant, nor P element remobilization experiments in order to obtain further P element insertion alleles at the *vmd2* locus.

Nonetheless, it is evident that the cloning of *vmd2* and the identification of the P insertion line EP (3) 3668 at its 5' region will greatly facilitate a functional analysis of *vmd2* in *Drosophila*.
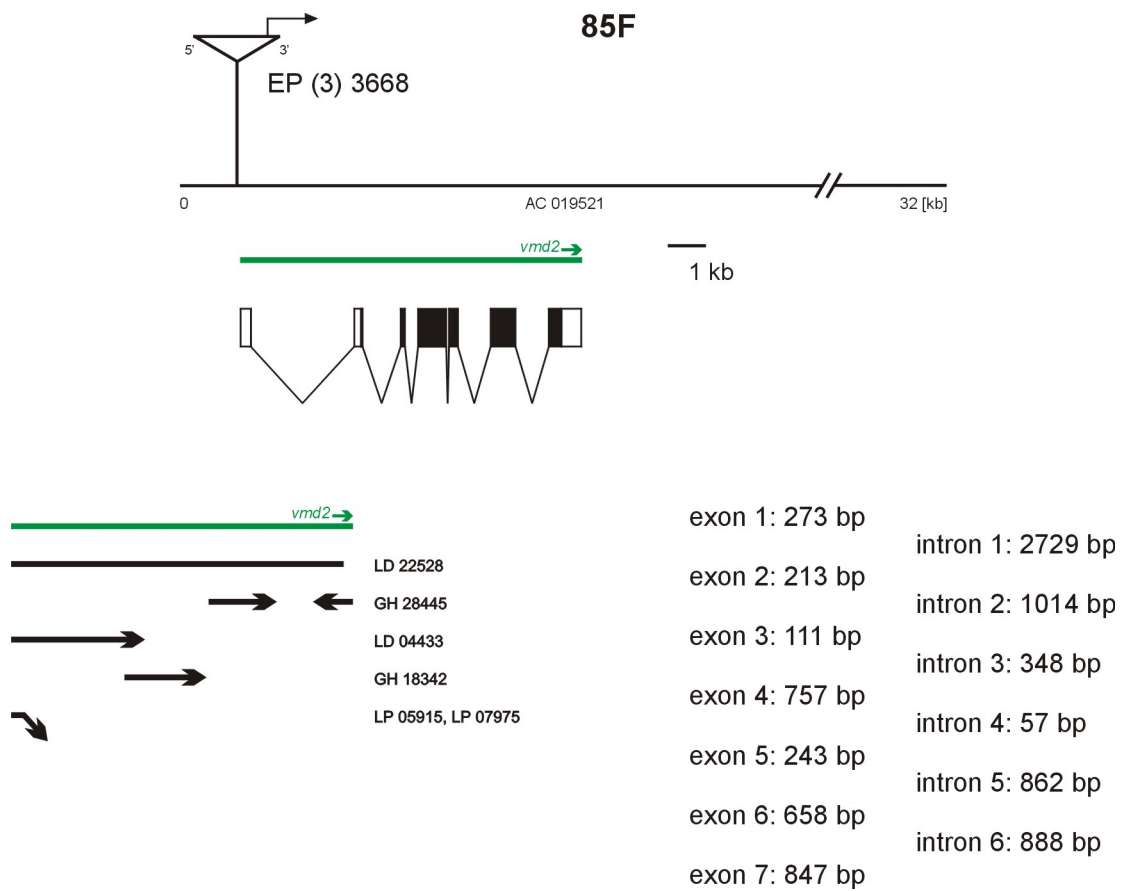
Figure 13. Molecular characterization of *Drosophila vmd2*.
The *vmd2* locus at 85 F is included within genomic sequence AC 019521 (black line). An EP element is inserted in 5'→3' orientation about 100 bp upstream of the known *vmd2* transcript (green line) in line EP (3) 3668. The *vmd2* exon-intron structure is shown with exons indicated by boxes and introns by thin lines. Filled boxes symbolize the *vmd2* coding region. Several EST clones exist for *vmd2*. Black lines indicate sequenced parts of these clones. LP 05915 and LP 07975 include intronic sequences indicated by a kinked line. *vmd2* exon and intron sizes are listed below.

4. *org-1* Genetics

4.1 EMS mutagenesis: Screen for new *C31* alleles

The molecular characterization of the *Drosophila* mutant *C31* revealed that a truncated I element is inserted within the *org-1* gene in this line, but not in several wild type strains (Porsch, 1997). We, therefore, surmised that *C31* would represent the first known *org-1* mutant and intended to isolate *org-1* mutations in a screen for new *C31* alleles, albeit this polymorphism does not provide a direct proof for the I element insertion to be responsible for the *C31* phenotype.

C31 is a X-linked, recessive mutation. It is manifested in hemizygous males and homozygous females in walking defects, structural aberrations in the central brain, and an altered wing posture (Strauss, 1995). In spite of the pleiotropy of this mutant, homozygous *C31/ C31* or deficiency-transheterozygotic *Df(1)RA2/ C31* females are viable and fertile (Strauss, 1995; Gert Pflugfelder, pers. comm.), allowing to directly screen the female offspring of mutagenized males and *C31* females.

Ethyl methanesulfonate (EMS), an alkylating agent that efficiently induces point mutations and, less frequently, chromosomal aberrations at random positions in the genome, was used as a mutagen (Grigliatti, 1986; Ashburner, 1989). *w* males (*y,w* males in round 1) were fed on a sucrose solution containing 25 mM EMS (50 mM EMS in round 4) and subsequently mated to *C31* virgins.
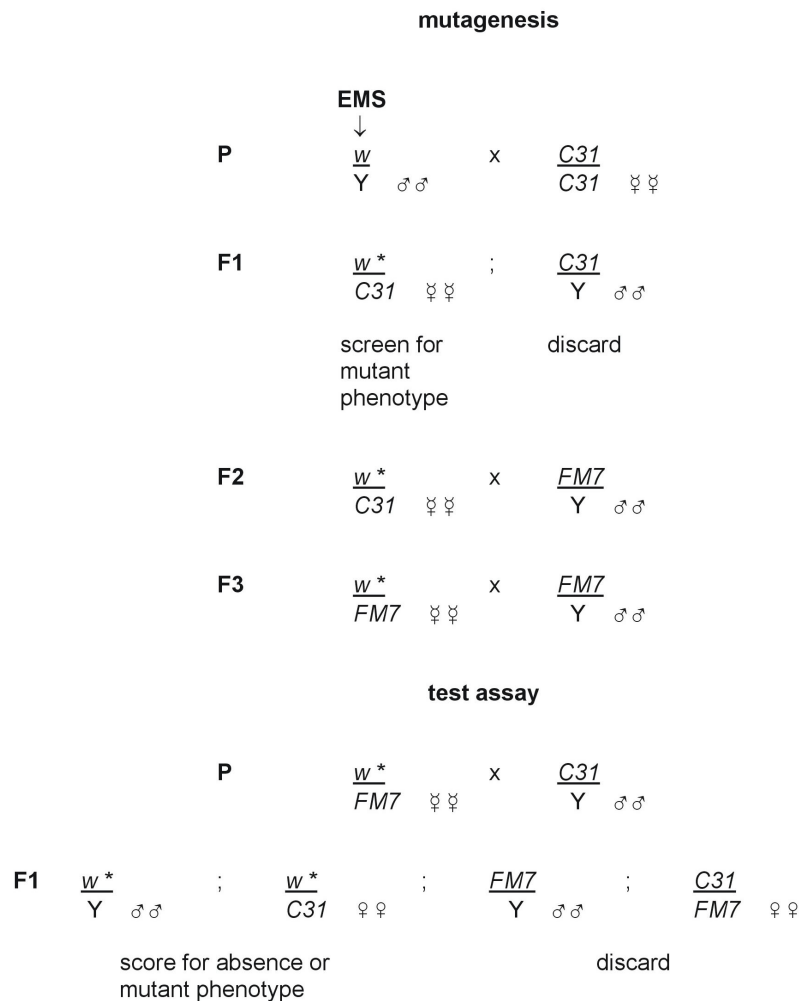
Figure 14. Crossing scheme for the isolation and assay of EMS-induced mutations.
* indicates an EMS-treated chromosome.

The *w*/ *C31* females among the offspring, transheterozygotic for a mutagenized X chromosome and the *C31* mutation, were collected as virgins, aged for several days, and finally screened for two visible phenotypes: the conspicuous "held-out" wing defect and an aberrant bristle pattern on the posterior head. The latter phene is not manifested in *C31* flies, but can be found in deficiency-transheterozygotic *Df(1)RA2/ C31* flies with high penetrance, where the postvertical bristles (PV) are frequently short and thickened and/ or the ocellar and interocellar bristles appear unordered or duplicated (Gert Pflugfelder, pers. comm.).

The scheme for the EMS mutagenesis is shown in Figure 14. The mutagenesis was carried out in 5 consecutive rounds (round 1-5). Their results are summarized in Table 4. About 44.500 F1 females were screened for the *C31* wing phenotype and/ or an affected head bristle pattern. A total of

207 candidates were isolated and mated to FM7 males. 135 balanced stocks (135/207 = 65,2%) could be obtained. These stocks were then assayed by crossing them to *C31* males. 12 stocks of interest were established and are listed in Table 5.

6 of these 12 lines, the mutants 7-1, 7-4, 10-1, 14-2, 14-3, and 41-3, all isolated from round 1, carry a dominant, X-chromosomal wing mutation. Their wings show a V-shaped posture with incised wing tips reminiscent of the *Drosophila* mutant *Notch* (*N*). Interestingly, these dominant mutations seem to interact genetically with *C31*, as their phenotypes are apparently enhanced by heterozygous *C31*.

mutagenesis	screened flies	isolated candidates	established stocks	new *C31* or *org-1* alleles
round 1 I/97	ca 4.000	69	6	0
round 2 I/98	ca 6.000	16	2	0
round 3 II/98	12.232	10	1	0
round 4 III/98	2.932	49	1	0
round 5 IV/98	19.317	63	2	0
rounds 1-5	ca 44.500	207	12	0

Table 4. Genetic data of the EMS mutagenesis.

mutant	stock	GOP stock number	description
7-1	7-1/FM7c	533	held-out wings distally notched, enhanced by *C31*?
7-4	7-4/FM7c	534	held-out wings distally notched, enhanced by *C31*?, recombination analysis suggests N
10-1	10-1/FM7c	535	held-out wings distally notched, enhanced by *C31*?
14-2	14-2/FM7c	536	held-out wings distally notched, enhanced by *C31*?
14-3	14-3/FM7c	537	held-out wings distally notched, enhanced by *C31*?
41-3	41-3/FM7c	538	held-out wings distally notched, enhanced by *C31*?
I-1	PV-1/TM3	636	PVs missing; posteriormost central tergite bristles missing, dominant
I-2	l(1) *Matze*/FM7a	611	all large dorsal head bristles missing, dominant
II-26	627/FM7a	627	recessive, semi-lethal, ocellar and interocellar bristles missing, 1st ocellus missing, dominant
III-44	PV-2/TM3	637	PVs missing, dominant; with high penetrance: distal endings of L2, 4, 5 missing
IV-42	PV-4/TM6	686	PVs missing; posteriormost central tergite bristles missing, dominant
IV-62	PV-3/FM7c	685	all large dorsal head bristles missing, dominant

Table 5. Established mutant stocks from the EMS mutagenesis.

It has been reported that the phenotypes of mild *N* alleles appear more intense when heterozygous with recessive wing mutants (Lindsley and Zimm, 1992). For line 7-4, a recombination mapping experiment was performed. The results place the mutation distally on the X chromosome, a chromosomal region, where the *N* locus at 3C 7-9 resides (Gert Pflugfelder, pers. comm.). These data suggest that 7-4, and possibly all other isolated wing mutants, might be *N* alleles.

The remaining 6 mutant lines all have dominant head bristle phenotypes. PV-1, PV-2, and PV-4 lack the postvertical bristles. They were mapped on the III chromosome. PV-3 and l(1) *Matze* flies do not have any large dorsal head bristles. In addition, the latter mutation, named after the author's haircut, is recessive lethal. The ocellar field is affected in the semi-lethal mutant 627, where the ocellar and interocellar bristles are missing as well as the median ocellus. 627, PV-3, and l(1) *Matze* map on the X chromosome. Besides their chromosomal mapping, these mutants have not been further characterized.

We monitored the effect of the EMS treatment in a control experiment depicted in Figure 15. *w* males were fed on a sucrose solution with or without 25mM EMS (50mM EMS in round 4) and subsequently mated to attX/ Y virgins. F1 females inherited their attX chromosome from the mother and the Y chromosome from the father, whereas F1 males have a maternal Y chromosome and are hemizygous for a paternal X chromosome.

If, therefore, one compares the number of sons from males with and without EMS treatment, one can estimate the mutagenic impact in a reduction of F1 males due to EMS-induced sex-linked recessive lethal mutations. The results of the EMS control experiment are given in Table 6. 1866 and 4780 sons were obtained from males with or without 25 mM EMS (rounds 1,2,3,5), respectively, so that EMS-treated flies produced 61% (1866/4780 = 39%) less male offspring than flies without 25mM EMS. In contrast, mutagenized males produced essentially the same number of daughters than males without EMS (2427:2429 = 1.00). The latter finding is not unexpected, as the *Drosophila* Y chromosome does not contain any essential gene, and indicates that EMS, at a concentration of 25mM, acted highly mutagenic, but not significantly noxious on treated flies. In round 4, EMS was administered to flies in an inadvertently elevated concentration of 50 mM. Interestingly, the doubled concentration of EMS was severely toxic to treated flies, as they barely produced any offsping (see Table 6).

The outcome of the control assay clearly demonstrates that our experimental procedure using 25 mM EMS profoundly induced mutations. Furthermore, the isolation of 12 mutant lines and the observation of several prominent mutants among the examined flies, such as *Curly*, corroborate that our mutagenesis was functional *per se*. We, however, failed to isolate any *org-1* or *C31* allele in our extensive screen for unknown reasons.
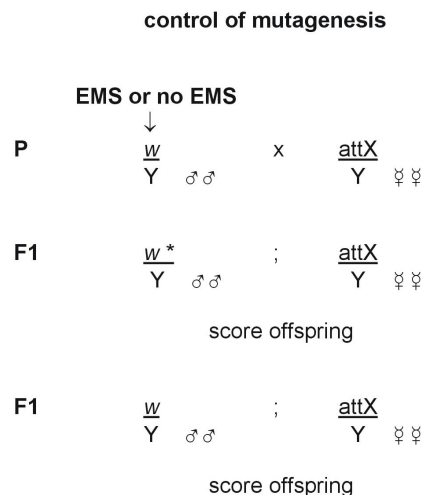
Figure 15. Crossing protocol to control the EMS mutagenesis.
* indicates an EMS-treated chromosome.

mutagenesis	EMS		no EMS		EMS	no EMS
	<u>attX</u> Y*	<u>w*</u> Y	<u>attX</u> Y	<u>w</u> Y	<u>attX</u> : <u>attX</u> Y* Y	<u>w*</u> : <u>w</u> Y Y
round 1 I/97	175	141	169	247	175:169 = 1.04	141:247 = 0.57
round 2 I/98	299	164	348	816	299:348 = 0.86	164:816 = 0.20
round 3 II/98	631	595	424	990	631:424 = 1.49	595:990 = 0.60
round 4 III/98	23	20	1134	1424	23:1134 = 0.02	20:1424 = 0.01
round 5 IV/98	1322	966	1488	2727	1322:1488 = 0.89	966:2727 = 0.35
rounds 1-5	2450	1886	3563	6204	2450:3563 = 0.69	1886:6204 = 0.30
rounds 1,2,3,5	2427	1866	2429	4780	2427:2429 = 1.00	1866:4780 = 0.39

Table 6. Control of the EMS mutagenesis.
* refers to a mutagen-treated chromosome.

4.2 Reverse genetic approaches: Screen for P element insertions at the *org-1* locus

The previous attempt to isolate *org-1* mutants as new *C31* alleles based on the hypothesis that the phenotype seen in *C31* flies is caused by an insertion in *org-1*. This approach, however, remained unsuccessful. No *org-1* nor new *C31* alleles could be obtained in a large-scale EMS mutagenesis, as described in chapter 4.1. Since we could not exclude the possibility that our failure is due to an idiosyncracy of *C31*, we intended not to rely on *C31* in further genetic experiments. Alternatively, we decided to continue our search for *org-1* mutants using a reverse genetic strategy, as we lacked a firm prediction for a screenable *org-1* phenotype.

The ultimate goal of this project is the isolation of a *Drosophila* mutant with a P element insertion at the *org-1* locus which interrupts the function of *org-1*.

4.2.1 Molecular screen for P element insertions at the *org-1* locus

In the course of this experiment, 540 viable, X-chromosomal P{lacW} element lines were screened for an integration at the *org-1* locus.

The used fly stocks were generated by Ulrich Schäfer and co-workers at the Jäckle laboratory, Göttingen. These lines represent the byproduct of a screen for lethal P element mutations on the X chromosome, the European contribution to the gene disruption project launched by the Berkeley *Drosophila* Genome Project (BDGP) (Peter *et al.,* 2002; Spradling *et al.*, 1995). The fly stocks were grouped in 54 batches of 10 lines each and genomic DNA was isolated from each pool. P element flanking genomic DNA was subsequently cloned by plasmid rescue and blotted on nylon membranes. These Southern blots, kindly provided by Thomas Raabe, were hybridized with overlapping restriction fragments isolated from cosmids 166H8 and 97G10. The probe comprises the restriction-mapped genomic region of about 59 kb at the *org-1* locus (Porsch, 1997). Two of the 54 pools, pools 9 and 31, strongly hybridized to the used probe. To identify the individual fly stocks responsible for the hybridization signals, the plasmid rescue experiment was separately repeated for all 20 lines from the two positive pools. A single positive line could be identified for both batches, line 9-7831/1 and 31-2756/1, respectively. Their P element insertion sites were mapped within the *org-1* region using subsequently smaller subsets of the isolated restriction fragments as probes. The hybridization data suggest similar P element insertion sites in both lines outside the *org-1* gene or its close proximity. Consistent results were gained from reciprocal hybridization experiments using the plasmid rescue products as probes on Southern blots with digested cosmid DNA. The P element insertion sites could be placed within small overlapping restriction fragments at the distal end of the restriction-mapped genomic region suggesting that the P elements in both lines are inserted about 35-40 kb downstream of the *org-1* transcription unit. Finally, *EcoR* I plasmid rescue clones were isolated for 9-7831/1 and 31-2756/1. Two types of clones could be obtained for 9-7831/1, whereas the plasmid rescues for 31-2756/1 were uniform. All different clones were sequenced. The sequences can be found on the accompanying CD-ROM [DNAseq/*org-1* genetics/plasmid rescues].

The results of several BLASTN searches revealed that in 9-7831/1 and 31-2756/1 P elements are in-

serted at chromosomal region 7E, approximately 36 and 38 kb downstream of the end of the *org-1* transcription unit, respectively. No predicted genes were found adjacent to these P elements. However, a BLASTN search with a 4 kb long query sequence that includes the two P element insertion sites (kb 295-299 in AE003443) identified the 7 EST clones RH09582.5', GH24113.5', GM09845.5', RH73791.5', RH38107.5', RH36953.5', and GM09770.5'. These EST clones are derived from two distinct transcripts. One transcript is represented by the EST clone RH09582 which 5' sequence can be aligned to AE003443 between 297 kb and 295,5 kb. RH09582 derives from a normalized head cDNA library indicating that this gene is transcribed in the adult head. The function of the encoded protein is unknown. The remaining 6, partially overlapping, EST sequences define a second transcript in proximity of the 2 P element insertions. Their 5' sequences align to AE003443 between 296,3 kb and 297,6 kb.

GM09770.5' (Table 7) has been chosen as a representative EST clone for this transcript. This EST sequence contains a 49 aa long 3' incomplete ORF and has no homology to any known protein sequence. The 6 EST clones derive from ovarian and adult head cDNA libraries.

The P element insertion sites are shown in Figure 16 in relation to *org-1* and genomic clones.

The analysis of a second plasmid rescue clone revealed an additional P{lacW} element insertion in line 9-7831/1. This transposon is located at 3C on the X chromosme, within the promoter region of the predicted gene *CG3603* (Table 7). *CG3603* encodes a putative oxidoreductase.

In summary, two P element insertions were discovered about 36-38 kb downstream of *org-1*, but no P element insertions could be identified within *org-1* or close-by sequences.

sequence	position in genomic sequence	cytological position	adjacent gene or EST clone
eco9-7831-1-cl17	kb 5,1 in AE003426	X; 3C	*CG3603*
eco9-7831-1-cl18	kb 298,6 in AE003443	X; 7E	GM09770.5'
eco31-2756-1	kb 296,2 in AE003443	X; 7E	RH09582.5' GM09770.5'

Table 7. P element insertion sites in lines 9-7831/1 and 31-2756/1.
The position within genomic sequences, their cytological position and adjacent genes or EST clones are listed. The sequences can be found on an accompanying CD-ROM [DNAseq/*org-1* genetics/plasmid rescues].

4.2.2 Characterization of lethal X-chromosomal P element lines at 7E-7F

In the following attempt to obtain an *org-1* insertion mutant, the *Drosophila* database flybase [http://flybase.bio.indiana.edu] was searched for all available P element insertion lines cytologically mapped to the *org-1* region at 7E-7F on the X chromosome. 19 lethal fly stocks with a single P{lacW} element insertion could be identified and were ordered from the Bloomington, IN, *Drosophila* stock center.

Genomic DNA was isolated from these lines in order to *in vitro* amplify P element flanking sequences using the inverse PCR (iPCR) technique (Ochman *et al.*, 1988; Sentry and Kaiser, 1994; Spradling *et al.*,1995). The obtained iPCR products were gel-purified and sequenced. Subsequently, BLASTN (Altschul *et al.*, 1990) searches with the P element flanking genomic sequences against all *Drosophila* sequences were performed to precisely determine the transposon insertion sites and the affected genes. The results of the sequence analysis are summarized in Table 8 and shown in Figure 16.
The sequences can be found on the accompanying CD-ROM [DNAseq/*org-1* genetics/lethalP].

This analysis revealed that 13 out of the 19 characterized lines have P element insertions within a 2,1 kb large genomic sequence (between kb 296,5 and 298,6 of AE003443), so that the vast majority of lethal P elements in 7E-7F is concentrated to this fragment. The evident preference for P element integrations at this site is further corroborated by two previously characterized lines 9-7831/1 and 31-2756/1 (chapter 4.2.1), as well as line

EP(X)1310 that also carry transposons at this P element hotspot. No gene has been predicted for this intervall, however, several EST clones of two nearby transcripts could be identified in a BLASTN search (Altschul *et al.*, 1990) using kb 295 - 299 of AE003443 as a query against all *Drosophila* EST sequences (see chapter 4.2.1). The two transcripts are represented by the EST clones RH09582 and GM09770. Their 5' sequences can be aligned to kb 297 – 295,5 or kb 296,3 – 297,6 in AE003443, respectively. Hence, the P elements in lines l(1) G0039, l(1) G0228, l(1) G0203, l(1) G0219, l(1) G0178, P{lacW} G0161b, l(1) G0332, l(1) G0166, l(1) G0356, l(1) G0295, l(1) G0376, l(1) G0425, and l(1) G0372 are all inserted within or close to two previously unpredicted genes and thereby presumably cause the lethal phenotype.

The lines l(1) G0099, l(1) G0488b, and l(1) G0413 all carry P element insertions at the *Neuroglian* (*Nrg*) locus. Their precise integration sites are at 101,6 kb, 115,7 kb, and 116,7 kb in AE003444, respectively, and they lie in the promoter region, the first intron, or the second exon of this gene. *Nrg* encodes a membrane-associated protein that functions in neuronal cell adhesion (Bieber *et al.*, 1989; Hortsch *et al.*, 1990). Amorphic *Nrg* alleles are embryonic lethal suggesting that the three P element insertions are functional *Nrg* null mutations.

The P transposon in l(1) G0095 is inserted at 140,8 kb in AE003444 and sits in the second exon of *CG12113*. This gene is predicted to code for a large protein of 1022 aa without significant similarity to any known protein.

l(1) G0071 and l(1) G0424 have single P element insertions at 5,4 kb and 5,8 kb in AE003444, respectively. Since no gene or transcript could be identified within a 6 kb sequence surrounding the transposon sites (kb 2 – 9 in AE003444), the cause of the lethality of these two lines remains unclear.

Interestingly, all 19 characterized lethal lines in 7E-7F have their P element inserted in the same 5' to 3' orientation with regard to the genomic sequences AE003443 and AE003444, whereas the P elements in lines 9-7831/1, 31-2756/1, and EP(X)1310 are inserted in the opposite orientation.

Although this study did not lead to the identification of a P element insertion within the *org-1* locus, several P elements on either side of *org-1* were identified. The closest P insertions are in line l(1) G0071 about 27 kb downstream of the *org-1* transcription unit and in line l(1) G0099 approximately 62 kb upstream of the putative *org-1* transcription

start site. These two P elements have been used in subsequent experiments described below. In addition, a hotspot for P element insertions could be discovered at about 37 kb downstream of *org-1*.

4.2.3 Local hop mutagenesis for P{lacW} insertions at the *org-1* locus

4.2.3.1 The generation of new X chromosomal P{lacW} insertion lines

In the course of previous attempts to isolate *org-1* insertion mutants, the P{lacW} lines l(1) G0071 and l(1) G0099 could be identified as so far nearest transposon insertions, residing about 27 kb downstream and 62 kb upstream of *org-1* (see chapters 4.2.1 and 4.2.2). As P elements preferentially transpose locally (Tower *et al.*, 1993; Zhang and Spradling, 1993), we considered these *org-1* flanking P elements as promising bases for local hops into *org-1* or its immediate vicinity.

The local hop *org-1* mutagenesis to be described consists of four separate, albeit simultaneously conducted, experiments (experiments A-D) using the two P element lines l(1) G0071 and l(1) G0099 and two transposase sources in all possible combinations. The experiments are: experiment A with G0099 and stock 404, experiment B with G0099 and stock 33, experiment C with G0071 and stock 404, and experiment D with G0071 and stock 33. The crossing protocol is shown in Figure 17. The parental crosses between the P element lines l(1) G0071 or l(1) G0099 and the jumpstarter lines 33, *w*; P[Δ2-3], *Sb/* TM3 *Ser*, or 404, *y w*; Δ2-3 TM3, *Sb/ Dr*, (these numbers refer to Gert Pflugfelder's fly stocks) combine the nonautonomous P{lacW} elements with an immobile, highly potent transposase source (Robertson *et al.*, 1988).

Figure 16. P element insertions at the org-1 locus (next page).
Genomic sequences are drawn as black lines and their accession numbers are given. P elements are shown as triangles with bars pointing to their integration sites. Lethal P{lacW} elements are in black along with their Bloomington stock numbers, the two viable P{lacW} elements 9-7831/1 and 31-2756/1 (chapter 4.2.1) in red, and the EP(X)1310 element in blue. Adjacent transcripts, EST clones, or predicted genes are drawn in green. Arrows indicate their directions. Cytological positions are given above the genomic sequences; distal is to the left, proximal is to the right. A scale bar is given.

P element insertions at the *org-1* locus
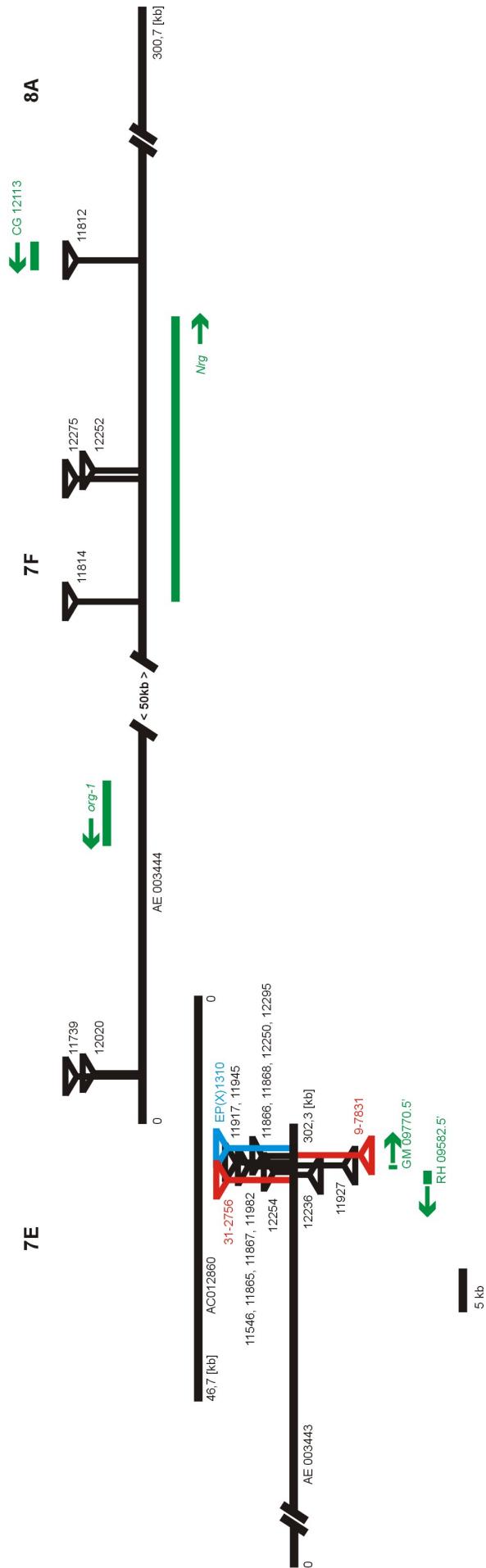
P element line	Bloomington stock number	annotated cytological position	position in genomic sequence	adjacent gene or EST clone
l(1) G0039	11546	7E5-6	kb 297,1 in AE003443	RH09582.5' GM09770.5'
l(1) G0071	11739	7E5-6	kb 5,4 in AE003444	none within 3kb
l(1) G0095	11812	7F1-4	kb 140,8 in AE003444	*CG12113*
l(1) G0099	11814	7F1-4	kb 101,6 in AE003444	*Nrg*
l(1) G0228	11865	7E3-6	kb 297,4 in AE003443	RH09582.5' GM09770.5'
l(1) G0203	11866	7E5-6	kb 298,6 in AE003443	GM09770.5'
l(1) G0219	11867	7E4-8	kb 297,4 in AE003443	RH09582.5' GM09770.5'
l(1) G0178	11868	7F	kb 298,6 in AE003443	GM09770.5'
P{lacW} G0161b	11917	7E5-8	kb 297,9 in AE003443	GM09770.5'
l(1) G0332	11927	7E	kb 297,6 in AE003443	GM09770.5'
l(1) G0166	11945	7E5-10	kb 297,8 in AE003443	GM09770.5'
l(1) G0356	11982	7E5-6	kb 297,4 in AE003443	RH09582.5' GM09770.5'
l(1) G0424	12020	7E5-11	kb 5,8 in AE003444	none within 3 kb
l(1) G0295	12236	7E	kb 296,5 in AE003443	RH09582.5' GM09770.5'
l(1) G0376	12250	7E5-10	kb 298,6 in AE003443	GM09770.5'
l(1) G0413	12252	7F	kb 116,7 in AE003444	*Nrg*
l(1) G0425	12254	7E	kb 297,4 in AE003443	RH09582.5' GM09770.5'
l(1) G0488b	12275	7F	kb 115,7 in AE003444	*Nrg*
l(1) G0372	12295	7E1-6	kb 298,5 in AE003443	GM09770.5'

Table 8. Characterization of lethal X-chromosomal P{lacW} element lines.

Females with both P element and transposase activity were collected as virgins and mated *en masse* to FM7 males. The offsping of the jumpcrosses was then genetically screened for new stable transposon insertions. For this, the lethality of the starter P elements allowed us to employ the "reversion-jumping" strategy (Tower *et al.*, 1993) for the identification of new insertion lines. All *white*[+] males among the offspring are revertants of the lethal phenotype (*i.e.* they have lost their starter transposon by precise excision) <u>and</u> contain a jumped P element as indicated by the remaining *miniwhite* P element marker gene.

A total of 1066 *white*[+] males without transposase gene were collected among approximately 73.750 screened males (1066/73750 = 1,5%) and mated to FM7 or *w* virgins to establish new stable insertion lines. These stocks were then analyzed for a X-linked inheritage of the *miniwhite* marker gene, and only 357 lines (357/1066 = 33,5%) with intrachromosomal transpositions were kept. 709 of the new insertion lines derived from a P element transposition to an autosome and were disposed (709/1066 = 66,5%). The genetic data of the local hop mutagenesis are summarized in Table 9.

local hop mutagenesis

experiment A

P $\quad$ $\dfrac{w,\ P\{lacW\}\ l(1)G0099}{FM7c}$ $\quad$ ♀♀ $\quad$ x $\quad$ $\dfrac{y,\ w}{Y}$; $\dfrac{\Delta 2\text{-}3,\ TM3,\ Sb}{Dr}$ $\quad$ ♂♂

F1 $\quad$ $\dfrac{w,\ P\{lacW\}\ l(1)G0099}{y,\ w}$; $\dfrac{\Delta 2\text{-}3,\ TM3,\ Sb}{III}$ $\quad$ ♀♀ $\quad$ x $\quad$ $\dfrac{FM7c}{Y}$ ♂♂

screen for *w*[+], *Sb*[+] ♂♂

F2 $\quad$ *w*[+], *Sb*[+] ♂. $\quad$ x $\quad$ FM7a ♀♀

screen for X-linked *w*[+]

experiment B

P $\quad$ $\dfrac{w,\ P\{lacW\}\ l(1)G0099}{FM7c}$ $\quad$ ♀♀ $\quad$ x $\quad$ $\dfrac{w}{Y}$; $\dfrac{\Delta 2\text{-}3,\ Sb}{TM3,\ Ser}$ ♂♂

F1 $\quad$ $\dfrac{w,\ P\{lacW\}\ l(1)G0099}{w}$; $\dfrac{\Delta 2\text{-}3,\ Sb}{III}$ $\quad$ ♀♀ $\quad$ x $\quad$ $\dfrac{FM7c}{Y}$ ♂♂

screen for *w*[+] ♂♂ with uniform eye color

F2 $\quad$ *w*[+] ♂ $\quad$ x $\quad$ FM7a ♀♀

screen for X-linked *w*[+]

Figure 17. Crossing protocol for the *org-1* local hop mutagenesis.
The crossing schemes are shown for experiments A and B. Experiments C and D correspond to experiments A and B, except that P{lacW} l(1)G0071 was used as the starter P element line.

experiment	screened ♂♂	isolated w^+; Δ2-3⁻ ♂♂	w^+ ♂♂, autosomal P{lacW}	w^+ ♂♂, X-linked P{lacW}	analyzed X-linked P{lacW} lines
A	ca 21.250	63	18	45*	39
B	ca 22.500	103	44	59*	55
C	ca 15.000	328	239	89	86
D	ca 15.000	572	408	164	163
A+B	ca 43.750	166	62	104*	94
C+D	ca 30.000	900	647	253	249
A-D	ca 73.750	1066	709	357	343

Table 9. Genetic data of the *org-1* local hop mutagenesis.
* indicates numbers that include lines with uncertain X-chromosomal P elements due to an initially inaccurate linkage analysis. The experiments are: experiment A with G0099 and stock 404, experiment B with G0099 and stock 33, experiment C with G0071 and stock 404, and experiment D with G0071 and stock 33. G0099 is the upstream element, G0071 the downstream element.

4.2.3.2 The molecular characterization of new X chromosomal P{lacW} insertion lines

343 of the 357 new X-chromosomal P element lines were further analyzed by molecular techniques (14 lines perished). Genomic DNA was isolated from these lines, digested with *Cfo* I or *Sau*3A I restriction endonucleases and self-ligated to allow the *in vitro* amplification of P element ends neighboring genomic sequences by iPCR (Ochman *et al.*, 1988; Sentry and Kaiser, 1994; Dalby *et al.*, 1995). 5' and 3' iPCR products were separated on agarose gels, blotted onto nylon membranes, and immobilized. Subsequently, the resulting Southern blots were successively probed with three overlapping genomic clones containing DNA from the *org-1* locus. The position of these genomic clones is shown in Figure 18 relative to *org-1*.

These clones include the two previously restriction-mapped cosmids 166H8 and 97G10 (Porsch, 1997). In addition, the BAC clone BACR17J10 could be identified in a detailed database analysis of all genomic clones around the *org-1* locus (Figure 19).
BACR17J10 contains about 173 kb genomic DNA surrounding *org-1* and was obtained from BACPAC Resources, Bruce Lyon Memorial Res. Lab, Oakland, CA, USA. Its identity was confirmed by PCR amplifications with *org-1* specific primer pairs as well as end-sequencing prior to the use in the hybridization experiments.

The putative P element insertion sites could finally be inferred from a comparative analysis of the data obtained from the hybridization series. For instance, if an iPCR product hybridizes with all three probes used, then the corresponding P element insertion can be placed within the intersection of the three clones, *i.e.* within the interval AE003444 kb 12-37 (see Figure 18). Hybridization signals for an iPCR product with the BACR17J10 and the cosmid 166H8 probes, but not the cosmid 97G10 probe, map the corresponding P element into the intersection of BACR17J10 and cosmid 166H8 and the complement of cosmid 97G10, *i.e.* within the interval AE003443 kb 297 – AE003444 kb 12, and so forth.

Table 10 summarizes the mapping data for the 343 P element insertion sites. These results are graphically presented in Figure 18.

For 307 of the 343 investigated lines (307/343 = 89,5%), the putative P element insertion sites could be placed in relation to the three genomic clones. The transposon insertion sites in the other 36 lines (36/343 = 10,5%) could not be analyzed, as no hybridization signals nor iPCR products with a flanking genomic sequence of at least 150 bp could be obtained, although, in such cases, the iPCR experiments were repeated with the other restriction enzyme as well. Previous controls determined a minimal sequence requirement of 150 bp for detectable hybridizations in our experiments (data not shown).

P-element insertion mutagenesis "Local Hop" at the *org-1* locus

8A

7F

7E

P[lacW]l(1)G0099

P[lacW]l(1)G0071

< 50kb >

Nrg

org-1

5 kb

cosmid 97G10

cosmid 166H8

BACR 17J10

Analyzed lines: 343

Not included within BAC clone: 185

Not interpretable yet: 36

295 297
9 + (8)

12 37
4 + (2)

49 168
5 + (1)

297 12
70 + (22)

37 49
(1)
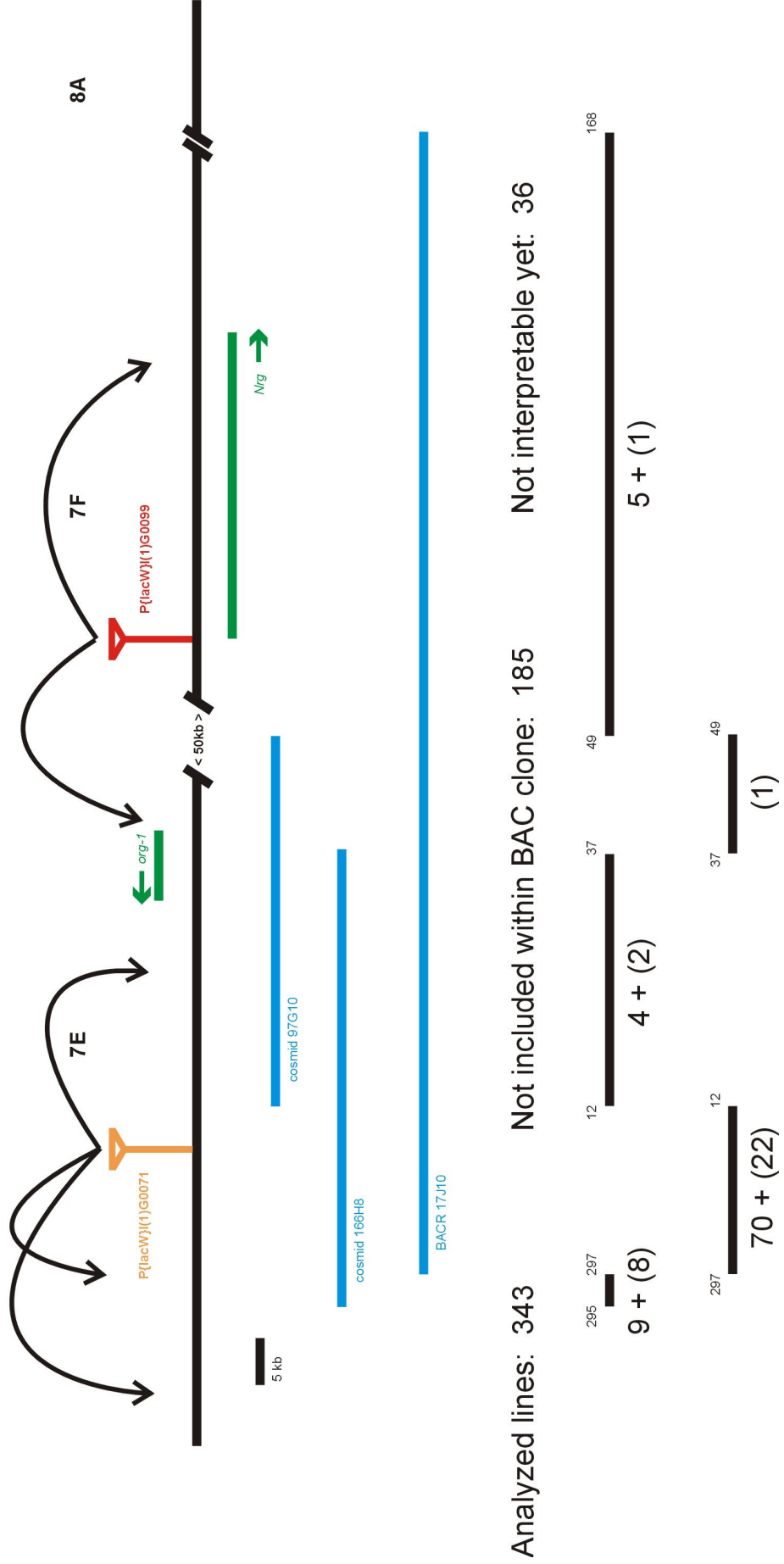
Figure 18. Local hop P element insertion mutagenesis at the *org-1* locus (previous page).
The genomic sequence around the *org-1* locus is indicated by the black line. Cytological positions are given above; distal is to the left, proximal is to the right. The starter P elements are shown as triangles with bars pointing to their intergration sites. Relevant genes are shown in green with arrows indicating their direction of transcription. Blue lines represent the genomic clones used as probes. The positions of new P insertions inferred from the hybridization experiments are summarized below. Intervals are indicated by black lines. The small numbers above the black bars give the coordinates within the genomic sequences AE003443 and AE003444 in kb. Large numbers below indicate the number of putative P integration events within the interval. Brackets indicate ambiguous cases.

In 185 lines (185/343 = 53,9%), the P elements lie outside of the 175 kb large interval AE003443 kb 295 – AE003444 kb 168 comprised by the three genomic clones. 109 lines (109/343 = 31,8%) bear a transposon insertion within AE003443 kb 295 – AE003444 kb 12. This interval contains the insertion site of the starting P element l(1) G0071 and the previously identified hotspot for P insertions (see chapter 4.2.2). 7 putative P element insertions (7/343 = 2%) are placed within two intervals spanning AE003444 kb 12–49. As the *org-1* locus at AE003444 kb 39–32 is included in this region, these 7 transposon insertions are of highest interest. Finally, 6 P insertions (6/343 = 1,8%) were found for the interval AE003444 kb 49–168. Thus, 13 potentially interesting local transpositions could be identified among 343 candidate lines in the hybridization experiments (13/343 = 3,8%). PCR products from 12 of these 13 lines could be gel-purified and sequenced in order to precisely locate the transposon insertion sites. In addition, several other P element lines were included in the sequencing project to assess the reliability of the hybridization-based mapping analysis.

Therefore, the exact positions of P elements with presumed insertion sites between AE003443 kb 297 and AE003444 kb 12 were determined for the lines 204, 382, 464, and 551. All these four lines bear a transposon within the predicted interval at about 5,4 kb, 4,6 kb, 7,2 kb, and 5,5 kb of AE003444, respectively, closely surrounding the starter P element l(1) G0071 site at 5,4 kb. The P element in line 464 is integrated upstream of the gene encoding TATA box-binding protein-related factor 2 (*Trf2*) (Rabenstein *et al.*, 1999) (Figure 20).

Hybridization data suggested putative P insertions within AE003444 kb 12-37 for the lines 204, 213, 266, 274, 543, and 599.
Indeed, lines 204, 274, and 543, carry P elements at about 21,4 kb , 21,4 kb, and 21, 7 kb, respectively (Figure 20). These three P elements lie about 10 kb downstream of the *org-1* gene and represent the *org-1* nearest insertions found in this analysis. They are inserted at the 5' regions of the two proposed genes *CG12125* and *CG1440*. The gene *CG1440* is predicted to encode a cysteine-type endopeptidase (Table 12). Line 213 carries a P element at about kb 125 of AE003442, some 500 kb distally from the starter transposon l(1) G0099 at chromosomal band 7C on the X chromosome. The P element is inserted within the first intron of the predicted gene *CG10777* encoding a putative RNA helicase.

experiment	analyzed X-linked P{lacW} lines	P{lacW} in interval AE003443 kb 295-297	P{lacW} in interval AE003443 kb 297 - AE003444 kb 12	P{lacW} in interval AE003444 kb 12-37	P{lacW} in interval AE003444 kb 37-49	P{lacW} in interval AE003444 kb 49-168	P{lacW} not in interval AE003443 kb 295 - AE003444 kb 168	not interpretable yet
A	39	(1)	0	0	0	2 + (1)	27	7
B	55	0	3	(1)	(1)	2	40	8
C	86	1 + (1)	23 + (8)	0	0	0	42	11
D	163	8 + (6)	44 + (14)	4 + (1)	0	1	76	10
A+B	94	(1)	3	(1)	(1)	4 + (1)	67	15
C+D	249	9 + (7)	67 + (22)	4 + (1)	0	1	118	21
A-D	343	9 + (8)	70 + (22)	4 + (2)	(1)	5 + (1)	185	36

Table 10. Mapping of P element insertion sites.
Numbers in brackets symbolize uncertain mappings deduced from weak hybridization signals.

Genomic clones at the *org-1* locus
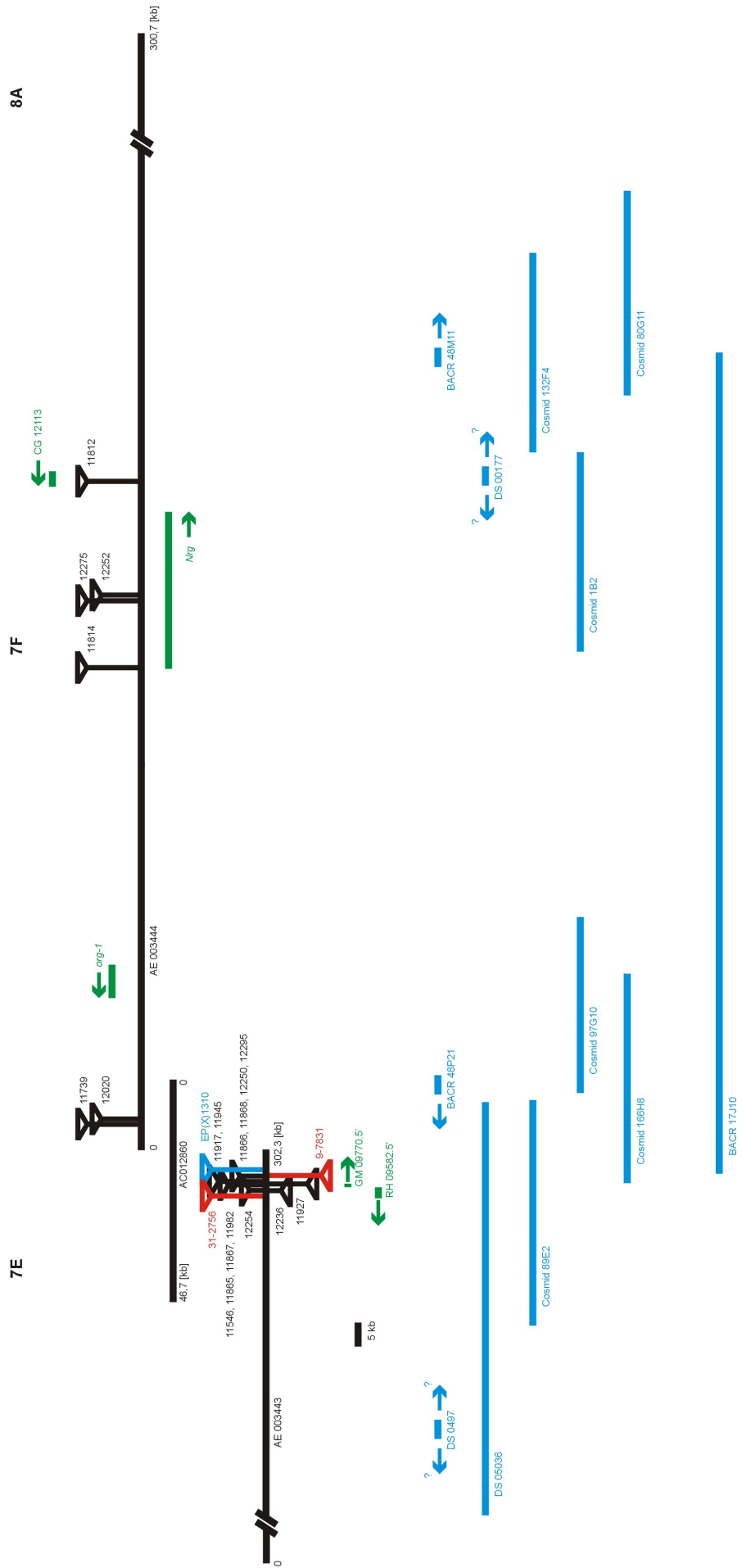
Figure 19. Genomic clones at the *org-1* locus (previous page).

All genomic clones from the chromosomal interval 7E-7F are shown in blue lines relative to genomic sequences around the *org-1* locus. Genomic sequence contigs are shown as black lines. Cytological positions, relevant genes and transcripts, and P element insertions are indicated as previously in Figures 16 and 18.

No insertion within the presumed interval could be found in line 266. Instead, its P element resides at 5,4 kb of AE003444 next to the starter P site.

For line 599, only a P element insertion within the 5' region of the *white* locus at 3C on the X chromosome was identified.

One putative P insertion was considered for AE003444 kb 37-49 due to weak hybridization signals for line 128 iPCR products. However, two determined P elements insertion sites are located on the second chromosome and a further genetic analysis demonstrated that line 128 lacks any X-chromosomal P element (Gert Pflugfelder, pers. comm.).

line	P{lacW} mapped into interval	sequenced iPCR product	cytological position	position within clone or genom. sequence	adjacent gene
(26)	AE003444 kb 49-168	LH26-Cfo3'-900bp	X;7E-7F	101,9-102,3 kb in AE003444	*Neuroglian*
82	AE003444 kb 49-168	LH82-Sau5`-1800bp	X;7E-7F	144,9-145,0 kb in AE003444	DNA topoisomerase (*CG11265*)
96	AE003444 kb 49-168	LH96-Sau5`-1300bp	X;7E-7F	175,6-175,7 kb in AE003444	*CG12659*
(128)	AE003444 kb 37-49	LH128-Sau3`-800bp LH128-Sau3`-1100bp	2R;55E-55F 2L;25C-25D	BACR27L09 BACR28N20	- *Msp300* (*CG18252*)
131	AE003444 kb 49-168	LH131-Cfo5'-700bp	X;7E-7F	101,6-101,7 kb in AE003444	*Neuroglian*
138	AE003444 kb 49-168	LH138-Sau5'-1200bp	X;7E-7F	101,1-101,3 kb in AE003444	*Neuroglian*
204	AE003443 kb 297-AE003444 12 and AE003444 kb 12 - 37	LH204-Sau3'-550bp LH204-Sau3'-1100bp LH204-Sau5'-1050bp	X;7E-7F X;7E-7F X;7E-7F	both at 5,4-5,7 kb in AE003444 21,4-21,6 kb in AE003444	- - *CG12125* -
(213)	AE003444 kb 12-37	LH213-Sau3'	X;7C	125-124,6 kb in AE003442	RNA helicase (*CG10777*)
266	AE003444 kb 12-37	LH266-Sau3'-500bp	X;7E-7F	5,4-5,6 kb in AE003444	-
(274)	AE003444 kb 12-37	LH274-Sau5'-900bp	X;7E-7F	21,4-21,6 kb in AE003444	*CG12125*
382	AE003443 kb 297-AE003444 12	LH382-Sau5'-1300bp	X;7E-7F	4,6-4,8 kb in AE003444	-
464	AE003443 kb 297-AE003444 12	LH464-Sau3'-350bp	X;7E-7F	7,2-7,3 kb in AE003444	*Trf2*
543	AE003444 kb 12-37	LH543-Cfo5'-500bp	X;7E-7F	21,7 kb in AE003444	cysteine-type endopeptidase (*CG1440*)
551	AE003443 kb 297-AE003444 12	LH551-Sau3'	X;7E-7F	5,5-5,6 kb in AE003444	-
599	AE003444 kb 12-37	LH599-Sau5'-700bp	X;3C	BACN33B1	*white* (*CG2759*)

Table 11. Summary of sequence data of relevant local transposition lines.

Local transposition lines in brackets indicate vague mappings deduced from weak hybridization signals. The determined DNA sequences are on an accompanying CD-ROM [DNAseq/*org-1* genetics/local hop].

P-element insertion mutagenesis
Summary of local transpositions
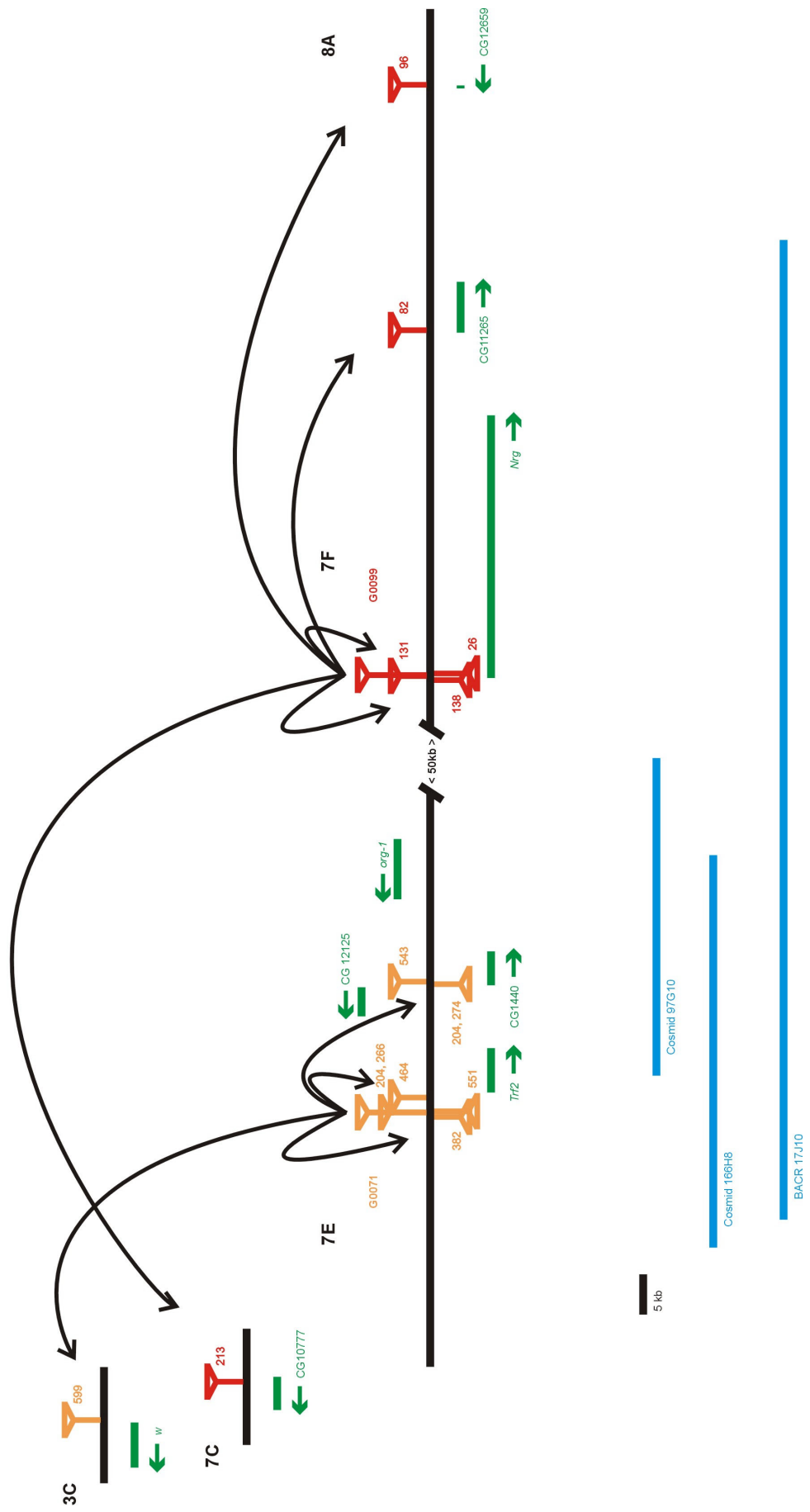
Figure 20. Summary of local transpositions (previous page).
P elements are drawn as triangles with bars pointing to their integration sites in the genomic sequence (black line). The two protruding triangles symbolize the starter P elements. l(1) G0071 and its derivatives are in orange, l(1) G0099 and descendants are in red. Adjacent genes are shown in green with arrows to indicate the direction of transcription. The cytological positions are given above the genomic sequences. Blue lines represent genomic clones used in the hybridization experiments. A 5 kb scale bar is shown.

Finally, P elements within AE003444 kb 49-168 were indicated for lines 26, 82, 96, 131, 138, and 172. Their approximate insertion sites are at 101,9 kb, 144,9 kb, 175,6 kb, 101,6 kb, and 101,1 kb of AE003444, respectively; the molecular analysis for line 172 remained elusive.

Hence, the transposon insertions in line 26, 131, and 138 are all in immediate vicinity to the site of the starter P element l(1) G0099 at 101,7 kb of AE003444 and are all associated with the *Neuroglian* gene.

Line 82 carries a P element in the 5' region of the predicted gene *CG11265* putatively encoding a DNA topoisomerase, while line 96 has a P insertion associated with the predicted gene *CG12659*.

Table 11 and Figure 20 summarize the sequencing results.

Interestingly, just as for the group of lethal P element lines (see chapter 4.2.2), the orientation of the P elements in the local transposition lines is far from being random. 14 characterized P elements are integrated in the same 5' to 3' orientation as the genomic sequences AE003443 and AE003444 or both starter P elements, whereas only line 213 has its transposon in the opposite orientation.

Taken together, 6 new genes between 7C and 8A could be associated with P{lacW} insertions, however, we failed to target the *org-1* locus. Although we could approach *org-1* by 16 kb, the *org-1*-closest P elements in lines 204, 274, and 543 are still 10 kb away from the 3' end of the gene under investigation.

4.3 Generation of deficiencies in 7E-7F

4.3.1 P element-mediated construction of precise deletions

Cooley *et al.* (1990) describe a P element-based method to efficiently generate precise deficiencies in the *Drosophila* genome. They succeeded in creating a desired deletion with P element insertion sites as deficiency endpoints, when they remobilized two P transposable elements within a progenitor strain. According to that, a temporarily present transposase catalyzes the excision of both P elements, resulting in two chromosomal breaks. The ligation of the chromosomal fragments by the endogenous DNA repair machinery would subsequently give rise to the desired deletion (Figure 21).
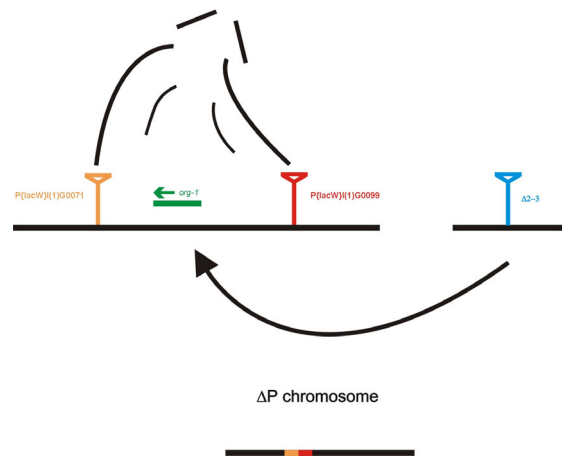
Figure 21. Constructing deletions at the *org-1* locus with defined endpoints (Cooley *et al.*, 1990).
Two P{lacW} elements, G0071 and G0099, flanking the *org-1* gene, are remobilized by a P element encoded transposase, Δ2-3. Upon excision of both P elements, the two chromosomal breaks are ligated resulting in a ΔP chromosome. The deficiency endpoints coincide with the P element insertion sites. Residual P element sequences frequently remain and are indicated on the ΔP chromosome (after Cooley *et al.*, 1990).

Among 19 recently characterized lethal P{lacW} insertion lines from the *org-1* containing chromosomal region 7E-7F, lines l(1) G0071 and l(1) G0099 were determined as the ones closest to *org-1* (see chapter 4.2.2). Their P elements are integrated about 27 kb downstream and 62 kb upstream of the 6 kb large *org-1* transcription unit.

Hence, using Cooley's technique with these two *org-1* flanking P elements, we should be able to generate precise deletions of approximately 95 kb that include the *org-1* locus.

Such a deficiency would be a highly valuable tool for *org-1* genetics.

This 95 kb large deletion would comprise between 16 and 20 transcription units. The maximal gene number is reckoned from 18 predicted genes plus two unpredicted transcripts that could be identified by EST clones. For four neighboring genes, all predicted to encode small transcripts, no EST clones were found that would confirm their existence (Table 12). 7 genes within the desired deletion have been cloned: *Trf2* encoding the general transcription factor TATA box-binding protein-related factor 2 (TRF2) (Rabenstein *et al.*, 1999),

org-1 that codes for a T-box transcription factor (Porsch *et al.*, 1998; this work), *Sptr* encoding a sepiapterin reductase (Seong *et al.*, 1998; Seong *et al.*, 2000), *Cp36* and *Cp38*, two genes that code for chorion proteins, *ovarian tumor* (*otu*), a gene essential for oogenesis (Geyer *et al.*, 1993 and references therein), and *Neuroglian* (*Nrg*) encoding an integral membrane glycoprotein that functions in cell adhesion (Bieber *et al.*, 1989; Hortsch *et al.*, 1990).

otu, *Cp36*, and *Nrg* are the only genes of the 95 kb interval for which mutants exits. Mutations in *otu* or *Cp36* cause female sterility, while *Nrg* null alleles are embryonic lethal.

Table 12 and Figure 22 summarize the data of genes within the desired deficiency interval.

gene/transcript molecular function	position within genomic sequence AE003444	EST clones	mutant alleles	references
Trf2 general transcription factor	8,2-13,7 kb	2 AT, 4 LD	-	Rabenstein *et al.*, 1999
CG12125	21,3-17,7 kb	1 AT, 1 LP, 1 GM	-	
CG1440 peptidase	21,6-25,8 kb	1 AT, 9 LD, 1 GH	-	
CG12123	26,9- 26,0 kb	1 GH, 2 SD	-	
org-1 regulatory transcription factor	38,8-32,7 kb	-	-	Porsch *et al.*, 1998
AT08049.5'	23,3-23,8 kb	1 AT	-	
CG15347	43,7-42,8 kb	1 LP	-	
Es2 enzyme	45,8-48,7 kb	1 SD, 5 LD, 1 LP, 1 AT, 1 GM, 1 GH	-	
Sptr sepiapterin reductase	48,7-47,6 kb	1 GH, 2 LD	-	Seong *et al.*, 1998 Seong *et al.*, 2000
CG12116	53,2-51,9 kb	1 LP, 11 GH	-	
CG15348	55,6-55,9 kb	-	-	
CG15349	56,3-57,1 kb	-	-	
CG15350	58,1-59,7 kb	-	-	
CG15351	61,3-62,3 kb	-	-	
Cp36 chorion protein	63,5-64,4 kb	-	4, female sterile	
Cp38 chorion protein	66,0-67,2 kb	-	-	
otu	73,0-69,6 kb	-	33, female sterile	Geyer *et al.*, 1993 and references therein
CG1521	74,0-100,1 kb	3 LD, 1 GH, 1 AT	-	
AT17449.5'	95,6-96,1 kb	1 AT	-	
Nrg cell adhesion protein	101,6-134,4 kb	1 GH	10, embryonic lethal	Bieber *et al.*, 1989; Hortsch *et al.*, 1990

Table 12. Genes within the designated deletion interval in 7E-7F.
Accumulated data of the genes within the desired deletion of 95 kb are presented. The number of EST clones is given for different clone sources separately. EST sources are: AT testis, GH adult head, GM ovary, LD embryo, LP larvae and pupae, SD Schneider cells.

genes at the *org-1* locus
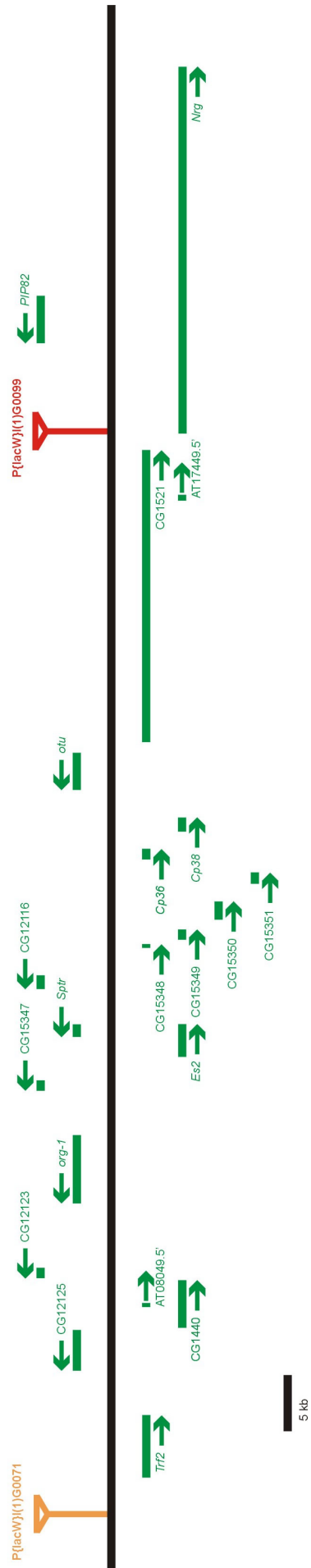
Figure 22. Genes at the *org-1* locus (previous page).
Genes and transcripts are drawn in green with arrows indicating the direction of transcription. The P elements G0071 (orange) and G0099 (red) are shown as triangles with bars pointing to their insertion sites.
Cytological positions are given above the genomic sequence (black line); distal is to the left, proximal is to the right. A 5 kb scale bar is given.

In their case study, Cooley *et al.* (1990) obtained precise deficiencies only, if they provided the P elements which mark the designated deletion breakpoints *in cis*. The first step in generating the desired deletion is therefore to construct a "deletion progenitor strain" (Cooley *et al.*, 1990) that has both *org-1* flanking P elements together on a recombinant chromosome. We expected to create deletions in 7E-7F by crossing a potent transposase gene into the recombinant line. Putative deficiencies should be identified by screening the offspring for new *C31* alleles.

4.3.2 Recombination of two *org-1* flanking P{lacW} elements

The generation of a precise deficiency according to Cooley *et al.* (1990) requires two P transposable elements at the future endpoints of a desired deletion in *cis* configuration. The construction of such a "deletion progenitor strain" (Cooley *et al.*, 1990) was therefore our first task in this project. We intended to bring together the two *org-1* flanking P elements l(1) G0071 and l(1) G0099 on a chromosome by meiotic recombination. Figure 23 shows the crossing scheme that led to the generation of that recombinant stock.

Since both X-chromosomal P element insertions are lethal, l(1) G0071 virgins were first crossed to *Df(1) GE202 ; Dp(1;2)* males. Males with the l(1) G0071 P element could be recovered due to presence of the duplication *Dp(1;2)* that rescues the lethality of the P insertion. These males were mated to l(1) G0099 virgins, and transheterozygous l(1) G0071/ l(1) G0099 females were obtained in the subsequent generation. Thus, the two lethal P insertions complement each other, as expected from their different insertion sites (see Figure 16; l(1) G0071 and l(1) G0099 are named 11739 and 11814 therein). Recombinant l(1) G0071, l(1) G0099 flies were expected among the progeny of transheterozygous virgins and FM7 males.

However, their identification turned out to be very intricate, because both P elements are of the same type, P{lacW}, containing identical *miniwhite* marker genes.

**generation of a "deletion progenitor strain":
recombination of P{lacW} l(1)G0071 and P{lacW} l(1)G0099**

P $\quad$ $\dfrac{\text{P\{lacW\} l(1)G0071}}{\text{FM7c}}$ $\qquad$ x $\qquad$ $\dfrac{\text{Df(1) GE202}}{Y}$; $\dfrac{\text{Dp(1;2)sn[+]}^{72d}}{\text{Dp(?;2)bw}^{D}}$
$\qquad\qquad\qquad\qquad$ ♀ ♀ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ♂ ♂

F1 $\quad$ $\dfrac{\text{P\{lacW\} l(1)G0071}}{Y}$; $\dfrac{\text{Dp(1;2)sn[+]}^{72d}}{\text{II}}$ $\quad$ x $\quad$ $\dfrac{\text{P\{lacW\} l(1)G0099}}{\text{FM7c}}$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ♂ ♂ $\qquad\qquad\qquad\qquad\qquad$ ♀ ♀

F2 $\quad$ $\dfrac{\text{P\{lacW\} l(1)G0071}}{\text{P\{lacW\} l(1)G0099}}$ $\qquad$ x $\qquad$ $\dfrac{\text{FM7c}}{Y}$
$\qquad\qquad\qquad\qquad\qquad$ ♀ ♀ $\qquad\qquad\qquad$ ♂ ♂

molecular screen for recombinants

F3 $\quad$ **stock** $\dfrac{\text{P\{lacW\} l(1)G0071, P\{lacW\} l(1)G0099}}{\text{FM7c}}$ $\qquad$ x $\qquad$ $\dfrac{\text{FM7c}}{Y}$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ♀ ♀ $\qquad\qquad\qquad\qquad$ ♂ ♂

Figure 23. Crossing protocol for the generation of a P{lacW} l(1) G0071, P{lacW} l(1) G0099 recombinant stock.

Nonetheless, the expression strength of their *mini-white* markers differs remarkably. While P{lacW} I (1) G0071 flies show a light orange eye pigmentation indicating a fairly weak *miniwhite* expression, P{lacW} I (1) G0099 flies have red eyes. We had hoped to detect recombinant flies with dark red eyes on account of an additive effect of both *mini-white* genes. Unfortunately, however, transheterozygous I(1) G0071/ I(1) G0099 females (and, thus, recombinant I(1) G0071, I(1) G0099/ FM7 females alike) could not be distinguished from I(1) G0099/ FM7 flies by their eye colors, since the strong expression of the P{lacW} I (1) G0099 marker itself simply overshadows the mildly expressed P{lacW} I (1) G0071 *miniwhite* gene.
Therefore, we could not phenotypically identify recombinant flies, but had to develop a molecular screen instead.

Three oligonucleotides, 772-rev1, 772-rev2, and 774-rev1 (corresponding to Gert Pflugfelder's stock numbers 772 and 774 for P{lacW} I (1) G0071 and P{lacW} I (1) G0099, respectively) that anneal to genomic sequences several hundred basepairs downstream of the P element insertion sites were ordered. These primers, in combination with the transposon-specific primer pry2 derived from sequences of the P{lacW} 3' end, would give rise to specific PCR products of 600 bp, 750 bp, and 550 bp, respectively (Figure 25). Furthermore, a PCR protocol was established to reliably amplify from crude DNA isolations of single G0071 or G0099 flies.

3000 red-eyed females were collected as virgins among the offspring of transheterozygous I(1) G0071/ I(1) G0099 mothers and FM7 males. These possibly recombinant flies were individually paired with FM7c males, and, upon the appearance of larvae, sacrificed to perform single fly DNA preparations.
The red eye color of the recombination candidates indicated the presence of the P{lacW} I (1) G0099 transposon. We, therefore, PCR genotyped them for the concomitant presence of the light orange P{lacW} I (1)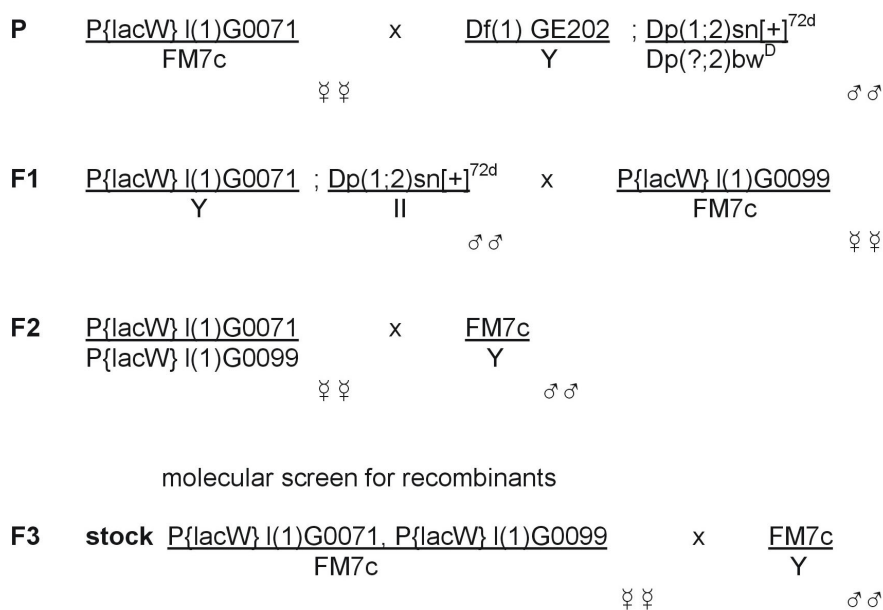 G0071. A total of 800 candidates were analyzed in pools of 3 flies using the 772-rev1 and pry2 primer pair. The expected PCR product of 600 bp was obtained with 3 pools. Their individual DNA preparations were then separately investigated by PCRs with the three primer pairs mentioned above. All three PCR products were obtained for a single fly DNA from each positive pool, whereas the other flies of these groups showed the amplificate for the G0099 P element only.
The recombinant lines are A16, D90, and G46. The three PCR amplificates for line A16 were gel-

purified, and DNA sequencing unambiguously demonstrated the presence of both P elements in this line.

The molecular screen was stopped after the identification of the three recombinants. Recombinants were observed with a frequency of 0,375 Centi-Morgan (cM) (3 recombinants/ 800 tested individuals), consistent with the theoretical recombination frequency of about 0,3 cM for a 100 kb interval at this cytological region (Poeck *et al.*, 1993; Ashburner, 1989).

The recombination line A16 was used as the "deletion progenitor strain" (Cooley *et al.*, 1990) in the jump-out experiments described below.

4.3.3 Jump out mutagenesis I: Isolation of new *C31* alleles

After the generation of a "deletion progenitor strain" (Cooley *et al.*, 1990) that carries two *org-1* flanking P elements *in cis*, we intended to induce precise deletions following the introduction of the potent transposase gene Δ2-3 (Robertson *et al.*, 1988). Our strategy was to identify putative deletions by screening their offspring for new *C31* alleles. *C31*, a recessive mutant isolated by Roland Strauss due to an aberrant walking behavior (Strauss and Trinath, 1996), was genetically mapped into 7E2-3 and 7F1-2 (Gert Pflugfelder and Roland Strauss, pers. comm.) and has been shown to contain an insertion within *org-1* that is absent in several wild type strains (Porsch, 1997). Among other defects, *C31* flies show a prominent wing "held-out" phenotype. We expected to detect the desired deletions as new *C31* alleles by scoring mutagenized flies for held-out wings.
The crossing scheme of this mutagenesis is shown in Figure 24.
Recombinant I(1) P{lacW} G0071, I(1) P{lacW} G0099/ FM7 females were crossed to males of Gert Pflugfelder's jumpstarter stock 404 to yield flies with both, the deletion progenitor chromosome and the Δ2-3 transposase gene. Such females were collected as virgins and were mated to *C31* males. Their progeny was scored for females with held-out wings by Gert Pflugfelder and co-workers. 89 virgins with held-out wings could be found and were individually paired with FM7i-pACT-GFP/ Y males to balance the putative Δ P{lacW} G0071, I(1) P{lacW} G0099 chromosome (hereafter: ΔP chromosome) in the next generation. However, since the selected candidates carried the putative deletion chromosome over the *C31* chromosome, recombination between these two chromosomes

may have occurred. We took this issue into account by making up to 7 individual stock crosses in parallel for each of the 34 candidates for which progeny could be obtained at all (34/89 = 38,2%). Thereby, we could establish 23 stocks (23/89 = 25,8%) that have their putative ΔP chromosome balanced (3 lines perished, for 8 lines only stocks with the *C31* chromosome were obtained).

12 of the 23 stocks contain a viable ΔP chromosome. All these 12 ΔP chomosomes lack a functional *white* gene, suggesting that the precise loss of both starter P elements led to a reversion of their lethal phenotype.

The remaining 11 stocks have lethal ΔP chromosomes. They were assayed by crossing the ΔP chromosomes over *C31*. 3 ΔP chromosomes complement *C31*, while the other 8 ΔP lines uncover the *C31* wing phenotype. Furthermore, Roland Strauss and co-workers investigated the brain anatomy and the walking behavior of the ΔP/ *C31* flies of these 8 lines, and could show that these 8 lines completely uncover the pleiotropic defects of *C31* (Roland Strauss, pers. comm.).

Hence, we have isolated 8 new *C31* alleles in our screen for precise deletions in 7E-7F.

jump out mutagenesis I: screen for new *C31* alleles

P	P{lacW} l(1)G0071, P{lacW} l(1)G0099 / FM7i-pAct-GFP ♀♀	x	*y, w* / Y	; Δ2-3, TM3, *Sb* / *Dr* ♂♂

F1	P{lacW} l(1)G0071, P{lacW} l(1)G0099 / *y, w* ♀♀	; Δ2-3, TM3, *Sb* / III	x	*C31* / Y ♂♂

screen for ♀♀ with *C31* wing held-out phenotype.
cross candidates individually to FM7i-pAct-GFP/Y ♂♂.

F2	ΔP{lacW} l(1)G0071, P{lacW} l(1)G0099 / *C31* ♀♀	; *Sb+* / III	x	FM7i-pAct-GFP / Y ♂♂

for each candidate, set up several individual stock crosses in parallel.

F3	**stock** ΔP{lacW} l(1)G0071, P{lacW} l(1)G0099 / FM7i-pAct-GFP ♀♀	x	FM7i-pAct-GFP / Y ♂♂

Figure 24. Crossing protocol for the jump out mutagenesis I: Screen for new *C31* alleles.

4.3.4 Molecular characterization of new *C31* alleles

As described above, we had expected to identify precisely generated deletions in 7E-7F as new *C31* alleles. 23 candidate stocks could be established among which 8 lines proved to be allelic to *C31*. These lines were molecularly analyzed for the expected deficiency. Single fly DNA preparations were performed with ΔP/ *C31* mothers, after they had passed the ΔP chromosome to their progeny. 5

diagnostic PCRs were designed to test for the presence or absence of genomic DNA at the *C31* insertion site as well as at the starter P element ends on the ΔP chromosomes (Figure 25).

Using primer pairs comprising the 1,3 kb large *C31* I element insertion, one would expect the large *C31* amplificate without the additional wild type product only, if the ΔP chromosome lacks the homologous sequences. Surprisingly, the wild type PCR product was obtained for all ΔP/ *C31* candidates investigated (Table 13; PCR, *C31* insertion site). In addition, all new *C31* alleles, *i.e.* lines 3,7,41, 49, 50,

67, 76, and 82, showed the PCR products for the designated endpoints of the desired deletion (Table 13; PCR, 3' P G0071 and 5' P G0099). These products can only be amplified, if both, the starter P element ends proximal to the desired deficiency and the neighboring genomic sequences are present suggesting that the new *C31* alleles do not carry the expected deletion.

To corrobotate this unanticipated finding, genomic DNA of ΔP/ *C31* flies was isolated, *BamH* I or *Sal* I restriction-digested, and blotted onto nylon membranes. The resulting Southern blots were hybrid-ized to isolated genomic fragments of the *org-1* locus (Figure 26). A 4,7 kb large *BamH* I/ *BamH* I fragment that includes the *C31* insertion site detected the wild-typic restriction fragment in addition to the 1,3 kb enlarged *C31* fragment in all investigated ΔP/ *C31* preparations. The 4 other probes recognized only wild-typic restriction fragments without length polymorphisms indicating that the ΔP chromosomes do not bear P element insertions nor deletion breakpoints within the investigated interval.

Characterization of ΔP candidates: Relevant primers
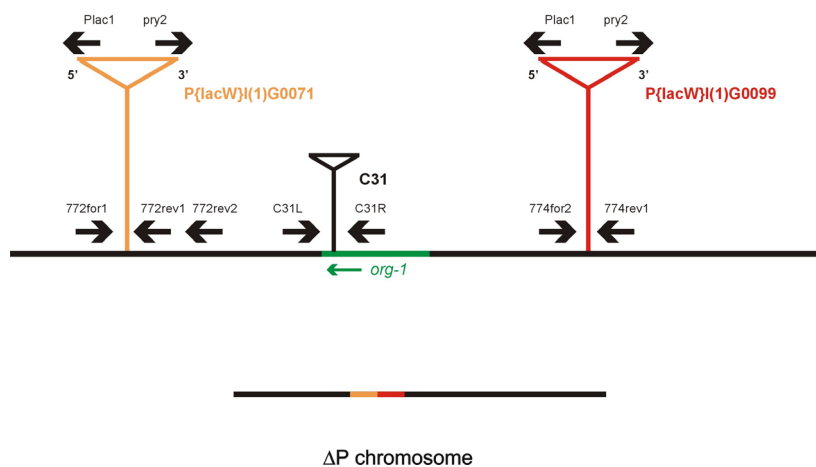
Figure 25. Characterization of ΔP candidates: Relevant PCR primers.

The results of the PCR analysis in combination with the hybridization experiments unambiguously show that the new *C31* alleles do not have the expected 95 kb deficiency. Furthermore, the assumed relation between *C31* and *org-1* appears highly questionable now, since no alterations at the *org-1* locus could be found in the new *C31* alleles.

What, then, causes the *C31* syndrome? Alternatively, *C31* is not a consequence of a mutation in *org-1*, but is caused by a gene defect at a closely linked locus instead.

A new focus for the mutation in *C31* is provided by the results of the PCR analysis described above. Three of the four starter P element ends could be amplified from all 8 new *C31* alleles (Table 13; PCR, 3' P G0071, 5' and 3' P G0099), whereas the 5' product for P G0071 was absent. This product, however, was obtained from the 3 lethal ΔP stocks that do not uncover *C31* (lines 17, 73, and 80; Table 13; PCR, 5' P G0071).

Thus, the *C31* alleles among the lethal ΔP lines correlate with the absence of the 5' P G0071 PCR product.

This PCR product can only be amplified, if both primer annealing sites are intact. It will not be formed, if the 5' end of the G0071 P element and/or the normal genomic sequence upstream of P G0071 are missing. Therefore, it seems plausible, that a deletion of genomic DNA distal to P G0071 is responsible for the lack of this PCR product and for the phenotype of the new *C31* alleles. If, in addition, such ΔP *C31* lines retained an intact 5' P{lacW} G0071 end, one would be able to amplify their present flanking genomic sequences by 5' iPCR.

Genomic DNA of ΔP flies was isolated, digested with *Cfo* I or *Sau3A* I restriction endonucleases, circularized, and subsequently used as template for 5' iPCRs. Multiple iPCR products were obtained for

	genetics		molecular analysis					Southern
			PCR					Southern
candidate	phenotype of ∆P	wings of ∆P/*C31* flies	5' P G0071	3' P G0071	5' P G0099	3' P G0099	*C31* insertion site	RFLP analysis
1	viable;*white⁻*	n.d.	-	-	-	-	wild type	n.d.
3 [3-5]	lethal	held-out	-	+	+	+	wild type	wild type
7 [7-4]	lethal	held-out	-	+	+	+	wild type	wild type
17 [17-2]	lethal	wild type	+	+	+	+	wild type	wild type
25	viable;*white⁻*	n.d.	-	-	-	-	wild type	n.d.
30	viable;*white⁻*	n.d.	-	-	-	-	wild type	n.d.
35	viable;*white⁻*	n.d.	-	-	-	-	n.d.	n.d.
37	viable;*white⁻*	n.d.	-	-	-	-	wild type	n.d.
38	viable;*white⁻*	n.d.	-	-	-	-	wild type	n.d.
41[41-5]	lethal	held-out	-	+	+	+	wild type	wild type
42	viable;*white⁻*	n.d.	-	-	-	-	wild type	n.d.
49 [49-5]	lethal	held-out	-	+	+	+	wild type	wild type
50 [50-2;50-3]	lethal	held-out	-	+	+	+	wild type	wild type
58	viable;*white⁻*	n.d.	-	-	-	-	wild type	n.d.
67 [67-1;67-4]	lethal	held-out	-	+	+	+	wild type	wild type
69	viable;*white⁻*	n.d.	-	-	n.d.	-	wild type	n.d.
73[73-5]	lethal	wild type	+	+	+	+	wild type	wild type
76[76-2]	lethal	held-out	-	+	+	+	wild type	wild type
80 [80-2;80-5]	lethal	wild type	+	+	+	+	wild type	wild type
81 [81-1]	viable;*white⁻*	wild type	-	-	-	-	wild type	n.d.
82	lethal	held-out	-	+	+	+	wild type	wild type
83	viable;*white⁻*	n.d.	-	-	-	-	n.d.	n.d.
A [A-1]	viable;*white⁻*	wild type	-	-	-	-	wild type	wild type

Table 13. Molecular and genetic characterization of ∆P *C31* candidates.
Candidate numbers in brackets are the names of established and kept stocks. n.d.: not determined. The presence or absence of the specific PCR product is indicated by "+" or "-".

most of the lines. These amplificates were gel-extracted and sequenced. Readable P element flanking genomic sequences were analyzed using BLASTN database searches. 2 P element transpositions (line 3-5, interchromosomal; 80-5, local into hot-spot), 3 transpositions within P elements (41-5, 76-2, 82-5), and 4 starter P G0099 products (41-5, 49-5, 76-2, 82-5) were identified. In contrast, the original 5' iPCR product of the P G0071 starter element, has not been found among the sequenced amplificates, consistent with the lack of the 5' PCR products for the ∆P *C31* lines. Instead, several P element flanking genomic sequences were obtained that map further distally to the P G0071 site (Figure 27).

For 3 lines, 49-5, 50-3, and 82-5, these genomic sequences map to AE003443 kb 253,9-254,1, about 53 kb distal to the P G0071 insertion site.

Line 67-1 has P element 5' neighboring genomic sequences about 124 kb further upstream at AE003443 kb 129,3, line 41-5 has P element 5' neighboring genomic sequences at AE003425 kb 145,5 kb on X chromosome at 3B-3C. Line 49-5 eventually has an additional P element transposition to AE003443 kb 232,5 that remains dubious, however, as only 14 internal bp of the 79 bp flanking genomic sequence could be aligned.

For a number of sequence reactions, genomic sequences could not be gained due to a poor sequence quality or, more frequently, because of a superposition of several sequences beyond the P element portion. The latter problem points to an inhomogeneous template DNA apparently caused by the concomitant gel-extraction of two or more different PCR products of the same size.

RFLP analysis of ΔP/C31 flies
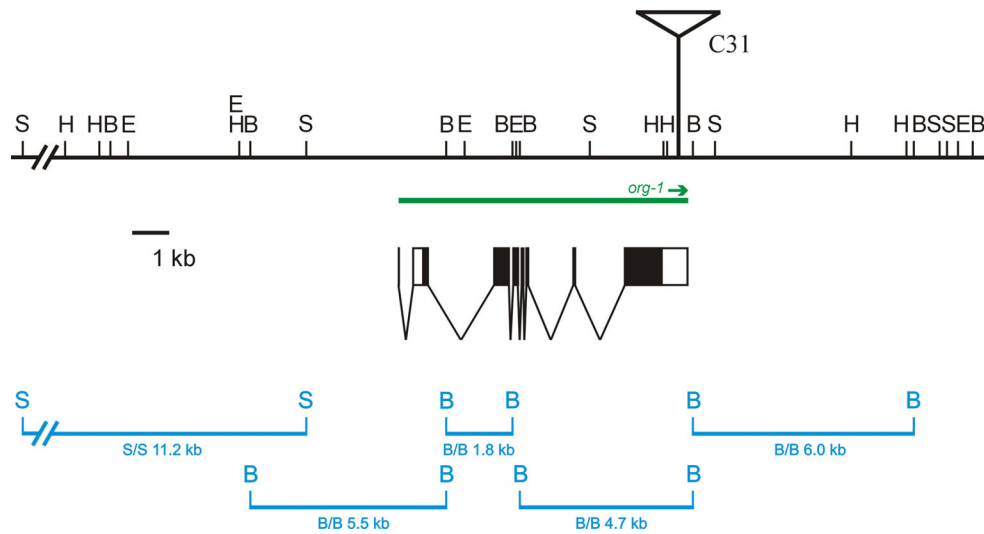
Figure 26. Characterization of ΔP candidates: RFLP analysis.
A restriction map of the *org-1* locus is shown with restriction sites specified for *BamH* I (B), *EcoR* I (E), *Hind* III (H), and *Sal* I (S). The *C31* I element is shown as triangle with a bar pointing to the integration site. The *org-1* transcription unit is shown in green. The exon-intron structure is given below. Exons are represented by boxes, introns by thin bars. Filled boxes represent the translated region of the transcript. Isolated restriction fragments of cosmid 97G10 that have been used as probes in the RFLP analysis are shown as blue lines at positions that correspond to the locations within the *org-1* map.

Table 14 summarizes the results of this sequencing analysis.

All obtained genomic sequences derived from P elements with the same orientation as P G0071. Therefore, they may result from local transpositions as well as from deletions distal to P G0071.

Some 55 kb genomic sequence distal to the P{lacW} G0071 insertion site was searched for genes or transcripts (Table 15). 4 predicted genes lie within this region. In addition, 3 previously unpredicted transcripts could be identified by EST clones. Two of them, represented by the EST clones RH09582.5' and GM09770.5' have previously been shown to be associated with lethal P element insertions (see chapter 4.2.2).
Most interestingly, the third transcript, identified by EST clone GH26370.5', is affected in ΔP *C31* lines 49-5, 50-3, and 82-5. These 3 lines contain P{lacW} sequences within the 5' region of GH26370.5' (see Figure 27). Regardless whether the P elements within this transcription unit are new insertions or a consequence of distal deletions from P{lacW} G0071, a defect in GH26370.5' coincides with the *C31* syndrome in lines 49-5, 50-3, and 82-5. This makes the transcription unit GH26370.5' a (new) candidate for *C31*.

In summary, the molecular characterization of ΔP *C31* stocks showed that the 8 new *C31* alleles do not contain the desired deletion in 7E-7F. Furthermore, it is likely that *C31* is not caused by aberrations in *org-1*, but by mutations distal to P G0071, possibly in the transcription unit GH26370.5'.

4.3.5 Jump out mutagenesis II: Generation of deletions at the *org-1* locus

Our first attempt to isolate precise deficiencies across the *org-1* locus by screening for new *C31* alleles remained unsuccessful, probably because *C31* is not uncovered by the intended deletion. Therefore, a different strategy was used to screen for deletion candidates when we repeated the initial jump out mutagenesis. Mutagenized flies were now scored for the loss of the *miniwhite* P element marker genes in spite of the observation by Cooley *et al.* (1990) that the vast majority (35/45 = 77,8%) of induced deletions retained a functional marker.

stock	iPCR	products	sequence analysis
3-5	5' *Cfo*	650 bp 800 bp 1200 bp	n.d.
3-5	5' *Sau*	600 bp 650 bp 1200 bp	kb 215,9-216 in AE003463 on chromosome 2 R overlaying sequence poor sequence quality
7-4	5' *Cfo*	550 bp 650 bp	n.d.
7-4	5' *Sau*	800 bp 900 bp	overlaying sequence overlaying sequence
41-5	5' *Cfo*	650 bp	kb 101,6-101,7 in AE003444; P G0099
41-5	5' *Sau*	1200 bp	kb 145,5 kb in AE003425 on X chromosome at 3B-C; P element inserted in P element
49-5	5' *Cfo*	550 bp 650 bp 700 bp	kb 254,1 in AE003443 kb 101,6-101,7 in AE003444; P G0099 overlaying sequence
49-5	5' *Sau*	500 bp 600 bp 1200 bp	short genomic sequence internal 14 of 79 bp genomic sequence:kb 232,5 in AE003443 poor sequence quality
50-3	5' *Cfo*	600 bp 700 bp	n.d.
50-3	5' *Sau*	500 bp 600 bp 1200 bp	kb 254,1 in AE003443 overlaying sequence overlaying sequence
67-1	5' *Cfo*	700 bp 900 bp 1200 bp	n.d.
67-1	5' *Sau*	800 bp 1100 bp	kb 129,3-129,4 in AE003443 overlaying sequence
76-2	5' *Cfo*	550 bp 650 bp 800 bp	overlaying sequence kb 101,6-101,7 in AE003444; P G0099 overlaying sequence
76-2	5' *Sau*	1300 bp 1600 bp	P element inserted in P element overlaying sequence
80-5	5' *Cfo*	650 bp 750 bp	n.d.
80-5	5' *Sau*	550 bp 650 bp	kb 297,5 in AE003443 overlaying sequence
82-5	5' *Cfo*	550 bp 650 bp	kb 253,9 in AE003443 kb 101,6-101,7 in AE003444; P G0099
82-5	5' *Sau*	1100 bp 1300 bp	kb 101,6-101,7 in AE003444; P G0099 P element inserted in P element

Table 14. DNA sequencing of 5' iPCR products of ΔP stocks.

All investigated stocks have a lethal ΔP chromosome. Stock numbers in bold symbolize *C31* alleles. n.d.: not determined.
overlaying sequence: sequences beyond the P element portion could not be read due to superimposed sequences.

gene/transcript molecular function	position within genomic sequence AE003443	EST clones	mutant alleles	references
GH26370.5'	254,0-252,4 kb	1 GH	-	-
CG1387	272,7-276,2 kb	10 AT, 1 GH	-	-
CG15345	285,2-284,4 kb	1 GM, 1 RE	-	-
CG10555	285,9-289,5 kb	1 GH, 11 LD, 7 RE, 1 RH, 1 SD	-	-
CG11190	294,2-290,4 kb	5 AT, 11 LD, 15 RE, 2 RH, 2 SD	-	-
RH09582.5'	297,0-295,5 kb	1 RH	6	this work
GM09770.5'	296,3-297,6 kb	1 GH, 2 GM, 3 RH	13, lethal	this work

Table 15. Genes distal to P{lacW} G0071
Accumulated data of genes within the interval AE003443 kb 250-AE003444 kb 8 are presented. The number of EST clones is given for different clone sources separately. EST sources are: AT testis, GH adult head, GM ovary, LD embryo, RE normalized embryo, RH normalized head, SD Schneider cells.

P element insertions or deletion endpoints in new *C31* alleles
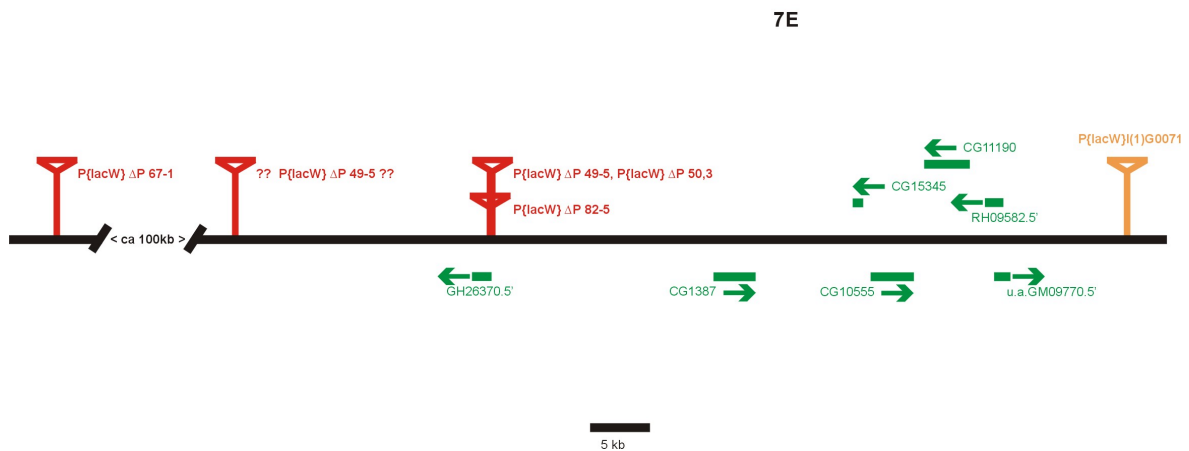
Figure 27. P element insertions or deletions endpoints in new *C31* alleles.
P elements from several ΔP *C31* lines map distal to P G0071 (orange triangle) and are shown as red triangles. Bars point to their insertion sites (if transpositions) or distal endpoints (if deletions from P G0071). Genes and transcripts distal to P{lacW} G0071 and within the sequence interval AE003443 kb 250-AE003444 kb 8 are drawn in green with arrows indicating the direction of transcription. Distal is to the left, proximal is to the right. A 5 kb scale bar is given.

The crossing scheme of the second jump out mutagenesis is shown in Figure 28. First, a transposase stock with a FM6, *white* chromosome was generated. FM6, *w*/ Y; Δ2-3, TM3, *Sb*/ *Dr* males were mated to line A16 l(1) P{lacW} G0071, l(1) P{lacW} G0099/ FM7 virgins to yield flies with a balanced A16 recombinant chromosome and the Δ2-3 transposase gene. Such females were collected as virgins and crossed to FM6, *w*/ Y males. Their progeny was screened for virgins with white or light orange eyes. Gert Pflugfelder performed the second half of the screening procedure and all subsequent genetic work in this experiment.

52 individuals could be found and were singly mated to FM6, *w*/ Y males. 43 stocks could be established (43/52 = 82,7%), of which 8 stocks con-

tained a lethal, *white*⁻ ΔP chromosome. 16 stocks had a viable, *white*⁻ ΔP chromosome and the remaining 19 lines had a lethal, *white*⁺ ΔP chromosome. All lethal and two of the viable *white*⁻ ΔP stocks were subsequently examined for deletions at the *org-1* locus using an RFLP analysis as described in chapter 4.3.4. The lines to be investigated were crossed to *C31*. Southern blots were made of *BamH* I or *Sal* I digested genomic DNA of ΔP/ *C31* flies and hybridized to the 4,7 kb large *BamH* I/ *BamH* I fragment that includes the *C31* insertion site (Figure 26). The 6,0 kb large *BamH* I/ *BamH* I fragment of the *C31* chromosome was obtained for all ΔP/ *C31* preparations, however, the concomitant wild type fragment lacked for lines 23, 24, 31, and 39. Hybridization experiments with the corresponding *Sal* I blots further confirmed that the homologous fragment is absent on these 4 ΔP chromosomes.

The 4,7 kb *BamH* I/ *BamH* I fragment comprises a major part of the *org-1* gene including 4 coding exons. Hence, we have isolated 4 mutants deficient in *org-1*.

Next, the 4 deletion mutants and 7 additional ΔP lines were checked for the presence or absence of genomic DNA at the starter P element ends using previously developed PCRs (see chapter 4.3.4 and Figure 25). Both deletion-proximal PCR products are absent in line 31 (Table 16; PCR, 3' P G0071 and 5' P G0099). Lines 23 and 24 lack the 3' P G0071 product, but may have retained intact 5' P G0099 ends. For the deletion mutant 39 as well as 7 additional ΔP stocks, amplificates in all 4 PCRs were obtained.

Thus, we could isolate 4 ΔP stocks in which *org-1* is at least partly deleted. The PCR analysis for the starter P element ends suggests that the size of the deleted segment may vary within the 4 deficiencies. Whereas ΔP lines 23, 24, and 31 putatively carry the designated 95 kb deficiency with its expected endpoints, line 39 must contain are more restricted internal deletion. Further molecular and genetic experiments, however, are required to fully characterize these new deletions.

jump out mutagenesis II: screen for loss of P element markers

P   *y, w* ; $\dfrac{\underline{\Delta2\text{-}3, \text{TM3}, Sb}}{Dr}$ x $\dfrac{\underline{\text{FM6}, w}}{Y}$; III

         ♀♀   ♂♂

F1 $\dfrac{\underline{\text{FM6}, w}}{X}$; $\dfrac{\underline{\Delta2\text{-}3, \text{TM3}, Sb}}{III}$ x $\dfrac{\underline{\text{FM6}, w}}{Y}$; III

         ♀♀   ♂♂

F2 $\dfrac{\underline{\text{P\{lacW\} l(1)G0071, P\{lacW\} l(1)G0099}}}{\text{FM7i-pAct-GFP}}$; III   x $\dfrac{\underline{\text{FM6}, w}}{Y}$; $\dfrac{\underline{\Delta2\text{-}3, \text{TM3}, Sb}}{Dr}$

         ♀♀   ♂♂

F3 $\dfrac{\underline{\text{P\{lacW\} l(1)G0071, P\{lacW\} l(1)G0099}}}{\text{FM6}, w}$; $\dfrac{\underline{\Delta2\text{-}3, \text{TM3}, Sb}}{III}$ x $\dfrac{\underline{\text{FM6}, w}}{Y}$

         ♀♀   ♂♂

screen for ♀♀ with *white* or light orange eyes.
cross candidates individually to FM6, w/Y ♂♂.

F4 $\dfrac{\underline{\Delta\text{P\{lacW\} l(1)G0071, P\{lacW\} l(1)G0099}}}{\text{FM6}, w}$; $\dfrac{\underline{Sb+}}{III}$ x $\dfrac{\underline{\text{FM6}, w}}{Y}$

         ♀♀   ♂♂

Figure 28. Crossing protocol for the jump out mutagenesis II: Screen for loss of P element markers.

| | genetics | | molecular analysis | | | | |
| | | | PCR | | | | Southern |
candidate	phenotype of ΔP	wings of ΔP/*C31* flies	5' P G0071	3' P G0071	5' P G0099	3' P G0099	RFLP analysis
1	viable;*white*⁻	wild type	n.d.	n.d.	n.d.	n.d.	wild type
2	viable;*white*⁻	wild type	n.d.	n.d.	n.d.	n.d.	wild type
3	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
4	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
5	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
6	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
7	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
8	lethal;*white*⁻	wild type	+	+	(+)	+	wild type
9	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
10	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
11	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
12	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
14	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
15	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
16	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
19	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
20	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
22	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
23	lethal;*white*⁻	wild type	+	-	(-)	+	**deletion**
24	lethal;*white*⁻	wild type	+	-	(+)	+	**deletion**
25	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
26	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
28	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
29	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
30	lethal;*white*⁻	wild type	+	(+)	(+)	(+)	wild type
31	lethal;*white*⁻	wild type	-	-	-	+	**deletion**
32	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
35	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
38	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
39	lethal;*white*⁻	wild type	+	+	+	+	**deletion**
40	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
41	lethal;*white*⁻	wild type	+	+	+	+	wild type
42	lethal;*w*⁺⁻	wild type	n.d.	n.d.	n.d.	n.d.	wild type
43	lethal;*w*⁺	wild type	(+)	+	+	+	wild type
45	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
46	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
47	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
48	lethal;*w*⁺	wild type	+	+	+	+	wild type
49	lethal;*w*⁺	wild type	n.d.	n.d.	n.d.	n.d.	wild type
50	lethal;*white*⁻	n.d.	+	+	(+)	(+)	wild type
51	viable;*white*⁻	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
52	lethal;*w*⁺	wild type	+	+	+	+	wild type

Table 16. Molecular and genetic characterization of ΔP candidates.

ΔP candidates with deletions in *org-1* are in bold. n.d.: not determined. The presence or absence of the specific PCR product is indicated by "+" or "-". Brackets symbolize vague PCR results.

4.3.6 The enhancer trap line MP8

The enhancer trap line MP8 [Matze Porsch, 8th fly of interest in screen] was isolated by Gert Pflugfelder as a byproduct of the P element jump out mutagenesis II due to its remarkable eye coloring. In MP8 flies, red pigments are restricted to the ventral part of the eye. Their P{lacW} insertion was genetically mapped on the II chromosome, and a homozygous stock has been established. MP8 genomic DNA was isolated to allow iPCR amplifications. 5' and 3' iPCRs were conducted with *Cfo* I or *Sau3A* I digested, self-ligated DNA resulting in a single product for each reaction. The 3' products for *Cfo* I and *Sau3A* I of 900 bp and 700 bp, respectively, were gel-purified and sequenced. Identical flanking genomic sequences were obtained. BLASTN searches with the P element neighboring sequence placed it within the 307 kb large genomic sequence AE0035778 on chromosome 2L. *sloppy paired 2* (*slp2*) could be identified as associated gene. *slp2* is a single exon gene and encodes a fork-head transcription factor (Grossniklaus *et al.*, 1992). It has originally been cloned by enhancer trapping. Two mutant alleles are described for *slp2*: a deficiency that removes *slp2* regulatory sequences and a deletion that lacks *slp2* and the neighboring *slp1* gene.

The P{lacW} element in MP8 is inserted in the *slp2* promoter region (Figure 29).
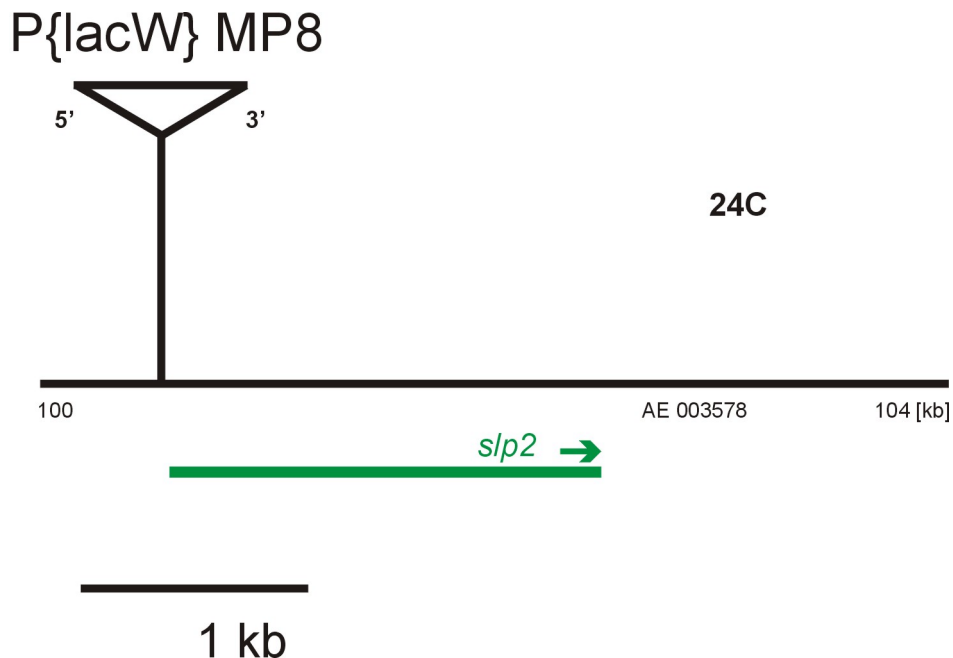

Figure 29. The enhancer trap MP8 P{lacW} insertion in *slp2*.
The P element is indicated as a triangle with a bar pointing to the insertion site along the genomic sequence (black line). The *slp2* gene is shown in green with an arrow indicating the transcriptional direction. The cytological position is given above the genomic sequence. A 1 kb scale bar is given.

5. Mapping determinants of functional specificity in OMB and ORG-1

5.1 Consequences of ectopic *omb* and *org-1* expression in *Drosophila* development

Several previous studies on T box genes revealed profound consequences on developmental processes in gain-of-function situations of these genes (*e.g.* Cunliffe and Smith, 1994; O'Reilly *et al.*, 1995; Grimm and Pflugfelder, 1996). We, therefore, were curious to know, whether ectopic *org-1* would influence the normal developmental programme as well and, if yes, how.

To address these questions, we generated UAS-*org-1* transgenic flies and ectopically expressed *org-1* using the Gal4/ UAS system (Brand and Perrimon, 1993). 5 different Gal4 driver lines were included in our analysis, *dpp*-Gal4-K54 (GOP stock # 530) (Staehling-Hampton *et al.*, 1994), E132-Gal4 (502) (Halder *et al.*, 1995), 30A-Gal4 (567) (Brand and Perrimon, 1993), GMR-Gal4 (786), and *omb*[P3]-Gal4 (55), each of which providing the yeast transcription factor Gal4 in a specific expression pattern. In flies transheterozygous for the UAS-*org-1* and Gal4 transgenes, Gal4 binds to the UAS promoter and cell-autonomously activates *org-1* within the domain of Gal4 expression.

dpp-Gal4/ UAS-*org-1*

decapentaplegic (*dpp*) encodes a secreted protein of the TGF-β family that functions as a morphogen in many developmental pathways in *Drosophila*. The used *dpp*-Gal4-K54 line expresses Gal4 in the pattern of *dpp* transcription during imaginal disc development. We observed that *dpp*-Gal4-K54 driven ectopic *org-1* expression severely interferes with the normal development of many organs and results in flies with a plethora of remarkable phenotypes (Figures 30 and 31).

The dorsal thorax of *dpp*-Gal4-K54/ UAS-*org-1* flies shows a profound, longitudinal cleft that separates the anterior scutum medially into two symmetrical halves. This cleft ends at about the center of the scutum, from where a tumourous-like outgrowth of unidentified tissue extends posteriorly and replaces all the posterior scutum and the scutellum (Figure 30 A,C,D). The focus of these defects is restricted to the notum, as the anterior dorsal abdomen appears unaltered.

A further conspicuous phenotype of *dpp*-Gal4-K54 induced ectopic *org-1* is manifested in the ventral abdomen. In wild type flies, only the dorsal abdominal segments, the tergites, are pigmented, but not the ventral sternites. Each abdominal tergite has a light brown color and contains a dark brown stripe at its posterior end (Figure 30 E). These stripes extend ventrally from the tergites into the sternites in *dpp*-Gal4/ UAS-*org-1* flies (Figure 30 F), where the contrast between the pale cuticle and the dark ectopic stripes gives the ventral abdomen of these flies a "zebra-like" pattern.

Furthermore, ectopic *org-1* causes extreme malformations of the antennae, all thoracic legs, and the wings.

The wild-typic antenna of *Drosophila* can be subdivided into 4-6 parts, according to different reference sources (Shorrocks, 1972; Casares and Mann, 2001, and references therein) (Figure 31 C). It consists of, from proximal to distal, the scape (first segment), the pedicel (second segment), the third segment, the basal cylinder, and the arista. The antennal structures distal to the pedicel are collectively referred to as the flagellum.
In the antenna, ectopic *org-1* induces a transformation of the flagellum into distal leg structures. In strong cases, the basal cylinder and the finely branched arista are completely replaced by tarsal structures (Figure 31 I, note the presence of a claw at the distal end). In addition, the third antennal segment contains ectopic bristles within its distal half. Thus, *dpp*-Gal4 driven *org-1* leads to a homeotic transformation of the distal antenna into corresponding leg structures, whereas the two proximal-most antennal segments remain unaffected.

The consequences of ectopic *org-1* on the development of the pro-, meso-, and metathoracic legs are qualitatively identical. The thoracic legs all have properly developed proximal segments, but suffer from shortened and thickened distal leg segments (Figure 31, compare J with M, N with O, and P with Q). Therefore, as for the antennal phenotype, the effect of ectopic *org-1* on distinct segments significantly differs and correlates with their relative position along the proximal-distal axis of the appendage. Whereas coxa and trochanter appear to be wild-typic, the femur and the tibia are short, and the tarsus is even further compressed to such an
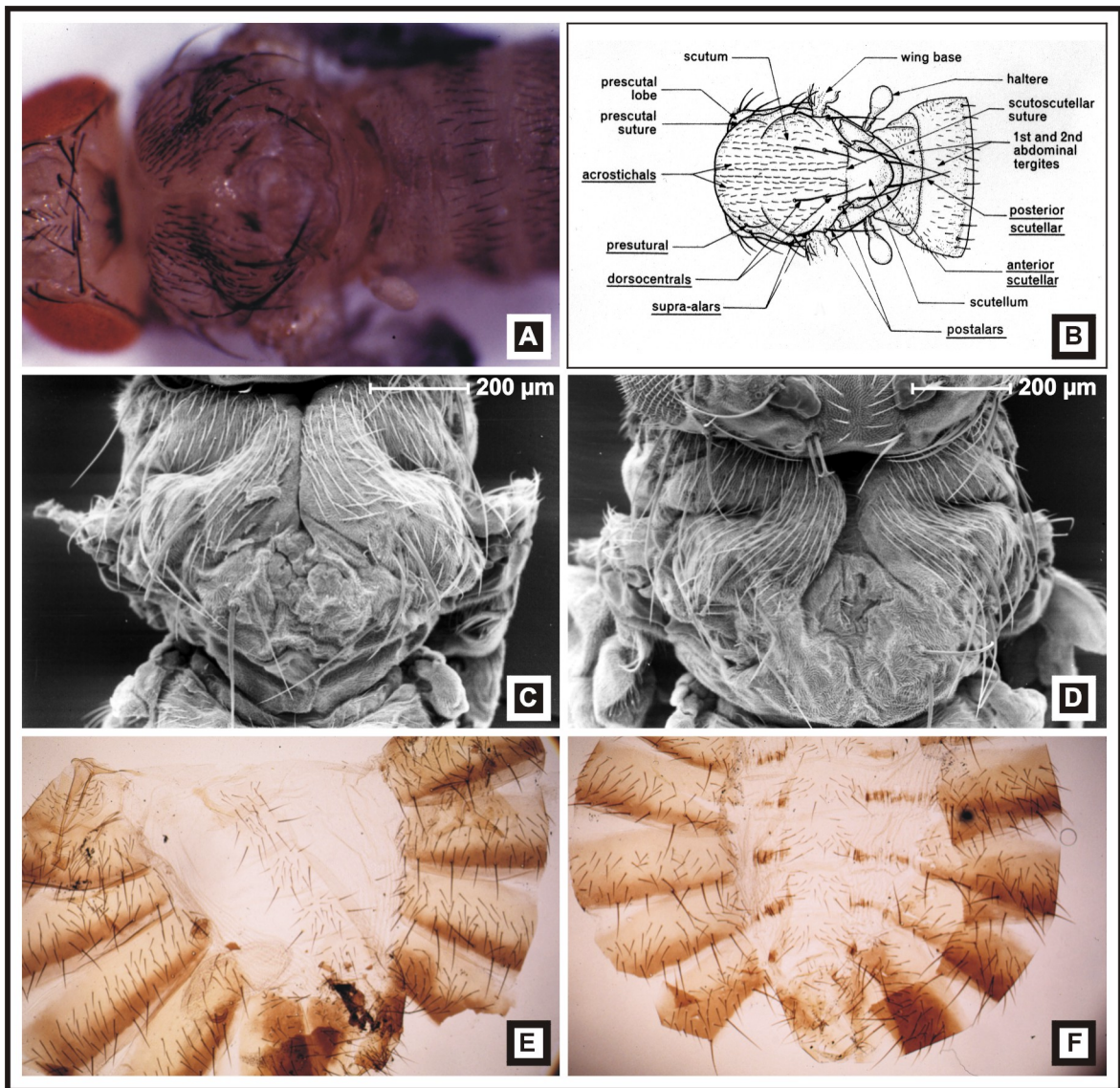
Figure 30. Consequences of *dpp*-Gal4-K54 driven ectopic *org-1* on thoracic and abdominal development.
A. Dorsal view on a *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] animal (50x magnification). A longitudinal cleft separates the dorsal scutum into two symmetrical halves. The posterior scutum and the scutellum are replaced by an unorganized outgrowth. B. Schematic drawing of the dorsal throrax and the anterior dorsal abdomen of *Drosophila* (Shorrocks, 1972). C and D. Scanning electron microscopic (SEM) pictures of the dorsal thorax of *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A3b] flies (100x magnification each). E and F. Preparations of the abdominal cuticle of a wild type (E) or a *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] (F) female. Note that the brown stripes of the abdominal pigmentation extend ventrally from the posterior ends of the tergites into the sternites (100x magnification each). All flies were grown at 25°C. The described phenotypes have also been observed with an untagged *org-1* transgene (not shown).

extend that the 5 tarsal subsegments can no longer be individually recognized. The residual tarsus is very hairy (Figure 31 O and Q), which might be an indication for the fusion of the tarsal subsegments.

We observed that the leg phenotypes are weakest in the prothoracic legs and strongest in the metathoracic legs of individual flies.

Ectopic *org-1* is manifested in the wing as well. Again, it is the distal portion of the wing that is un-

folded or fused, whereas its proximal region is less severely affected (Figure 31 D,E).

Thus, *dpp*-Gal4-K54 driven ectopic *org-1* has profound consequences on the development of distal segments of various appendages.

Most astoundingly, in spite of all these phenotypes described above, these flies free themselves from their pupal cases and live for several days, if saved from dehydration.
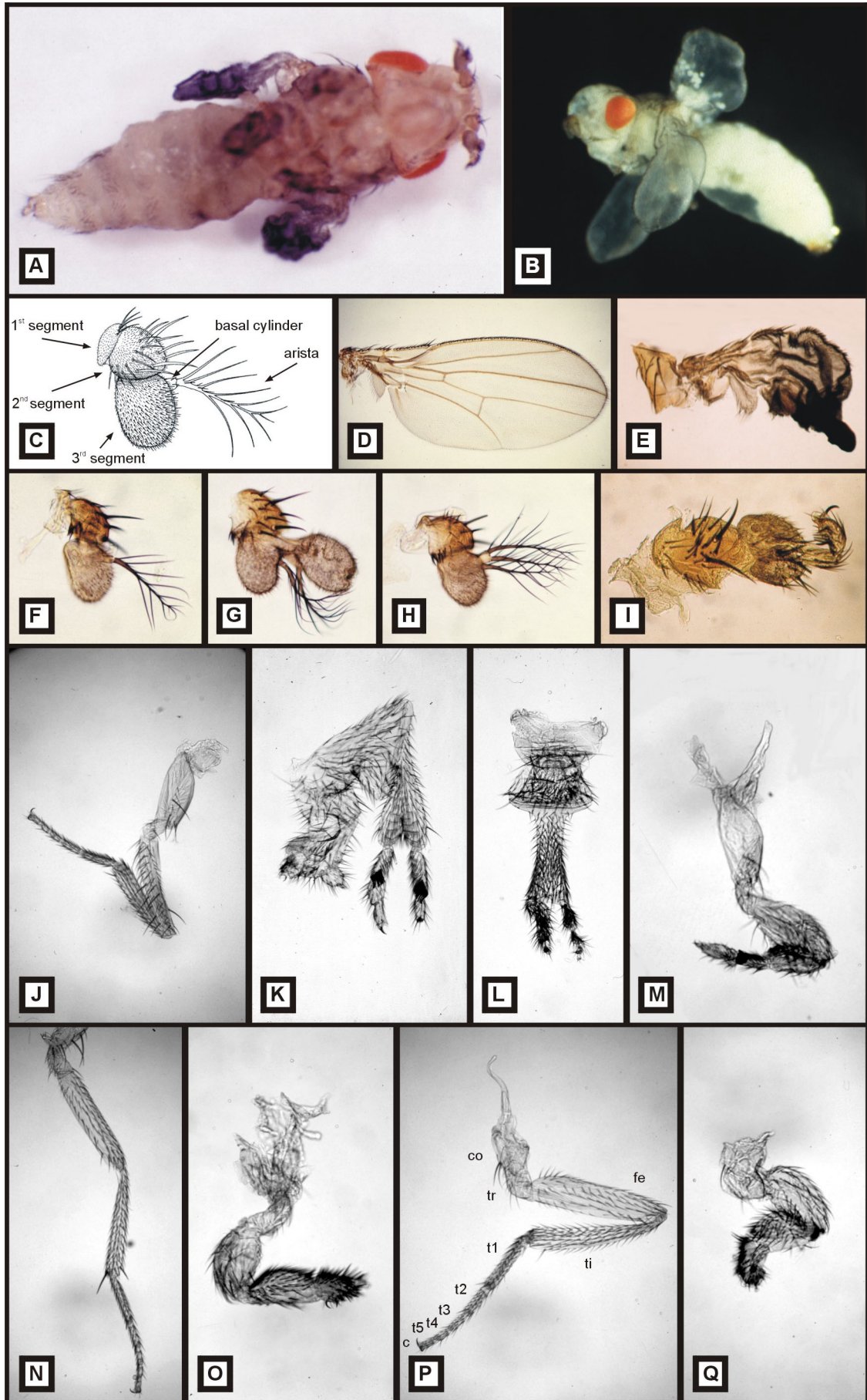
Figure 31. Consequences of ectopic *org-1* and *omb* on appendage development (previous page).
A. Habitus of a young *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] female showing antenna to leg transformations, stunted legs, and vestigial wings. The ectopic pigmentation on the ventral abdomen (Figure 30 F) is barely visible at that age (25x magnification). B. Habitus of a pharate adult *dpp.blk1*-Gal4; UAS-*omb* fly with an ectopic pair of wings (data taken from Grimm and Pflugfelder, 1996; Grimm, 1997). C. Schematic drawing of a wild-typic antenna (slightly modified after P.Bryant, flybase [http://flybase.bio.indiana.edu]). D and E. Wing preparations of a wild type male (D, 50x magnification) and a *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] female (E, 100x magnification). F-I. Preparations of antennae. F. wild type (125x magnification). G and H. antennae of animals with a heat shock induced expression of an *hsp70-omb* transgene showing a bifurcation in the third antennnal segment (G) or a triplicated arista (H) (125x magnification each) (F-H, data taken from Grimm, 1997). I. antenna of a *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] female. The third antennal segment contains ectopic bristles at its distal part. The arista is transformed into tarsal structures with a claw at the distal end (250x magnification). J-M. Prothoracic legs. J. wild type female (100x magnification). K and L. *dpp.blk1*-Gal4; UAS-*omb* flies. K. Deformed male leg with a bifurcated distal tibia. L. Fused pair of prothoracic legs. (K and L, data taken from Grimm, 1997). M. Prothoracic leg of a *dpp*-Gal4-K54/ UAS HA-*org-1*NTC-HA [A1] male. Distal leg segments are short and thickened (125x magnification). N and O. Mesothroacic leg of a wild type female (M, 100x magnification) or a *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] female (125x magnification). P and Q. Metathoracic leg of a wild type female (P, 100x magnification) or a *dpp*-Gal4-K54/ UAS-HA-*org-1*NTC-HA [A1] female (160x magnification). The 5 leg segments are, from proximal to distal: coxa (co), trochanter (tr), femur (fe), tibia (ti), and tarsus which is subdivided into 5 tarsal subsegments (t1-t5) and a pair of claws (c). Distal leg segments in O and Q are extremely compressed, when compared to the wild-typic legs in N and P.
All flies were grown at 25°C. The described phenotypes of ectopic *org-1* have also been observed with an untagged *org-1* transgene.

Like ectopic *org-1*, ectopic *omb* leads to remarkable changes of normal developmental pathways in *Drosophila* as well. The consequences of ectopic *omb* expression have already been intensely studied previously (Grimm and Pflugfelder, 1996; Grimm, 1997). *dpp*-Gal4/ UAS-*omb* flies are late pupal lethal, and, when rescued from their pupal cases, show an ectopic pair of wings and largely reduced eyes (Figure 31 B; Figure 32 D; Grimm and Pflugfelder, 1996; Grimm, 1997). Furthermore, ectopic *omb* may result in duplications of distal antennal or distal leg segments (Figure 31 G,H,K). Therefore, the ectopic expression of *org-1* and *omb* affects distal appendages differently, with *org-1*

causing stunted or transformed segments, while *omb* leads to duplications instead.

In summary, when we compare the consequences of *dpp*-Gal4 driven ectopic *org-1* with those of ectopic *omb*, we find that the induced phenotypes are qualitatively different for these related genes. Distinct effects of ectopic *omb* or *org-1* have also been obtained with other Gal4 driver lines and are described below.

GMR-Gal4/ UAS-*org-1*; GMR-Gal4/ UAS-*omb*

Gal4 expression in line GMR-Gal4 is driven under control of a *glass* enhancer, providing strong expression in the eye imaginal disc in all cells behind the morphogenetic furrow (Ellis *et al.*, 1993). The consequences of *org-1* expression on differentiating photoreceptor cells are comparably weak. The regular arrangement of the ommatidia is partly disturbed which leads to a rough appearance of such eyes. However, eyes of GMR-Gal4/ UAS-*org-1* flies retain their overall ommatidial organization (Figure 32 B). In contrast, the eyes of flies with GMR-Gal4 driven *omb* are highly degenerated (Figure 32 C). Ommatidial structures are lost and the eye pigments are diffusely spread across the eye field and are eventually concentrated at its margins. In addition, the eye size is reduced in its anterior-posterior axis. The strong impact of GMR-Gal4 driven *omb* on eye development, however, is not unanticipated. Previously used different lines with Gal4 expression during eye development gave strong eye phenotypes as well. For instance, ectopic *omb* driven by *sev*E-Gal4 results in pharate adults with similarily disorganized eyes (Grimm, 1997). *dpp*-Gal4/ UAS-*omb* leads to a severe reduction of the eye size (Figure 32 D) or even to the complete absence of the eye (Grimm, 1997). These data are consistent with the proposed role of *omb* during eye development, where *omb* functions as an "anti-eye" gene to delimit the field of the future eye (Chao *et al.*, in prep.).

30A-Gal4/ UAS-*org-1*; 30A-Gal4/ UAS-*omb*

30A-Gal4 expresses Gal4 within the blade and hinge regions of the wing imaginal disc and has a marginal expression domain within the antennal imaginal disc (Brand and Perrimon, 1993).
30A-Gal4/ UAS-*org-1* flies have with 100% penetrance lacquered, held-out wings. Strong UAS-*org-1* responder lines were semi-lethal with 30A-Gal4.
No transheterozygous 30A-Gal4/ UAS-*omb* flies could be obtained.
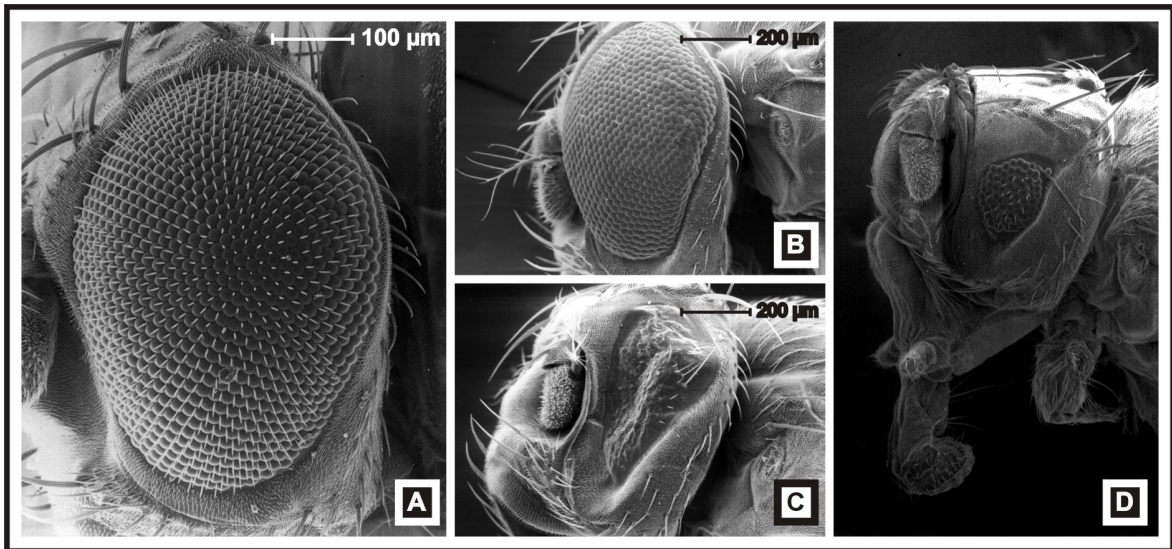
Figure 32. Consequences of ectopic *org-1* and *omb* on eye development.
A. SEM picture of a wild type eye, lateral view. The compound eye of *Drosophila* consists of a regular arrangement of about 800 ommatidia. A 100 µm scale bar is given (150x magnification, photograph courtesy of Dr. Doris Kretschmar). B. Lateral view of an eye of a GMR-Gal4/ UAS-HA-*org-1*NTC-HA [A1] fly. The regular pattern of the ommatidia is disturbed, especially noticeable at the posterior part of the eye. C. Lateral view of an eye of a GMR-Gal4/ UAS-*omb* fly. The ommatidia are completely degenerated and the eye field is reduced in its anterior-posterior axis. 200 µm scale bars are given in B and C (150x magnification each). D. Strongly reduced eye size in a *dpp*-Gal4-K54/ UAS-*omb* fly (data taken from Grimm, 1997).

E132-Gal4/ UAS-*org-1*; E132-Gal4/ UAS-*omb*

The line E132-Gal4 shows expression of Gal4 in discrete regions of various imaginal discs (Halder *et al.*, 1995). Strong UAS-*org-1* lines were lethal in combination with E132-Gal4, while weak responder lines gave E132-Gal4/ UAS-*org-1* flies with held-out wings and distal antenna to leg transformations similar to those of *dpp*-Gal4/ UAS-*org-1* flies. In addition, a small number (~10%) of flies manifested an ectopic outgrowth below the wings (data not shown).
E132-Gal4/ UAS-*omb* flies were lethal.

omb^P3^-Gal4/ UAS-*org-1*; *omb*^P3^-Gal4/ UAS-*omb*

The expression of both UAS-*org-1* and UAS-*omb* using *omb*^P3^-Gal4 was lethal.

Taken together, our ectopic expression experiments demonstrate (i) that *org-1* is capable of altering various developmental pathways such as appendage or trunk development, and (ii) that comparable gain of function situations of *org-1* and *omb* have different phenotypic consequences. The latter finding raises the question for the molecular deter-

minants of functional specificity in OMB and ORG-1 and will be addressed below.

Finally, it is important to note that we observed differences in the expressivity and penetrance of the phenotypes not only between different UAS-*org-1* transgenic lines (see chapter 5.3), but also when crosses with individual lines were repeated. Therefore, even subtle changes in parameters that influence the rearing conditions such as temperature, moisture, or food composition may significantly alter the outcome of Gal4/ UAS crosses. Therefore, in comparative experiments, lines were reared in parallel.

5.2 Identification of the determinants of functional specificity within OMB and ORG-1

Previous experiments, in which we ectopically expressed *omb* or *org-1* during imaginal disc development, revealed that both genes strongly disturb various developmental programmes that determine the morphology of the adult fly. We found, however, that the phenotypical consequences of *omb* and *org-1* were different in comparable gain-of-function situations. Such qualitative differences became so far most obvious in eye development, where *omb* counteracts eye formation, while *org-1*

leaves the developing eye nearly unaffected (see Figure 32). The consequences of ectopic *omb* and *org-1* significantly differed in trunk or appendage development, too (chapter 5.1).

These observations raise the question where within the OMB and ORG-1 protein sequences their functional specificity is encoded.

omb and *org-1* code for putative T-box transcription factors with an about 190 amino acids (aa) large, centrally located DNA binding motif, the T domain. Outside the T domains, no significant sequence similarities between OMB and ORG-1 nor to any other known protein exist.

To begin to address the question where within these two proteins specificity determinants reside, we conceptionally subdivided the proteins into three parts: the T domain, the portion N-terminal of the T domain, and the remaining sequences C-terminal of the T domain, hereafter referred to as "N domain" and "C domain", respectively (in spite of the fact that these sequences not necessarily represent functional units). We intend to determine the relevance of these domains for the functional specificity of OMB and ORG-1.

Our experimental procedure comprises (i) the cloning of a series of chimeric *omb-org-1* transgenes containing all possible OMB and ORG-1 domain compositions, (ii) the generation of transgenic fly stocks, (iii) the determination of the relative expression strength of individual transgenic lines, and (iv) the assay, in which the different constructs are tested *in vivo* with lines of similar strong transgene expression using Gal4 drivers that give distinguishable phenotypes for *omb* or *org-1*, such as GMR-Gal4 or *dpp*-Gal4-K54.

Although this strategy might represent only a first step towards the identification of specificity-relevant amino acids or peptide motifs,

the experimental outcome may already provide us a hint to how functional specificity between T-box proteins is provided mechanistically.

Accordingly, functional specificity of transcription factors can be obtained, if these have either target gene specificity (*i.e.* they regulate a distinct set of target genes) or function differently on a similar or identical set of downstream genes (*i.e.* they act as transcriptional activator or repressor). These possibilities are simplified and are not mutually exclusive.

If target gene specificity exists for OMB and ORG-1, their functional specificities could be explained by differences in their DNA binding characteristics that enable these proteins to use distinct enhancers. For that case, the crucial specificity determinants are expected to lie within the DNA binding domains.

Conversely, transcription factors that regulate identical target genes may do so by binding to the same regulatory DNA sequences and, thus, may have very similar DNA binding characteristics. Their functional specificities may then result from differences in the mode of trancriptional regulation or may be conferred by interacting accessory proteins. It is conceivable that, in such cases, the molecular determinants of specificity may also reside outside the T domain.

5.2.1 Molecular cloning of *omb-org-1* constructs

An *omb-org-1* domain swap project (DSP) was set up to map specificity determinants within OMB and ORG-1 *in vivo* using chimeric transgenes. Therefore, both proteins were conceptionally subdivided into three parts. The homologous DNA binding motif, the T domain (or T-box), is centrally located within both proteins and flanked by large N-terminal and a C-terminal domains (Figure 33).
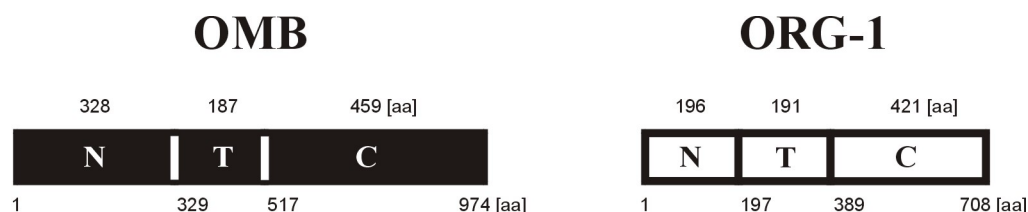
Figure 33. Domain structure of OMB and ORG-1.
The OMB and ORG-1 proteins are conceptionally subdivided into a N-terminal domain (N), a central T domain (T), and C-terminal domain (C). Numbers above the boxes indicate domain sizes, numbers below the boxes give the relative position within the proteins.

The extent of the T domains in OMB and ORG-1 was defined according to the X-ray structure of the *Xenopus* Brachyury T domain bound to its target DNA (Müller and Herrmann, 1997; for an alignment of T domain sequences, see Porsch *et al.*, 1998). The T domains of OMB and ORG-1 comprise 187 and 191 aa, respectively, and show 60,8% aa identity. Outside their T domains, however, these proteins have no significant sequence similarities.

A set of oligonucleotides was designed by Gert Pflugfelder allowing to PCR amplify the N-, T-, and C domains from the *omb* and *org-1* cDNAs (Figure 34; Table 17). Unique restriction sites were added to the primers, so that the amplified domains could be cloned into pKS and, subsequently, be used as modules from which chimeric transgenes could be assembled.

OMB-ORG-1 domain swap primers: A

OMB
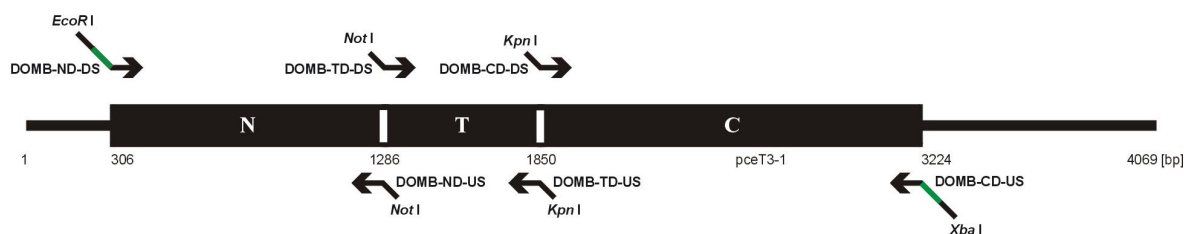
DOMB-ND-DS: *Drosophila* OMB N domain downstream primer

```
5' ACACAGAATTCAAA ATG GAG CAG AAG CTG ATC TCC GAG GAG GAC CTG AAC AGA TAC GAC GTC CAG GAG 3'
       EcoR I                                      2                               7
             met glu gln lys leu ile ser glu glu asp leu asn arg tyr asp val gln glu
```

DOMB-ND-US: *Drosophila* OMB N domain upstream primer

```
   gly val val asp asp pro ala ala al
   323                 328    Not I
3' CCG CAG CAG CTA CTA GGG CGC CGG CG TGTGT 5'
```

DOMB-TD-DS: *Drosophila* OMB T domain downstream primer

```
5' ACACA GCG GCC GCT AAG GTC ACG CTG GAG GGC 3'
         Not I   329                   334
      ala ala ala lys val thr leu glu gly
```

DOMB-TD-US3: *Drosophila* OMB T domain upstream primer3

```
   phe arg asp thr gly ala gly thr
   511                 516   Kpn I
3' AAA GCA CTA TGA CCA CGG CCA TGG TGTGT 5'
```

DOMB-CD-DS: *Drosophila* OMB C domain downstream primer

```
5' ACACA GGT ACC GGC AAG CGG GAA AAG AAT 3'
         Kpn I 517                   522
      gly thr gly lys arg glu lys asn
```

DOMB-CD-US: *Drosophila* OMB C domain upstream primer

```
   gly gly thr asp gln met glu gln lys leu ile ser glu glu asp leu asn ***
   970             974                                          Xba I
3' CCG CCA TGC CTA GTC TAC CTC GTC TTC GAC TAG AGG CTC CTC CTG GAC TTG ACT AGATCTTGTGT 5'
```

OMB-ORG-1 domain swap primers: B

ORG-1
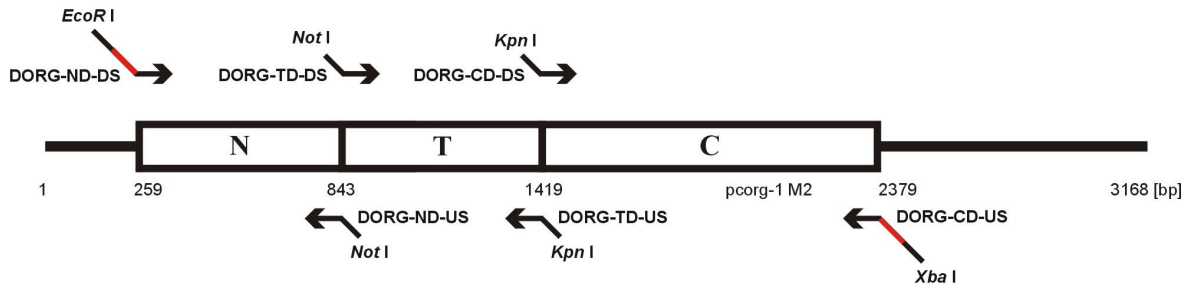
DORG-ND-DS: *Drosophila* **ORG-1 N domain downstream primer**

```
5' ACACAGAATTCAAA ATG TAC CCC TAC GAT GTG CCC GAT TAC GCC ACG CAC CTG ATG GGC CCC 3'
      EcoR I                                                  2                    7
              met tyr pro tyr asp val pro asp tyr ala thr his leu met gly pro
```

DORG-ND-US: *Drosophila* **ORG-1 N domain upstream primer**

```
   pro ser leu ala gln ala ala ala al
   191              196    Not I
3' GGT AGC GAC CGG GTC CGG CGC CGG CG TGTGT 5'
```

DORG-TD-DS: *Drosophila* **ORG-1 T domain downstream primer**

```
5' ACACA GCG GCC GCT ATT GTG GTG CTG GAG ACG 3'
         Not I     197              202
         ala ala ala ile val val leu glu thr
```

DORG-TD-US: *Drosophila* **ORG-1 T domain upstream primer**

```
   phe arg asp asp gly thr gly thr
   383              388    Kpn I
3' AAA GCC CTA CTA CCG TGG CCA TGG TGTGT 5'
```

DORG-CD-DS: *Drosophila* **ORG-1 C domain downstream primer**

```
5' ACACA GGT ACC AAC GAT GTA ACC ACT GGC 3'
         Kpn I 389              394
         gly thr asn asp val thr thr gly
```

DORG-CD-US: *Drosophila* **ORG-1 C domain upstream primer**

```
   ile asp leu val pro arg tyr pro tyr asp val pro asp tyr ala ***
   703              708                                  Xba I
3' TAT CTA GAC CAC GGC GCG ATG GGG ATG CTA CAC GGG CTA ATG CGG ACT AGATCTTGTGT 5'
```

Figure 34. Design of DSP primers (previous pages).
Annealing positions of primers used to amplify individual domains are shown within the OMB (A) and ORG-1 (B) open reading frames. The primer sequences are given below along with the encoded protein sequences. Numbers indicate codon positions within the OMB or ORG-1 open reading frames. Restriction sites used for cloning are underlined. Artificially introduced amino acids at the T domain borders are shown in bold. The Cavener consensus sequences required for an efficient initiation of translation in *Drosophila* (Cavener, 1987) are marked as black boxes within the DOMB-ND-DS and DORG-1-ND-DS primers. Within oligonucleotides DOMB-CD-US and DORG-1-CD-US, *dam* recognition sequences overlapping with the *Xba* I sites are shown in bold and italics. The MYC epitope (Evan *et al.*, 1985) and HA epitope (Wilson *et al.*, 1984) sequences are in green and red, respectively.

As the *omb* and *org-1* sequences did not contain identical restriction sites that coincide with the T domain ends, *Not* I and *Kpn* I sites were artificially introduced at the 5' and 3' end of the T domains, respectively. The N domain downstream primers contained terminal *EcoR* I sites, while the C domain upstream primers supplied *Xba* I sites. Thereby, composite *omb-org-1* chimeric genes could be directly cloned *EcoR* I-*Not* I-*Kpn* I-*Xba* I under control of the Gal4 UAS promoter into the

germline transformation vector pUAST via *EcoR* I/ *Xba* I.

MYC and HA epitope tags were added to the N- and C-terminal domains of OMB and ORG-1, respectively, in order to make the chimeric proteins detectable for available monoclonal antibodies (mab) (Evan *et al.*, 1985; Wilson *et al.*, 1984).

Our experiment comprises the analysis of a total of 12 transgenes (Figure 35). 8 chimeric *omb-org-1* constructs containing the OMB and ORG-1 N-, T-, and C domains in all possible combinations make up the core experiment. An additional 4 full-length OMB and ORG-1 constructs, with and without epitope tags, serve as controls to monitor, if the artificially introduced amino acids at the domain borders or the added epitopes somehow influence the characterics of wild type OMB or ORG-1.

Full-length, untagged *omb* transgenic flies already existed (Grimm, 1997) and were included in our experiment. The other 11 constructs were cloned and transformed into w^{1118} flies as described below.

reaction name	product	primer pair	restriction sites	size [bp]
rxn1	MYC-omb N	DOMB-ND-DS, DOMB-ND-US	*EcoR* I, *Not* I	1062
rxn2	MYC-omb NT	DOMB-ND-DS, DOMB-TD-US3	*EcoR* I, *Kpn* I	1606
rxn3	MYC-omb NTC-MYC	DOMB-ND-DS, DOMB-CD-US	*EcoR* I, *Xba* I	3016
rxn4	omb T	DOMB-TD-DS, DOMB-TD-US3	*Not* I, *Kpn* I	586
rxn5	omb TC-MYC	DOMB-TD-DS, DOMB-CD-US	*Not* I, *Xba* I	1999
rxn6	omb C-MYC	DOMB-CD-DS, DOMB-CD-US	*Kpn* I, *Xba* I	1432
rxn7	HA-org-1 N	DORG-ND-DS, DORG-ND-US	*EcoR* I, *Not* I	634
rxn8	HA-org-1 NT	DORG-ND-DS, DORG-TD-US	*EcoR* I, *Kpn* I	1215
rxn9	HA-org-1 NTC-HA	DORG-ND-DS, DORG-CD-US	*EcoR* I, *Xba* I	2206
rxn10	org-1 T	DORG-TD-DS, DORG-TD-US	*Not* I, *Kpn* I	601
rxn11	org-1 TC-HA	DORG-TD-DS, DORG-CD-US	*Not* I, *Xba* I	1591
rxn12	org-1 C-HA	DORG-TD-DS, DORG-TD-US	*Kpn* I, *Xba* I	1012

Table 17. Summary of PCR products using *omb-org-1* DSP primers.
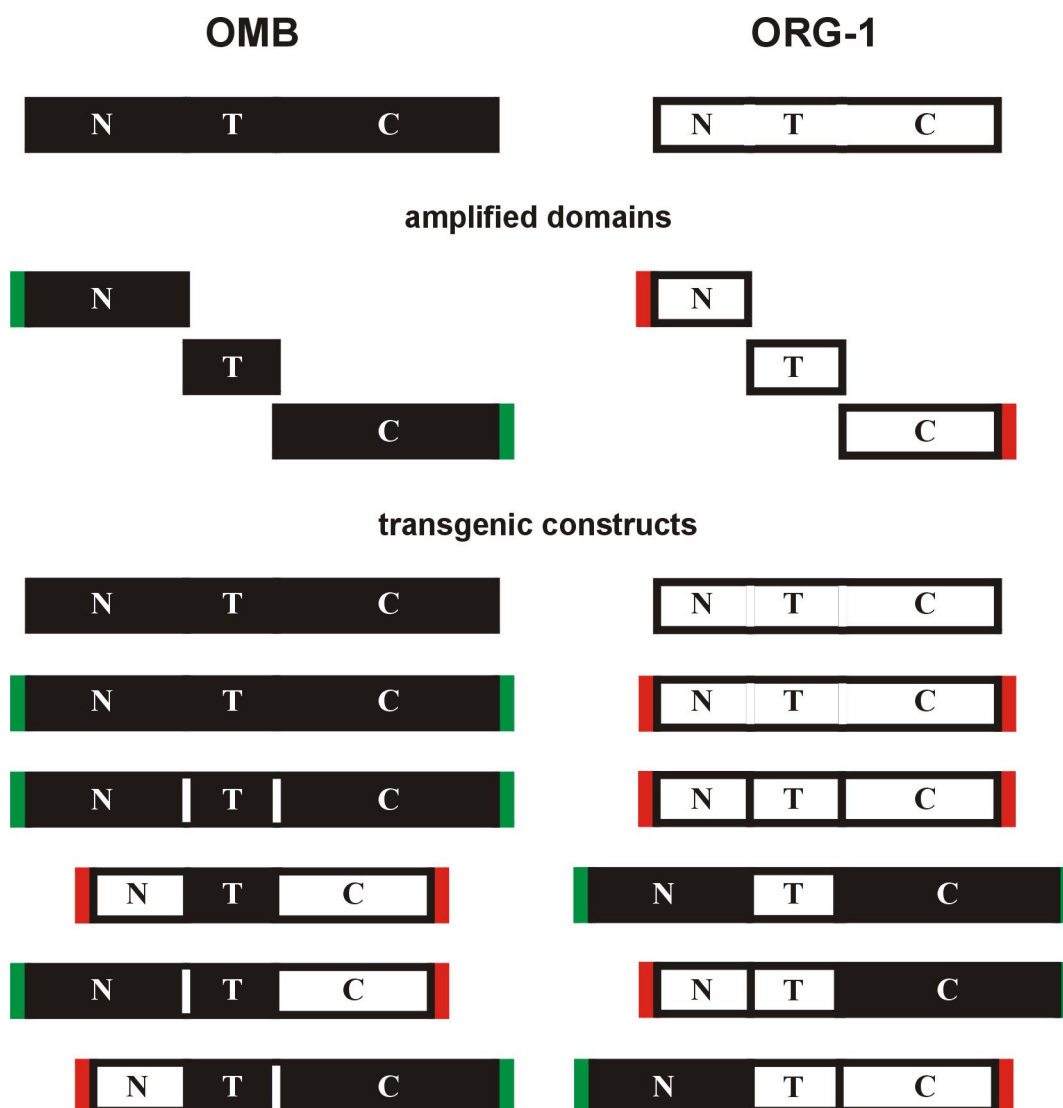
Figure 35. Summary of transgenic constructs.
Individual domains of OMB (black box) and ORG-1 (white box) were amplified and used as modules to build chimeric transgenic constructs. In addition, continuous, full-length OMB and ORG-1 constructs with or without terminal MYC (green) or HA (red) tags, respectively, were included as controls. Vertical bars within boxes symbolize discontinuous proteins assembled from single domains.

5.2.1.1 Cloning of isolated domains of *omb* and *org-1*

As a prerequisite for the cloning of *omb-org-1* chimeric genes, the N-, T-, and C domains of OMB and ORG-1 were first amplified from the cDNAs and individually cloned into pKS. The isolated domains are supplied with appropriate restriction sites at their ends, so that they could be used as modules from which chimeric proteins with any OMB or ORG-1 domain composition could be assembled.

The MYC-tagged OMB N domain was amplified from the *omb* cDNA pceT3-1 using *Pfu* DNA polymerase and the primer combination DOMB-ND-DS

and DOMB-ND-US. Terminal *EcoR* I and *Not* I sites were introduced by linker PCR, allowing the cloning of the *EcoR* I and *Not* I digested amplificate into pKS *EcoR* I/ *Not* I. Sequencing of clone rxn1-cl1 revealed a frameshift mutation within this amplificate (bp 1146 of the comb sequence was deleted).

Figure 36. Cloning of isolated domains of OMB (next page).
Strategies that gave rise to clones of error-free OMB domains are shown. Black boxes symbolize OMB sequences, green boxes MYC epitopes. Relevant restriction sites are shown in black, green or red, and a mutation site within an initial clone in blue. See text for further details.

omb C-MYC
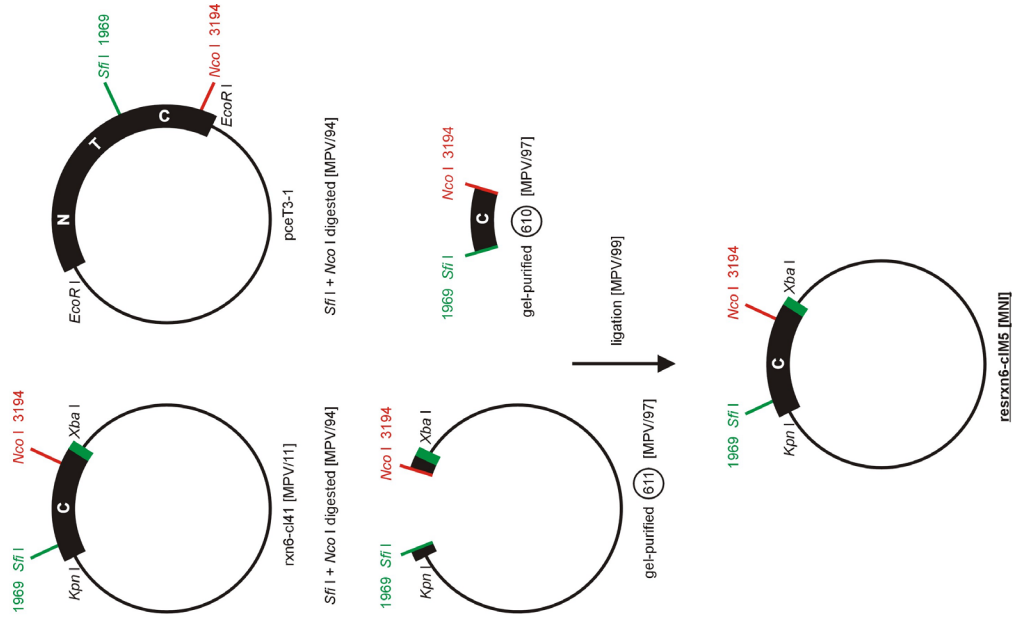
omb T
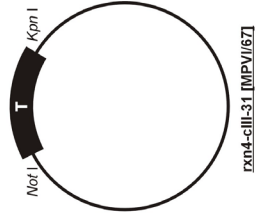
MYC-*omb* N
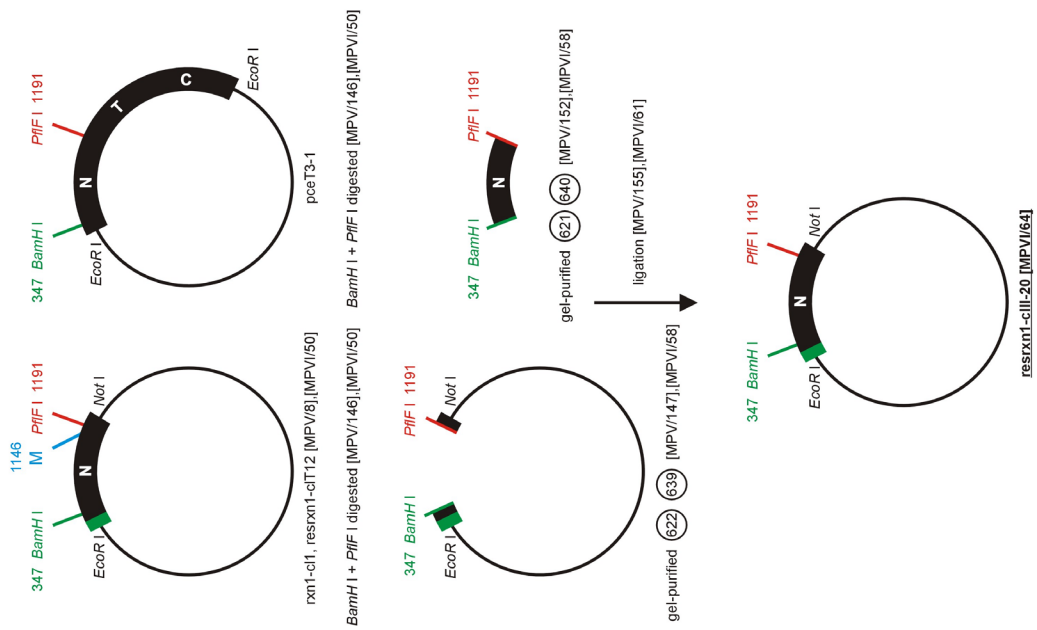
Therefore, a *BamH* I/ *PflF* I fragment of rxn1-cl1 that includes the mutation site was replaced by the homologous fragment of pceT3-1. This rescue cloning approach had to be repeated once to obtain clone resrxn1-clII-20. Sequencing of resrxn1-clII-20 confirmed the cloning of the authentic OMB N domain (Figure 36).

The OMB T domain was *Pfu* PCR amplified from pceT3-1 with primers DOMB-TD-DS and DOMB-TD-US3 that provide terminal *Not* I and *Kpn* I cloning sites, respectively. The restriction-digested PCR product was cloned into pKS *Not* I/ *Kpn* I. Clone rxn4-clII-31 was shown to contain the genuine OMB T domain sequence (Figure 36).

The MYC-tagged OMB C domain was *Pfu* PCR amplified using the primer pair DOMB-CD-DS and DOMB-CD-US. These primers added terminal *Kpn* I and *Xba* I sites, respectively, so that the amplificate could be cloned into pKS *Kpn* I/ *Xba* I. Among several partially sequenced clones, clone rxn6-cl41 was without mutations within the sequenced region. As we frequently observed sequence alterations within initial clones for other constructs, we preventively exchanged the previously unsequenced part of rxn6-cl41 for the corresponding *Sfi* I/ *Nco* I restriction fragment of pceT3-1. The integrity of the resulting clone resrxn6-clM5 was verified by sequencing (Figure 36).

The HA-tagged ORG-1 N domain was *Pfu* amplified from the *org-1* cDNA pcorg-1M2-cl10 using the primer pair DORG-ND-DS and DORG-ND-US. The primer-encoded *EcoR* I and *Not* I restriction sites enabled us to clone the digested amplificate into pKS via *EcoR* I/ *Not* I. Sequencing of the initial clone rxn7-cl51 revealed a 30 bp large insertion of unknown origin within the DORG-ND-DS primer in addition to a nonsense mutation within the amplificate (**G**GA → **T**GA transversion at bp 391 of the *org-1* cDNA sequence). These mutations could be rescued by replacing an *EcoR* I/ *PflF* I fragment of rxn7-cl51 with the corresponding fragment of clone rxn9-cl27 (Figure 37). The resulting clone resrxn7-clI4 was sequenced to prove the successful correction.

The ORG-1 T domain was amplified by linker PCR using DORG-TD-DS and DORG-TD-US and cloned into pKS *Not* I/ *Kpn* I. Sequencing of the initial clone rxn10-clO5 revealed two nucleotide transversions that both cause missense mutations

(C**C**C [Pro] → C**A**C [His] and TT**C** [Phe] → TT**A** [Leu] at bp 1085 and 1308 of the *org-1* cDNA sequence, respectively). A *Bgl* II/ *Nco* I restriction fragment that contains both mutation sites was excised from clone rxn10-clO5 and replaced with the homologous fragment of pcorg-1M2. The two mutation sites were rescued in the resulting clone resrxn10-clX4, however, it subsequently turned out, that DNA preparations from this rescue clone were inhomogeneous (a fraction of the plasmids lacked bp 1003 of the *org-1* cDNA sequence). Therefore, the plasmid preparation from resrxn10-clX4 was retransformed and a clean clone, resrxn10-clIX-9, containing an authentic ORG-1 T domain, could be isolated (Figure 37).

Finally, the HA-tagged ORG-1 C domain was *Pfu* PCR amplified from pcorg-1M2-cl10 using the primer combination DORG-TD-DS and DORG-TD-US. The terminal *Kpn* I and *Xba* I sites of the amplificate allowed its directed cloning into pKS *Kpn* I/ *Xba* I. As for most of the other constructs, sequencing of the initial clone rxn12-cl37 uncovered a mutation within the amplified domain as well. A missense mutation, caused by a single nucleotide transversion (G**G**A [Gly] → G**T**A [Val] at bp 1653 of the *org-1* cDNA sequence) was rescued by exchanging a *BssH* II/ *Nde* I fragment of rxn12-cl37 for the corresponding fragment of pcorg-1M2 (Figure 37). The integrity of the resulting clone resrxn12-clF1 was verified by sequencing.

5.2.1.2 Cloning of continuous *omb* and *org-1* transgenes

Cloning of corg-1M2 into pUAST

After several attempts had failed to directly subclone the *org-1* cDNA from pcorg-1M2-cl10 into pUAST via *EcoR* I (only clones with the *org-1* cDNA inserted in the wrong orientiation relative to the Gal UAS promoter were obtained), an alternative cloning strategy was developed and successfully employed (Figure 38).

The *org-1* cDNA was released from pcorg-1M2-cl10 by an *EcoR* I digestion and separated from the pKS *EcoR* I fragment of similar size by a concomitant *Sca* I incubation. The gel-purified corg-1M2 *EcoR* I fragment was cloned into pKS via *EcoR* I again. Several clones were obtained, including pcorg-1M2-clU6 (5'-3'), that now have the *org-1* cDNA inserted in 5' → 3' orientation in relation to the T3 promoter, and, thus, contain the *org-1* cDNA in the opposite direction as in the original clone
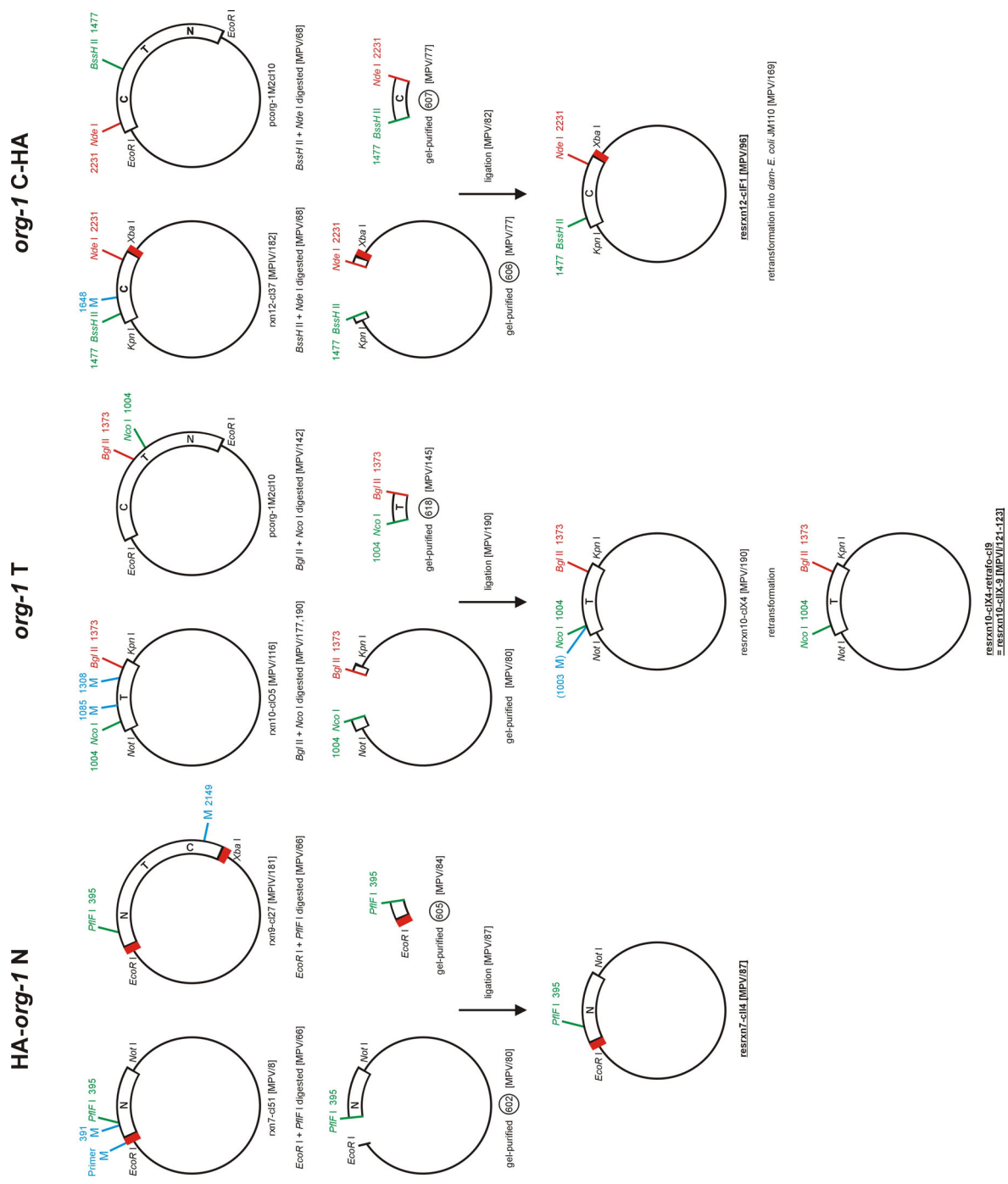
Figure 37. Cloning of isolated domains of ORG-1.
Strategies that gave rise to clones of error-free ORG-1 domains are shown. White boxes symbolize ORG-1 sequences, the red boxes HA epitopes. Relevant restriction sites are shown in black, green or red, and mutation sites within initial clones in blue. See text for further details.

pcorg-1M2-cl10. This subcloning enabled us to clone the *org-1* cDNA directed into pUAST via *EcoR* I/ *Xba* I (Figure 38).

Cloning of HA-*org-1*NTC-HA into pUAST

The full-length ORG-1 with an HA tag on either terminus was amplified by linker PCR using *Pfu* polymerase and primer DORG-ND-DS in combination with DORG-CD-US. This amplificate was cloned into pKS via *EcoR* I/ *Xba* I and the resulting clone, rxn9-cl27, was sequenced at both ends. A single nucleotide transition was detected within the

cloning of corg-1M2 into pUAST

pcorg-1M2cl10

EcoR I + *Sca* I digested [MPV/67]

gel-purified (600) [MPV71]

ligation [MPV/120]

pcorg-1M2-clU6 (5'-3') [MPV/156]

Xba I + *Sca* I + partially *EcoR* I digested [MPV/167,172]

gel-purified (632) [MPV185]

Gal UAS

pUAST

ligation [MPV/189]

Gal UAS

pUAST

corg-1M2pUAST-clV3 [MPVI/5]

transformed into *w1118* embryos [MPVI/20]
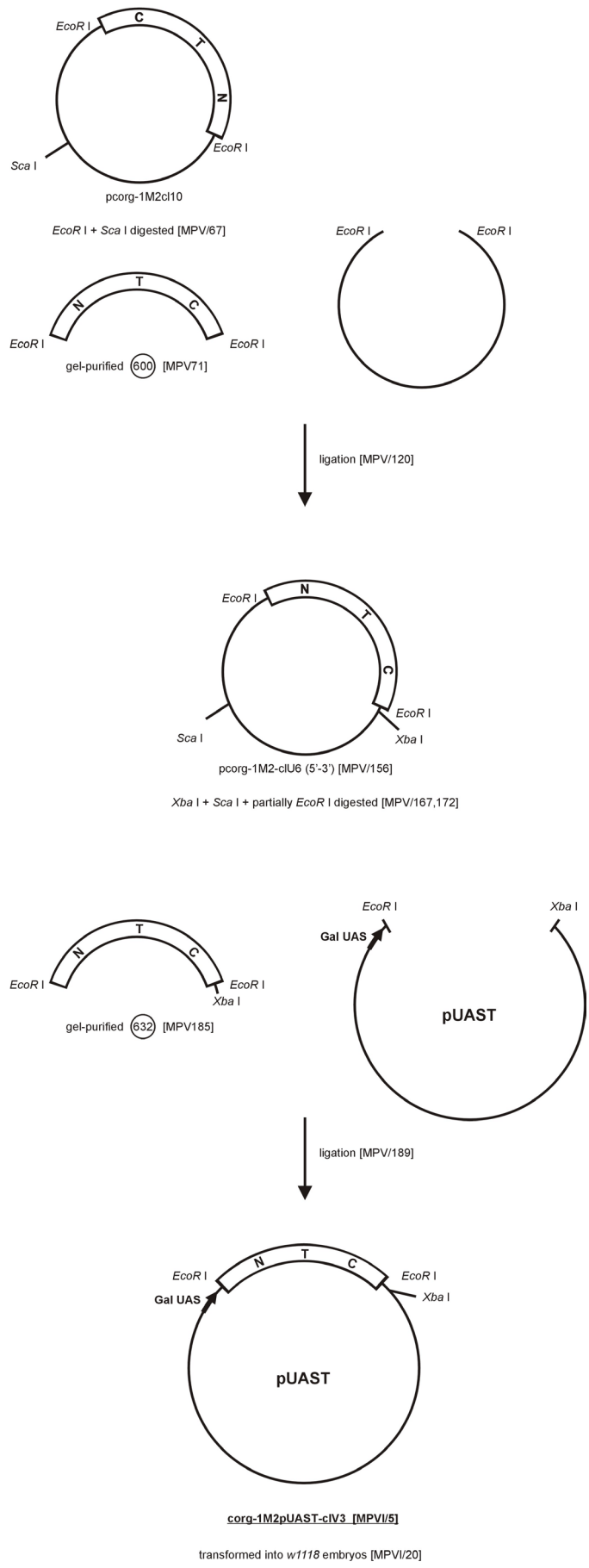
Figure 38. Cloning of corg-1M2 into pUAST (previous page).

The successful cloning strategy is shown according to previous figures. See text for further details.

sequenced region of the amplificate, changing a **G**CT [Ala] codon at position bp 2149 of the *org-1* cDNA sequence to an **A**CT [Thr]. Therefore, the internal part of the amplificate including the site of the determined nucleotide substitution was excised as a large *Nde* I/ *PflF* I fragment and replaced with the homologous fragment of pcorg-1M2-cl10 to yield clone resrxn9-clJ1 (Figure 39). Sequencing across the *Nde* I and *PflF* I cloning sites confirmed the correction of the mutation at bp 2149 and, thus, the successful rescue operation. Next, the HA-*org-1*NTC-HA construct had to be subcloned *EcoR* I/ *Xba* I into pUAST.

Unfortunately, we found that the *Xba* I site in resrxn9-clJ1 was resistant to cleavage due to *dam* methylation. The site-specific methylase Dam encoded by the *dam* gene catalyzes the transfer of a methyl group from S-adenosylmethionine to the N^6 position of adenine in the sequence GATC (Marinus and Morris, 1973; Geier and Modrich, 1979). Since we inadvertently designed the primers DOMB-CD-US and DORG-CD-US both with a T<u>GA</u> stop codon juxtaposed to the <u>TC</u>TAGA *Xba* I recognition site (see Figures 34 A,B), a *dam* recognition site overlaps the *Xba* I recognition site in all of our constructs containing an OMB or ORG-1 C-terminus. The *Xba* I cloning sites within those constructs were protected from being blocked by the use of *dam* deficient *E. coli* host cells.

Therefore, plasmid resrxn9-clJ1 was retransformed into *dam⁻ E. coli* B8 cells, before the HA-*org-1*NTC-HA construct could be released from resrxn9-clJ1 as an *EcoR* I/ *Xba* I fragment and subcloned into pUAST *EcoR* I/ *Xba* I (Figure 39).

Cloning of MYC-*omb*NTC-MYC into pUAST

The intact, N- and C-terminally MYC-tagged, OMB construct was amplified from the *omb* cDNA using primers DOMB-ND-DS and DOMB-CD-US. This amplificate, however, could not be cloned, although the apparently toxic impact of OMB constructs on bacterial host cells was considered and mild cloning conditions were applied (Roth, 1991; Grimm, 1997). The MYC-*omb*NTC-MYC construct was then cloned from three components of already ex-

isting plasmids instead. A *Pst* I/ *Sac* I fragment of rxn1-cl1 provided the N-terminal sequences and the majority of the pKS sequences, a *Sfi* I/ *Sac* I fragment of resrxn6-clM5 the C-terminal part of the construct and residual vector sequences, and a *Pst* I/ *Sfi* I fragment of pceT3-1 contributed the core fragment of the OMB sequences (Figure 40). Due to the *dam* methylated *Xba* I recognition site, the pKS construct resrxn3-clY1 had to be retransformed into *dam⁻ E. coli* SCS110 prior to subcloning of the MYC-*omb*NTC-MYC construct into pUAST via *EcoR* I/ *Xba* I.

5.2.1.3 Cloning of chimeric *omb-org-1* transgenes

The chimeric *omb-org-1* transgenes were assembled from the 6 single domains of OMB and ORG-1. The individual N-, T-, and C domains were excised from verified clones by double digestion with *EcoR* I and *Not* I, *Not* I and *Kpn* I, and *Kpn* I and *Xba* I, respectively (see chapter 5.2.1.1 and Figures 36 and 37). Both C domain constructs were first retransformed into *dam⁻ E. coli* host cells to circumvent the inhibition of methylated *Xba* I sites. The released domains were gel-purified and served as modules to construct all 8 possible OMB-ORG-1 proteins with any N-, T-, and C domain composition.
The transgenes HA-*org-1*N+*omb*T+*org-1*C-HA and HA-*org-1*N+*omb*T+C-MYC were both cloned into pKS *EcoR* I/ *Xba* I prior to subcloning into pUAST. However, since this intermediate step was rather laborious due to the required retransformation into *dam⁻ E. coli* hosts and, on the other hand, the direct cloning of the single domains into pUAST proved to be straightforward and highly efficient, all remaining chimeric *omb-org-1* transgenes were cloned straight into pUAST *EcoR* I/ *Xba* I (Figure 41).

All transgenic pUAST constructs were checked by analytical restriction digests and sequencing across the cloning sites prior to their use in germline transformations.

Figure 39. Cloning of HA-*org-1*NTC-HA into pUAST (next page).

cloning of HA-*org-1* NTC-HA into pUAST

Nde I + *Pfl*F I digested [MPV/67]

rxn9-cl27 [MPIV/181]

pcorg-1M2cl10

Nde I + *Pfl*F I digested [MPV/67]

gel-purified (603) [MPV/80]

gel-purified (604) [MPV/80]

ligation [MPV/86]

resrxn9-clJ1 [MPV/91]

retransformation into *dam- E. coli* B8:
resrxn9-clJ1-retrafo-cl6 [MPV/100]

EcoR I + *Xba* I digested [MPV/117]

gel-purified (614) [MPV117]

pUAST

ligation [MPV/123]

pUAST

org-1NTCpUAST-clP5 [ANI/41]

transformed into *w1118* embryos [MPV/171]

cloning of MYC-*omb* NTC-MYC into pUAST

rxn1-cl1 [MPV/8]

Pst I + *Sac* I digested [MPV/179]

pceT3-1

Pst I + *Sfi* I digested [MPV/179]

resrxn6-clM5 [MNI]

Sfi I + *Sac* I digested [MPV/179]

gel-purified ⑥③⓪ [MPV/183]

gel-purified [MPV/189]

gel-purified ⑥③① [MPV/183]

ligation [MPV/191]

resrxn3-clY1 [MPV/191]

retransformation into *dam- E. coli* SCS 110 [MPVI/95]

EcoR I + *Xba* I + *Dra* I digested [MPVI/116]

pUAST

gel-purified ⑥⑤⑦ [MPV/116]

ligation [ABI/39]

pUAST

ombNTCpUAST-clX-10 [MPVI/122]

transformed into *w1118* embryos [MPVI/129]

Figure 40. Cloning of MYC-*omb*NTC-MYC into pUAST (previous page).

omb-org-1 chimeric constructs directly cloned into pUAST

org-1N+T+CpUAST-cIXI-2 [MPVI/127]

transformed into *w1118* embryos [MPVI/131]

org-1N+T+ombCpUAST-cIXII-1 [MPVII/121]

transformed into *w1118* embryos [MPVII/175]

ombN+T+org-1CpUAST-cIVI-11 [MPVI/98]

transformed into *w1118* embryos [MPVI/139]

ombN+T+CpUAST-cIVIII-1 [MPVI/109]

transformed into *w1118* embryos [MPVI/139]

ombN+org-1T+ombCpUAST-cIXI-10 [MPVI/127]

transformed into *w1118* embryos [MPVI/131;139]

ombN+org-1T+CpUAST-cIVI-19 [MPVI/98]

transformed into *w1118* embryos [MPVI/131;139]

omb-org-1 chimeric constructs subcloned from pKS into pUAST

org-1N+ombT+org-1CpUAST-cIVIII-30 [MPVI/110]

transformed into *w1118* embryos [MPVI/124;131]

org-1N+ombT+CpUAST-cIVIII-21 [MPVI/110]

transformed into *w1118* embryos [MPVI/124]
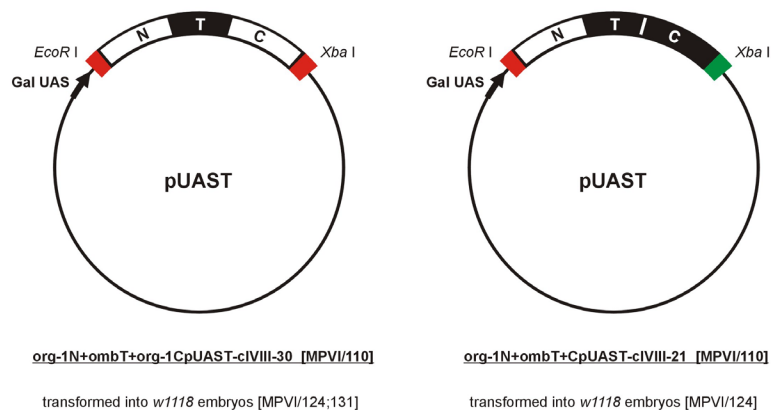
Figure 41. Cloning of *omb-org-1* chimeric transgenes into pUAST.
Black boxes symbolize OMB sequences, white boxes ORG-1 domains. The HA and MYC epitopes are shown as red and green boxes, respectively. Relevant restriction sites are given.

5.2.2 Generation of *omb-org-1* transgenic flies

As described above, 11 *omb, org-1*, or *omb-org-1* chimeric transgenes were cloned into the P element transformation vector pUAST under control of a Gal4 UAS promoter. These pUAST constructs were co-injected with pUChsπΔ2-3 helper vector (Rio and Rubin, 1985) into w^{1118} embryos in order to obtain germline transformations (Santamaria, 1986, Spradling, 1986). Injected flies were mated to w^{1118} flies and transformants could be identified by the presence of the *white*[+] marker of the pUAST vector. The transgenes were then chromosomally mapped by segregation analysis and, if homozygotically viable, made homozygous.

At least 6 independent transgenic lines could be established for each construct after one or two transformation procedures except for construct HA-*org-1*N+T+*omb*C-MYC (Table 18). Three initial at-tempts to transform this transgene remained unsuccessful, although different DNA preparations of the sequenced construct were used. After the cloning of this particular construct had been repeated (Figure 41), it could be transformed into w^{1118} flies without further complications.

The results of the germline transformations are summarized in Table 18.

To test, if the generated transgenic lines are functional, at least a subset of stocks of each transformation series was crossed to Gal4 driver lines that caused profound phenotypes with UAS-*omb* and UAS-*org-1* transgenes (see chapter 5.1). Transformants for all 11 constructs were found to be functionally active (Table 18).

series	construct	transformed clone	number of independent transformants	functionally active?	reference
A	N T C	org-1NTCpUAST-clP5 [ANI/41]	7 lines	yes	MPV/171
B	N T C	corg-1M2pUAST-clV3 [MPVI/5]	10 lines	yes	MPVI/20
C	N T C	org-1N+T+ombCpUAST-clXII-1 [MPVII/121]	8 lines	yes	MPVI/124, 131, 139; MPVII/175
D	N T C	org-1N+ombT+org-1CpUAST-clVIII-30 [MPVI/110]	10 lines	yes	MPVI/124,131
E	N T C	org-1N+ombT+CpUAST-clVIII-21 [MPVI/110]	6 lines	yes	MPVI/124
F	N T C	ombN+T+CpUAST-clVIII-1 [MPVI/109]	13 lines	yes	MPVI/139
G	N T C	ombN+T+org-1CpUAST-clVI/11 [MPVI/98]	19 lines	yes	MPVI/139
H	N T C	ombNTCpUAST-clX-10 [MPVI/122]	8 lines	yes	MPVI/129
I	N T C	ombN+org-1T+CpUAST-clVI-19 [MPVI/98]	8 lines	yes	MPVI/131,139
J	N T C	org-1N+T+CpUAST-clXI-2 [MPVI/127]	9 lines	yes	MPVI/131
K	N T C	ombN+org-1T+ombCpUAST-clXI-10 [MPVI/127]	17 lines	yes	MPVI/131,139

Table 18. Summary of the generation of transgenic flies.
Multiple references indicate corresponding numbers of trials to transform the given construct.

5.3 Determination of the expression strength of individual transgenic lines

The laborious cloning procedures and the generation of transgenic flies provided us the key instuments to conduct our main experiment, in which we express the different chimeric *omb-org-1* UAS-constructs with suitable Gal4 driver lines. Previous to that, however, we were concerned about differences in the expression strength of individual transgenic lines. When we, for instance, tested all 7 independent transformants of series A in parallel with the same Gal4 drivers, we obtained a whole spectrum of phenotypic severity, although all these lines contain the identical UAS-HA-*org-1*NTC-HA transgene (chapter 5.3.2). Conceivably, different genomic insertion sites of a given transgene determine whether an individual line behaves weakly, moderately or strongly with a given Gal4 driver. The strength of individual transgenic lines is thereby not necessarily correlated with the expressivity of their w^+ marker (eye colors of transformants differed from pale-orange to brick-red wild type) (Klemenz *et al.*, 1987, and data not shown; but see Grimm, 1997).

Since our aim was to compare various transgenic constructs for qualitative differences, excluding different transgene quantities due to position effects, we first developed a system in which we can identify individual lines with similar transgene expression levels.

5.3.1 Establishing a detection system for transgene expression

The following method to determine the relative expression strength of different UAS-transgenes has been developed in cooperation with Martin Roth, Würzburg. It comprises three parts: a heat shock (hs) induceable *hsp70*-Gal4 transgene, several UAS responder lines to be examined, and antibodies against *Drosophila* SAP47 (anti-SAP47, mab nc46/1, Reichmuth *et al.*, 1995) and against the transgene-encoded protein, in our cases anti-HA (mab 12CA5) or anti-MYC (mab 1-9E10.2) (Evan *et al.*, 1985; Wilson *et al.*, 1984).

The UAS-transgenic lines are crossed to *hsp70*-Gal4 flies and transheterozygous descendants are exposed to a single hs (45 min at 37°C) that induces the ubiquitous expression of Gal4 and, subsequently, the activation of the UAS-transgenes. At a certain time after hs, the flies are decapitated and head homogenates are made. Samples of the head extracts are then analyzed on Western blots by simultaneously incubating with anti-SAP47 and anti-HA (or anti-MYC) in order to detect the amount of induced transgene-encoded protein in relation to a reference protein, SAP47.

In initial experiments, we examined the kinetics of transgene induction in our system. 5 individual lines from two transgenic series were chosen for this analysis: lines A1, A2a, and A4a, all containing an UAS-HA-*org-1*NTC-HA transgene, and lines F1b and F9 each with an UAS-MYC-*omb*N+T+C-MYC transgene. Head extracts were made before and 0 to 24 h after hs and processed as described (Material and Methods). The resulting Western blots display the time curves of transgene expression for these 5 lines (Figure 42 A-E).

All lines studied show a strong induction of the transgenic proteins in response to the thermal treatment. The amount of synthesized protein reaches a maximum at about 4-10 h after hs and remains elevated for at least further 14 h. The kinetics of transgene activation, however, differs between the lines tested. Line A1 (Figure 42 A), for instance, shows already a high level of HA-tagged ORG-1 expression at the end of the hs that further increases to be maximal between 4-7 h after hs. It then declines to a moderate level and remains constant until the end of the experimental observation. The other lines investigated, however, are less responsive, having a latency of about 3 hours, before the onset of transgenic protein synthesis becomes detectable (Figure 42 B-E). Their transgene expression gradually inclines and stays high until at least 24 h after hs. Furthermore, the lines F1b and F9 (Figure 42 D and E) have a weak basal activity of the transgenes, as their protein products are already visible prior to the thermal shock.

Thus, the five individual transgenic lines significantly differ in the onset, responsiveness, kinetics, and basal activity of their *hsp70*-Gal4 induced transgene activation.

It is conceivable that these differences convey the individual strength of UAS responder lines.

5.3.2 Determination of the relative expression strength of individual transgenic lines

Based on the observation that transgene-encoded proteins were detectable within 3 h to the hs in our time-curve experiments described above, we decided to systematically analyze all UAS-transgenic
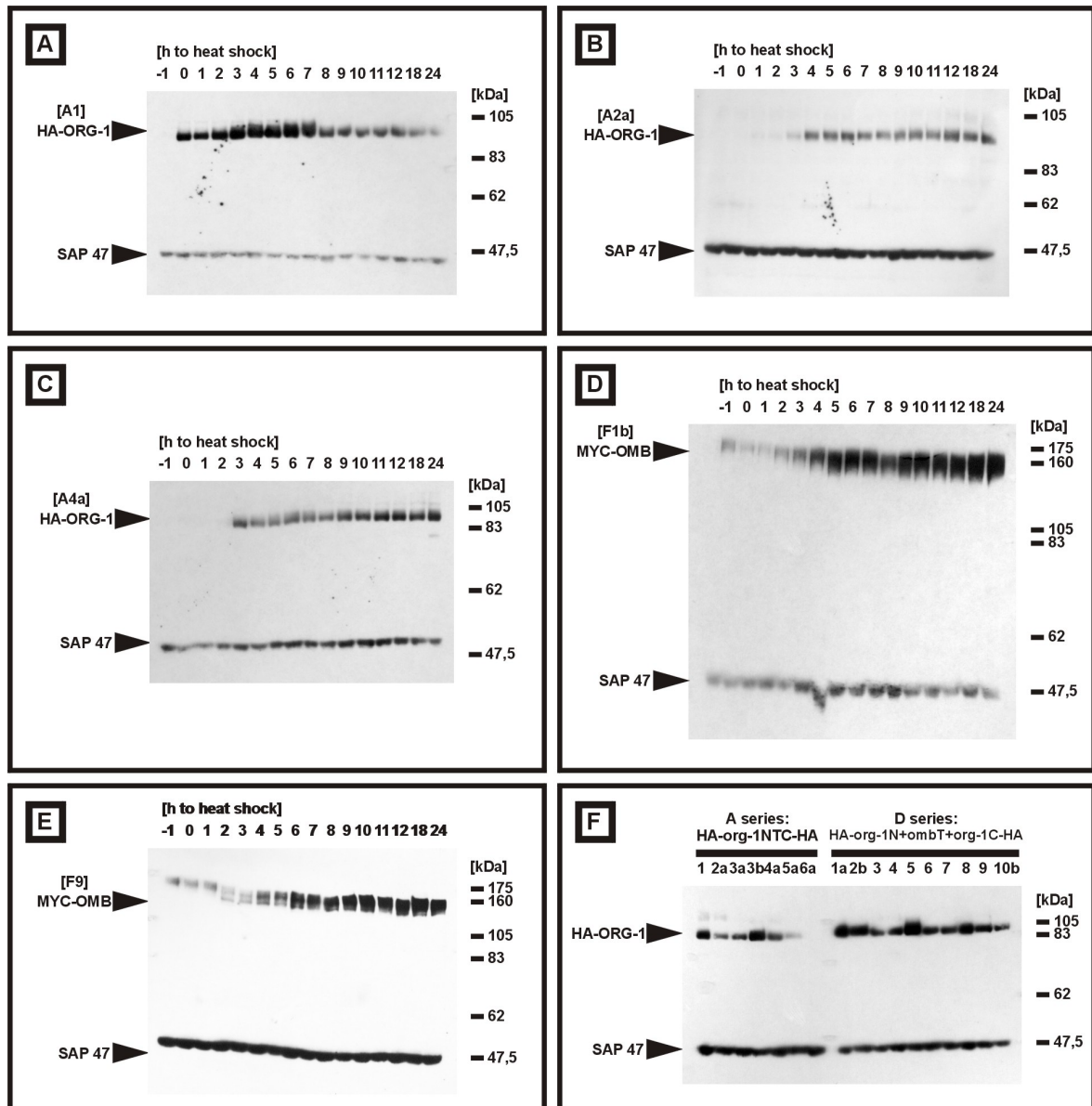
Figure 42. Determination of the expression strength of UAS-transgenic lines.
A-E. Western blots showing the transgene expression of individual UAS lines in response to a single hs. A. *hsp70*-Gal4/ UAS-HA-*org-1*NTC-HA [A1]. B. *hsp70*-Gal4/ UAS-HA-*org-1*NTC-HA [A2a]. C. *hsp70*-Gal4/ UAS-HA-*org-1*NTC-HA [A4a]. D. *hsp70*-Gal4/ UAS MYC-*omb*N+T+C-MYC [F1b]. E. *hsp70*-Gal4/ UAS MYC-*omb*N+T+C-MYC [F9]. F. Relative expression strength of individual UAS-transgenic flies of series A and D 4 h after hs.

lines for their transgene expression levels at 3 h and 4 h after the end of the hs.

Accordingly, head extracts of the transformation series were prepared and tested on Western blots in groups to initially determine relative differences among lines with an identical UAS-transgene. Figure 42 F shows a Western blot with samples from series A and D prepared 4 h after hs.

Like on the initial blots, two signal bands are present as well: The lower band derives from SAP47, an abundant synapse associated protein of 47 kDa (Reichmuth *et al.*, 1995), that appears with similar intensity throughout the lanes of a given blot (see

also Figure 42 A-E). Since SAP47 expression is not influenced by the heat shock treatment (see Figure 42 A-E), it may serve to control for a comparable sample load in our experiments. The upper signal at about 95 kDa, however, detects very different amounts of the transgene-encoded proteins HA-*org-1*NTC-HA (A series) and HA *org-1*N+*omb*T-*org-1*C-HA (D series) for individual lines (Figure 42 F). We next investigated the 7 transgenic lines of series A in parallel with three Gal4 drivers, *dpp*-Gal4-K54, 30A-Gal4 and E132-Gal4. Transheterozyous descendants of all lines manifested phenotypes consistent with those described previously

(see chapter 5.1 and Figures 30 and 31), but strikingly differed in expressivity and penetrance among each other. The phenotypic severity of individual lines thereby correlates with their expression strength seen in the Western blot experiment (Table 19, Figure 42 F). Line A3b gave the strongest phenotypes, followed by A1 (strong), A3a, A4a and A2a (moderate) and A5a (weak), whereas A6a represents a very weak responder line.

In the course of this work, the relative expression strength of individual lines was determined for the transformation series A, D, F, H, J, and K that all have transgenic constructs with two identical epitope tags (series A,D,J HA tags, series F,H,K MYC tags) (Table 20).

Future experiments will be required to complete this analysis and ought to include: (i) The examination of the relative expression strength within the remaining groups C, E, G, and I (all containing constructs with both an HA and a MYC tag), (ii) the determination of equally strong responder lines among series with the same epitope composition (*i.e.* by testing subsets of these series on identical blots), (iii) the determination of incubation conditions for anti-HA and anti-MYC with which both monoclonal antibodies give comparable signal intensities (the use of different tags for OMB and

ORG-1 is disadvantageous for this analysis and has historical reasons [these primers were also used to study the formation of OMB and ORG-1 heterodimers, where both proteins had to be made distinguishable]), and ultimately (iv) the determination of equally strong responder lines among all series (*i.e.* by the incubation a single Western blot containing lines from all series with simultaneously anti-SAP47, anti-HA, and anti-MYC).

5.4 Consequences of the ectopic expression of *omb-org-1* chimeric transgenes

Although our analysis for different responder lines with an about equal expression strength is still incomplete, we were curious to preliminarily study the effects of the various constructs *in vivo*.

Therefore, at least a subset of UAS lines from all transformation series was crossed in parallel to GMR-Gal4 flies. Their transheterozygous offspring was examined for an *org-1*- or *omb*-like eye phenotype (see Figure 32). The results of these experiments are summarized in Table 21.

Gal4-driver	A1	A2a	A3a	A3b	A4a	A5a	A6a
30A-Gal4	semi-lethal 100% held-out, *lac* wings	viable 100% held-out, *lac* wings	reduced viability 100% held-out, *lac* wings	semi-lethal 100% held-out, *lac* wings	reduced viability 100% held-out, *lac* wings	viable 100% held-out, *lac* wings	viable 15% held-out, *lac* wings
E132-Gal4	lethal	95% aristapedia	lethal	lethal	lethal	100% aristapedia	lethal
dpp-Gal4	reduced viability aristapedia outgrowth at scutellum	viable thickened arista scutellar bristle defect	reduced viability mild aristapedia scutellar bristle defect	semi-lethal aristapedia stunted legs scutellar defects	reduced viability mild aristapedia scutellar bristle defect	viable thickened arista	viable
hs-Gal4 expression strength	+++	++	++	+++++	++/+++	+	0

Table 19. Correlation of phenotypic severity with expression strength of individual HA-*org-1*NTC-HA (A series) transgenic lines.
Phenotypes of Gal4/ UAS HA-*org-1*NTC-HA transgenic lines are given. Percentage values indicate phenotypic penetrance. *lac*: lacquered wings. The relative expression strengths of individual lines are taken from Table 20 and were determined in Western blot experiments (see text for details).

A series: HA-org-1NTC-HA

	3h	4h
A1	++++	+++
A2a	++	++
A3a	++	++
A3b	++++	+++++
A4a	++	++/+++
A5a	0	+
A6a	0	0

D series: HA-org-1N+ombT+org-1C-HA

	3h	4h
D1a	+++++	++++/+++++
D2b	++++	++++/+++++
D3	++++	+++
D4	++++	++/+++
D5a	+++++	+++++
D6	+	+++/++++
D7	++	+++/++++
D8a	++	++++
D9b	+	++++
D10b	++++	++/+++

F series: MYC-ombN+T+C-MYC

	3h	4h
F1b	+++	+++
F3a	0	0
F5	+++	+++
F6b	+++	+
F7b	++/+++	+
F9	++	+
F10b	++	0
F11a	+++/++++	+++
F13a	++	+
F14a	† [++++]	† [+++]
F14b	n.d.	+++
F15a	+++/++++	n.d.
F15b	+++	+++

H series: MYC-ombNTC-MYC

	3h	4h
H1a	† [++]	† [++++]
H1d-1	† [++]	† [++++]
H2a	† [++]	† [+++]
H2d-1	† [++++]	† [+++]
H2e-2	++++	+++
	† [++++]	† [+++]
H3a	+	+
	† [+]	† [+/++]
H3d-1	+++	++++
	† [+/++]	† [+]
H3d-2	++++	++++
	† [+/++]	† [++]

J series: HA-org-1N+T+C-HA

	3h	4h
J1a	+++/++++	++
J1c	+++/++++	++
J2a	++++	++
J2b	+++/++++	+
J3a	+++++	+++/++++
J4b	+++	+++/++++
J5	+++	++++
J6	+++	+
J7	n.d.	n.d.

K series: MYC-ombN+org-1T+ombC-MYC

	3h	4h
K1	++++	++
K2b	+	++
K3a	0	+
K3b	+	+
K4a	+	0
K4b	n.d.	n.d.
K5	++++	++++
K6	+++	+++
K8b	n.d.	n.d.
K9b	+++	+++
K10a	n.d.	n.d.
K11b	+++	+++
K12b	+++	+++
K13a	+	+++
K13b	+++	++
K14b	+++	+++
K15a	n.d.	+++

Table 20. Relative expression strength of individual UAS-transgenic lines.
Head extracts of hsp70-Gal4/ UAS-transgenic flies were made 3h or 4h after hs and analyzed on Western blots as shown in Figure 42 F. The relative expression strength of individual lines was determined within transformation series and rated (0: no expression detectable - +++++: very strong expression; †: semi-lethal or lethal at 25°C, [ ]: relative expression strength at 18°C instead, n.d.: not determined. The strongest responder line of each series is shown in bold and underlined.

First, we observed that the effects of untagged ORG-1 (B series), HA-tagged continuous ORG-1 (A series), or HA-tagged assembled ORG-1 (which contains several artificially introduced amino acids at the T domain borders, J series) are comparable among each other. All these ORG-1 constructs led to overall intact, albeit roughened eyes (see Figure 32 B). Analogous to that, untagged OMB (GOP# 255, Grimm, 1997), MYC-tagged full-length (H series) or assembled OMB (F series) severely interfered with eye development, consistent with previous observations (Figure 32 C). Viable offspring manifested highly degenerated ommatidia, a diffuse or lost eye pigmentation, as well as a reduced eye size. In addition, line F14a and most H lines were pupal lethal in combination with GMR-Gal4.

Thus, these control experiments show that neither the introduced internal cloning sites nor the terminal epitope tags affect the functional specificity of ORG-1 or OMB.

Next, we tested the behavior of *omb-org-1* chimeric constructs in subsequent GMR-Gal4 experiments, in which induced phenotypes were assessed to be *omb* specific or *omb*-like, if the affected eyes showed degenerated ommatidia in conjunction with a diffuse or lost pigmentation.

We found that all chimeric *omb-org-1* constructs behaved *omb*-like, except for HA-*org-1*N+*omb*T+C-MYC which functioned *org-1*-like (Table 21).

These data demonstrate that all OMB domains are individually capable of changing an otherwise ORG-1 protein to function *omb*-like during eye development, suggesting that all three OMB domains carry independent specificity determinants.
The T domain constitutes thereby the most relevant part in OMB, since it is sufficient to confer HA-*org-1*N+*omb*T+*org-1*C-HA OMB specificity. The OMB N- and C domains are critical parts for conferring OMB specificity, too. They direct the chimeric proteins MYC-*omb*N+*org-1*T+C-HA and HA-*org-1*N+T+*omb*C-MYC to behave *omb*-like. Interest-

ingly, the specificity determinants of both domains function additively, as the effects of GMR-Gal4 driven MYC-*omb*N+*org-1*T+*omb*C-MYC are enhanced to become comparable to those of ectopic OMB itself.

Unexpectedly, however, the OMB T- and C domains along with the *ORG-1* N domain have an *org-1*-like specificity, although both OMB domains mediate *omb* specificity *per se*. Thus, the ORG-1 N domain appears to suppress the specificity deter-

minants within the OMB T- and C domains by an unknown mechanism.

Taken together, the major determinants of OMB's functional specificity reside within the OMB T domain, but additional specificity-relevant sequences within its N- and C-terminal domains exist. The distinct determinants function additively, with the caveat that OMB specificity determinants are suppressed in HA-*org-1*N+*omb*T+C-MYC for unknown reason.

series	construct	phenotype	analyzed line	omb/org-1 like?
B	N T C	rough eyes	**B8a**, B1b, B4a, B9	org-1
A	N T C	rough eyes (Figure 4.1-3 B)	**A3b**[+++++], A1[+++], A5a	org-1
J	N T C	slightly rough eyes	J3a [+++/++++], J1a, J1c	org-1
	N T C	semi-lethal, ommatidia degenerated, diffuse eye pigmentation, reduced eye size (Figure 4.1-3 C)	GOP # 255	**omb**
H	N T C	puppal lethal semi-lethal, ommatidia degenerated, diffuse eye pigmentation, reduced eye size	**H1a** † [++++], H series, H3d-2[++++]	**omb**
F	N T C	puppal lethal ommatidia degenerated, diffuse eye pigmentation, reduced eye size	**F14a** † [+++], F1b[+++], F9	**omb**
I	N T C	some ommatidia degenerated, rough eyes, diffuse eye pigmentation, eye size wild type	I4, I1d	**(omb)**
D	N T C	ommatidia degenerated, diffuse eye pigmentation, reduced eye size	**D5a**[+++++], D series	**omb**
C	N T C	ommatidia degenerated, diffuse eye pigmentation, eye size wild type	C3a, C1b, C2a, C2b,C3c, C4b	**(omb)**
G	N T C	ommatidia degenerated, diffuse eye pigmentation, reduced eye size	G13a, G1b, G15a	**omb**
K	N T C	semi-lethal, ommatidia degenerated, diffuse eye pigmentation, reduced eye size	K1 [++], K4b [n.d.], K13b [++]	**omb**
E	N T C	rough eyes, reduced eye size	**E2e**, E series	(org-1)

Table 21. Consequences of the ectopic expression of *omb-org-1* chimeric transgenes.

6. Discussion

6.1 *C31*, the initial *org-1* mutant candidate, is probably caused by mutations in a distal locus

A major goal of this work was to investigate the role of *org-1* in *Drosophila* development. Fundamental to this aim is the isolation and analysis of *org-1* mutant flies, in which the consequences of an impaired or absent *org-1* function are displayed in the mutant phenotypes. A starting point for our search for *org-1* mutants was provided by *C31*, a recessive, pleiotropic *Drosophila* mutant that was deficiency-mapped to the *org-1* cytological region at 7E-7F on the X chromosome (Strauss, 1995). Molecular analysis revealed that *C31* carries an insertion of a 5' truncated retrotransposable I element within the 3' untranslated region of the *org-1* transcript (Porsch, 1997; chapter 3.2). The I element insertion was absent in five investigated wild type strains (Porsch, 1997) strongly suggesting that it might be responsible for the *C31* syndrome and, thus, that *C31* would represent the first mutant *org-1* allele. Based on this assumption, we ran a large-scale EMS mutagenesis in which we screened about 44.500 mutagenized individuals for new *C31* alleles using the visible *C31* "held-out" wing phenotype and/ or head bristle pattern defects that are found in deficiency-transheterozygotic *Df(1)RA2/ C31* flies. We, however, failed to isolate any *C31* or *org-1* allele in this experiment, although the mutagenesis was functional *per se* and one roughly estimates to obtain 1 hit in 2000 to 5000 screened individuals for most loci (Kevin Moses, pers. comm.). Why, then, did our mutagenesis remain unsuccessful? In principle, two possible explanations for our failure exist: (i) we either have not induced new *C31* alleles at all, or (ii) we induced new *C31* mutations, but could not establish fly stocks from those.

A comparison of frequencies with which mutant stocks were obtained from *C31* candidates with either held-out wings or head bristle phenotypes indicates that flies with held-out wings are significantly less viable and/ or fertile. Only 50,9% of the isolated held-out mutants produced offspring (55/112 = 49,1% perished), while 78,3% of the head bristle mutants propagated (5/23 = 21,7% died without offspring) (Table 22). We observed that flies with held-out wings frequently stick in the food medium and perish. Reduced fitness of flies

with held-out wings has also been observed in a subsequent genetic experiment in which we intended to detect precise deletions spanning the *org-1* locus as *C31* alleles. In this P element based approach, only 38,2% of *C31* mutant candidates with held-out wings propagated (34/89 = 38,2%; chapter 4.3.3), while in a comparable experiment in which deletion candidates were screened for the absence of the flanking P element *miniwhite* markers, 82,7% of the isolated w^- flies gave offspring (43/52 = 82,7%; chapter 4.3.5). It is conceivable that the selective disadvantage of the held-out wing phenotype contributed to the failure of our EMS mutagenesis.

Interestingly, however, the second *C31* mutagenesis was successful and led to the isolation of 8 new *C31* alleles, demonstrating that a screen for the conspicuous wing phenotype was suitable to isolate *C31* alleles. Since the only (recognizable) difference in both *C31* mutagenesis was the nature of the mutagen, it is suggestive that distinct characteristics of chemical versus transposon mutagenesis were responsible for the different outcome in both experiments.

EMS is a potent alkylating agent that primarily causes G:C to A:T transitions and, less frequently, chromosomal aberrations at random positions in the genome. Its relatively easy handling, its efficiency and a relatively low toxicity to flies make EMS the most commonly used chemical mutagen in *Drosophila* genetics (Grigliatti, 1986; Ashburner, 1989). However, EMS has a considerable drawback, too, as a large proportion of EMS-induced mutants are mosaics. In EMS screens for visible phenotypes, this frequently leads to the isolation of mutant individuals that do not transmit the mutation to the F2, because only the affected tissue is mutant in such F1 flies, but not their germline (Grigliatti, 1986; Ashburner, 1989; and references therein). A case study by Jenkins (1967) demonstrated that only about one-third of EMS-induced *dumpy* (*dp*) F1 mutants transmitted the *dp* mutation to the F2. Thus, it is conceivable that our EMS mutagenesis induced *C31* mutations, but we failed to establish those because a large proportion of new mutations was not transmissible.

In the second *C31* mutagenesis, we remobilized two *org-1* flanking P elements and had expected to identify precisely generated deletions in 7E-7F as new *C31* alleles. Indeed, 8 new *C31* alleles could be isolated. Surprisingly, however, the molecular characterization of the new *C31* lines revealed that these mutant chromosomes do not contain the expected 95 kb deletion, nor large internal deletions or P element insertions within the designated dele-

tion interval including the *org-1* locus. This data strongly suggested that the *C31* syndrome is not caused by mutations in *org-1*. Subsequently generated deletion mutants lacking at least a major part of the *org-1* ORF failed to uncover *C31* and, thus, unambiguously demonstrated that *org-1* is not associated with *C31*.

The molecular analysis of the new *C31* alleles included a series of PCR experiments in which we tested for the presence of the starter P element ends and/ or neighboring genomic sequences. PCR products for the P element ends proximal to the desired deficiency and for the 3' end of the proximal P element G0099 were obtained for all 8 new *C31* alleles. However, the 5' end of the distal P element G0071 could not be amplified from any of these lines and, thus, correlated with the *C31* phenotype (chapter 4.3.4). Since this PCR product can only be amplified, if the primer sites within the 5' end of P G0071 and the normal upstream genomic sequence are both present, we surmised that mutations distal to P element G0071 might be responsible for *C31* and, therefore, amplified P element neighboring genomic sequences by 5' inverse PCR. Consistent with the absence of the 5' PCR product for P G0071 in the new *C31* alleles, the original 5' inverse PCR product was not obtained for P element G0071, but several other P element flanking genomic sequences instead that map further distally to the P G0071 site. 5' P element flanking genomic sequences in the lines 49-5, 50-3, and 82-5 all mapped within 200 bp to AE003443 at kb 253,9-254,1, about 53 kb distal to the P G0071 insertion site. These P element flanking sequences align to the 5' end of a transcription unit that is represented by the EST clone GH26370.5'. Since the three genomic sequences

all derive from P elements with the same orientation as P G0071, they either may have been amplified from locally transposed P elements or may result from deletions distal to P G0071. It is currently not yet clear, which of both hypothesis holds true.

The characteristic feature of P elements to frequently transpose into the 5' region of genes argues in favor of local transpositions. However, these new P element insertions do not explain, why we lacked the 5' PCR product for P G0071 in the new *C31* lines. Moreover, if one postulates that the P transpositions into the GH26370.5' transcription unit would be responsible for the *C31* phenotype, one would have to assume the tight correlation (8/8 cases) of the *C31* syndrome with the absence of the 5' PCR product for P G0071 to be solely coincidential.

A deletion distal to P G0071, on the other hand, could explain why we did not obtain the 5' PCR amplificate for P element G0071. It appears to be puzzling, however, why the three putative deletion endpoints coincide within 200 bp at the 5' region of GH26370.5'.

A relatively straightforward experiment should clarify this issue. Accordingly, one would cross a P element line with an insertion between AE003443 kb 253,9-254,1 and the P G0071 site, *e.g.* line 31-2756/1 (chapter 4.2.1) to the ΔP *C31* lines and collect the transheterozygotic P{lacW} 31-2756/1/ ΔP *C31* daughters. Genomic DNA of these flies would be prepared, restriction-digested and blotted. The resulting Southern blot is subsequently hybridized to a probe that recognizes a genomic fragment which includes the P element insertion site, and the resulting signal bands are then analyzed for the presence or absence of the wild type fragment.

mutagenesis	total candidates	candidates with				
		wing defect	heldout wings *	head bristle defect	wing AND head bristle defect	head bristle defect only §
round 1 I/97	25/69	25/69	24/53	0/8	0/8	0
round 2 I/98	5/16	4/12	3/4	4/14	3/10	1/4
round 3 II/98	6/10	6/10	5/7	5/8	5/8	0
round 4 III/98	21/49	17/35	11/26	10/22	6/8	4/14
round 5 IV/98	18/63	18/57	12/22	8/28	8/22	0/5
rounds 1-5	75/207 = 0.36	70/183 = 0.38	55/112 = 0.49	27/80 = 0.34	22/56 = 0.39	5/23 = 0.22

Table 22. Loss of induced mutant candidates.
The number of mutants that could not be established as fly stocks is given per number of isolated candidates. * classified as flies with both wings in held out posture. § classified as head bristle phene without any concomitant wing defect, however, eventually with additional defects.

If one obtains a wild type signal, the corresponding ΔP *C31* line can not contain the about 53 kb distal deletion. Conversely, if the wild type genomic fragment would not be recognized, the ΔP *C31* line carries a distal deletion and the P element neighboring sequences in AE003443 kb 253,9-254,1 presumably indicate the distal endpoints of such deletions in lines 49-5, 50-3, and 82-5.

The currently most likely *C31* candidate gene is GH26370.5', because this transcription unit is affected regardless, if a P element transposition or a deletion has occured.
In the case of a distal deletion, however, additional four predicted genes (*CG1387*, *CG10555*, *CG15345* and *CG11190*) are *C31* candidate loci, too. Two previously unpredicted transcripts represented by RH09582.5' and GM09770.5' which also lie within the eventual 53 kb distal deletion interval have already been associated with lethal P element insertions that did not uncover *C31* and, thus, can be excluded.

Taken together, our data unambiguously demonstrate that *C31* is not an *org-1* mutant and suggest that *C31* is probably caused by a mutated distal locus, possibly GH26370.5'.

We, finally, addressed the question for the origin of the I element insertion in *C31*. I elements are non LTR (long terminal repeats) retrotransposable elements underlying the inducer-reactive (I-R) hybrid dysgenesis, a genetic system in *Drosophila* characterized by the frequent occurence of sterile or mutant offspring among the progeny of appropriately crossed inducer (I) and reactive (R) strains (Fawcett *et al.*, 1986). I strains are thereby determined by the presence of functional I elements (Kidwell, 1983). Transposition of I factors takes place at high frequency in the ovaries of female offspring of crosses between I males and females of R strains (Sezutsu *et al.*, 1995). Since *C31* was isolated from a mutagenesis in a wild type Berlin background, it seemed puzzling to us, how the EMS treatment should have induced an I element transposition into *org-1*.

When we molecularly analyzed the newly generated ΔP *C31* alleles in a RFLP analysis for deletions or P element insertions within the *org-1* locus, several wild type strains were used as controls including wild type Canton S and wild type Berlin stocks from Roland Strauss and Prof. Heisenberg. We confirmed for wild type Canton S and Berlin [Roland Strauss] our previous observation that both lines do not contain the *C31* I insertion (hybridization data not shown; Porsch, 1997). Most surpris-

ingly, however, we obtained the *C31* signal in addition to the wild type signal for wild type Berlin [Heisenberg] implying that this stock is heterogeneous for the *C31* I insertion within *org-1*.

The wild type Berlin stock [Roland Strauss] originally derived from wild type Berlin [Heisenberg] and possibly flies from both sources were used for the EMS mutagenesis from which *C31* was isolated (Roland Strauss, pers. comm.).
The current wild type Berlin stock [Roland Strauss] was established as attX stock from a small number of wild type Berlin [Heisenberg] males. It is conceivable that a founder effect led to the elimination of the I element containing *org-1* allele within the wild type Berlin attX stock [Roland Strauss] making this stock homogeneous for the I factor-free *org-1* allele.
Unfortunately, the wild type Berlin stock [Heisenberg] was not included in our initial RFLP analysis (Porsch, 1997), so that we erroneously concluded that the I insertion within *org-1* is a polymorphism specific to *C31*.

Our recent observations strongly suggest now that the *C31* I factor within *org-1* derives from the wild type Berlin [Heisenberg] stock which itself is heterogeneous for this I element insertion.

6.2 Reverse genetic approaches to mutate *org-1*

After we had excluded that *C31* is an *org-1* mutant as initially anticipated, we applied a reverse genetic strategy to isolate *org-1* mutants, since we had no reliable prediction for an *org-1* mutant phenotype.
At that point of time, *Drosophila* genetics still lacked a tool for the targeted disruption of cloned genes by homologous recombination, while in other eukaryotic model organisms, such as yeast or mouse, a high recombination frequency or the usage of embryonic stem cells, respectively, made it feasible to routinely knock-out genes of interest by gene replacement approaches (Thomas and Capecchi, 1987; Rothstein, 1991; Engels, 2000). A targeted gene knock-out in *Drosophila* by homologous recombination has only been described recently (Rong and Golic, 2000; Rong and Golic, 2001).

Previous reverse genetic approaches in the fly aimed to identify P transposable element insertions within the gene locus under investigation. According to that, we tried to associate a P element insertion with *org-1*. Therefore, 540 viable X-chromosomal P insertion lines were screened for

an insertion site within genomic clones containing *org-1*. Two positive lines were found to carry their P elements 36 kb and 38 kb downstream of the *org-1* transcription unit (chapter 4.2.1). We subsequently analyzed all available *Drosophila* P insertion lines cytologically mapped to 7E-7F for their precise insertion sites. A total of 19 lethal lines were characterized. These lines were generated in a massive gene disruption project of the BDGP that extensively targets *Drosophila* genes with transposon insertions (Spradling *et al.,* 1995; Peter *et al.,* 2002). Since both genetically investigated *Drosophila* T-boxes genes, *omb* and *byn*, encode essential functions, and since many T-box mutants manifest profound embryonic phenotypes, it seems plausible to consider lethality for *org-1* null alleles, too. Unfortunately, none of the 19 P lines had an insertion within *org-1*. Instead, 13 of the 19 lines carry an insertion within a 2 kb large P element hotspot about 37 kb distal to *org-1*. Thus, although large-scale attempts of the BDGP such as the EST project (Rubin *et al.,* 2000) or the gene disruption project (Spradling *et al.,* 1995; Peter *et al.,* 2002) greatly facilitate the cloning and functional analysis of many *Drosophila* genes, the study of some transcripts still requires individual efforts. *org-1* certainly belongs to the latter category: no EST clone or P element has hitherto been found for *org-1*. Conversely, a number of EST clones and a semilethal EP insertion line could simply be identified for the *Drosophila vmd2* gene by searching databases (chapter 3.5).

Further *org-1* genetic experiments employed two *org-1* flanking P{lacW} insertions, l(1) G0071 and l(1) G0099, being inserted 27 kb downstream and 62 kb upstream of *org-1*, respectively. As P elements have an intrinsic tendency to preferentially transpose to nearby sites, we performed local hop experiments using l(1) G0071 and l(1) G0099 to target P insertions to *org-1*.

Prior to this project, we took into consideration that P elements frequently jump into the 5' end of genes by mapping the putative 5' region of the *org-1* transcription unit with a 5' RACE experiment.

The discrepancy of the putative *org-1* transcript size on Northern blots of about 3800 nt with the length of the 3' complete cDNA corg-1M2 of 3168 bp suggested that the full-length *org-1* transcript might extend up to 400 bp further at the 5' end of corg-1M2. 5' RACE products extended the known *org-1* exon 1 by 49 bp suggesting that the complete *org-1* transcript is only moderately longer than corg-1M2 and that the *org-1* transcription start site appears to be in vicinity of the present exon 1, although a search of upstream genomic sequences

for conserved promoter elements remained unsuccessful (chapter 3.1.2).

The P{lacW} elements l(1) G0071 and l(1) G0099 were only remobilized in the female germline due to the lethality of both starter transposons. We made use of the X-chromosomal P element lethality to screen with the "reversion-jumping" strategy for new, stable transposition lines (Tower *et al.,* 1993). All F1 males containing a *miniwhite* P element marker are revertants of the lethal transposon and carry a jumped P element and were collected. The "reversion-jumping" search strategy thereby efficiently filtered off the flies in which the starter P element remained or did not transpose. This screening method, however, was not without a drawback. The strict selection for viable insertions excludes the isolation of P element mutants that disrupt *org-1*, if *org-1* encodes an essential function. Nonetheless, we followed this strategy and hoped to associate a P{lacW} element with *org-1*, since even P insertions outside, but in the vicinity of *org-1* would be valuable for a subsequent generation of deletions by imprecise P element excision or for the targeting of *org-1* by gene replacement (Gloor *et al.,* 1991) or a second local hop experiment. Moreover, neighboring P{lacW} insertions may serve as *org-1* enhancer traps in which the expression pattern of *lacZ* reflects the expression of *org-1*.

357 new, stable X-chromosomal P element lines were established from a total number of 1066 transposition lines (1066-357 = 709 interchromosomal jumps) that were selected from about 73.750 screened males. 166 transposition lines derived from the *org-1* upstream P element G0099 (166/43.750 screened males = 0,38%), 900 transpositions were obtained from the downstream G0071 P element (900/30.000 = 3%) (Table 9). Hence, the frequency with which transpositions were obtained was almost 8-fold higher for G0071 than for G0099 (3/0,38 = 7,89).
343 of the 357 intrachromosomal transpositions were analyzed for P element insertions within large genomic clones containing the *org-1* locus. 13 lines with potential P insertions within *org-1* or nearby sequences could be identified and were precisely located. 6 new genes could be associated with P elements, however, unfortunately not *org-1*. The *org-1* closest transposons (lines 204, 274, and 543) inserted between AE003444 kb 21,4 – 21,7. We, therefore, approached *org-1* in our local hop mutagenesis by 16 kb, but still are 10 kb away from the *org-1* transcription unit (chapter 4.2.3.2).

All precisely determined P element integration sites lie around the 5' end of genes, exept line 213 that has a P insertion in the first intron of the predicted gene CG10777. A strong preference of P element insertions for the 5' region of genes has been reported (Spradling *et al.,* 1995).

Interestingly, 14 characterized P elements integrated in the same 5' to 3' orientation (5' P element end points distal), while only line 213 carries the P element in the opposite orientation. Moreover, all 19 investigated lethal P{lacW} elements in 7E-7F were also inserted in the 5' to 3' orientation and only two viable P{lacW} and an EP transposons are integrated in the unfavored direction. It is tempting to speculate that the nearby hotspot has an influence on the orientation with which P elements integrate. Conceivably, the hotspot magnetizes P elements to preferentially integrate at a given locus in a favored orientation.

The distribution of the insertion sites within the new P element lines confirmed previous observations that P element preferentially transpose locally (Tower *et al.,* 1993; Zhang and Spradling, 1993). 357 of the 1066 selected transpositions were intrachromosomal (357/1066 = 33,5%), 709 transpositions to autosomal sites were obtained (709/1066 = 66,5%). As the X chromosome makes up about one-fifth of the *Drosophila* genome, a significantly larger portion of P transpositions remained on the starter chromosome.

Furthermore, 105 of the 249 local hops that derived from G0071 were mapped to the interval AE003443 kb 295 – AE003444 kb 12 (105/249 = 42,2%). Thus, 42,2% of the intrachromosomal transpositions from G0071 inserted around the starter P element site at AE003444 kb 5,4 (Table 10), while only 2% and 0,4% of the transpositions were found within AE003444 kb 12 – 49 and kb 49 – 168, respectively. 47,4% of the G0071 derivatives inserted elsewhere on the X chromosome (118/249 = 47,4%), 8,4% could not be determined (21/249 = 8,4%).

Of the proximal G0099 starter element derived X-chromosomal transpositions, 5,3% jumped within the interval AE003444 kb 49 – 168 that includes the G0099 insertion site at AE003444 kb 101,6 (5/94 = 5,3%). The frequencies of transpositions into the regions AE003444 kb 12 – 49 and AE003443 kb 295 – AE003444 kb 12 were 2,1% (2/94 = 2,1%) and 4,3% (4/94 = 4,3%), respectively. 71,3% of the G0099 derived P elements lie elsewhere on the X (67/94 = 71,3%), 16% could not be localized (15/94 = 16%).

Mapping of the local transpositions revealed that a large number of P elements hopped into the interval AE003443 kb 295 – AE003444 kb 12 (Figure 18). Most of the lethal P elements in 7E-7F were also found to be concentrated to this region. It is well known that P elements do not randomly insert within the genome, but prefer some loci to others. For example, the *singed* gene is a favored site for P integration that occur at frequencies of about 10^{-2}, whereas the *vestigial* locus is hit with a rate of less than 10^{-6} (Engels, 1996 and references therein). Although there is no evidence that any loci are completely protected from P insertions, some genes are elusive of P insertions due to practical limitations for the required sample size (Engels, 1996). The *Alcohol dehydrogenase* (*Adh*) locus, for instance, proved to be highly resistant to P element mutagenesis, as no insertions in *Adh* have been recovered despite extensive screening (Kidwell, 1987).

The failure to isolate P transposons within the *org-1* gene indicates that *org-1* is a locus not easily accessible to P element insertions.

6.3 Generation of *org-1* deficiencies

We then followed a P element-based method in order to generate precise deletions spanning the *org-1* gene. Cooley *et al.* (1990) reported that remobilization of two P elements in *cis* configuration frequently induces deletions with the P insertion sites as deficiency breakpoints. Accordingly, we recombined the *org-1* flanking elements P{lacW} l(1) G0071 and P{lacW} l(1) G0099, mobilized both P elements and expected to obtain about 95 kb large deletions lacking the genomic sequence between both P elements. This 95 kb deficiency would comprise 16-20 genes including *org-1*.

We assessed this deletion to be a highly valuable tool for our work for several reasons: (i) a number of chemically induced mutants are known for the interval 7E-7F (*e.g.* Lefevre and Watkins, 1986). In contrast to P elements, chemical mutagens frequently cause much more subtle molecular damage, *e.g.* point mutations, that are difficult to detect or clone. The designated deletion, however, would allow us to identify *org-1* mutant candidates among established mutant stocks simply by complementation analysis. (ii) if no *org-1* allele would be found among existing mutants, the deletion would provide the opportunity to launch an exhaustive mutagenesis over this deficiency. (iii) Mutations in *TBX1*, the mammalian homolog of *org-1*, have been shown to be mainly responsible for DiGeorge Syndrome

(Jerome and Papaioannou, 2001; Lindsay *et al.,* 2001; Merscher *et al.,* 2001). DiGeorge patients frequently have large deletions that include the *TBX1* locus (Scambler, 2000). Hence, our deletion would provide a similar situation in the fly. (iv) the precise deficiency would help to clone *C31* and, in particular, should clarify the status of *org-1* for *C31*.

Mutants existed for three genes of the expected deletion interval: *otu* and *Cp36* are female sterile, whereas *Nrg* null alleles are embryonic lethal. Since their phenotypes were inappropriate to screen for the deficiency and since Cooley *et al.* (1990) observed in their case study that 77,8% of the isolated deletion mutants retained at least one of the flanking P element markers, we initially did not screen for the absence of the P element markers but used *C31* instead with which we expected to detect the desired deletion as new *C31* chromosomes by scoring mutagenized individuals for held-out wings.

8 new *C31* alleles were isolated in this screen. However, molecular analysis revealed that they do not contain the desired deletion. We failed to isolate the 95 kb deletion, because *C31* is not uncovered by this deficiency and apparently lies distal to it.

The P element mobilization was repeated and we screened for the loss of the flanking P element *miniwhite* marker genes in spite of their frequent retention on generated deletion chromosomes (Cooley *et al.,* 1990). Four deletion chromosomes were obtained. RFLP analysis showed that all four deletions lack at least an internal 4,7 kb large *BamH* I fragment containing four coding *org-1* exons. ΔP lines 23, 24 and 31 appear to carry the designated 95 kb deletion, while line 39 must contain a more restricted internal deletion, since PCR products were obtained for the deletion-proximal P element ends and neighboring genomic DNA (Table 16).

6.4 Further proceedings in *org-1* functional analysis

Although the four ΔP lines 23, 24, 31, and 39 were shown to represent *org-1* deletion mutants and can, thus, directly be used for a complementation analysis to identify *org-1* candidate genes, an accurate determination of the deletion extents is required for a reliable interpretation of any genetic data obtained with these lines.

All four lines give rise to the 3' PCR product for the P element G0099 which, therefore, defines the proximal limit of the deletions. Likewise, the 5' PCR product for P G0071 was obtained for ΔP lines 23, 24, and 39, restricting the deletions in these lines proximal to P G0071. ΔP 31 lacked this PCR product and, thus, may have a deletion extended further distally. To characterize the deletion interval more precisely, the amplification of the residual genomic sequences was attempted using long range PCR with pairs of primers annealing distally to G0071 and proximally to G0099 but remained unsuccessful (Angela Bahlo and Gert Pflugfelder, pers. comm.).

An alternative strategy would take usage of viable P element insertions within the designated interval for a RFLP analysis. In particular, P insertions close to the designated deletion breakpoints (*i.e.* the G0071 and G0099 insertion sites) should be included which were isolated in the *org-1* local hop experiment (Figure 20). Accordingly, one crosses *e.g.* local hop lines 464 (2 kb proximal to G0071), 543 (16 kb proximal to G0071), and 138 (0,5 kb distal to G0099) to the ΔP lines and selects the ΔP/ P{lacW} transheterozygotes among the offspring. Genomic DNA is isolated, digested, and blotted. The resulting Southern blots are then hybridized to probes that recognize restriction fragments that include the P element insertion sites, and are subsequently analyzed for the presence (indicating no deletion) or absence (indicating a deletion) of the wild type signal.

The ΔP lines may also be further characterized genetically by testing for the complementation of the female sterility of *Cp36* and *otu* mutants.

The next step in *org-1* genetics should be a complementation assay of the ΔP lines with the numerous chemically induced mutants in 7E-7F (*e.g.* Lefevre and Watkins, 1986). *org-1* mutant candidates may then be molecularly analyzed for mutations in *org-1*.

The functional analysis of *org-1* also demands a study of the *org-1* expression pattern during *Drosophila* development. Since all T-box mutants described so far revealed phenotypes in body areas in which the given T-box genes are expressed during development, the *org-1* expression pattern might allow to draw conclusions for the phenotypic spectrum of *org-1* mutants. Moreover, knowledge on the *org-1* expression pattern will certainly help to identify downstream target genes regulated by *org-1* and might provide clues to the factors which control the expression of *org-1* itself.

org-1 expression studies have not been carried out so far. However, a His-ORG-1$_{17-708}$ fusion protein

was expressed in *E. coli*, purified, and subsequently used to raise ORG-1 antisera in mice and rabbit (chapter 3.3). ORG-1 antisera were immunoreactive on Western blots with recombinant ORG-1 and *Drosophila* protein extracts but have hitherto not been used for immunohistochemistry to determine the expression pattern of ORG-1 in *Drosophila*. Likewise, thorough RNA *in situ* hybridization experiments have not been done yet.

6.5 *org-1* gain-of-function phenotypes

We investigated the consequences of *org-1* in gain-of-function situations using the Gal4/ UAS system with five different Gal4 lines, all directing ectopic *org-1* expression during imaginal development. Among the Gal4 lines tested, we found the most conspicuous phenotypes in *dpp*-Gal4-K54/ UAS-*org-1* flies. These animals displayed severe malformations in the body trunk and the appendages including a split notum, a tumorous outgrowth in place of the scutellum, ectopic pigmentation on the ventral abdomen, homeotic antenna-to-leg transformations, stumpy legs, and vestigial wings (Figures 30 and 31).

Thus, ectopic activation of *org-1* during imaginal disc development revealed for *org-1* the capacity to predominate or interfere in various tissues that give rise to distinct body parts of the adult fly. This ability has not been completely unexpected for *org-1*, since ectopic expression of other T-box proteins was previously shown to cause profound disturbances in normal development, too (*e.g.* Cunliffe and Smith, 1994; O'Reilly *et al.,* 1995; Grimm and Pflugfelder, 1996).

How ectopic *org-1* induces these phenotypes is currently not known yet and will certainly require the identification and study of *org-1* downstream target genes that mediate the deleterious effects of ectopic ORG-1. Analysis of identified T-box target genes revealed that a large proportion of those encode key developmental regulators such as signaling molecules or transcription factors (Table 1). Therefore, it seems imaginable that *org-1* downstream targets similarly include important developmental control genes, too, that when misregulated in ectopic *org-1* situations, conflict with proper developmental programmes.

As judged from the displayed phenotypes, possible candidates for *org-1* downstream effectors might include the genes *Distal-less* (*Dll*) and/ or *spineless* (*ss*) (Gorfinkiel *et al.,* 1997; Duncan *et al.,* 1998). *Dll* is a selector gene required for the identity and

growth of all ventral appendages in *Drosophila* (Gorfinkiel *et al.*, 1997). The expression of *Dll* in the central region of the leg or antennal imaginal discs is activated by the juxtaposition of *wingless* (*wg*) and *decapentaplegic* (*dpp*) expressing cells (Días-Benjumea *et al.*, 1994; Campbell and Tomlinson, 1995). It has been proposed that the expression of *Dll* is required for the formation of the proximodistal (P/D) axis of the limb (Días-Benjumea *et al.*, 1994; Campbell and Tomlinson, 1995). Lack of *Dll* function during limb development frequently results in the loss of distal appendage segments, whereas overexpression of *Dll* within its endogenous domain in the leg disc induces a duplication of the P/D axis that results in leg duplications (Gorfinkiel *et al.*, 1997). We observed that the most proximal segments of the antennae and legs of *dpp*-Gal4-K54/ UAS-*org-1* flies are normal but that more distal segments of both the appendages are increasingly stronger affected. Therefore, it seems possible that ectopic *org-1*, directly or indirectly, downregulates *Dll* during appendage development and thereby influences P/D axis formation or maintenance which, in consequence, leads to defects in the distal part of appendages.

In *Drosophila*, antennae, mouthparts, legs, external genitalia, and analia are regarded as a series of homologous, ventral appendages (Casares and Mann, 2001; Mann and Morata, 2000 and references therein). This work only described the phenotypes of *dpp*-Gal4-K54/ UAS-*org-1* flies in antennae and legs, but not in other ventral appendages. If, however, the defects of the ectopic activation of *org-1* in antennae and legs are due to a misregulation of *Dll*, it appears to be likely that other ventral appendages in *dpp*-Gal4-K54/ UAS-*org-1* are affected as well, since *Dll* is required in those, too. A thorough survey of *dpp*-Gal4-K54/ UAS-*org-1* flies indeed revealed malformations of mouthparts and genitalia (Gert Pflugfelder, pers. comm.) adding support for the idea that *Dll* expression is influenced by ectopic *org-1*.

Dll codes for a homeodomain transcription factor and *Dll* is required for the expression of *ss* in the distal portion of the antennal imaginal disc and the tarsal region of each leg disc (Duncan *et al.*, 1998). Interestingly, loss-of-function mutations in *ss* cause a deletion of the tarsal segmentation (segments 2-4 and a part of segment 1 are deleted) and manifest a homeotic transformation of the distal antenna into leg, known as *aristapedia* (Duncan *et al.*, 1998). Thus, *spineless-aristapedia* mutants resemble the antennal and tarsal phenotypes of *dpp*-Gal4-K54/ UAS-*org-1* flies suggesting that a repression of *ss*

expression, possibly via a downregulation of *Dll*, is responsible for those.

The putative repression of *Dll* and *ss* by ectopic *org-1* can be experimentally tested by comparing the expression of *Dll* and *ss* in leg and antennal imaginal discs of *dpp*-Gal4-K54/ UAS-*org-1* flies with their wild type expression patterns.

Again, knowlegde of the endogenous expression pattern of *org-1* will facilitate the identification of *org-1* target genes and will allow a better interpretation of the observed *org-1* gain-of-function phenotypes.

6.6 Mapping of functional specificity determinants in OMB and ORG-1

As described above, ectopic expression of *org-1* during imaginal disc development gives rise to adult flies with a plethora of morphological abnormalities. When we compared the consequences of ectopic *org-1* with those of *omb* in similar gain-of-function situations, we found that both genes led to marked but qualitatively different developmental defects raising the question for how functional specificity is achieved in the homologous OMB and ORG-1 proteins. The issue of T-box specificity is further illustrated by the antagonistic functions of Tbx4 and Tbx5, two of the most closely related T-box factors, in vertebrate limb development. *Tbx5* is expressed in the forelimb bud and controls the differentiation into wings, whereas *Tbx4* mRNA is predominantly found in the hindlimb bud selecting leg identity (Rodriguez-Esteban *et al.,* 1999; Takeuchi *et al.,* 1999; Logan and Tabin, 1999). In this study, we addressed the question of functional specificity of T-box proteins by mapping specificity determinants within the OMB and ORG-1 proteins.

We, therefore, conceptionally subdivided both proteins into the central T-box DNA binding domain, a N-terminal domain, and a C-terminal domain, and investigated the relevance of these domains for functional specificity *in vivo* by ectopically expressing chimeric *omb-org-1* transgenes.
UAS-transgenes containing OMB and ORG-1 domains in all possible combinations were expressed using GMR-Gal4 and were investigated for their effects on eye development. We previously found that GMR-Gal4 driven *omb* expression leads to a degeneration of the photoreceptor cells, while GMR-Gal4/ UAS-*org-1* flies retain intact, albeit slightly roughened eyes.
All transgenes that contained any OMB domain produced flies with degenerated eyes, implying that all three domains in OMB contribute to its specific-

ity. The T domain seems to comprise the strongest specificity determinants of OMB, since in an otherwise ORG-1 context, the T-box of OMB resulted in eyes comparable to those seen with a full-length *omb* transgene. Likewise, MYC-*omb*N+T+*org-1*C-HA and MYC-*omb*N+*org-1*C+*omb*C-MYC transgenes showed *omb* specificity. Surprisingly, however, GMR-Gal4 driven expression of the HA-*org-1*+*omb*T+C-MYC transgene did not cause a degeneration of photoreceptor cells and therefore was classified *org-1*-like, although the reduced eye size observed in these flies is an *omb*-like phene. We had anticipated *omb* specificity for this construct, as chimeric transgenes containing a single OMB T or C domain were sufficient to confer those *omb* specificity.

This unexpected finding might be explained by different expression strengths of the investigated transgenes. Dependency of the expression of a transgene on its insertion site in the genome is known in *Drosophila* as "position effects" (Heslip and Hodgetts, 1994 and references therein). Since we want to compare the chimeric *omb-org-1* constructs for qualitative differences only, we developed a method to determine the relative expression strengths of different UAS-transgenes by which we can exclude quantitative effects due to differences in transgene expression strength. We used this detection system to measure the expression level of several transgenic series, but have not completed this analysis yet in the course of this work. I described in chapter 5.3.2 how this analysis might be completed and took the issue of different transgene expression levels into consideration by using a number of independent insertion lines for each chimeric transgene within the GMR-Gal4 expression experiment. It remains possible, however, that all tested lines for the construct HA-*org-1*N+*omb*T+C-MYC differ in their expression strength from the other transgenic lines, thereby causing the conflicting *org-1*-like phenotype.

Gert Pflugfelder continued this project by systematically analyzing all lines containing *omb-org-1* chimeric transgenes with *dpp*-Gal4-K54.
Importantly, although individual lines revealed differences in phenotypic expressivity (due to position efffects), phenotypes were qualitatively consistent within transgenic series. The *dpp*-Gal4 experiment confirmed that all three OMB domains contribute to functional specificity. Interestingly, transgene HA-*org-1*N+*omb*T+C-MYC that we tentatively assessed in the GMR-Gal4 expression experiment to be *org-1*-like unambigously displayed *omb* specificity with *dpp*-Gal4 (Gert Pflugfelder, pers. comm.).

The ectopic expression experiments with GMR-Gal4 and *dpp*-Gal4 might be complemented by using additional Gal4-driver(s) that should produce *omb* and *org-1* specific phenotypes, *e.g. Dll*-Gal4 (Gorfinkiel *et al.,* 1997). Moreover, this work would be further improved, if one could demonstrate an about equal expression strength of the different *omb-org-1* chimeric transgenes that are used in the comparative analysis. Furthermore, nuclear localization should been assayed for the different OMB-ORG-1 chimeric proteins, since we do not know, whether the nuclear localization signal(s) in OMB and ORG-1 reside within the same part of both proteins. This control experiment can be performed by staining tissues with transgene expression with the monoclonal antibodies directed against the HA or MYC epitopes.

The results of GMR-Gal4 and *dpp*-Gal4 driven ectopic expression of chimeric *omb-org-1* transgenes revealed that all domains in OMB (and ORG-1?) contribute to functional specificity. The T domain was thereby identified as the major specificity determinant in OMB. Conlon *et al.* (2001) investigated the relevance of the T-boxes of Xbra, VegT, and Eomesodermin for specificity by expressing T-box fusion proteins (comprising a central T domain flanked by the Gal4 DNA binding domain and the VP16 activation domain) in early *Xenopus* embryos and monitoring target gene expression. Similar to our results, they revealed that the specificity of *Xbra*, *VegT*, and *Eomesodermin* resides, to a large part, in the T-box domains.

The T domain is an unusually large DNA binding motif that in OMB and ORG-1 comprises 187 and 191 aa, respectively. Different T-box domains appear to have overall similar *in vitro* DNA binding abilities, because all T domain factors tested, representing the five T-box subfamilies, recognized the palindromic T consensus binding site *in vitro* (Kispert and Herrmann, 1993; Papapetrou *et al*., 1997; Grimm and Pflugfelder, in prep.; Sinha *et al*., 2000; Carreira *et al*., 1998; Carlson *et al*., 2001; Bruneau *et al*., 2001; Ghosh *et al*., 2001; Papapetrou *et al*., 1999; Hsueh *et al*., 2000). Individual T domain proteins, however, differ in more subtle DNA binding characteristics such as dimerization tendency or the preference for certain arrangements of binding sites in regard to spacing or orientation (Kispert and Herrmann, 1993; Grimm and Pflugfelder, in prep.; Sinha *et al*., 2000). It is conceivable that these differences in DNA binding are responsible for the recognition of different enhancer elements and contribute to target gene specificity of T-box proteins *in vivo*. The replacement of a single aa residue (Asn) of VegT and Eomesoder-

min that is predicted to have DNA contact, with the corresponding aa of Xbra (Lys 149) was sufficient to change the expression of target genes of VegT and Eomesodermin to resemble that of Xbra (Conlon *et al.,* 2001), implying that DNA binding characteristics of the T domain are critical for T-box specificity.

Similar observations have been made with homeodomain (HD) proteins, another large family of transcription factors. Like T-box proteins, homeotic selector proteins are transcriptional regulators with a conserved DNA binding motif, the HD, and little sequence identity in regions outside the DNA binding domain. Homeotic proteins also recognize *in vitro* similar DNA sequences (Ekker *et al.,* 1994). A number of domain swap studies analogous to our experimental design were carried out with pairs of homeotic *Drosophila* proteins including Antennapedia, Sex combs reduced, Ultrabithorax, and Deformed (*e.g.* Lin and McGinnis, 1992; Chan and Mann, 1993; Zeng *et al.,* 1993; Furukubo-Tokunaga *et al.,* 1993). These studies identified the HD DNA binding motif as the most important part of homeotic selector proteins in determining functional specificity.

Our work suggests that sequences in OMB and ORG-1 outside the T-box motif contribute to functional specificity, too, indicating that some mechanism(s) other than DNA binding are involved in determining specificity. Distinct modes of transcriptional regulation might represent such a possible mechanism. The majority of T-box factors has been found to function as transcriptional activators, however, a few transcriptional repressors exist among members of the T-box family, too. Interestingly, dominant repressor domains have been identified in TBX3 and TBX2, the putative human orthologs of OMB. It is therefore imaginable, that OMB functions as a transcriptional repressor as well. If this holds true and ORG-1, as most other T-box proteins, would be a transcriptional activator, than OMB and ORG-1 might elicit distinct effects even on common target genes.

Another possibility for how non T-box sequences in OMB and ORG-1 account for specificity is provided by the presence of distinct protein-protein interaction domains. Specific interacting partners of OMB and ORG-1 may thereby influence the transcriptional regulation or the DNA binding behavior of both proteins differently. Interestingly, cofactor-mediated transactivation has recently been described for the T-box protein Tbr-1 (Hsueh *et al.,* 2000).

Use of specific DNA binding partners is also not without precedents for T-box proteins. The promoters of the *POMC* and *Nppa* genes both contain a

single T-box binding element juxtaposed to a HD binding site and require the cooperative binding of a T domain and a HD transcription factor, (Tpit and Pitx1 and Tbx5 with Nkx2-5 on the *POMC* and *Nppa* promoters, respectively) for transcriptional activation (Lamolet *et al.,* 2001; Hiroi *et al.,* 2001). The T-box and HD proteins bind their contiguous target sites on both promoters as heterodimers in tandem and form ternary protein-protein-DNA complexes (Lamolet *et al.,* 2001; Hiroi *et al.,* 2001). It is likely that the interaction with the HD factor is mainly required to increase the affinity of Tpit and Tbx5 for their DNA target sites (see Introduction). The interacting domains of Tbx5 and Nkx2-5 have been molecularly mapped within both proteins. The N-terminal 90 aa including 28 aa of the T domain are required in Tbx5 to interact with the HD of Nkx2-5 (Hiroi *et al.,* 2001) demonstrating that sequences outside the T domain may be involved in protein-protein interactions required for the activation of specific target genes.

The *omb-org-1* domain swap experiment taught us that specificity determinants in OMB and ORG-1 are not restricted to the T-boxes but are encoded in all parts of the proteins and suggests that mechanisms other than DNA binding specificity contribute to the functional specificity of T-box transcription factors. Which mechanisms, however, underly T-box specificity in general and the specificity of OMB and ORG-1 in particular, remains unanswered yet and will require future molecular investigation.

7. References

Agulnik, S. I., Garvey, N., Hancock, S., Ruvinsky, I., Chapman, D. L., Agulnik, I., Bollag, R., Papaioannou, V., and Silver, L. M. (1996). Evolution of mouse T-box genes by tandem duplication and cluster dispersion. Genetics *144*, 249-254.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-10.

Ashburner, M. (1989). *Drosophila*. A laboratory handbook (Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

Ausubel, F. M., Brent, R., Kingston, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994). Current protocols in molecular biology: Green Publishing Associates, Wiley-Interscience).

Bamshad, M., Lin, R. C., Law, D. J., Watkins, W. S., Krakowiak, P. A., Moore, M. E., Fransceschini, P., Lala, R., Holmes, L. B., Gebuhr, T. C., Bruneau, B. G., Schinzel, A., Seidman, J. G., Seidman, C. E., and Jorde, L. B. (1997). Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. Nat. Genet. *16*, 311-315.

Bamshad, M., Le, T., Watkins, W. S., Dixon, M. E., Kramer, B. E., Roeder, A. D., Carey, J. C., Root, S., Schinzel, A., Van Maldergem, L., Gardner, R. J., Lin, R. C., Seidman, C. E., Seidman, J. G., Wallerstein, R., Moran, E., Sutphen, R., Campbell, C. E., and Jorde, L. B. (1999). The spectrum of mutations in TBX3: Genotype/Phenotype relationship in ulnar-mammary syndrome. Am. J. Hum. Genet. *64*, 1550-62.

Basson, C. T., Bachinsky, D. R., Lin, R. C., Levi, T., Elkins, J. A., Soults, J., Grayzel, D., Kroumpouzou, E., Traill, T. A., Leblanc-Straceski, J., Renault, B., Kucherlapati, R., Seidman, J. G., and Seidman, C. E. (1997). Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. Nat. Genet. *15*, 30-34.

Bausenwein, B., Wolf, R., and Heisenberg, M. (1986). Genetic dissection of optomotor behavior in Drosophila melanogaster. Studies on wild-type and the mutant optomotor-blind. J. Neurogenetics *3*, 87-109.

Best, F. (1905). Über eine hereditäre Maculaffection: Beiträge zur Vererbungslehre. Z. Augenheilkd. *13*, 199-212.

Bieber, A. J., Snow, P. M., Hortsch, M., Patel, N. H., Jacobs, J. R., Traquina, Z. R., Schilling, J., and Goodman, C. S. (1989). Drosophila neuroglian: a member of the im-
munoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. Cell *59*, 447-60.

Bollag, R. J., Siegfried, Z., Cebra-Thomas, J. A., Garvey, N., Davison, E. M., and Silver, L. M. (1994). An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T locus. Nat. Genet. *7*, 383-389.

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.

Braybrook, C., Doudney, K., Marcano, A. C., Arnason, A., Bjornsson, A., Patton, M. A., Goodfellow, P. J., Moore, G. E., and Stanier, P. (2001). The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia. Nat. Genet. *29*, 179-83.

Brook, W. J. and Cohen, S. M. (1996). Antagonistic interactions between wingless and decapentaplegic responsible for dorsal-ventral pattern in the Drosophila leg. Science *273*, 1373-1377.

Brown, N. H. and Kafatos, F. C. (1988). Functional cDNA libraries from Drosophila embryos. J. Mol. Biol. *203*, 425-437.

Bruneau, B. G., Nemer, G., Schmitt, J. P., Charron, F., Robitaille, L., Caron, S., Conner, D. A., Gessler, M., Nemer, M., Seidman, C. E., and Seidman, J. G. (2001). A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell *106*, 709-21.

Bulfone, A., Smiga, S. M., Shimamura, K., Peterson, A., Puelles, L., and Rubenstein, J. L. R. (1995). T-Brain-1: a homolog to Brachyury whose expression defines molecularly distinct domains within the cerebral cortex. Neuron *15*, 63-78.

Bulfone, A., Wang, F., Hevner, R., Anderson, S., Cutforth, T., Chen, S., Meneses, J., Pedersen, R., Axel, R., and Rubenstein, J. L. (1998). An olfactory sensory map develops in the absence of normal projection neurons or GABAergic interneurons. Neuron *21*, 1273-82.

Campbell, G. and Tomlinson, A. (1995). Initiation of the proximodistal axis in insect legs. Development *121*, 619-628.

Carlson, H., Ota, S., Campbell, C. E., and Hurlin, P. J. (2001). A dominant repression domain in Tbx3 mediates transcriptional repression and cell immortalization: relevance to mutations in Tbx3 that cause ulnar-mammary syndrome. Hum. Mol. Genet. *10*, 2403-13.

Carreira, S., Dexter, T. J., Yavuzer, U., Easty, D. J., and Goding, C. R. (1998). Brachyury-related transcription fac-

tor Tbx2 and repression of the melanocyte-specific TRP-1 promoter. Mol. Cell. Biol. *18*, 5099-108.

Casares, F. and Mann, R. S. (2001). The ground state of the ventral appendage in Drosophila. Science *293*, 1477-80.

Casey, E. S., O'Reilly, M. A., Conlon, F. L., and Smith, J. C. (1998). The T-box transcription factor Brachyury regulates expression of eFGF through binding to a non-palindromic response element. Development *125*, 3887-94.

Cavener, D. R. (1987). Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucl. Acids. Res. *15*, 1353-1361.

Chan, S. K. and Mann, R. S. (1993). The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. Genes Dev. *7*, 796-811.

Chapman, D. L., Garvey, N., Hancock, S., Alexiou, M., Agulnik, S., Gibson-Brown, J. J., Cebra-Thomas, J., Bollag, R. J., Silver, L. M., and Papaioannou, V. E. (1996). Expression of the T-box family genes, *Tbx1-Tbx5*, during early mouse development. Devel. Dynam. *206*, 379-390.

Chapman, D. L. and Papaioannou, V. E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. Nature *391*, 695-7.

Chen, J., Zhong, Q., Wang, J., Cameron, R. S., Borke, J. L., Isales, C. M., and Bollag, R. J. (2001). Microarray analysis of Tbx2-directed gene expression: a possible role in osteogenesis. Mol. Cell Endocrinol. *177*, 43-54.

Chieffo, C., Garvey, N., Gong, W., Roe, B., Zhang, G., Silver, L., Emanuel, B. S., and Budarf, M. L. (1997). Isolation and characterization of a gene from the DiGeorge chromosomal region homologous to the mouse Tbx1 gene. Genomics *43*, 267-277.

Collins, F. S. and Weissman, S. M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. Proc. Natl. Acad. Sci. USA *81*, 6812-6816.

Conlon, F. L., Sedgwick, S. G., Weston, K. M., and Smith, J. C. (1996). Inhibition of Xbra activation causes defects in mesodermal patterning and reveals autoregulation of Xbra in dorsal mesoderm. Development *122*, 2427-2435.

Conlon, F. L. and Smith, J. C. (1999). Interference with brachyury function inhibits convergent extension, causes apoptosis, and reveals separate requirements in the FGF and activin signalling pathways. Dev. Biol. *213*, 85-100.

Conlon, F. L., Fairclough, L., Price, B. M., Casey, E. S., and Smith, J. C. (2001). Determinants of T box protein specificity. Development *128*, 3749-58.

Cooley, L., Thompson, D., and Spradling, A. (1990). Constructing deletions with defined endpoints in Drosophila. Proc. Natl. Acad. Sci. USA *87*, 3170-3173.

Cunliffe, V. and Smith, J. C. (1992). Ectopic mesoderm formation in Xenopus embryos caused by widespread expression of a Brachyury homologue. Nature *358*, 427-30.

Cunliffe, V. and Smith, J. C. (1994). Specification of mesodermal pattern in Xenopus laevis by interactions between Brachyury, noggin and Xwnt-8. EMBO J. *13*, 349-59.

Dalby, B., Pereira, A. J., and Goldstein, L. S. (1995). An inverse PCR screen for the detection of P element insertions in cloned genomic intervals in Drosophila melanogaster. Genetics *139*, 757-66.

Di Gregorio, A. and Levine, M. (1999). Regulation of Citropomyosin-like, a Brachyury target gene in the ascidian, Ciona intestinalis. Development *126*, 5599-609.

Díaz-Benjumea, F. J., Cohen, B., and Cohen, S. M. (1994). Cell interaction between compartments establishes the proximal-distal axis of *Drosophila* legs. Nature *372*, 175-179.

Dobrovolskaia-Zavadskaia, N. (1927). Sur la mortification spontaneé de la queue chez la souris nouveau-neé et sur l´existence d´un caractère heréditaire ´non-viable´. C.R. Hébd. Séances Soc. Biol. *97*, 114-116.

Duncan, D. M., Burgess, E. A., and Duncan, I. (1998). Control of distal antennal identity and tarsal development in Drosophila by spineless-aristapedia, a homolog of the mammalian dioxin receptor. Genes Dev. *12*, 1290-303.

Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E., and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. EMBO J. *13*, 3551-60.

Ellis, M. C., O'Neill, E. M., and Rubin, G. M. (1993). Expression of Drosophila glass protein and evidence for negative regulation of its activity in non-neuronal cells by another DNA-binding protein. Development *119*, 855-65.

Emanuel, B. S., Budarf, B. S., and Scambler, P. J. (1998). The genetic basis of conotruncal heart defects: the chromosome 22q11.2 deletion. In Heart Development, N. Rosenthal and R. Harvey, eds.: Academic Press).

Engels, W. R. (1996). P elements in Drosophila. In Transposable Elements, H. Saedler and A. Gierl, eds. (Berlin: Springer-Verlag), pp. 103-123.

Engels, W. R. (2000). Reversal of fortune for Drosophila geneticists? Science *288*, 1973-5.

Evan, G. I., Lewis, G. K., Ramsay, G., and Bishop, J. M. (1985). Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. Mol. Cell. Biol. *5*, 3610-6.

Fawcett, D. H., Lister, C. K., Kellett, E., and Finnegan, D. J. (1986). Transposable elements controlling I-R hybrid dysgenesis in D. melanogaster are similar to mammalian LINEs. Cell *47*, 1007-15.

Funke, B., Epstein, J. A., Kochilas, L. K., Lu, M. M., Pandita, R. K., Liao, J., Bauerndistel, R., Schuler, T., Schorle, H., Brown, M. C., Adams, J., and Morrow, B. E. (2001). Mice overexpressing genes from the 22q11 region deleted in velo-cardio-facial syndrome/DiGeorge syndrome have middle and inner ear defects. Hum. Mol. Genet. *10*, 2549-56.

Furukubo-Tokunaga, K., Flister, S., and Gehring, W. J. (1993). Functional specificity of the Antennapedia homeodomain. Proc. Natl. Acad. Sci. U S A *90*, 6360-4.

Geier, G. E. and Modrich, P. (1979). Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease. J. Biol. Chem. *254*, 1408-13.

Geyer, P. K., Patton, J. S., Rodesch, C., and Nagoshi, R. N. (1993). Genetic and molecular characterization of P element-induced mutations reveals that the Drosophila ovarian tumor gene has maternal activity and a variable null phenotype. Genetics *133*, 265-78.

Ghosh, T. K., Packham, E. A., Bonser, A. J., Robinson, T. E., Cross, S. J., and Brook, J. D. (2001). Characterization of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome. Hum. Mol. Genet. *10*, 1983-94.

Gibson-Brown, J. J., Agulnik, S. I., Chapman, D. L., Alexiou, M., Garvey, N., Silver, L. M., and Papaioannou, V. E. (1996). Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. Mech. Dev. *56*, 93-101.

Gibson-Brown, J. J., Agulnik, S. I., Silver, L. M., Niswander, L., and Papaioannou, V. E. (1998a). Involvement of T-box genes Tbx2-Tbx5 in vertebrate limb specification and development. Development *125*, 2499-509.

Gibson-Brown, J. J., S, I. A., Silver, L. M., and Papaioannou, V. E. (1998b). Expression of T-box genes Tbx2-Tbx5 during chick organogenesis. Mech. Dev. *74*, 165-9.

Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., and Engels, W. R. (1991). Targeted gene replacement in Drosophila via P element-induced gap repair. Science *253*, 1110-7.

Gluecksohn-Schoenheimer, S. (1938). The development of two tailless mutants in the house mouse. Genetics *23*, 573-584.

Gluecksohn-Schoenheimer, S. (1944). The development of normal and homozygous *brachy* (*T/T*) mouse embryos in the extraembryonic coelom of the chick. Proc. Natl. Acad. Sci. USA *30*, 134-140.

Gorfinkiel, N., Morata, G., and Guerrero, I. (1997). The homeobox gene Distal-less induces ventral appendage development in Drosophila. Genes Dev. *11*, 2259-71.

Griffin, K. J., Amacher, S. L., Kimmel, C. B., and Kimelman, D. (1998). Molecular identification of spadetail: regulation of zebrafish trunk and tail mesoderm formation by T-box genes. Development *125*, 3379-88.

Grigliatti, T. (1986). Mutagenesis. In *Drosophila*. A practical approach, D. B. Roberts, ed. (Oxford: IRL Press), pp. 39-58.

Grimm, S. and Pflugfelder, G. O. (1996). Control of the gene *optomotor-blind* in *Drosophila* wing development by *decapentaplegic* and *wingless*. Science *271*, 1601-1603.

Grimm, S. (1997). Das T-Box Protein Optomotor-blind von Drosophila melanogaster: DNA-Bindungseigenschaften und die Rolle in der Imaginalentwicklung: Dissertation. Universität Würzburg.

Grossniklaus, U., Pearson, R. K., and Gehring, W. J. (1992). The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. Genes Dev. *6*, 1030-51.

Halder, G., Callaerts, P., and Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science *267*, 1788-1792.

Harland, R. and Gerhart, J. (1997). Formation and function of Spemann's organizer. Annu. Rev. Cell. Dev. Biol. *13*, 611-67.

Hatcher, C. J., Kim, M. S., Mah, C. S., Goldstein, M. M., Wong, B., Mikawa, T., and Basson, C. T. (2001). TBX5 transcription factor regulates cell proliferation during cardiogenesis. Dev. Biol. *230*, 177-88.

He, M., Wen, L., Campbell, C. E., Wu, J. Y., and Rao, Y. (1999). Transcription repression by Xenopus ET and its human ortholog TBX3, a gene involved in ulnar-mammary syndrome [published erratum appears in Proc. Natl. Acad. Sci. USA 1999 Nov 9;96(23):13589]. Proc. Natl. Acad. Sci. USA *96*, 10212-7.

Heindel, U. (1998). Untersuchungen zur Omb Funktion bei *Drosophila melanogaster*: Diplomarbeit. Universität Würzburg.

Heisenberg, M. (1972). Behavioural diagnostics: a way to analyze visual mutants in *Drosophila*. In Information Processing in the Visual System, R. Wehner, ed. (Berlin: Springer Verlag), pp. 265-268.

Heisenberg, M. and Götz, K. G. (1975). The use of mutations for the partial degradation of vision in *Drosophila melanogaster*. J. Comp. Physiol. *98*, 217-241.

Heisenberg, M., Wonneberger, R., and Wolf, R. (1978). optomotor-blind[H31] - a Drosophila mutant of the lobula plate giant neurons. J. Comp. Physiol. *124*, 287-296.

Herrmann, B. G., Labeit, S., Poustka, A., King, T. R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. Nature *343*, 617-622.

Heslip, T. R. and Hodgetts, R. B. (1994). Targeted transposition at the vestigial locus of Drosophila melanogaster. Genetics *138*, 1127-35.

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R., and Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. Nat. Genet. *28*, 276-80.

Hofmeyer, K. (1996). Analyse regulatorischer DNA-Sequenzen des *optomotor-blind* Gens von *Drosophila melanogaster*: Diplomarbeit. Universität Würzburg.

Hofmeyer, K. (2001). The optic lobe regulatory region of the *Drosophila melanogaster* gene *optomotor-blind*: Dissertation. Universität Würzburg.

Horb, M. E. and Thomsen, G. H. (1997). A vegetally localized T-box transcription factor in Xenopus eggs specifies mesoderm and endoderm and is essential for embryonic mesoderm formation. Development *124*, 1689-1698.

Hortsch, M., Bieber, A. J., Patel, N. H., and Goodman, C. S. (1990). Differential splicing generates a nervous system-specific form of Drosophila neuroglian. Neuron *4*, 697-709.

Hsueh, Y. P., Wang, T. F., Yang, F. C., and Sheng, M. (2000). Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2 [see comments]. Nature *404*, 298-302.

Hyde, C. E. and Old, R. W. (2000). Regulation of the early expression of the Xenopus nodal-related 1 gene, Xnr1. Development *127*, 1221-9.

Isaac, A., Rodriguez-Esteban, C., Ryan, A., Altabef, M., Tsukui, T., Patel, K., Tickle, C., and Izpisua-Belmonte, J.

C. (1998). Tbx genes and limb identity in chick embryo development. Development *125*, 1867-75.

Isaacs, H. V., Pownall, M. E., and Slack, J. M. (1994). eFGF regulates Xbra expression during Xenopus gastrulation. EMBO J. *13*, 4469-81.

Jacobs, J. J., Keblusek, P., Robanus-Maandag, E., Kristel, P., Lingbeek, M., Nederlof, P. M., van Welsem, T., van de Vijver, M. J., Koh, E. Y., Daley, G. Q., and van Lohuizen, M. (2000). Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19(ARF)) and is amplified in a subset of human breast cancers. Nat. Genet. *26*, 291-299.

Jenkins, J. B. (1967). The induction of mosaic and complete dumpy mutants in Drosophila melanogaster with ethyl methanesulfonate. Mutat. Res. *4*, 90-2.

Jensen, S., Gassama, M. P., and Heidmann, T. (1994). Retrotransposition of the Drosophila LINE I element can induce deletion in the target DNA: a simple model also accounting for the variability of the normally observed target site duplications. Biochem. Biophys. Res. Comm. *202*, 111-9.

Jerome, L. A. and Papaioannou, V. E. (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat. Genet. *27*, 286-91.

Kidwell, M. G. (1983). Evolution of hybrid dysgenesis determinants in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA *80*, 1655-9.

Kidwell, M. G. (1987). A survey of success ratesusing *P*-element mutagenesis in *Drosophila melanogaster*. Dros. Inf. Serv. *66*, 81-86.

Kimmel, C. B., Kane, D. A., Walker, C., Warga, R. M., and Rothman, M. B. (1989). A mutation that changes cell movement and cell fate in the zebrafish embryo. Nature *337*, 358-62.

Kirsch, I. R., Green, E. D., Yonescu, R., Strausberg, R., Carter, N., Bentley, D., Leversha, M. A., Dunham, I., Braden, V. V., Hilgenfeld, E., Schuler, G., Lash, A. E., Shen, G. L., Martelli, M., Kuehl, W. M., Klausner, R. D., and Ried, T. (2000). A systematic, high-resolution linkage of the cytogenetic and physical maps of the human genome. Nat. Genet. *24*, 339-40.

Kispert, A. and Herrmann, B. G. (1993). The Brachyury gene encodes a novel DNA binding protein. EMBO J. *12*, 3211-3220.

Kispert, A. and Herrmann, B. (1994). Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. Dev. Biol. *168*, 179-193.

Kispert, A., Herrmann, B. G., Leptin, M., and Reuter, R. (1994). Homologs of the mouse Brachyury gene are involved in the specification of posterior terminal structures in Drosophila, Tribolium, and Locusta. Genes Dev. *8*, 2137-2150.

Kispert, A., Koschorz, B., and Herrmann, B. G. (1995a). The T protein encoded by Brachyury is a tissue-specific transcription factor. EMBO J. *14*, 4763-4772.

Kispert, A., Ortner, H., Cooke, J., and Herrmann, B. G. (1995b). The chick Brachury gene: developmental expression pattern and response to axial induction by localized activin. Dev. Biol. *168*, 406-415.

Klemenz, R., Weber, U., and Gehring, W. J. (1987). The white gene as a marker in a new P-element vector for gene transfer in Drosophila. Nucleic Acids Res. *15*, 3947-59.

Kofron, M., Demel, T., Xanthos, J., Lohr, J., Sun, B., Sive, H., Osada, S., Wright, C., Wylie, C., and Heasman, J. (1999). Mesoderm induction in Xenopus is a zygotic event regulated by maternal VegT via TGFbeta growth factors. Development *126*, 5759-70.

Kornberg, T. B. (1993). Understanding the homeodomain. J. Biol. Chem. *268*, 26813-6.

Kusch, T., Storck, T., Walldorf, U., and Reuter, R. (2002). Brachyury proteins regulate target genes through modular binding sites in a cooperative fashion. Genes Dev. *16*, 518-29.

Lamolet, B., Pulichino, A. M., Lamonerie, T., Gauthier, Y., Brue, T., Enjalbert, A., and Drouin, J. (2001). A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. Cell *104*, 849-59.

Lefevre, G. and Watkins, W. (1986). The question of the total gene number in Drosophila melanogaster. Genetics *113*, 869-895.

Li, Q. Y., Newbury-Ecob, R. A., Terrett, J. A., Wilson, D. I., Curtis, A. R. J., Yi, C. H., Gebuhr, T., Bullen, P. J., Robson, S. C., Strachan, T., Bonnet, D., Lyonnet, S., Young, I. D., Raeburn, J. A., Buckler, A. J., Law, D. J., and Brook, J. D. (1997). Holt-Oram syndrome is caused by mutations in *TBX5*, a member of the *Brachyury (T)* gene family. Nat. Genet. *15*, 21-29.

Liberatore, C. M., Searcy-Schrick, R. D., and Yutzey, K. E. (2000). Ventricular expression of tbx5 inhibits normal heart chamber development. Dev. Biol. *223*, 169-80.

Lin, L. and McGinnis, W. (1992). Mapping functional specificity in the Dfd and Ubx homeo domains. Genes Dev. *6*, 1071-81.

Lindsay, E. A., Botta, A., Jurecic, V., Carattini-Rivera, S., Cheah, Y. C., Rosenblatt, H. M., Bradley, A., and Baldini, A. (1999). Congenital heart disease in mice deficient for the DiGeorge syndrome region. Nature *401*, 379-83.

Lindsay, E. A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., Jurecic, V., Ogunrinu, G., Sutherland, H. F., Scambler, P. J., Bradley, A., and Baldini, A. (2001). Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. Nature *410*, 97-101.

Lindsley, D. L. and Zimm, G. G. (1992). The genome of *Drosophila melanogaster* (San Diego: Academic Press, Inc.).

Logan, M., Simon, H. G., and Tabin, C. (1998). Differential regulation of T-box and homeobox transcription factors suggests roles in controlling chick limb-type identity. Development *125*, 2825-35.

Logan, M. and Tabin, C. J. (1999). Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity [see comments]. Science *283*, 1736-9.

Lustig, K. D., Kroll, K. L., Sun, E. E., and Kirschner, M. W. (1996). Expression cloning of a Xenopus T-related gene (Xombi) involved in mesodermal patterning and blastopore lip formation. Development *122*, 4001-4012.

Mann, R. S. and Morata, G. (2000). The developmental and molecular biology of genes that subdivide the body of Drosophila. Annu. Rev. Cell. Dev. Biol. *16*, 243-71.

Marinus, M. G. and Morris, N. R. (1973). Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. J. Bacteriol. *114*, 1143-50.

Marquardt, A., Stohr, H., Passmore, L. A., Kramer, F., Rivera, A., and Weber, B. H. (1998). Mutations in a novel gene, VMD2, encoding a protein of unknown properties cause juvenile-onset vitelliform macular dystrophy (Best's disease). Hum. Mol. Genet. *7*, 1517-25.

Melnyk, A. R., Weiss, L., Van Dyke, D. L., and Jarvi, P. (1981). Malformation syndrome of duplication 12q24.1 leads to qter. Am. J. Med. Genet. *10*, 357-65.

Merscher, S., Funke, B., Epstein, J. A., Heyer, J., Puech, A., Lu, M. M., Xavier, R. J., Demay, M. B., Russell, R. G., Factor, S., Tokooya, K., Jore, B. S., Lopez, M., Pandita, R. K., Lia, M., Carrion, D., Xu, H., Schorle, H., Kobler, J. B., Scambler, P., Wynshaw-Boris, A., Skoultchi, A. I., Morrow, B. E., and Kucherlapati, R. (2001). TBX1 is responsible for cardiovascular defects in velo-cardio- facial/DiGeorge syndrome. Cell *104*, 619-29.

Müller, C. W. and Herrmann, B. G. (1997). Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. Nature *389*, 884-888.

O'Reilly, M. A., Smith, J. C., and Cunliffe, V. (1995). Patterning of the mesoderm in Xenopus: dose-dependent and synergistic effects of Brachyury and Pintallavis. Development *121*, 1351-9.

Ochman, H., Gerber, A. S., and Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics *120*, 621-3.

Ohuchi, H., Takeuchi, J., Yoshioka, H., Ishimaru, Y., Ogura, K., Takahashi, N., Ogura, T., and Noji, S. (1998). Correlation of wing-leg identity in ectopic FGF-induced chimeric limbs with the differential expression of chick Tbx5 and Tbx4. Development *125*, 51-60.

Papaioannou, V. E. (2001). T-box genes in development: from hydra to humans. Int. Rev. Cytol. *207*, 1-70.

Papapetrou, C., Edwards, Y. H., and Sowden, J. C. (1997). The T transcription factor functions as a dimer and exhibits a common human polymorphism Gly-177-Asp in the conserved DNA-binding domain. FEBS Letters *409*, 201-206.

Papapetrou, C., Putt, W., Fox, M., and Edwards, Y. H. (1999). The human TBX6 gene: cloning and assignment to chromosome 16p11.2. Genomics *55*, 238-41.

Paxton, C., Zhao, H., Chin, Y., Langner, K., and Reecy, J. (2002). Murine Tbx2 contains domains that activate and repress gene transcription. Gene *283*, 117-24.

Peter, A., Schottler, P., Werner, M., Beinert, N., Dowe, G., Burkert, P., Mourkioti, F., Dentzer, L., He, Y., Deak, P., Benos, P. V., Gatt, M. K., Murphy, L., Harris, D., Barrell, B., Ferraz, C., Vidal, S., Brun, C., Demaille, J., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Borkova, D., Minana, B., Kafatos, F. C., Bolshakov, S., Siden-Kiamos, I., Papagiannakis, G., Spanos, L., Louis, C., Madueno, E., de Pablos, B., Modolell, J., Bucheton, A., Callister, D., Campbell, L., Henderson, N. S., McMillan, P. J., Salles, C., Tait, E., Valenti, P., Saunders, R. D., Billaud, A., Pachter, L., Klapper, R., Janning, W., Glover, D. M., Ashburner, M., Bellen, H. J., Jackle, H., and Schafer, U. (2002). Mapping and identification of essential gene functions on the X chromosome of Drosophila. EMBO Rep. *3*, 34-8.

Pflugfelder, G. O., Schwarz, H., Roth, H., Poeck, B., Sigl, A., Kerscher, S., Jonschker, B., Pak, W. L., and Heisenberg, M. (1990). Genetic and molecular characterization of the optomotor-blind gene locus in Drosophila melanogaster. Genetics *126*, 91-104.

Pflugfelder, G. O., Roth, H., Poeck, B., Kerscher, S., Schwarz, H., Jonschker, B., and Heisenberg, M. (1992a). The lethal(1)optomotor-blind gene of Drosophila melanogaster is a major organizer of optic lobe development: isolation and charcterization of the gene. Proc. Natl. Acad. Sci. USA *89*, 1199-1203.

Pflugfelder, G. O., Roth, H., and Poeck, B. (1992b). A homology domain shared between Drosophila optomotor-blind and mouse Brachyury is involved in DNA binding. Biochem. Biophys. Res. Comm. *186*, 918-925.

Pflugfelder, G. O. and Heisenberg, M. (1995). *optomotor-blind* of *Drosophila melanogaster*: A neurogenetic approach to optic lobe development and optomotor behaviour. Comp. Biochem. Physiol. *110A*, 185-202.

Poeck, B., Balles, J., and Pflugfelder, G. O. (1993). Transcript identification in the *optomotor-blind* locus of *Drosophila melanogaster* by intragenic recombination mapping and PCR-aided sequence analysis of lethal point mutations. Mol. Gen. Genet. *238*, 325-332.

Porsch, M. (1997). Molecular characterization of the *optomotor-blind related gene-1* from *Drosophila melanogaster*: Diplomarbeit. Universität Würzburg.

Porsch, M., Hofmeyer, K., Bausenwein, B. S., Grimm, S., Weber, B. H. F., Miassod, R., and Pflugfelder, G. O. (1998). Isolation of a Drosophila T-box gene closely related to human TBX1. Gene *212*, 237-248.

Rabenstein, M. D., Zhou, S., Lis, J. T., and Tjian, R. (1999). TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. Proc. Natl. Acad. Sci. USA *96*, 4791-6.

Rao, Y. (1994). Conversion of a mesodermalizing molecule, the Xenopus Brachyury gene, into a neuralizing factor. Genes Dev. *8*, 939-47.

Rashbass, P., Cooke, L. A., Herrmann, B. G., and Beddington, R. S. (1991). A cell autonomous function of Brachyury in T/T embryonic stem cell chimaeras. Nature *353*, 348-51.

Reichmuth, C., Becker, S., Benz, M., Debel, K., Reisch, D., Heimbeck, G., Hofbauer, A., Klagges, B., Pflugfelder, G. O., and Buchner, E. (1995). The sap47 gene of Drosophila melanogaster codes for a novel conserved neuronal protein associated with synaptic terminals. Brain Res. Mol. Brain Res. *32*, 45-54.

Rio, D. C. and Rubin, G. M. (1985). Transformation of cultured Drosophila melanogaster cells with a dominant selectable marker. Mol. Cell. Biol. *5*, 1833-8.

Robertson, H. M., Preston, C. R., Phillis, R. W., Johnson-Schlitz, D. M., Benz, W. K., and Engels, W. R. (1988). A stable genomic source of P element transposase in Drosophila melanogaster. Genetics *118*, 461-70.

Rodriguez-Esteban, C., Tsukui, T., Yonei, S., Magallon, J., Tamura, K., and Izpisua Belmonte, J. C. (1999). The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity [see comments]. Nature *398*, 814-8.

Rong, Y. S. and Golic, K. G. (2000). Gene targeting by homologous recombination in Drosophila. Science *288*, 2013-8.

Rong, Y. S. and Golic, K. G. (2001). A targeted gene knockout in Drosophila. Genetics *157*, 1307-12.

Rorth, P. (1996). A modular misexpression screen in Drosophila detecting tissue-specific phenotypes. Proc. Natl. Acad. Sci. USA *93*, 12418-22.

Roth, H. (1991). Molekulare Analyse des omb Gen locus von Drosophila melanogaster: Dissertation. Universität Würzburg.

Roth, M. (1998). Die Rolle der Decapentaplegic-Proteoglykanbindungsdomänen in der Morphogenese von Drosophila melanogaster: Diplomarbeit. Universität Würzburg.

Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol *194*, 281-301.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S., Gong, F., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J., Li, Z., Hynes, R. O., Jones, S. J., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Vosshall, L. B., Zhang, J., Zhao, Q., Zheng, X. H., Zhong, F., Zhong, W., Gibbs, R., Venter, J. C., Adams, M. D., and Lewis, S. (2000). Comparative genomics of the eukaryotes. Science *287*, 2204-15.

Russ, A. P., Wattler, S., Colledge, W. H., Aparicio, S. A., Carlton, M. B., Pearce, J. J., Barton, S. C., Surani, M. A., Ryan, K., Nehls, M. C., Wilson, V., and Evans, M. J. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature *404*, 95-9.

Ruvinsky, I. and Silver, L. M. (1997). Newly identified paralogous groups on mouse chromosomes 5 and 11 reveal the age of a T-box cluster duplication. Genomics *40*, 262-6.

Ruvkun, G. and Hobert, O. (1998). The taxonomy of developmental control in Caenorhabditis elegans. Science *282*, 2033-41.

Ryan, K., Garret, N., Mitchell, A., and Gurdon, J. B. (1996). Eomesodermin, a key early gene in Xenopus mesoderm differentiation. Cell *87*, 989-1000.

Saito, D., Yonei-Tamura, S., Kano, K., Ide, H., and Tamura, K. (2002). Specification and determination of limb identity: evidence for inhibitory regulation of Tbx gene expression. Development *129*, 211-20.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: A laboratory manual, 2nd Edition (Cold Spring Habor, New York: Cold Spring Habor Laboratory Press).

Santamaria, P. (1986). Injecting eggs. In *Drosophila*. A practical approach, D. B. Roberts, ed. (Oxford: IRL Press), pp. 159-173.

Scambler, P. J. (2000). The 22q11 deletion syndromes. Hum. Mol. Genet. *9*, 2421-6.

Schinzel, A. (1987). Ulnar-mammary syndrome. J. Med. Genet. *24*, 778-781.

Schulte-Merker, S., Ho, R. K., Herrmann, B. G., and Nüsslein-Volhard, C. (1992). The protein product of the zebrafish homologue of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo. Development *116*, 1021-1032.

Schulte-Merker, S., Hammerschmidt, M., Beuchle, D., Cho, K. W., De Robertis, E. M., and Nusslein-Volhard, C. (1994). Expression of zebrafish goosecoid and no tail gene products in wild-type and mutant no tail embryos. Development *120*, 843-52.

Schulte-Merker, S. and Smith, J. C. (1995). Mesoderm formation in response to Brachyury requires FGF signalling. Curr. Biol. *5*, 62-7.

Sentry, J. W. and Kaiser, K. (1994). Application of inverse PCR to site-selected mutagenesis of Drosophila. Nucleic Acids Res. *22*, 3429-30.

Seong, C., Kim, Y. A., Chung, H. J., Park, D., Yim, J., Baek, K., Park, Y. S., Han, K., and Yoon, J. (1998). Isolation and characterization of the Drosophila melanogaster cDNA encoding the sepiapterin reductase. Biochim. Biophys. Acta. *1443*, 239-44.

Seong, C., Baek, K., and Yoon, J. (2000). Structure, chromosomal localization, and expression of the Drosophila melanogaster gene encoding sepiapterin reductase. Gene *255*, 357-61.

Sezutsu, H., Nitasaka, E., and Yamazaki, T. (1995). Evolution of the LINE-like I element in the Drosophila melanogaster species subgroup. Mol. Gen. Genet. *249*, 168-78.

Shorrocks, B. S. (1972). Invertebrate Types: Drosophila (London: Ginn & Company Limited).

Simon, H.-G., Kittappa, R., Khan, P. A., Tsilfidis, C., Liversage, R. A., and Oppenheimer, S. (1997). A novel family of T-box genes in urodele amphibian limb development and regeneration: candidate genes involved in vertebrate forelimb/hindlimb patterning. Development *124*, 1355-1366.

Singer, J. B., Harbecke, R., Kusch, T., Reuter, R., and Lengyel, J. A. (1996). Drosophila brachyenteron regulates gene activity and morphogenesis in the gut. Development *122*, 3707-3718.

Sinha, S., Abraham, S., Gronostajski, R. M., and Campbell, C. E. (2000). Differential DNA binding and transcription modulation by three T-box proteins, T, TBX1 and TBX2. Gene *258*, 15-29.

Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D., and Herrmann, B. G. (1991). Expression of a Xenopus homolog of Brachyury (T) is an immediate-early resonse to mesoderm induction. Cell *67*, 1-20.

Smith, J. C. (2001). Making mesoderm--upstream and downstream of Xbra. Int. J. Dev. Biol. *45*, 219-24.

Spradling, A. C. (1986). P-element mediated transformation. In *Drosophila*. A practical approach, D. B. Roberts, ed. (Oxford: IRL Press), pp. 175-198.

Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T., and Rubin, G. M. (1995). Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. Proc. Natl. Acad. Sci. USA *92*, 10824-10830.

Staehling-Hampton, K., Jackson, P. D., Clark, M. J., Brand, A. H., and Hoffmann, F. M. (1994). Specificity of bone morphogenetic protein-related factors: cell fate and gene expression changes in Drosophila embryo induced by decapentaplegic but not 60A. Cell Growth & Differentiation *5*, 585-593.

Stennard, F., Carnac, G., and Gurdon, J. B. (1996). The Xenopus T-box gene, Antipodean, encodes a vegetally localised maternal mRNA and triggers mesoderm formation. Development *122*, 4179-4188.

Stott, D., Kispert, A., and Herrmann, B. G. (1993). Rescue of the tail defect of Brachyury mice. Genes Dev. *7*, 197-203.

Strauss, R. (1995). A screen for EMS-induced X-linked locomotor mutants in *Drosophila melanogaster*. J. Neurogenetics *10*, 53-54.

Strauss, R. and Trinath, T. (1996). Is walking in a straight line controlled by the central complex? Evidence from a new *Drosophila* mutant. In Göttingen neurobiology report 1996, N. Elsner and H.-U. Schnitzler, eds. (Stuttgart New York: Georg Thieme Verlag), pp. 135.

Szabo, S. J., Kim, S. T., Costa, G. L., Zhang, X., Fathman, C. G., and Glimcher, L. H. (2000). A novel transcription factor, T-bet, directs Th1 lineage commitment [In Process Citation]. Cell *100*, 655-69.

Tada, M., Casey, E. S., Fairclough, L., and Smith, J. C. (1998). Bix1, a direct target of Xenopus T-box genes, causes formation of ventral mesoderm and endoderm. Development *125*, 3997-4006.

Tada, M. and Smith, J. C. (2000). Xwnt11 is a target of Xenopus Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. Development *127*, 2227-38.

Takabatake, Y., Takabatake, T., and Takeshima, K. (2000). Conserved and divergent expression of T-box genes Tbx2-Tbx5 in Xenopus. Mech. Dev. *91*, 433-437.

Takahashi, H., Mitani, Y., Satoh, G., and Satoh, N. (1999). Evolutionary alterations of the minimal promoter for notochord-specific Brachyury expression in ascidian embryos. Development *126*, 3725-34.

Takeuchi, J. K., Koshiba-Takeuchi, K., Matsumoto, K., Vogel-Hopker, A., Naitoh-Matsuo, M., Ogura, K., Takahashi, N., Yasuda, K., and Ogura, T. (1999). Tbx5 and Tbx4 genes determine the wing/leg identity of limb buds [see comments]. Nature *398*, 810-4.

Tamura, K., Yonei-Tamura, S., and Belmonte, J. C. (1999). Differential expression of Tbx4 and Tbx5 in Zebrafish fin buds. Mech. Dev. *87*, 181-4.

Technau, U. and Bode, H. R. (1999). HyBra1, a Brachyury homologue, acts during head formation in Hydra. Development *126*, 999-1010.

Thomas, K. R. and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-12.

Tower, J., Karpen, G. H., Craig, N., and Spradling, A. C. (1993). Preferential transposition of Drosophila P elements to nearby chromosomal sites. Genetics *133*, 347-59.

Trofatter, J. A., Long, K. R., Murrell, J. R., Stotler, C. J., Gusella, J. F., and Buckler, A. J. (1995). An expression-independent catalog of genes from human chromosome 22. Genome Res. *5*, 214-24.

Vaughan, C. J. and Basson, C. T. (2001). Molecular determinants of atrial and ventricular septal defects and patent ductus arteriosus. Am. J. Med. Genet. *97*, 304-9.

Wattler, S., Russ, A., Evans, M., and Nehls, M. (1998). A combined analysis of genomic and primary protein structure defines the phylogenetic relationship of new members if the T-box family. Genomics *48*, 24-33.

Wilkinson, D. G., Bhatt, S., and Herrmann, B. G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation [see comments]. Nature *343*, 657-9.

Wilson, I. A., Niman, H. L., Houghten, R. A., Cherenson, A. R., Connolly, M. L., and Lerner, R. A. (1984). The structure of an antigenic determinant in a protein. Cell *37*, 767-78.

Wilson, V., Rashbass, P., and Beddington, R. S. (1993). Chimeric analysis of T (Brachyury) gene function. Development *117*, 1321-31.

Wilson, V., Manson, L., Skarnes, W. C., and Beddington, R. S. (1995). The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. Development *121*, 877-86.

Woollard, A. and Hodgkin, J. (2000). The caenorhabditis elegans fate-determining gene mab-9 encodes a T-box protein required to pattern the posterior hindgut. Genes Dev. *14*, 596-603.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Genes Dev. *13*, 3185-90.

Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A., and Scott, M. P. (1993). Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity. Development *118*, 339-52.

Zhang, P. and Spradling, A. C. (1993). Efficient and dispersed local P element transposition from Drosophila females. Genetics *133*, 361-73.

Zhang, J. and King, M. L. (1996). Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. Development *122*, 4119-4129.

8. Acknowledgements

Most of all I owe my family, and especially, my parents for providing me a fabulous education, support, and love from which everything I have has its roots and which was fundamental to this piece of work, too. Next, my profound gratitude goes to Prof. Gert O. Pflugfelder who thoroughly advised my thesis and taught me scientific work by innumerable sessions with which he not only led me through the maze of *org-1* and *C31* genetics (thereby saving me from my worst nightmares), but also helped me to find solutions to the numerous problems that emerged in the course of the *omb-org-1* domain swap project. His confidence in my work demonstrated by the allowance to independently and freely conduct my experiments is greatly acknowlegded. Moreover, Gert's friendly and modest personality together with his delightful sense of humour made my time in his laboratory very joyful. His broad interest in many areas of biological research greatly stimulated my work and certainly contributed to the XXL size of this thesis. I would also like to thank Prof. Martin Heisenberg and Prof. Erich Buchner for teaching me and for encouragement throughout. Although a favorite for this job, Prof. Georg Krohne was willing to assess my thesis as well, and beyond that trained me at the SEM. Prof. Bernhard Weber's immediate interest in analyzing human *TBX1* is greatly acknowledged.

I am very thankful to Björn Brembs who next to my parents certainly is the person to whom I can talk about everything and who, with all his wits and talents, never failed to provide appropriate help whenever urgently needed in academic or private issues. Not to mention our cheerful sportive rivalries or the jolly trips to Kiel with the university handball team.

Handball was my major source of recreation from my graduate studies and, in particular during the first half of my work, kept my mood in balance with the traps and pitfalls of my scientific work. I would like to thank all my team-colleagues and friends from the DJK Waldbüttelbrunn handball team. I am especially thankful to our coach Manni Wirth. His philosophy, a team has to overcome desperate defeats to become a good one helped me to see odd results as a challenge. Many thanks also go to Judith Gerlach for her warm friendship and for taking care of my physical condition after I retired from handball.

My deep gratitude also goes to Martin Roth. Martin helped me to develop the detection system of UAS-transgene expression, and beyond that, had a great impact on my work as a constant help with all his smartness. Without him and his Pottsau fellow Sebastian Scharf, my daily work would not have been as half as funny as it was.

It was a great pleasure to me to introduce Silvia Roth to our laboratory. Silvia provided outstanding and diligent technical assistance and greatly supported the local-hop experiments and the Western blot analysis.

In the course of my work I could give five students a first training in laboratory work. Matthias Nowak, Ali Nowrouzi, Daniel Horbelt, Angela Bahlo and Thomas Hendel all were brave, talented students that, each in his/ her own way, contributed to my work. Among these, Angie deserves a special credit. She continued her excellent practical course on the *omb-org-1* swaps and now, together with Sabine Schulze, is a successor as one wishes to have, both being charming, intelligent, and with positive attitudes throughout.

Reinhard Wolf, Andreas Eckart, Konrad Öchsner, Christian Leipold, Roman Ernst, Björn Brembs, Antonio Prado, Martin Roth and others were a great help in solving technical problems.

Finally I would like to thank all my friends and collegues, past and present, of the Department of Genetics and Neurobiology which made my stay highly enjoyable.

9. Curriculum Vitae

Matthias Heinrich Porsch

personal data: born on June, 16th 1971 in Würzburg
single
German nationality

school education:

1977-1981   elementary school, Waldbüttelbrunn
1981-1990   high school Riemenschneider-Gymnasium, Würzburg
June 1990   high school graduation "Abitur": "B" (2,2)

military service:

07/1990 – 06/1991   mandatory military service, Hammelburg

university education:

09/1991-05/1997   Bavarian Julius-Maximilians University, Würzburg
studies of Biology

09/1991-09/1993   undergraduate studies. Intermediate examination: "A"
graduate studies. Majors: Biochemistry and Animal Physiology
08/1994-08/1995   DAAD scholar
studies at the State University of New York at Albany, USA
May 1997   diploma examination: "A"

08/1997-05/2002   PhD student at the
Department of Genetics and Neurobiology (Prof. Heisenberg)

March 2000   EMBO Workshop: "Master control genes in Development and Evolution"
Ascona, Switzerland

PhD thesis: "OMB and ORG-1: Homologous *Drosophila*
T-box proteins with functional specificity" (Advisor: Prof. Pflugfelder)

special qualifications:

software: MS-Office, Corel Draw, Photoshop
foreign languages: fluent in English
basic skills in French

hobbies: handball (player, youth team's coach, referee), dancing, outdoor activities

Würzburg, June, 4th 2002   Matthias Porsch

10. List of publications

original papers

Porsch M.; Hofmeyer K.; Bausenwein B; Grimm S.; Weber B.H.F.; Miassod R.; Pflugfelder G.O. (1998): Isolation of a *Drosophila* T-box gene closely related to human *TBX1*. Gene 212: 237-248.

Kretzschmar D.; Poeck B.; Roth H.; Ernst R.; Keller A.; Porsch M.; Strauss R.; Pflugfelder G.O. (2000): Defective pigment granule biogenesis and aberrant behavior by mutations in the *Drosophila* AP-3b adaptin gene *ruby*. Genetics 155: 213-223 .

published meeting abstracts

Porsch M.; Hofmeyer K.; Bausenwein B; Miassod R.; Pflugfelder G.O. (1997): *optomotor-blind-related-gene-1*, a new *Drosophila* T-box gene paralogous to *optomotor-blind*. In: Elsner N.; Waessle H.(eds) Göttingen Neurobiology Report 1997. Georg Thieme Verlag Stuttgart New York: 194

Pflugfelder G.O.; Porsch M.; Trinath T.; Strauss R. (2000): Function of the Drosophila T-domain protein org-1: developmental and behavioural analysis.
Europ. J. Neurosci. 2000 12 (Suppl. 11): 248

papers in preparation

Porsch M.; Bahlo A.; Schulze S.; Pflugfelder G.O. (2002): Mapping determinants of functional specificity in the *Drosophila* T-box proteins OMB and ORG-1.

11. Summary

Members of the T-box gene family encode tran-scription factors that play key roles during embry-onic development and organogenesis of inverte-brates and vertebrates. The defining feature of T-box proteins is an about 200 aa large, conserved DNA binding motif, the T domain. Their importance for proper development is highlighted by the dra-matic phenotypes of T-box mutant animals. My thesis was mainly focused on two *Drosophila* T-box genes, *optomotor-blind* (*omb*) and *optomotor-blind related 1* (*org-1*), and included (i) a genetic analysis of *org-1* and (ii) the identification of molecular de-terminants within OMB and ORG-1 that confer functional specificity.

(i) Genetic analysis of *org-1* initially based on a re-cently isolated behavioral *Drosophila* mutant, *C31*. *C31* is a X-linked, recessive mutant and was defi-ciency-mapped to 7E-F, the cytological region of *org-1*. This pleiotropic mutant is manifested in sev-eral walking defects, structural aberrations in the central brain, and a "held-out" wing posture. Mo-lecular analysis revealed that *C31* contains an in-sertion of a 5' truncated retrotransposable I ele-ment within the 3' untranslated transcript of *org-1*, suggesting that *C31* might represent the first mu-tant *org-1* allele. Based on this hypothesis, we screened about 44.500 F1 female offspring of EMS mutagenized males and *C31* females for the "held-out" wing phenotype, but failed to isolate any *C31* or *org-1* mutant, although this mutagenesis was functional *per se*. Since we could not exclude the possibility that our failure is due to an idiosyncracy of *C31*, we intended not to rely on *C31* in further genetic experiments and followed a reverse genetic strategy which aimed to isolate P element inser-tions within *org-1*.

All available P element lines cytologically mapping to 7E-7F were characterized for their precise inser-tion sites. 13 of the 19 analyzed lines had P ele-ment insertions within a hot-spot about 37 kb downstream of *org-1*. No P element insertions within the *org-1* locus could be identified, but sev-eral P element insertions were determined on ei-ther side of *org-1*. The nearest insertions, 27 kb downstream and 62 kb upstream of the *org-1* tran-script were used for several local-hop experiments, in which we associated 6 new genes with P inser-tions, but failed to target the *org-1* locus. The clos-est P elements are still 10 kb away from *org-1*. Subsequently, we employed *org-1* flanking P ele-ments to induce precise deletions in 7E-F spanning the *org-1* locus.

The *org-1* flanking P elements were brought to-gether on a recombinant chromosome by meiotic recombination. Remobilization of P elements in *cis* configuration frequently results in deletions with the P element insertion sites as deficiency endpoints.

In a first attempt, we expected to identify putative deficiencies by screening for new *C31* alleles. 8 new *C31* alleles could be isolated. The new *C31* chromosomes, however, did not carry the desired deletion. Molecular analysis indicated that *C31* is not caused by aberrations in *org-1*, but by muta-tions in a distal locus, possibly in a transcription unit 80 kb downstream of *org-1*.

We repeated the remobilization of the P elements in the deletion progenitor strain and screened for the absence of P element markers. 4 lethal chro-mosomes could be isolated with a deletion of the *org-1* locus.

(ii) The consequences of ectopic *org-1* were ana-lyzed using UAS-*org-1* transgenic flies and a num-ber of different Gal4 driver lines. Misexpression of *org-1* during imaginal development interfered with the normal development of many organs and re-sulted in flies with a plethora of phenotypes. These include a homeotic transformation of distal antenna (flagellum) into distal leg structures, a strong size reduction of the legs along their proximo-distal axis, and stunted wings. Moreover, the dorsal thorax of *dpp*-Gal4/ UAS-*org-1* flies show a profound, longi-tudinal cleft that separates the anterior scutum medially into two symmetrical halves. The posterior scutum and the scutellum are replaced by a tumor-ous-like outgrowth.

Like ectopic *org-1*, ectopic *omb* leads to dramatic changes of normal developmental pathways in *Drosophila* as well. *dpp*-Gal4/ UAS-*omb* flies are late pupal lethal and show an ectopic pair of wings and largely reduced eyes. Furthermore, ectopic *omb* may result in duplications of distal antennal or leg segments.

GMR-Gal4 driven ectopic *omb* expression in the developing eye causes a degeneration of the pho-toreceptor cells, while GMR-Gal4/ UAS-*org-1* flies have intact eyes.

Hence, ectopic *org-1* and *omb* induce profound phenotypes that are qualitatively different for these homologous genes.

To begin to address the question where within OMB and ORG-1 the specificity determinants re-side, we conceptually subdivided both proteins into three domains and tested the relevance of these domains for functional specificity *in vivo*. The single domains were cloned and used as modules to assemble all possible *omb-org-1* chimeric trans-

genes. A method was developed to determine the relative expression strength of different UAS-transgenes, allowing to compare the various transgenic constructs for qualitative differences only, excluding different transgene quantities. Analysis of chimeric *omb-org-1* transgenes with the GMR-Gal4 driver revealed that all three OMB domains contribute to functional specificity.

12. Zusammenfassung

Die Mitglieder der T-box Genfamilie kodieren Transkriptionsfaktoren mit Schlüsselrollen in der Embryogenese und der Organentwicklung von Invertebraten und Vertebraten. Charakteristisch für T-box Proteine ist der Besitz einer T Domäne, eines ungefähr 200 Aminosäuren großen, homologen DNA Bindungsmotivs. Die Relevanz dieser Proteine in vielen Entwicklungsprozessen zeigt sich deutlich in den dramatischen Phänotypen von Tieren mit Mutationen in T-box Genen. Die vorliegende Arbeit konzentrierte sich vor allem auf das Studium von zwei *Drosophila* T-box Genen, *optomotor-blind* (*omb*) und *optomotor-blind related 1* (*org-1*) und beinhaltet (i) eine genetische Analyse der *org-1* Gens und (ii) die Identifikation der molekularen Determinanten innerhalb OMB und ORG-1, die den verwandten Proteinen ihre funktionelle Spezifität verleihen.

(i) Die genetische Analyse des *org-1* Gens stützte sich anfänglich auf die *Drosophila* Verhaltensmutante *C31*. *C31* ist eine X-gekoppelte, rezessive Mutation und war mittels Defizienzen in den zytologischen Bereich 7E-7F kartiert worden, in dem sich auch *org-1* befindet. Die pleiotrope Mutante *C31* zeigt Defekte im Laufverhalten, strukturelle Veränderungen im Zentralkomplex des Fliegengehirns und eine Flügelfehlstellung. Eine Molekularanalyse ergab, daß *C31* eine Insertion eines 5' verkürzten I Retrotransposons innerhalb des 3' untranslatierten *org-1* Transkripts enthält und ließ vermuten, daß *C31* das erste mutante *org-1* Allel darstellen könnte. Dieser Hypothese folgend durchsuchten wir ca 44.500 F1 Weibchen, die der Kreuzung von EMS mutagenisierten Männchen mit *C31* Weibchen abstammten, auf den *C31* Flügelphänotyp, konnten allerdings keine *org-1* oder *C31* Mutante isolieren, obwohl unsere Mutagenese *per se* funktional war. Da wir nicht ausschließen konnten, daß unser Scheitern durch eine Eigentümlichkeit der *C31* Mutante verursacht wurde, basierten wir weitere genetische Experimente nicht mehr auf *C31* und verfolgten stattdessen eine revers-genetische Stra-

tegie mit dem Ziel, P Element Insertionen im *org-1* Gen zu isolieren.

Alle verfügbaren Fliegenlinien mit P Elementen in 7E-7F wurden molekular charakterisiert und ihre Integrationsstellen präzise bestimmt. 13 der 19 analysierten Linien trugen ihre P Element Insertionen in einem *hot-spot* ungefähr 37 kb distal zu *org-1*. Keine P Element Insertion innerhalb des *org-1* Gens konnte gefunden werden, jedoch wurden mehrere P Elemente auf beiden Seiten von *org-1* identifiziert. Die beiden *org-1* nächsten Insertionen befanden sich 27 kb distal und 62 kb proximal zur *org-1* Transkriptionseinheit und wurden für mehrere *local-hop* Experimente verwendet, in denen wir 6 neue Gene mit P Insertionen assoziieren konnten, jedoch nicht *org-1*. Die *org-1* nähesten P Elemente befinden sich noch ca 10 kb entfernt von *org-1*.

Nachfolgend wurden zwei *org-1* flankierende P Elemente verwendet, um präzise Deletionen über den *org-1* Genlokus zu erzeugen.
Zwei *org-1* flankierende P Elemente wurden zunächst mittels meiotischer Rekombination auf einem Chromosom vereinigt. Die Remobilisierung von P Elementen in *cis* Anordnung führt häufig zu Deletionen mit den P Element Insertionsstellen als Defizienz-Endpunkten. In einem ersten Versuch erwarteten wir mutmaßliche Defizienzen als neue *C31* Allele zu identifizieren. Acht neue *C31* Allele konnten isoliert werden. Zu unserer Überraschung trugen diese neuen *C31* Chromosomen aber nicht die gewünschte Deletion. Weitere molekulare Analysen ergaben, daß *C31* nicht durch Mutationen im *org-1* Gen verursacht wird, sondern durch Mutationen in einem distalen Gen, möglicherweise in einer Transkriptionseinheit 80 kb entfernt von *org-1*.
Wir wiederholten die P Element Remobilisierung, suchten nun aber auf Verlust der P Element-Marker nach Defizienzen. Vier lethale Chromosomen konnten isoliert werden, die eine Deletion über das *org-1* Gen tragen.

(ii) Die Konsequenzen einer ektopischen Expression von *org-1* wurden mit Hilfe von UAS-*org-1* transgenen Fliegen und einer Reihe Gal4 Treiberlinien studiert. Mißexpression von *org-1* während der Imaginalentwicklung stört die normale Entwicklung in vielen Organen und führt zu Fliegen mit einer Vielzahl von Phänotypen. Diese beinhalten eine homeotische Transformation distaler Antennensegmente in distale Beinstrukturen, stark verkürzte Beine und verkrüppelte Flügel. Desweiteren weisen *dpp*-Gal4/ UAS-*org-1* Fliegen eine tiefe Spalte auf dem dorsalen Thorax auf, die das anteriore Scutum in zwei symmetrische Hälften teilt. Ein tumorartiger

Auswuchs ersetzt in diesen Tieren das posteriore Scutum und das Scutellum.

Ebenso wie ektopische *org-1* Expression bewirkt auch die ektopische Expression von *omb* eine dramatische Veränderung des normalen Entwicklungsprogramms. *dpp*-Gal4/ UAS-*omb* Fliegen sind puppal lethal und weisen ein ektopisches Flügelpaar und verkleinerte Augen auf. Zusätzlich führt ektopisches *omb* zu Duplikationen von distalen Antennen- oder Beinsegmenten. GMR-Gal4 getriebene ektopische *omb* Expression in der Augenentwicklung verursacht eine Degeneration der Photorezeptorzellen, während GMR-Gal4/ UAS-*org-1* Tiere intakte Augen besitzen.

Die ektopische Expression von *omb* und *org-1* verursacht also jeweils deutliche, jedoch qualitativ sehr unterschiedliche Phänotypen für die homologen Gene.

Um zu bestimmen, wo sich innerhalb der OMB und ORG-1 Proteine die Spezifitätsdeterminanten befinden, haben wir beide Proteine konzeptionell in drei Domänen unterteilt und die Bedeutung der einzelnen Domänen für funktionelle Spezifität mit Hilfe von chimären *omb-org-1* Transgenen *in vivo* untersucht.

Eine Methode zur Bestimmung der relativen Expressionsstärke von unterschiedlichen UAS-Transgenen wurde etabliert, so daß verschiedene Transgene auf rein qualitative Unterschiede verglichen werden können und sich quantitative Effekte ausschließen lassen. Die Analyse der chimären *omb-org-1* Transgene mit der GMR-Gal4 Treiberlinie ergab, daß alle drei OMB Domänen zur funktionellen Spezifität von OMB beitragen.

13. Erklärung

Erklärungen gemäß § 4 Absatz 3 der Promotionsordnung der Fakultät für Biologie der Bayerischen Julius-Maximilians-Universität Würzburg vom 15. März 1999.

1. Ich erkläre ehrenwörtlich, die vorliegende Dissertation selbständig angefertigt zu haben und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt zu haben.

2. Ich erkläre desweiteren, daß die vorliegende Arbeit weder in gleicher noch ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

3. Ich erkläre weiterhin, daß ich früher keine akademischen Grade erworben habe oder zu erwerben versucht habe.

Würzburg, den 4. Juni 2002   Matthias Porsch

14. Appendix

Matze's Fliegenliste

Matze Stock Box I
org-1,omb, DmVmd2 Genetik

attX C31	GOP stock# 185
w, C31	verifiziert über RFLP analyse
G116	EMS mutante, kann partiell nicht C31 komplementieren
attX G116	s.o.
9-7831/1	Göttinger P-Linie mit 2 P-Element-Insertionen, eine 35 kb downstream org-1
31-2756/1	Göttinger P-Linie mit P-Element-Insertion, ca 35 kb downstream von org-1
GOP-Linie 255	UAS-omb II
GOP-Linie 256	UAS-omb III
EP 3668	P-Element-Insertion im Dm VMD2 Gen
P772	P-Element-Insertion 7 E/F; Blo# : 11739
A16	P772,P774/ FM7c Rekombinationschromosom mit 2 org-1 flankierenden P Elementen
D90	P772,P774/ FM7c Rekombinationschromosom mit 2 org-1 flankierenden P Elementen
G46	P772,P774/ FM7**a** Rekombinationschromosom mit 2 org-1 flankierenden P Elementen
lawc^{P1}	ct^n, lawc^{P1}/ FM7 GOP stock# 779
lawc	lawc^{EF520}/ FM7 GOP stock# 780
D149	Df(1)RA2/ FM7
[MP8	Enhancer Trap, ventral eye expression. Homozygot (II.)]

Matze Stocks A Box
w1118 transformiert mit pUAST EcoRI-HA org-1NTC HA-XbaI

A1	II.	homozygot
A2a	wohl II.	homozygot
A3a	II., rezessiv lethal	über Gla
A3b	III., rezessiv lethal	über TM3
A4a	II.	homozygot
A5a	X	homozygot
A6a	II., rezessiv lethal	über Gla

Matze Stocks B Box
w1118 transformiert mit pUAST EcoRI-corg-1M2-XbaI

B1b	III., rezessiv lethal	über TM3	**watch P loss**
B2a	III.	homozygot	
B3b	II., rezessiv lethal	über Gla	
B4a	II.	homozygot	
B5b	III.	homozygot	
B7b	III.	homozygot	
B8a	II.	über Gla	
B9a	III., rezessiv lethal	über TM1	
B11b	II.	über Gla	
B12b	III., rezessiv lethal	über TM1	

Matze Stock Box IV
Gal4-Treiber

X35		omb Enhancer Trap line GOP stock 82
dpp-Gal4 K54		GOP stock 530
E132-Gal4		GOP stock 502
GMR-Gal4		GOP stock 786
omb-Gal4	y w ombP3/FM7	GOP stock 55
30A-Gal4		GOP stock 567
hs-Gal4 (III)		GOP stock 796
UAS-lacZ (III)		GOP stock 297
twi-Gal4 w;twi-Gal4;twi-Gal4		Michael Bate lab via Bone 31.8.2000 see"twi:a
myogenic switch in Drosophila"(M.Bate)		
Act-Gal4(III)	y,w; Act-Gal4/TM6B, Tb	Bloomington #3954, Yash Hiromi
Act-Gal4(II)	y,w; Act-Gal4/CyO, y+	Bloomington #4414, Yash Hiromi
CyO/II; dpp-Gal4 K54/TM3		GOP stock 530 nun über TM3
II/II; dpp-Gal4 K54/TM3		GOP stock 530 nun über TM3, ohne CyO
chromosom		

Matze Stocks D Box
w1118 transformiert mit pUAST EcoRI-HA org-1N+ombT+org-1C HA-Xbal

D1a	II.	homozygot	
D2b	II.	homozygot	
D3	II.	homozygot	
D4	III., lethal	über TM3	
D5a	III	homozygot	
D6	III	homozygot	
D7	II.	homozygot	**watch P loss**
D8a	III., lethal	über TM3	
D9b	III.	homozygot	
D10b	X	homozygot	**watch P loss**

Matze Stocks E Box
w1118 transformiert mit pUAST EcoRI-HA org-1N+ombT+ombC MYC-Xbal

E1	II.	homozygot
E2a, rot	III., rezessiv lethal	über TM3
E2d, hellorange	III.	über TM3
E2e	II.	homozygot
E3	III.	homozygot
E4b	II.	homozygot

Matze Stocks C Box
w1118 transformiert mit pUAST EcoRI-HA org-1N+org-1T+ombC MYC-Xbal

C1b	III	homozygot
C2a	II., lethal	über Gla
C2b	X	homozygot
C3a	III., lethal	über TM3
C3c	II	homozygot
C4b	III., lethal	über TM3
C5b	III., lethal	über TM3
C6a	III	homozygot

Matze Stocks F Box
w1118 transformiert mit pUAST EcoRI-MYC ombN+T+C MYC-Xbal

F1b	III., lethal	über TM3
F3a	III.	homozygot und über TM3
F5	III., lethal	über TM3
F6b	III.	homozygot
F7b	II., semilethal	über Gla
F9	II., lethal	über Gla
F10b	X, lethal	über FM7c
F11a	III.	homozygot
F13a	II., lethal	über Gla
F14a	II.	homozygot
F14b	II., lethal	über Gla
F15a	II.	homozygot
F15b	II., lethal	über Gla

Matze Stocks G Box
w1118 transformiert mit pUAST EcoRI-MYC ombN+T+org-1C HA-Xbal

G1a	III., lethal	über TM3
G2b	III., semilethal	über TM3
G3a	II., lethal	über Gla
G3b	III., semilethal	ber TM3
G4b	III., semilethal	über TM3
G5a	II., lethal	über Gla
G6	III., semilethal	über TM3
G7a	II.	homozygot
G7b	III.	homozygot
G9a	II., lethal	über Gla
G10	III., lethal	über TM3
G11a	II., lethal	über Gla
G11b	III., lethal	über TM3
G12b	II., semilethal	über Gla
G13a	II., lethal	über Gla
G14b	III., lethal	über TM3
G15a	X	homozygot
G15b	III., lethal	über Gla
G16	III., lethal	über TM3

Matze Stocks H Box
w1118 transformiert mit pUAST EcoRI-MYC ombNTC MYC-XbaI

H1a	II., semilethal	über Gla	
H1d-1	II.	homozygot	
H2a	II., semilethal	über Gla	
H2d-1	X	homozygot	
H2e-2	II.	homozygot	
H3a	II.	homozygot	**watch P loss**
H3d-1	II.	homozygot	
H3d-2	X	homozygot	**watch P loss**

Matze Stocks I Box
w1118 transformiert mit pUAST EcoRI-MYC ombN+org-1T+org-1C HA-XbaI

I1b, rot	III., semilethal	über TM3
I1c, orange	III., semilethal	über TM3
I1d	II., semilethal	über Gla
I1e, gelb	III., semilethal	über TM3
I2b	III., lethal	über TM3
I3a	II.	homozygot
I3b	III., lethal	über TM3
I4	II.	homozygot

Matze Stocks J Box
w1118 transformiert mit pUAST EcoRI-HA org-1N+T+C HA-XbaI

J1a, rot	III.	homozygot
J1c, orange	III.	homozygot
J2a	III., semilethal	über TM3
J2b	III.	homozygot
J3a	II.	homozygot
J4b	X	homozygot
J5	III.	homozygot
J6	X	Stocks etablieren
J7	III., lethal	über TM3

Matze Stocks K Box
w1118 transformiert mit pUAST EcoRI-MYC ombN+org-1T+ombC MYC-XbaI

K1	II.	homozygot
K2b	III., lethal	über TM3
K3a	III., semilethal	über TM3
K3b	II., lethal	über Gla
K4a	III.	über TM3. homozygot muß wiederholt werden
K4b	II., lethal	über Gla
K5	III.	homozygot
K6	III.	homozygot
K8b	III., lethal	über TM3
K9b	III., lethal	über TM3
K10a	III., lethal	über TM3
K11b	III., lethal	über TM3
K12b	III.	homozygot
K13a	III., lethal	über TM3
K13b	X	homozygot
K14b	X	homozygot
K15a	II.	homozygot

Matze Stocks LH Box
Sammlung von relevanten P{lacW} Stämmen aus org-1 local hop Mutagenese

LH-1-26	P{lacW} Insertion in 7E-7F
LH-1-82	P{lacW} Insertion in 7F
LH-1-96	P{lacW} Insertion in 7F-8A
LH-1-131	P{lacW} Insertion in 7E-7F
LH-1-138	P{lacW} Insertion in 7E-7F
LH-1-204	P{lacW} Insertion in 7E
LH-1-213	P{lacW} Insertion in 7C
LH-1-266	P{lacW} Insertion in 7E
LH-1-274	P{lacW} Insertion in 7E-7F
LH-1-464	P{lacW} Insertion in 7E
LH-1-543	P{lacW} Insertion in 7E-7F
LH-1-599	P{lacW} Insertion in 3C (white)

Matze's Oligoliste

name	sequence	length [bp]	purpose	supplier
cosLf	TTC CTG AGG CTG GAC G	16	PCR primer für cos mapping probe	Gibco
cosLr	CGG GTT TTC GCT ATT T	16	PCR primer für cos mapping probe	Gibco
cosRf	CCG CCC GTA ACC TGT C	16	PCR primer für cos mapping probe	Gibco
cosRr	CTG TAA GCG GAT GCC G	16	PCR primer für cos mapping probe	Gibco
GP1-5R	TCC CAG CAG CAG CAA CTC CA	20	seq primer aus Pudong customs sequ. von pcorg-1	Pudong
GP1-1R	TGT TGT TGT TGC TGA TGC TG	20	seq primer aus Pudong customs sequ. von pcorg-1	Pudong
GP1-3U	CAA GCC GCT GCT GTT GTC CA	20	seq primer aus Pudong customs sequ. von pcorg-1	Pudong
GP1-6U	TGC GAT GAA CTG GAA TTG TG	20	seq primer aus Pudong customs sequ. von pcorg-1	Pudong
GP1-4R	CCG CCT GGT GGG CTA GCT GC	20	seq primer aus Pudong customs sequ. von pcorg-1	Pudong
GP1-2U	CAA TAG CCA CCA TTC GCC GT	20	seq primer aus Pudong customs sequ. von pcorg-1	Pudong
GP9-1U	CAT GTG GGT GGG TGG CTG GA	20	seq primer aus Pudong customs sequ. von pcorg-9	Pudong
BB4R1	CTA CGT GCT TTC TGC CCC	18	seq. primer für gen. frag aus org1 region, MP	Gibco
BB4L1	TTG TGC TGT CCC TTG AAC	18	seq. primer für gen. frag aus org1 region, MP	Gibco
BB7R1	TTC AGA ATT TTC AAA GTG CAA	21	seq. primer für gen. frag aus org1 region, MP	Gibco
BB7L1	CCT GCC TCT TCA TCT CCA	18	seq. primer für gen. frag aus org1 region, MP	Gibco
BB4-T7m.rev	TAG TAG TGG TTC TGC CCG	18	seq. primer to cover gap zw. B/B4 und B/B7 in org1-cos3; MP	Gibco
BB7-M13m.rev	GCT AGA CGA AGT AGG TTA	18	seq. primer to cover gap zw. B/B4 und B/B7 in org1-cos3; MP	Gibco
BB4L2	GCA GGC TGC AGT GAA GCA	18	seq. primer für gen. frag aus org1 region, MP	Gibco
BB4R2	TTT AAA GTA AAT CGA GTT TTG	21	seq. primer für gen. frag aus org1 region, MP	Gibco
BB7L2	CCA AAA AGG CCG CGA C	16	seq. primer für gen. frag aus org1 region, MP	Gibco
BB7R2	GAG ATG GAG ATA AAG TGC TAT	21	seq. primer für gen. frag aus org1 region, MP	Gibco
corg-1F	TGT GTA GTC GAC ACC AGC CTA ACG GAC T	28	expression subcloning of org1 cDNA into pET vector	Gibco
corg-1R	TGT GTT GTC GAC CCT TCT CTT CCC AGT T	28	expression subcloning of org1 cDNA into pET vector	Gibco
BB7-R3	TGC GTT TTC AGT TTC ACA	18	seq. primer für gen. frag aus org1 region, MP	Gibco
BB7mL1	CCG TAG CAG GGG GAG G	16	seq. primer für gen. frag aus org1 region, MP (m=Mitte)	Gibco
BB7mR1	TTG GCC GCA CTC ATC A	16	seq. primer für gen. frag aus org1 region, MP (m=Mitte)	Gibco
BB7-R4	AAA CAT TTT TTA ATG GGC GA	20	org-1 Projekt	Gibco
EE1-R1	CAA GCA AGA AGC AGC ACG	19	org-1 Projekt	Gibco
EE1-L1	CTT TAG GTA GAC AGC CGC C	19	org-1 Projekt	Gibco
C31L	AGA AAG GAA GCC TTT AAT TTT CGC C	25	org-1 Projekt	Gibco
C31R	ATA ATA TCC CCG AAA GCC ACA TAG C	25	org-1 Projekt	Gibco
corg1-L1	GAC GAA GTA CCC TGA CTC AG	20	org-1 Projekt	Gibco
BB7-ml11	CGT CGT CAT CGC CCT CCT	18	org-1 Projekt	Gibco
BB7-r5	GTG GCG CCC ATG CCA TAG	18	org-1 Projekt	Gibco
BB7-mr2	CTG CTC CAT CTG GTA GCC	18	org-1 Projekt	Gibco
BB7-mr3	GTT GCT GCT TCA TGC GCG	18	org-1 Projekt	Gibco
C31-5end	AAA AAG ATC GAA TCC TGC ATT TA	23	org-1 Projekt	Gibco
BB7-rev	GGT GAG CTT CAA TTT GTC AA	20	org-1 Projekt	Gibco

name	sequence	length [bp]	purpose	supplier
BB4-for	CAG TGA AGC CCA ATA GCC AT	20	org-1 Projekt	Gibco
corg1-5'end	TCC GCC GGT CAT GTA CGG C	19	seq primer corg1M2, org-1 locus	Gibco
exon5-F	CTC CAG CAA TCC ATT CGC C	19	seq primer org-1 locus	Gibco
exon5-R	TGC CAT CAT CCC GAA AGC C	19	seq primer org-1 locus	Gibco
corg1M2-5'	TTG ATT ATT GCA CCC ACA CAC	21	seq primer org-1 locus	Gibco
C31-5'2	TTT TTC TAG CTT AGG ACG TAA AT	23	seq primer C31 insertion	Gibco
org1exon1-R	CAG CGC TTG ACG GAA CGG AT	20	seq primer org-1 locus	Gibco
org1exon1-L	ATC CGT TCC GTC AAG CGC TG	20	seq primer org-1 locus	Gibco
bb8rev	CCA TGA CGC ACC TGA TGG GC	20	seq primer org-1 locus	Gibco
hstbx1-F	TGG TGG AGG GGA AGG CCG AC	20	PCR hstbx1	Gibco
hstbx1-R	GCT TGT CGA GGA CAC GAT TTG C	23	PCR hstbx1	Gibco
tbx1ex7-F	AGG TCG GGT GGC CCA GGC TGC A	22	PCR amplification human TBX1	Gibco
tbx1ex7-R	AGG CGG ATC AGG GCG GCG CCT G	22	PCR amplification human TBX1	Gibco
tbx1ex6-F	AGC CCC ACC GCT GGA GCT GAT TCC	24	PCR amplification human TBX1	Gibco
tbx1ex6-R	TAC ACC CGC TTT TCC AGA GGC GTT G	25	PCR amplification human TBX1	Gibco
tbx1ex5-F	GCC CTC TGG GTT CAC CTC CAC ATG C	25	PCR amplification human TBX1	Gibco
tbx1ex5-R	ACT CGA GGC CTT GGG GGA CAC CGG	24	PCR amplification human TBX1	Gibco
tbx1ex4-F	AAG GGG GGC TGC CTT CCA CCA GC	23	PCR amplification human TBX1	Gibco
tbx1ex4-R	CGC CAC TTT CCA GGG TGC CCT CC	23	PCR amplification human TBX1	Gibco
tbx1ex3-F	GAG GAG AAA CGC ACG CGG GCG G	22	PCR amplification human TBX1	Gibco
tbx1ex3-R	CAG CCC TGG CGG CAG CAC GTG G	22	PCR amplification human TBX1	Gibco
org1ex8-F	TGA GCA GCA TTC GCT GTA GAT TTA AGC	27	PCR amplification org-1 locus	Gibco
org1ex8-R	ATC GCT TTT CCC CCT GCC ATT TC	23	PCR amplification org-1 locus	Gibco
org1ex7-F	GGC CCG CCA GCT TGT CGG CAT G	22	PCR amplification org-1 locus	Gibco
org1ex7-R	CCA CAT TTT CGC AGA CAA ATG AAA TCG C	28	PCR amplification org-1 locus	Gibco
org1ex3-6-F	CCA CTT TAA CCC GCA CTG TAA CAC CA	26	PCR amplification org-1 locus	Gibco
org1ex3-6-R	GCC GCT TCG AAT GCC AAG ATG AAC	24	PCR amplification org-1 locus	Gibco
org1ex2-F	GTT GGC GTT GGA TTT TCG CAC CAC	24	PCR amplification org-1 locus	Gibco
org1ex2-R	ACT CCG CGG TAG AGT TGC CTA ATC C	25	PCR amplification org-1 locus	Gibco
in1-rev	CGA ACT GAT TAT CCC CGT A	19	seq primer org-1 locus	Gibco
corg-1RHA	TGT GTT GTC GAC TTA GGC GTA GTC TGG GAC GTC GTA TGG GTA GCG CGG CAC CAG ATC T	58	expression cloning of HA-tagged ORG-1	Gibco
in1rev2	GAT CTG ATT TCG AAT GCG	18	seq primer org-1 locus	Gibco
corg1-F2	TGT GTA GTC GAC ACG CAC CTG ATG GGC C	28	expression cloning of full-length ORG-1	Gibco
in1rev3	CCT TGG ATC GCA CCG ACA T	19	seq primer org-1 locus	Gibco
BB4-preF	GTC CGT ATC GCT GCC AAC CG	20	seq primer for org-1 constructs	Gibco
BB7-mL0	CAT GCG TAC CAG CGA CAG GT	20	seq primer for org-1 constructs	Gibco
DOMB-ND-DS	ACA CAG AAT TCA AAT GGA GCA GAA GCT GAT CTC CGA GGA GGA CCT GAA CAG ATA CGA CGT CAG GAG	68	TPSO (T-protein swap oligo) Myc tagged downstream primer (mit EcoRI site) für OMB N-terminus N-domain-downstream	Gibco
DOMB-ND-US	TGT GTG CGG CCG CGG GAT CAT CGA CGA CGC C	31	TPSO (T-protein swap oligo) primer (mit NotI site) für OMB N-terminus N-domain-upstream	Gibco

name	sequence	length [bp]	purpose	supplier
DOMB-TD-DS	ACACA GCG GCC GCT AAG GTC ACG CTG GAG GGC	32	TPSO (T-protein swap oligo) primer (mit NotI site) für OMB T-domain T-domain-downstream	Gibco
DOMB-TD-US	TGTGT GGT ACC GGC ACC AGT ATC ACG CAA	29	TPSO (T-protein swap oligo) primer (mit KpnI site) für OMB T-domain T-domain-upstream	Gibco
DOMB-CD-DS	ACACA GGT ACC GGC AAG CGG GAA AAG AAT	29	TPSO (T-protein swap oligo) primer (mit KpnI site) für OMB C-terminus C-domain-downstream	Gibco
DOMB-CD-US	TGTGTTCTAGA TCA GTT CAG GTC CTC CTC GGA GAT CAG CTT CTG CTC CAT CTG ATC CGT ACC GCC	65	TPSO (T-protein swap oligo) primer (mit Xba site und Myc tag) für OMB C-terminus C-domain-upstream	Gibco
DORG-ND-DS	ACACAGAATTCAAA ATG TAC CCC TAC GAT GTG CCC GAT TAC GCC ACG CAC CTG ATG GGC CCC	62	TPSO (T-protein swap oligo) primer (mit Xba site und Myc tag) für OMB C-terminus C-domain-upstream	Gibco
DORG-ND-US	TGTGTGCGGCCGC GGC CTG GGC CAG CGA TGG	31	TPSO (T-protein swap oligo) primer (mit Xba site und Myc tag) für OMB C-terminus C-domain-upstream	Gibco
DORG-TD-DS	ACACA GCG GCC GCT ATT GTG GTG CTG GAG ACG	32	TPSO (T-protein swap oligo) primer (mit Xba site und Myc tag) für OMB C-terminus C-domain-upstream	Gibco
DORG-TD-US	TGTGT GGT ACC GGT GCC ATC ATC CCG AAA	29	TPSO (T-protein swap oligo) primer (mit Xba site und Myc tag) für OMB C-terminus C-domain-upstream	Gibco
DORG-CD-DS	ACACA GGT ACC AAC GAT GTA ACC ACT GGC	29	TPSO (T-protein swap oligo) primer (mit Xba site und Myc tag) für OMB C-terminus C-domain-upstream	Gibco
DORG-CD-US	TGTGTTCTAGA TCA GGC GTA ATC GGG CAC ATC GTA GGG GTA GCG CGG CAC CAG ATC TAT	59	TPSO (T-protein swap oligo) primer (mit Xba site und Myc tag) für OMB C-terminus C-domain-upstream	Gibco
	GTG AGT GAT TGT GTG TGG GTG C	22	ORG-1 RACE CONTROL für Kontrollamplifikation in org-1 5' RACE	Gibco
org-1back1	GTG AGC ATC ACA TAT CCA GCC	21	org-1 5'RACE primer	Gibco
org-1forward 1	ACG TTA TTC GCG CAG TTG TCA	21	org-1 5'RACE primer	Gibco
org-1back2	TGT GTA GTC GAC ATT CCA ACG ATC ACG CTC GCC	33	org-1 5'RACE primer	Gibco
org-1forward 2	ACA CAT GTC GAC TTG CAC CCA CAC ACA ATC ACT	33	org-1 5'RACE primer	Gibco
Dm-vmd2-r1	TAA ATG TGA AGA GTG GAA CT	20	Seq Primer für vmd2 ESTclones	Gibco
Dm-vmd2-r2	ATG TAT CGA TTG TGA TGA CC	20	Seq Primer für vmd2 ESTclones	Gibco
Dm-vmd2-mr1	TGA GGT CAA CTG GAT GGT GG	20	Seq Primer für vmd2 ESTclones	Gibco
Dm-vmd2-mr2	AGT GGC ATT CAC TTC TCA GC	20	Seq Primer für vmd2 ESTclones	Gibco
Dm-vmd2-ml1	AGT ACG AAT ACA CCG CCA GG	20	Seq Primer für vmd2 ESTclones	Gibco
Dm-vmd2-l1	CTA GGT GTC CTT CAA CTG CC	20	Seq Primer für vmd2 ESTclones	Gibco
Dm-vmd2-r3	CAG TGT GGA CTC CTC ATC AG	20	Seq Primer für vmd2 ESTclones	Gibco
Dm-vmd2-mr3	CGG GAG CTG GAA CCT CTG GA	20	Seq Primer für vmd2 ESTclones	Gibco

name	sequence	length [bp]	purpose	supplier
Dm-vmd2-ml0	ATC ATG CGT CCT CGC TCA TC	20	Seq Primer für vmd2 ESTclones	Gibco
Domb-TD-US2 (s. 304)	TGT GTG GTA CCG GCA CCA GTA TCA CGC AA	29	MP domain swap	MWG
org-1IN1-for	TCT AGG CAT CCA GAG ATC GCA TTC G	25	MP 5´ RACE project	MWG
org-1IN1-rev	AAA TCC AAC GCC AAC CTT TCG CAG T	25	MP 5´ RACE project	MWG
org-1EX1-for	CCT ACG CTC AAT TCT GCG CGG CAT A	25	MP 5´ RACE project	MWG
org-1EX1-rev	CCG CCG AAT GAG AAC AAC TAC TGC T	25	MP 5´ RACE project	MWG
Pry1	CCT TAG CAT GTC CGT GGG GTT TGA AT	26	inverse PCR am 3´ Ende von placW (aus BDGP) (up)	MWG
Pry2	CTT GCC GAC GGG ACC ACC TTA TGT TAT T	28	inverse PCR am 3´ Ende von placW (aus BDGP) (down)	MWG
pUAST-down	5'- AAA TCA ACT GCA ACT ACT GAA -3	21	Primer fuer Seq und PCR von pUAST cloning site	MWG
pUAST-up	5'- TCT CTG TAG GTA GTT TGT CCA -3'	21	Primer fuer Seq und PCR von pUAST cloning site	MWG
Dm-vmd2-r0	5'- TTG TGT AAG AAG TTC GGC GG -3'	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-mr4	5'- CAG CCG AGA GAC AGT GGA GA -3'	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-mr5	5'- TGA AAT CGG AGG ACG CCA TC -3'	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-ml-1	5'- ATC TGA GCA GCA ATT TGA GA -3	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-ml-2	5'- TTG ATT CAC GGC ACG CAA GC -3'	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-l2	5'- TAG GGA CTT GGA GCT CTC GC -3'	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-l3	5'- TCA TTG GCT CTC ATG GAT GT -3'	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-l4	5'- GAT AGA GTT CCC GAA ACG CT -3'	20	Seq Primer Dm-vmd2	MWG
Dm-vmd2-l5	5'- CGA TCA GAT AGG ACA CCT GA -3'	20	Seq Primer Dm-vmd2	MWG
DOMB-TD-US3	5'- TGT GTG GTA CCG GCA CCA GTA TCA CGA AA -3'	29	PCR Primer fuer domain swap project	MWG
corg-1M2-1	5'- GTC CAT CAT GCA AAT GTA GG -3'	20	Seq Primer corg-1M2	MWG
corg-1M2-2	5'- TTG AAG CTC ACC AAT AAC CA -3'	20	Seq Primer corg-1M2	MWG
corg-1M2-3	5'- TCC GGC AAT TCA CCC GAC TT -3'	20	Seq Primer corg-1M2	MWG
corg-1M2-4	5'- TTA CGG TTC GGC TGC ACA TC -3'	20	Seq Primer corg-1M2	MWG
corg-1M2-5	5'- CCG ACG CAT GTT TCC CAC GT-3'	20	Seq Primer corg-1M2	MWG
omb-TD-L	5'- ATT TAG CCT TGG CAT CCA GTC – 3'	21	Seq Primer DSP omb/org-1 chimeric constructs	MWG
org-1-TD-R	5'- ACG AGC ACT CCA GCC ACT TTC – 3'	21	Seq Primer DSP omb/org-1 chimeric constructs	MWG
org-1-TD-L	5'- GCA AAT GTA GGT GGC ATG TGG – 3'	21	Seq Primer DSP omb/org-1 chimeric constructs	MWG

corg-1M2 Primerkarte

```
      org1exon1-R                          ORG-1 RACE CONTROL   ━━
      ━━━━━━━━━━━━━━━━▶
      CCACCAGCGCTTGACGGAACGGATGTTGCACAGTGAGTGAGTGAGTGAGTGAATGAGTGA
   1  ---------+---------+---------+---------+---------+---------+   60
      GGTGGTCGCGAACTGCCTTGCCTACAACGTGTCACTCACTCACTCACTTACTCACT
            ◀━━━━━━━━━━━━━━━        org1exon1-L

      ━━━━━━━━━━━━━━━▶
      GTGATTGTGTGTGGGTGCAATAATCAAATTTGACAACTGCGCGAATAACGTTCGAGATCA
  61  ---------+---------+---------+---------+---------+---------+  120
      CACTAACACACACCCACGTTATTAGTTTAAACTGTTGACGCGCTTATTGCAAGCTCTAGT
          ◀━━━━━━━━  ◀━━━━━━━━━━━━    ◀━━━━━━━━━━━━
           org-1forward2         corg1M2-5'      org-1forward1

      AAACAAATTGTGCCGGAATAAATGGAATAAATGGGCCAAGAACTCGCTGCGATTTCCTTT
 121  ---------+---------+---------+---------+---------+---------+  180
      TTTGTTTAACACGGCCTTATTTACCTTATTTACCCGGTTCTTGAGCGACGCTAAAGGAAA

      org-1back1                              org-1back2
      ━━━━━━━━━━━━━━━━━━▶
      CTTCCACTGTGAGCATCACATATCCAGCCATCATATATATACATATATATATATATTCCA
 181  ---------+---------+---------+---------+---------+---------+  240
      GAAGGTGACACTCGTAGTGTATAGGTCGGTAGTATATATATGTATATATATATATAAGGT
          corg1-F2          DORG-ND-DS
          bb8rev     ━━━━━━━━━━━━━━━━━▶
      ━━━━━━━━━━━━━━━━━━━━▶
      ACGATCACGCTCGCCATGACGCACCTGATGGGCCCCACTGAGTGCGCCGGCGCCATGATG
 241  ---------+---------+---------+---------+---------+---------+  300
      TGCTAGTGCGAGCGGTACTGCGTGGACTACCCGGGGTGACTCACGCGGCCGCGGTACTAC

   1                     M  T  H  L  M  G  P  T  E  C  A  G  A  M  M   15
                 corg-1F
                 ━━━━━━━━━━━━━━━▶
      ACCACCACATCCATGCAGTTCCTGGACACCAGCCTAACGGACTACAACTGCTATGGCAAC
 301  ---------+---------+---------+---------+---------+---------+  360
      TGGTGGTGTAGGTACGTCAAGGACCTGTGGTCGGATTGCCTGATGTTGACGATACCGTTG

  16   T  T  T  S  M  Q  F  L  D  T  S  L  T  D  Y  N  C  Y  G  N   35

      GACTACTGGACATCGCCGTACATGACCGGCGGACTCAGTCCCATGAAGCAGATCGAAGCC
 361  ---------+---------+---------+---------+---------+---------+  420
      CTGATGACCTGTAGCGGCATGTACTGGCCGCCTGAGTCAGGGTACTTCGTCTAGCTTCGG
                        ◀━━━━━━━━━━━━━━━    corg1-5'end
  36   D  Y  W  T  S  P  Y  M  T  G  G  L  S  P  M  K  Q  I  E  A   55

      TGCATCCAAACGGCTGGCAAGGATCGCAGCTCGTACAAGCCGCTGGAGCAGATCGATGCC
 421  ---------+---------+---------+---------+---------+---------+  480
      ACGTAGGTTTGCCGACCGTTCCTAGCGTCGAGCATGTTCGGCGACCTCGTCTAGCTACGG

  56   C  I  Q  T  A  G  K  D  R  S  S  Y  K  P  L  E  Q  I  D  A   75

      AAATTGGCGGACATCGAGACGCACAGTACAGGCAGCACTGGCACCGCGAACAGCAACAGC
 481  ---------+---------+---------+---------+---------+---------+  540
      TTTAACCGCCTGTAGCTCTGCGTGTCATGTCCGTCGTGACCGTGGCGCTTGTCGTTGTCG

  76   K  L  A  D  I  E  T  H  S  T  G  S  T  G  T  A  N  S  N  S   95

      AGCACCAGCAGCATCTCGAATCCCAGTTGCCCGGATCAGTCGTCGTCGTCGTCATCGTCG
 541  ---------+---------+---------+---------+---------+---------+  600
      TCGTGGTCGTCGTAGAGCTTAGGGTCAACGGGCCTAGTCAGCAGCAGCAGCAGTAGCAGC

  96   S  T  S  S  I  S  N  P  S  C  P  D  Q  S  S  S  S  S  S  S  115
```

```
          BB4-preF
        ┌──────────────────────────►
        TCCGTATCGCTGCCAACCGATTATGCCGGCGTACACAGTGAAGCCTCGATGGCACCAACA
601     ---------+---------+---------+---------+---------+---------+     660
        AGGCATAGCGACGGTTGGCTAATACGGCCGCATGTGTCACTTCGGAGCTACCGTGGTTGT

116     S   V   S   L   P   T   D   Y   A   G   V   H   S   E   A   S   M   A   P   T     135


        GCCGGCGGCACGGCAGTCACAACGACATCAGCTGGCGGAGTCAGTGCATCCACCGCGTCC
661     ---------+---------+---------+---------+---------+---------+     720
        CGGCCGCCGTGCCGTCAGTGTTGCTGTAGTCGACCGCCTCAGTCACGTAGGTGGCGCAGG

136     A   G   G   T   A   V   T   T   T   S   A   G   G   V   S   A   S   T   A   S     155


        AAAAAGTTCAAGGGACAGCACAAAAAAGACAACAACAGTGCGGAGAACGGTACAGTGAAG
721     ---------+---------+---------+---------+---------+---------+     780
        TTTTTCAAGTTCCCTGTCGTGTTTTTTCTGTTGTTGTCACGCCTCTTGCCATGTCACTTC
        ◄──────────────────    ┌────────┐
                               │ BB4L1  │
156     K   K   F   K   G   Q   H   K   K   D   N   S   A   E   N   G   T   V   K     175


        CCCAATAGCCATAATATCAGCAAAGGTGAATCGGAGCCAGTGCATCCATCGCTGGCCCAG
781     ---------+---------+---------+---------+---------+---------+     840
        GGGTTATCGGTATTATAGTCGTTTCCACTTAGCCTCGGTCACGTAGGTAGCGACCGGGTC
                                              ◄──────────────────
176     P   N   S   H   N   I   S   K   G   E   S   E   P   V   H   P   S   L   A   Q     195
            ┌────────────┐
            │ DORG-TD-DS │
        ═══════════════════════►   ┌────────────┐
                                   │ OrTDExO-DS │
        GCCATTGTGGTGCTGGAGACGAAGGCGCTGTGGGATCAGTTCCATGCCCAGGGCACCGAA
841     ---------+---------+---------+---------+---------+---------+     900
        CGGTAACACCACGACCTCTGCTTCCGCGACACCCTAGTCAAGGTACGGGTCCCGTGGCTT
                   ┌────────────┐
                   │ DORG-ND-US │
196     A   I   V   V   L   E   T   K   A   L   W   D   Q   F   H   A   Q   G   T   E     215
                      ┌────────────┐
                      │ corg-1M2-5 │──────────────────►
        ATGATCATCACCAAGACGGGCCGACGCATGTTTCCCACGTTTCAGGTGAGGATCGGTGGT
901     ---------+---------+---------+---------+---------+---------+     960
        TACTAGTAGTGGTTCTGCCCGGCTGCGTACAAAGGGTGCAAAGTCCACTCCTAGCCACCA

216     M   I   I   T   K   T   G   R   R   M   F   P   T   F   Q   V   R   I   G   G     235


        TTGGATCCACATGCCACCTACATTTGCATGATGGACTTTGTGCCCATGGATGACAAACGC
961     ---------+---------+---------+---------+---------+---------+     1020
        AACCTAGGTGTACGGTGGATGTAAACGTACTACCTGAAACACGGGTACCTACTGTTTGCG
        ┌──────────┐◄────────◄───────────────    ┌────────────┐
        │org-1-TD-L│                             │ corg-1M2-1 │
236     L   D   P   H   A   T   Y   I   C   M   M   D   F   V   P   M   D   D   K   R     255


        TATCGCTACGCCTTTCACAACTCCTGCTGGGTGGTGGCTGGCAAGGCGGATCCCATTTCC
1021    ---------+---------+---------+---------+---------+---------+     1080
        ATAGCGATGCGGAAAGTGTTGAGGACGACCCACCACCGACCGTTCCGCCTAGGGTAAAGG

256     Y   R   Y   A   F   H   N   S   C   W   V   V   A   G   K   A   D   P   I   S     275


        CCGCCCAGGATTCATGTGCATCCCGACTCGCCAGCCGTCGGCTCCAATTGGATGAAGCAG
1081    ---------+---------+---------+---------+---------+---------+     1140
        GGCGGGTCCTAAGTACACGTAGGGCTGAGCGGTCGGCAGCCGAGGTTAACCTACTTCGTC

276     P   P   R   I   H   V   H   P   D   S   P   A   V   G   S   N   W   M   K   Q     295
                      ┌────────────┐
                      │ corg-1M2-2 │──────────────────►
        ATCGTGTCCTTTGACAAATTGAAGCTCACCAATAACCAGCTGGACGAAAATGGACATATC
1141    ---------+---------+---------+---------+---------+---------+     1200
        TAGCACAGGAAACTGTTTAACTTCGAGTGGTTATTGGTCGACCTGCTTTTACCTGTATAG

296     I   V   S   F   D   K   L   K   L   T   N   N   Q   L   D   E   N   G   H   I     315
```

```
      ATTCTGAACTCCATGCATCGCTACCAGCCGCGTTTCCATCTGGTTTATCTGCCACCGAAG
1201  ---------+---------+---------+---------+---------+---------+  1260
      TAAGACTTGAGGTACGTAGCGATGGTCGGCGCAAAGGTAGACCAAATAGACGGTGGCTTC

 316  I  L  N  S  M  H  R  Y  Q  P  R  F  H  L  V  Y  L  P  P  K   335
```

[org-1-TD-R ──────────▶]

```
      AACGCCTCCTTGGATGAGAACGAGCACTCCAGCCACTTTCGCACTTTCATCTTTCCGGAA
1261  ---------+---------+---------+---------+---------+---------+  1320
      TTGCGGAGGAACCTACTCTTGCTCGTGAGGTCGGTGAAAGCGTGAAAGTAGAAAGGCCTT

 336  N  A  S  L  D  E  N  E  H  S  S  H  F  R  T  F  I  F  P  E   355
```

[exon5-F ───]

```
      ACGAGCTTTACGGCCGTAACTGCCTACCAGAATCAGCGGGTGACACAGCTGAAGATCTCC
1321  ---------+---------+---------+---------+---------+---------+  1380
      TGCTCGAAATGCCGGCATTGACGGATGGTCTTAGTCGCCCACTGTGTCGACTTCTAGAGG

 356  T  S  F  T  A  V  T  A  Y  Q  N  Q  R  V  T  Q  L  K  I  S   375
```

[──────────▶] [DORG-CD-DS] [───────────▶]

```
      AGCAATCCATTCGCCAAAGGCTTTCGGATGATGGCACCAACGATGTAACCACTGGCGGT
1381  ---------+---------+---------+---------+---------+---------+  1440
      TCGTTAGGTAAGCGGTTTCCGAAAGCCCTACTACCGTGGTTGCTACATTGGTGACCGCCA

 376  S  N  P  F  A  K  G  F  R  D  D  G  T  N  D  V  T  T  G  G   395
```

[◀────── exon5-R] [◀── DORG-TD-US] [OrTDExO-US]

```
      GGCAGCAGCATGTCCTCCATGAGTCACGAAAGTCAGGCGCGCATGAAGCAGCAACAGCAG
1441  ---------+---------+---------+---------+---------+---------+  1500
      CCGTCGTCGTACAGGAGGTACTCAGTGCTTTCAGTCCGCGCGTACTTCGTCGTTGTCGTC

 396  G  S  S  M  S  S  M  S  H  E  S  Q  A  R  M  K  Q  Q  Q  Q   415
```

```
      CAACAGCAGCAGCAGCAGCAGCAACTGCAGCAGCAACAGCAACAGCAGCAGCAACTC
1501  ---------+---------+---------+---------+---------+---------+  1560
      GTTGTCGTCGTCGTCGTCGTCGTTGACGTCGTCGTTGTCGTTGTCGTCGTCGTTGAG

 416  Q  Q  Q  Q  Q  Q  Q  Q  Q  L  Q  Q  Q  Q  Q  Q  Q  Q  Q  L   435
```

```
      AAGGAGCGAACGGCAGCAACCAGCAACTTTGGCCTAAGTTGCAGCGAACTGGCCATTGAG
1561  ---------+---------+---------+---------+---------+---------+  1620
      TTCCTCGCTTGCCGTCGTTGGTCGTTGAAACCGGATTCAACGTCGCTTGACCGGTAACTC

 436  K  E  R  T  A  A  T  S  N  F  G  L  S  C  S  E  L  A  I  E   455
```

```
      CAACAGCAGCAGCAGCAACAGCAACAGGGAGTTCTGCAGCTACCGGCCACGCCCTCCAGC
1621  ---------+---------+---------+---------+---------+---------+  1680
      GTTGTCGTCGTCGTCGTTGTCGTTGTCCCTCAAGACGTCGATGGCCGGTGCGGGAGGTCG

 456  Q  Q  Q  Q  Q  Q  Q  Q  Q  G  V  L  Q  L  P  A  T  P  S  S   475
```

[corg-1M2-3 ──────────▶]

```
      AGCTCCACCTCCGGCAATTCACCCGACTTGCTGGGCTACCAGATGGAGCAGCAACTGCAA
1681  ---------+---------+---------+---------+---------+---------+  1740
      TCGAGGTGGAGGCCGTTAAGTGGGCTGAACGACCCGATGGTCTACCTCGTCGTTGACGTT

 476  S  S  T  S  G  N  S  P  D  L  L  G  Y  Q  M  E  Q  Q  L  Q   495
```

[◀────── BB7-mr2]

```
      CAGCAACACCAACAGCAGCAGCAACAGCAACACCAGTCCCAGCAGCAACATCTCCACCAG
1741  ---------+---------+---------+---------+---------+---------+  1800
      GTCGTTGTGGTTGTCGTCGTCGTTGTCGTTGTGGTCAGGGTCGTCGTTGTAGAGGTGGTC

 496  Q  Q  H  Q  Q  Q  Q  Q  Q  Q  H  Q  S  Q  Q  Q  H  L  H  Q   515
```

```
     CAACACCAGGCTAACCAGCAACAATCGCTGCTCCAACAGAGCCAGAATCACACGCAATAT
1801 ---------+---------+---------+---------+---------+---------+ 1860
     GTTGTGGTCCGATTGGTCGTTGTTAGCGACGAGGTTGTCTCGGTCTTAGTGTGCGTTATA

516  Q  H  Q  A  N  Q  Q  Q  S  L  L  Q  Q  S  Q  N  H  T  Q  Y   535
               BB7-mL0
     GGCAGCTATCATCATGCATACCAGGCACAGGTGCAGTCGCATCCCCTAACGCCGCACTCC
1861 ---------+---------+---------+---------+---------+---------+ 1920
     CCGTCGATAGTAGTACGTATGGTCCGTGTCCACGTCAGCGTAGGGGATTGCGGCGTGAGG

536  G  S  Y  H  H  A  Y  Q  A  Q  V  Q  S  H  P  L  T  P  H  S   555

     AGCAGCTCCGCATCCCCGCCAGCAACTGCTGCGCCGGGAGCAAGTGCAGCAACAGCAGCA
1921 ---------+---------+---------+---------+---------+---------+ 1980
     TCGTCGAGGCGTAGGGGCGGTCGTTGACGACGCGGCCCTCGTTCACGTCGTTGTCGTCGT

556  S  S  S  A  S  P  P  A  T  A  A  P  G  A  S  A  A  T  A  A   575
                     BB7mL1
     GTAGCAGCAGCAGCAGCAGCAGCAGCCGTAGCAGGGGGGAGGAGCAGGAGCAGGCGGA
1981 ---------+---------+---------+---------+---------+---------+ 2040
     CATCGTCGTCGTCGTCGTCGTCGTCGGCATCGTCCCCCTCCTCGTCCTCGTCCGCCT

576  V  A  A  A  A  A  A  A  A  A  V  A  G  G  G  A  G  A  G  G   595

     GCAACGTCAGCCACACAAGTGATGAGTGCGGCCAATATCTATTCGAGCATTGGACAACCG
2041 ---------+---------+---------+---------+---------+---------+ 2100
     CGTTGCAGTCGGTGTGTTCACTACTCACGCCGGTTATAGATAAGCTCGTAACCTGTTGGC

                                           BB7mR1
596  A  T  S  A  T  Q  V  M  S  A  A  N  I  Y  S  S  I  G  Q  P   615

     TATGCCCAGGAGCAGAGCAACTTTGGTGCGATCTACCATCATAATGCTGCTGCCGCAGCG
2101 ---------+---------+---------+---------+---------+---------+ 2160
     ATACGGGTCCTCGTCTCGTTGAAACCACGCTAGATGGTAGTATTACGACGACGGCGTCGC

616  Y  A  Q  E  Q  S  N  F  G  A  I  Y  H  H  N  A  A  A  A  A   635

     GCCGCTCACTATCATCATGGTCATGCCCACGGCCACGCCCATAGCCATGCGCACGCCCAC
2161 ---------+---------+---------+---------+---------+---------+ 2220
     CGGCGAGTGATAGTAGTACCAGTACGGGTGCCGGTGCGGGTATCGGTACGCGTGCGGGTG

636  A  A  H  Y  H  H  G  H  A  H  G  H  A  H  S  H  A  H  A  H   655

     GCCCATGGACCATATGCCAGCGCCTACGACAAGCTGAAGGTGTCGCGTCATGCGGCAGCT
2221 ---------+---------+---------+---------+---------+---------+ 2280
     CGGGTACCTGGTATACGGTCGCGGATGCTGTTCGACTTCCACAGCGCAGTACGCCGTCGA

656  A  H  G  P  Y  A  S  A  Y  D  K  L  K  V  S  R  H  A  A  A   675
                           corg-1M2-4
     GCCGCCTATGGCATGGGCGCCACCTATCCAAGTTTTTACGGTTCGGCTGCACATCACCAG
2281 ---------+---------+---------+---------+---------+---------+ 2340
     CGGCGGATACCGTACCCGCGGTGGATAGGTTCAAAAATGCCAAGCCGACGTGTAGTGGTC
                           BB7-r5
676  A  A  Y  G  M  G  A  T  Y  P  S  F  Y  G  S  A  A  H  H  Q   695

     ATGATGCGACCGAATAGCTACATAGATCTGGTGCCGCGCTAAGGGAGCAGCAACTTGGAA
2341 ---------+---------+---------+---------+---------+---------+ 2400
     TACTACGCTGGCTTATCGATGTATCTAGACCACGGCGCGATTCCCTCGTCGTTGAACCTT
                                                       corg-1RHA
696  M  M  R  P  N  S  Y  I  D  L  V  P  R  *                    709
                                               DORG-CD-US
```

```
      GAGAAGGATTTCGGATTTCGGATTTCGGATACTCTATGGAATTAACTGCACTTACACTTG
2401  ---------+---------+---------+---------+---------+---------+  2460
      CTCTTCCTAAAGCCTAAAGCCTAAAGCCTATGAGATACCTTAATTGACGTGAATGTGAAC


      CCTGTAAAAATGATTGTAAAATCCAAACTTAGACTACGTCATCTATAGCCAAAGCTATAC
2461  ---------+---------+---------+---------+---------+---------+  2520
      GGACATTTTTACTAACATTTTAGGTTTGAATCTGATGCAGTAGATATCGGTTTCGATATG


      ATATACATATATGTGTAAATCTCATGCCAAAGATTCGTTCTAAAATCAAGAATCTATTTC
2521  ---------+---------+---------+---------+---------+---------+  2580
      TATATGTATATACACATTTAGAGTACGGTTTCTAAGCAAGATTTTAGTTCTTAGATAAAG


      ┌──────┐ ──────────────────────────────▶
      │ C31L │
      └──────┘
      CAAAGTTTAGAAAGGAAGCCTTTAATTTTCGCCCATTAAAAAAATGTTTTAACAAAACAAA
2581  ---------+---------+---------+---------+---------+---------+  2640
      GTTTCAAATCTTTCCTTCGGAAATTAAAAGCGGGTAATTTTTTACAAAATTGTTTTGTTT
                        ◀────────────────────────  ┌───────┐
                                                    │ BB7-R4│
                                                    └───────┘

      AACATAACTAAGCTTAAGCCAAAACTATAATAACAGGAATTATTTTTTAGCAAGCTTAAT
2641  ---------+---------+---------+---------+---------+---------+  2700
      TTGTATTGATTCGAATTCGGTTTTGATATTATTGTCCTTAATAAAAAATCGTTCGAATTA


      TTTTAAGCATTCAAATTCATTCTTTCGCGAAACATTTGGAATTTGGAGCGATTTGATTCT
2701  ---------+---------+---------+---------+---------+---------+  2760
      AAAATTCGTAAGTTTAAGTAAGAAAGCGCTTTGTAAACCTTAAACCTCGCTAAACTAAGA


      TGATTTTAGAATCAATTTCAAGTATTAGCAGCCAGAAAACCAAAAATAAATGCAACAAGT
2761  ---------+---------+---------+---------+---------+---------+  2820
      ACTAAAATCTTAGTTAAAGTTCATAATCGTCGGTCTTTTGGTTTTTATTTACGTTGTTCA

                                               ┌────────┐ ──────
                                               │ C31-5'2│
                                               └────────┘
      ATTACAAGTATTTCTACATACAAAAATTACCATTAAAAGTTAAAATATTTTTTTTTTTCT
2821  ---------+---------+---------+---------+---------+---------+  2880
      TAATGTTCATAAAGATGTATGTTTTTAATGGTAATTTTCAATTTTATAAAAAAAAAAAGA


      ──────────────────────────▶
      AGCTTAGGACGTAAATTTTATTGATTTGTGTGAAACTGAAAACGCATAAAACATTTCGGT
2881  ---------+---------+---------+---------+---------+---------+  2940
      TCGAATCCTGCATTTAAAATAACTAAACACACTTTGACTTTTGCGTATTTTGTAAAGCCA
                          ◀────────────────────────  ┌───────┐
                                                      │ BB7-R3│
                                                      └───────┘

      GTAAACTGTAGTGTAATTTTAATATACATATTATTATTATTATTTTTTTTTTTTTGCTTAA
2941  ---------+---------+---------+---------+---------+---------+  3000
      CATTTGACATCACATTAAAATTATATGTATAATAATAATAATAAAAAAAAAAAAACGAATT


      CACTCTAGGTTTTTTTTTTCTATGTAAATACAAGTACATATGTATGTCGCTATATATATAT
3001  ---------+---------+---------+---------+---------+---------+  3060
      GTGAGATCCAAAAAAAAAAGATACATTTATGTTCATGTATACATACAGCGATATATATATA
```

```
      ATATATATATATATATTTAAGAACTGCAACAGTTTCAAGCAATAAAAACAAAGAAAATTT
3061  --------+---------+---------+---------+---------+---------+   3120
      TATATATATATATATAAATTCTTGACGTTGTCAAAGTTCGTTATTTTTGTTTCTTTTAAA
```

```
      TAAACCGAAACTCTAGCAAACAGAAGCATAAATTAACCAAAAAAAAAA
3121  --------+---------+---------+---------+-------   3168
      ATTTGGCTTTGAGATCGTTTGTCTTCGTATTTAATTGGTTTTTTTTTT
```