Development 110, 161–171 (1990)
Printed in Great Britain © The Company of Biologists Limited 1990

bicoid mRNA localization signal: phylogenetic conservation of function and RNA secondary structure

PAUL M. MACDONALD

Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA

Summary

Transcripts of the *bicoid* (*bcd*) gene are localized to the anterior pole of the *Drosophila* oocyte, thereby allowing formation in the embryo of an anteroposterior gradient of the *bcd* protein morphogen. We previously showed that a 630 nucleotide portion of the 3′ noncoding region of the *bcd* mRNA is necessary for this localization, and is sufficient to confer anterior localization on a heterologous transcript. Here I have used a comparative analysis to begin to more precisely define the *cis*-acting mRNA localization signal. The *bcd* genes from six additional *Drosophila* species were cloned, and DNA of the 3′ noncoding regions sequenced. Three of these regions were tested interspecifically for mRNA localization in *D. melanogaster* and each functioned correctly; these regions must therefore contain the *cis*-acting signal. The primary sequences, which are up to 50% divergent from the *D. melanogaster* gene, show patchy homology throughout most of the region. Interestingly, all seven species can potentially form a large stereotypic secondary structure. This structure is a likely candidate for the localization signal and can be used for the rational design of mutations to test that possibility.

Key words: mRNA localization, *bicoid*, *Drosophila*, phylogenetic comparison.

Introduction

In *Drosophila*, three separate patterning activities dictate the initial organization of the body plan along the anteroposterior axis. Anterior body parts including head and thoracic regions are under the control of the anterior system, the posterior system is responsible for the abdominal region, and a terminal system specifies the ends of the embryo. Each system includes several components; one is the patterning activity, and many or all of the additional functions serve to localize that activity (reviewed in Nüsslein-Volhard *et al.* 1987). Localization is a critical process since mislocalization of any activity, either anterior, posterior, or terminal, has been shown to reorganize the body pattern (Lehmann and Nüsslein-Volhard, 1986; Frohnhöfer and Nüsslein-Volhard, 1987; Frohnhöfer *et al.* 1986; Klinger *et al.* 1988; Strecker *et al.* 1989; Wharton and Struhl, 1989).

For the anterior system, the patterning activity is the *bicoid* (*bcd*) protein (Frohnhöfer and Nüsslein-Volhard, 1986), which is distributed in a concentration gradient along the anteroposterior axis, peaking at the anterior pole (Driever and Nüsslein-Volhard, 1988*a*). *bcd* protein has sequence-specific DNA-binding activity (Driever and Nüsslein-Volhard, 1989), and activates transcription of subordinate gene(s) in a concentration-dependent fashion (Driever and Nüsslein-Volhard, 1989; Struhl *et al.* 1989; Driever *et al.* 1989). Thus the gradient of *bcd* protein can specify multiple positions along the body axis by specifying different domains of gene activation. Given such a mechanism, formation of the *bcd* protein gradient is of obvious importance. The events underlying *bcd* gradient formation can be divided into two stages. In the first, *bcd* mRNA is deposited specifically at the anterior pole of the oocyte (Frigerio *et al.* 1986) and so can act as a point source for the synthesis of *bcd* protein, which begins only after fertilization, during embryogenesis (Driever and Nüsslein-Volhard, 1988*a*). In the second stage, suitable rates of *bcd* protein synthesis, diffusion from the source and *bcd* protein degradation can collectively lead to formation of the observed gradient (Crick, 1970; Driever and Nüsslein-Volhard, 1988*a*). These latter events are presumably under the control of general cellular constraints and mechanisms, since all mutants known to affect *bcd* localization specifically do so at the level of mRNA distribution (Berleth *et al.* 1988; Stephenson *et al.* 1988).

One mechanism for localizing *bcd* mRNA at the anterior end of the oocyte is suggested by the spatial organization within the ovarioles. Each egg chamber contains a cluster of nurse cells connected by cytoplasmic bridges to the anterior pole of a single oocyte (King, 1970). Thus, after synthesis in the nurse cells, movement of *bcd* mRNA into the oocyte could allow for anterior localization *via* trapping by a receptor at the site of entry (Berleth *et al.* 1988). This receptor need not be localized to a region within the oocyte, and must

only be absent from the nurse cells to avoid premature anchoring. The receptor could bind directly to *bcd* mRNA, or alternatively, to a specific RNP complex preformed in the nurse cells by association of another factor with *bcd* mRNA. In either case, the *bcd* mRNA must at some point be recognized by the localization machinery, suggesting the existence of a binding site or 'localization signal' within the mRNA. Indeed, we previously demonstrated that a large portion of the 3′ noncoding region of the *bcd* mRNA was necessary for its localization, and was sufficient for anterior localization of a heterologous mRNA. Removal of 100–150 nucleotides (nt) from either end of this 630 nt segment eliminated localization activity. These observations suggested that the localization signal might consist of two or more discrete elements located within the large region, or alternatively that the entire region was required, perhaps in the form of a large RNA secondary structure (Macdonald and Struhl, 1988).

The goal of this work is to more precisely define the localization signal. One general approach to this type of problem is to systematically mutate the DNA *in vitro* either in short regions (linker-scanning; McKnight and Kingsbury, 1982) or at single nucleotides (Myers *et al.* 1986), and then test each mutant in a functional assay. Factors that weigh against the use of this method here are the large size of the *bcd* mRNA localization signal, which would require the construction and testing of many mutants, and the considerable time and effort required for the functional assay of each of the many mutants. A different approach is to identify related sequences that perform the same function, and ask what features are shared. The importance of common sequences or structures could then be tested in subsequent experiments by the creation and functional assays of a more manageable number of specific mutations. One way to obtain the necessary related sequences follows from the observation that some *cis*-acting elements in DNA and RNA molecules have been conserved during evolution, and display activity when tested for function interspecifically (Nomura *et al.* 1968; Corces *et al.* 1981; Mitsialis and Kafatos, 1985; Fischer and Maniatis, 1986). Taking this approach, I have isolated and analyzed the *bcd* genes from six additional *Drosophila* species. I find that each of several 3′ noncoding regions tested for localization activity interspecifically in *D. melanogaster* functions correctly. Interestingly, each has the potential to form an extensive, stereotypical secondary structure. Taken together with our earlier results, these observations are consistent with the notion that the *cis*-acting element responsible for anterior localization of *bcd* mRNA is, or is a part of, the secondary structure.

Materials and methods

Isolation and characterization of bcd *genes from various* Drosophila *species*

Lambda phage libraries of genomic DNA from *D. heteroneura* and *D. picticornis* were provided by John Hunt, and a genomic *D. virilis* library was provided by Ron Blackman. To construct the *D. simulans*, *D. teissieri* and *D. sechellia* libraries, genomic DNA of each species (a gift from Gail Simmons) was partially digested with *Sau*3A, 10–15 kb fragments were excised from an agarose gel, ligated to *Bam*HI-digested DASH (Stratagene) and packaged into phage particles *in vitro*. All libraries were screened at reduced stringency (McGinnis *et al.* 1984) with a ^{32}P labeled *Bgl*II–*Eco*RV restriction fragment from the *D. melanogaster bcd* gene (nucleotides 3726–4098 in Berleth *et al.* 1988). Phage isolated in this screen were characterized by standard methods, and restriction fragments were subcloned into pEMBL vectors (Dente *et al.* 1983) for standard dideoxy sequencing. All sequences have been submitted to the GenBank database. Accession numbers are: *D. teissieri*, M32121; *D. virilis*, M32122; *D. simulans*, M32123; *D. sechellia*, M32124; *D. heteroneura*, M32125; *D. picticornis*, M32126.

RNA structure predictions

Structures were predicted using the computer program of Zuker (1989), following the instructions of Jaeger *et al.* (1990). In general, the 5 most optimal structures within 10 % of the minimal free energy were analyzed. During initial predictions, the helix III structure was consistently present and was sometimes removed from the sequences to simplify the analysis. To prevent the appearance of only slightly different structures, the distance value was varied, usually at 3 % to 5 % of total sequence length. The many structural motifs were compared manually to identify those present in all species.

Construction and analysis of modified and hybrid bcd *genes*

Hybrid and modified *bcd* genes are all derivatives of construct BBB of Macdonald and Struhl (1988). BBB consists of an 8 kb genomic *Eco*RI fragment inserted into a modified C20 transformation vector (Rubin and Spradling, 1983). Upon introduction into the germline, BBB displays full wild-type *bcd* activity and a single copy is sufficient to rescue bcd^- flies. Hybrid genes were constructed by replacing an *Eco*RV–*Xba*I restriction fragment from the 3′ noncoding region of the *D. melanogaster* gene (nucleotides 4098–4883 in Berleth *et al.* 1988) with equivalent fragments from the *D. simulans*, *D. teissieri* and *D. virilis* genes; these hybrid genes (and flies transformed with the genes) are designated 1257, 1259, and 1258, respectively. The modified *bcd* gene is construct BBT of Macdonald and Struhl (1988), in which an *Mlu*I–*Stu*I restriction fragment from the 3′ noncoding region of the *D. melanogaster* gene (nucleotides 4105–4730 in Berleth *et al.* 1988) has been replaced by a *D. melanogaster* tubulin a1 gene fragment containing the polyadenylation signal (Theurkauf *et al.* 1986).

These genes were introduced into the *D. melanogaster* genome by P element-mediated transformation (Rubin and Spradling, 1982). Balanced stocks were generated and transformant lines with second chromosome insertions (1257–2, 1258–3 and 1259–2) were used to construct stocks heterozygous for both bcd^{E1} (on the third chromosome) and the transgene. From these stocks animals homozygous for the bcd^{E1} mutation were collected and placed in cages. Embryos were harvested and *bcd* and *hb* proteins detected by immunohistochemical staining (Macdonald and Struhl, 1986) using rat anti-*hb* (Struhl *et al.* 1989) and rat anti-*bcd* antisera

Results

bicoid *genes from various* Drosophila *species*

Seven *Drosophila* species were chosen for comparison. The *D. melanogaster bcd* gene has already been cloned (Frigerio *et al.* 1986) and characterized structurally (Berleth *et al.* 1988). Of the additional six, *D. simulans*, *D. sechellia* and *D. teissieri* provide examples of species closely related to *D. melanogaster*, having diverged between 2.5–17 million years ago (Cariou, 1987). The more distantly related examples were *D. virilis* and two Hawaiian species, *D. heteroneura* and *D. picticornis* (all diverged from *D. melanogaster* 40–60 million years ago (Beverley and Wilson, 1984; 1985)). The *bcd* gene from each of the six additional *Drosophila* species was isolated by screening genomic phage libraries as described in Materials and methods. For each gene, the general organization was established by hybridization to various portions of the *D. melanogaster* gene, and the segment containing the carboxyterminal coding region and flanking 3′ noncoding region was sequenced. The aligned sequences are presented in Fig. 1, with the more closely related genes grouped together to facilitate presentation and discussion of RNA secondary structure (below).

For all genes the DNA sequence of the carboxyterminal coding region is highly conserved, and the deduced protein sequences are almost identical, showing only conservative substitutions and in one species the deletion of a single amino acid. The sequences diverge immediately after the stop codon, with only slight similarity to the *D. melanogaster* sequence retained among the distantly related Hawaiian species and *D. virilis*. After an AT-rich 'spacer' region present only in the three distant species, regions of conserved sequence reappear. These form a pattern repeated over much of the next 600 nt, in which small blocks of sequences shared by most or all species are interspersed with regions of dissimilarity (within the dissimilar regions the closely related species are frequently identical). In the region corresponding to the last 150 nt of the *D. melanogaster* mRNA, the sequences diverge enough to preclude any meaningful alignment of genes of all species, except for one short segment where 13 of 14 nt are highly conserved.

Conservation of bcd *mRNA localization function*

To test for conservation of localization function, the 3′ noncoding regions from the *D. simulans*, *D. teissieri* and *D. virilis bcd* genes were incorporated into hybrid *bcd* genes, introduced into the *D. melanogaster* germ line by P element transformation, and assayed for mRNA localization activity. These three species span a broad range of evolutionary distance from *D. melanogaster*. In each hybrid gene, an otherwise normal *D. melanogaster bcd* gene is altered by removing a restriction fragment containing the segment previously shown to be both necessary for localization of the *bcd* mRNA and sufficient to confer anterior localization on a heterologous mRNA (Macdonald and Struhl, 1988), and replacing it with the equivalent *bcd* gene restriction fragment from one of the other species (see Fig. 1). Thus all hybrid genes encode a normal *D. melanogaster bcd* protein. For the localization assay, one or two copies of the hybrid genes were introduced into bcd^{E1} flies by appropriate genetic crosses, and the patterns of *bcd* and *hunchback* (*hb*) proteins were visualized by immunohistochemistry. The bcd^{E1} allele lacks all *bcd* function, and makes no detectable *bcd* antigen (although it, like all other published *bcd* alleles, makes normal levels of mRNA). Thus the hybrid transgene is the only source of functional and seropositive *bcd* product. Since the *bcd* protein gradient requires localization of *bcd* mRNA to form its normal pattern (Berleth *et al.* 1988; Driever and Nüsslein-Volhard, 1988*b*), gradient formation serves as a measure of RNA localization. *hb* protein was also visualized. Zygotic expression of *hb* in a broad anterior zone is under the direct control of *bcd* (Tautz, 1988; Driever and Nüsslein-Volhard, 1989; Struhl *et al.* 1989; Driever *et al.* 1989). Because the posterior boundary of this zone shifts in a measurable fashion in response to changes in the distribution of *bcd* (Struhl *et al.* 1989), it provides a sensitive assay for correct localization of *bcd*.

The results of the assays are shown in Fig. 2. Wild-type embryos display the *bcd* and *hb* protein patterns due to normal *bcd* mRNA localization. All of the hybrid *bcd* genes show *bcd* (Fig. 2C,E) and *hb* (Fig. 2D,F) protein patterns that are almost identical to wild type, indicating that each retains normal or nearly normal mRNA localization function (data not shown for *D. simulans*). In BBT; bcd^{-} embryos, which express a *bcd* gene lacking the localization signal (Macdonald and Struhl, 1988), the *bcd* protein is no longer localized (Fig. 2G) and *hb* expression is correspondingly altered (Fig. 2H). For completeness, the hybrid genes were also tested for their ability to rescue the bcd^{-} lethality (the regulative properties of the embryo (see for example Driever and Nüsslein-Volhard, 1988*b*) make this assay much less sensitive than those used above). Not surprisingly, a single copy of any of the hybrid *bcd* genes is sufficient to rescue bcd^{-} embryos, while the BBT gene fails to rescue.

Conservation of bcd *mRNA secondary structure among all species examined*

Conservation of mRNA localization function must be due to the conservation of a *cis*-acting element in the *bcd* 3′ noncoding region. This element could consist of two or more short conserved sequences, of which there are many (Fig. 1); it could be a single large folded RNA structure; or it could be one or more small recognition domains whose presentation depends on the formation of a particular folded structure. Both of the latter two models require that the *bcd* mRNA 3′ noncoding regions of the different species (at least those shown to be functional in *D. melanogaster*) be able to form a common secondary structure. To consider this possibility, the RNA folding program of Zuker (1989) was applied independently to each sequence. On the basis of minimization of free energy of folding, this program predicts both optimal and suboptimal secondary struc-

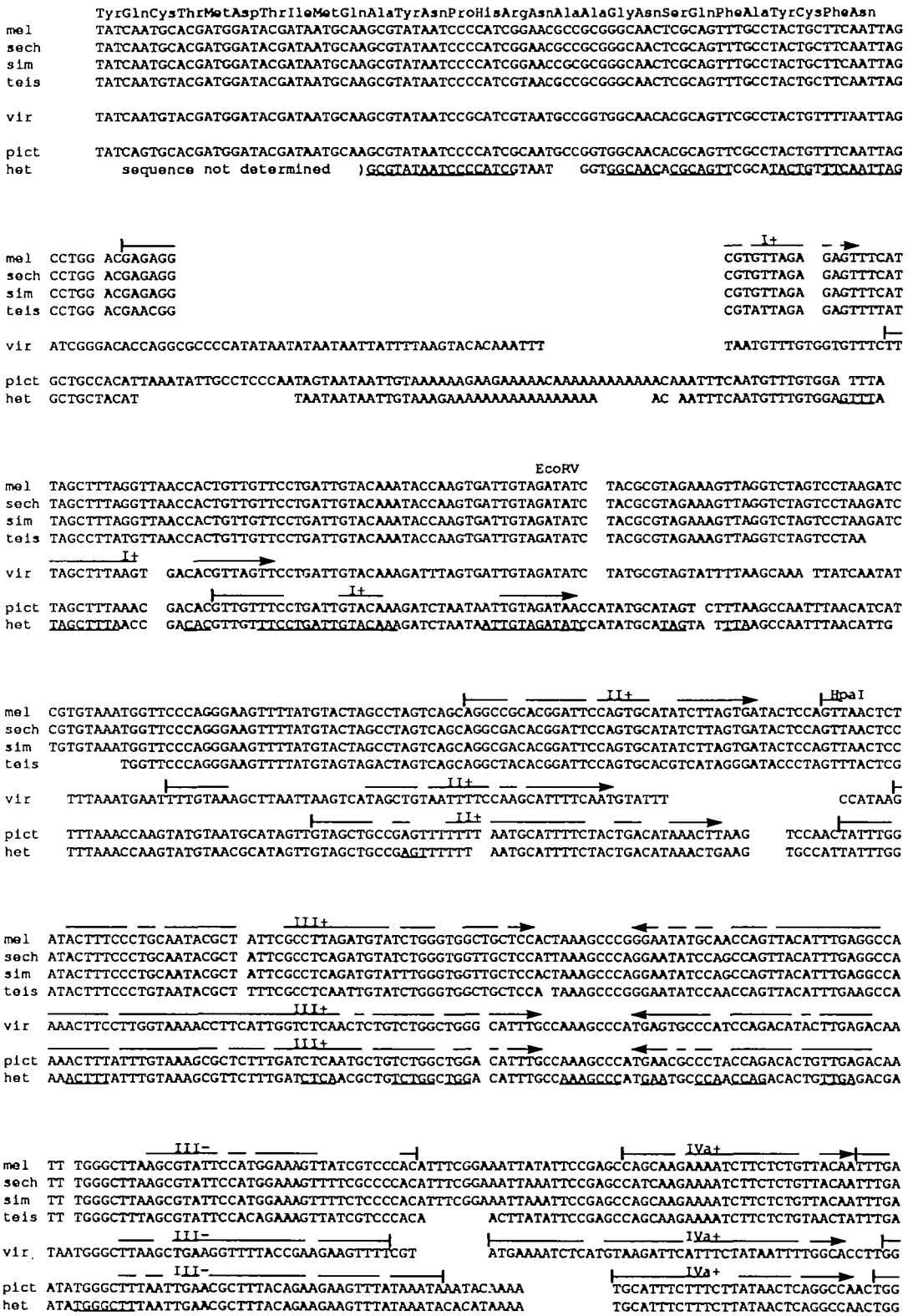
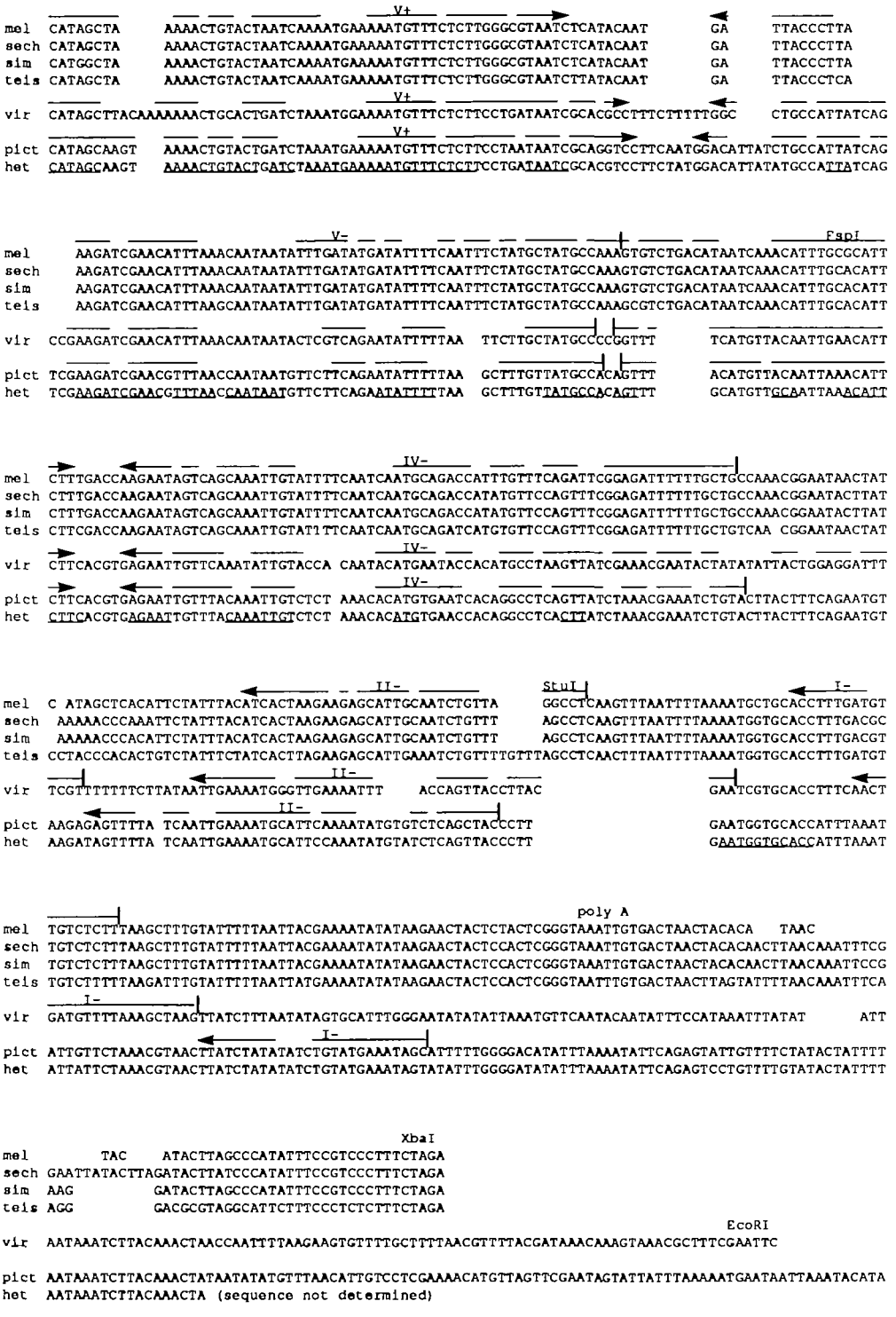


Fig. 1. Aligned DNA sequences from the *bcd* genes of different *Drosophila* species. The sequences include the last 87 nt of the coding region and extend just past the position where the *D. melanogaster* mRNA is polyadenylated. Species are arranged with closely related species in separate tiers. Sequences were initially aligned with the Gap program of the University of Wisconsin Genetics Computer Group software package (Devereux *et al.* 1984). As this program does not allow simultaneous alignment of multiple sequences, the individual alignments were assembled and manually adjusted to provide



the best overall fit (a better alignment may exist). Blocks (3 or more continuous nt) of highly conserved (identity in 6 of 7 genes) sequences are underlined below the entire group. Interrupted arrows above the individual tiers indicate paired nucleotides present in each designated stem of the predicted RNA secondary structures (Fig. 3). Selected restriction sites present in the *D. melanogaster* sequence are indicated. Notable domains are: the smallest fragment with ability to confer anterior localization on a heterologous mRNA in *D. melanogaster* embryos, *Eco*RV–*Xba*I; deletions into that fragment that abolish localization activity, from the 5′ side to the *Hpa*I site and from the 3′ side to the *Fsp*I site; fragments removed/inserted in the construction of interspecies hybrid *bcd* genes, *Eco*RV–*Xba*I (the *D. virilis Eco*RI site was used in that hybrid). Abbreviations: mel, *melanogaster*; sech, *sechellia*; sim, *simulans*; teis, *teissieri*; vir, *virilis*; pict, *picticornis*; het, *heteroneura*.

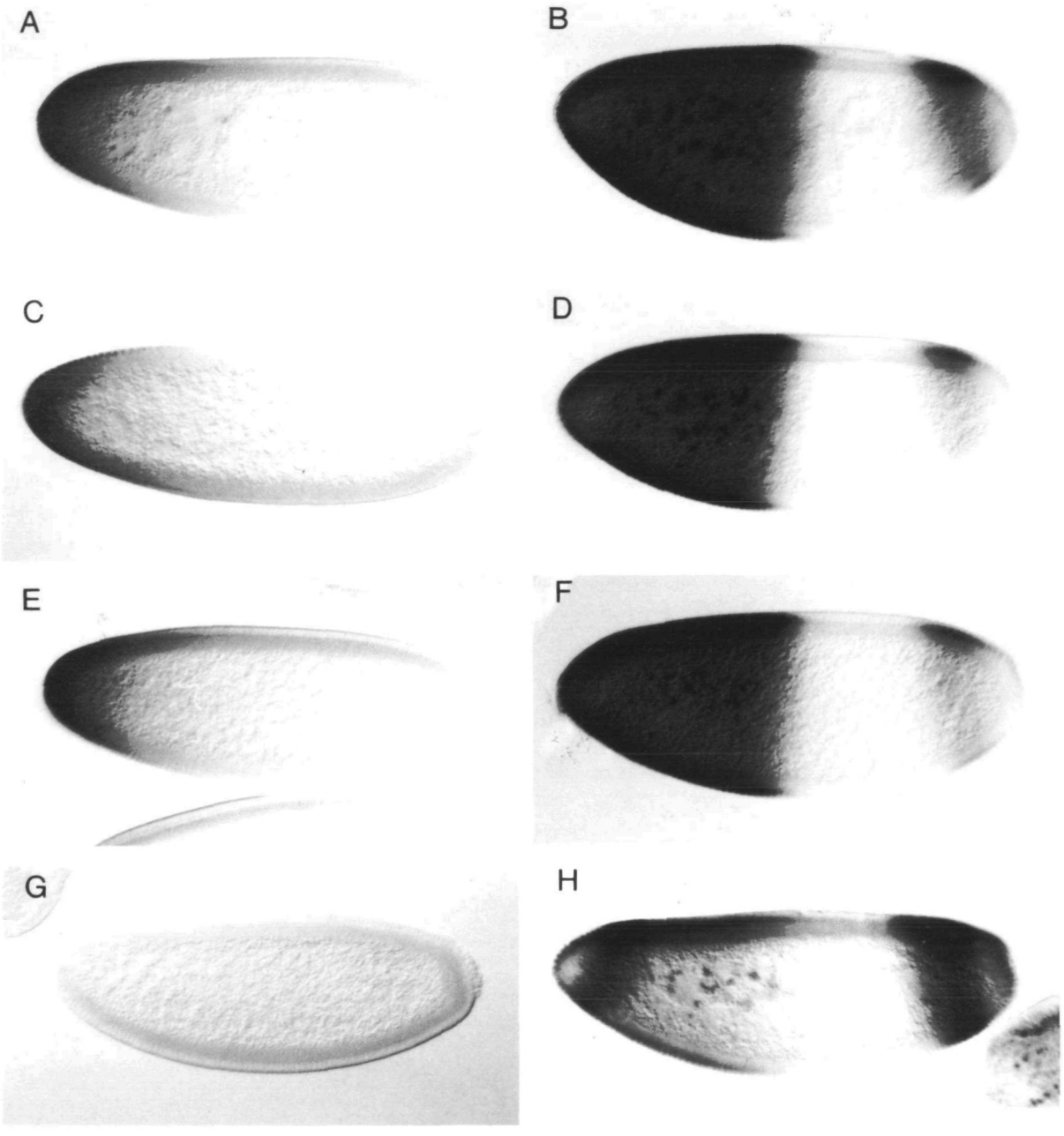


Fig. 2. Phylogenetic conservation of the localization signal in *bcd* mRNAs of other *Drosophila* species. Embryos in the left column (A,C,E,G) show the distribution of *bcd* protein, and embryos in the right column (B,D,F,H) show the anterior zone of *hb* protein expression. (A and B) wild-type (ry^{506}; the host for transformation) embryos. (C and D, E and F) Embryos from bcd^{-} mothers expressing hybrid bcd^{+} genes; localization signals are from *D. teissieri* and *D. virilis*, respectively. Both *bcd* and *hb* patterns are nearly identical to wild-type (the posterior boundary of the anterior *hb* expression domain is shifted slightly to the anterior (3–5%) in embryos expressing the hybrid genes). (G and H) Embryos from bcd^{-} mothers expressing the BBT gene, which has wild-type *bcd* coding sequences but lacks the localization signal. The *bcd* protein cannot be detected with the antibody (G); the BBT mRNA is dispersed through the embryo (Macdonald and Struhl, 1988) and so the protein is not expected to be localized. The low level of *bcd* protein is insufficient to activate *hb* expression in the normal pattern. Instead, the duplication of posterior *hb* expression which occurs in bcd^{-} embryos is seen (H). The pattern of *hb* expression is very sensitive to the copy number of the BBT gene; two copies are frequently enough to establish a novel situation, in which the duplicated posterior zone of *hb* expression is superimposed on a broad but weak anterior zone of *hb*. It is not known which of the *hb* mRNAs, maternal or zygotic, is responsible for this anterior zone of expression.

tures, thereby allowing multiple folding motifs to be generated. These motifs were then compared to see if similar structures could form by pairing of equivalent regions of the 3′ noncoding regions of each species (as would be expected if they were derived from a common ancestor). A single stereotypic structure emerged from

such a comparison, and it differs considerably from the structure previously predicted from the *D. melanogaster* sequence alone (Macdonald and Struhl, 1988). The *D. melanogaster* and *D. virilis* versions of the current structure are presented in Fig. 3. In addition, all the major stems are indicated above the sequences of each group of closely related species in Fig. 1.

This *bcd* RNA structure is conveniently divided into six sections, including five stems and one variable region, as indicated in Fig. 3. Structures were also predicted for each variable region, but because analysis of the folding possibilities for the seven species has not yet revealed a single shared pattern, none is shown. As seen in Fig. 1, stem-loop structures III, IV and V of all species form by pairing of aligned (and similar) sequences, while the individual arms of stems I and II are not aligned in the primary sequences. For many of the stems of the closely related species there are both the compensatory changes typically found in comparisons of highly conserved RNA secondary structures (e.g. Noller and Woese, 1981; James *et al.* 1988), as well as changes that create or disrupt a base pair. Thus any evolutionary support that the stereotypic structure does form *in vivo* comes primarily from its persistent presence as a complete unit, rather than from strong support for each individual stem-loop. Additional evidence favoring the structure is provided by the locations of several small gaps or insertions in the *D. teissieri* gene, relative to the *D. melanogaster* gene. None of the three changes significantly alters the overall structure. One gap (12 nt) is within the variable region, one (7 nt) lies between stems III and IV, and the insertion (5 nt) enlarges a nucleotide bulge at the base of stem II.

Notably, the structure common to all species encompasses most of the 3′ noncoding portion of the *bcd* mRNA and is largely contained within the *D. melanogaster* segment previously shown to confer localization activity to a heterologous mRNA. As indicated in Fig. 1, the smallest segment that retained this activity includes all of stems II–V, and the loss of stem I does not markedly alter the rest of the structure. However, the partial deletions that eliminate localization activity both extend further into the structure. One removes all of the one strand of stem II as well as almost all of one strand of stem IV, while the other removes all of the other strand of stem II. This good correlation between localization function and integrity of the secondary structure, taken together with the possibility of forming a similar structure in all species examined, supports models of localization signal organization in which a secondary structure is involved.

Discussion

The initial anteroposterior organization of the *Drosophila* body plan arises from the actions of three separate patterning systems, each with a localized activity. For the anterior system, this activity is the *bcd* protein, whose correct localization requires prelocalization of *bcd* mRNA at the anterior pole of the embryo. We previously demonstrated that a *cis*-acting element of the *D. melanogaster bcd* mRNA was responsible for anterior localization, and that localization activity was contained within a 630 nt segment encompassing most of the 3′ noncoding region. Here I describe the isolation of the *bcd* genes from several additional *Drosophila* species, and show that the 3′ noncoding regions from each of three such genes tested can localize the *bcd* mRNA in *D. melanogaster*. Since these regions must therefore contain the localization signal, they can be used to identify conserved sequences or structures whose role in localization may subsequently be tested by the construction and functional analysis of appropriate mutants.

Primary sequences that are highly conserved among all the various *bcd* 3′ noncoding regions have been indicated in Fig. 1; limiting the comparison to only those genes demonstrated to have localization activity in *D. melanogaster* does not markedly change the overall picture. Most of the highly conserved sequences lie between the *Hpa*I and *Fsp*I restriction sites. Additional sequences are required for localization (from *Eco*RV to *Hpa*I, and from *Fsp*I to *Stu*I; Macdonald and Struhl, 1988), suggesting that the necessary level of sequence conservation is low, or that these regions may serve a general role with limited sequence constraints (as suggested below). One region of high sequence conservation, just upstream from the *Eco*RV site, is outside of the minimal localization element. This region may be conserved for reasons not involving mRNA localization. Alternatively, this region may be essential for correct mRNA localization, but since our previous assay for mRNA localization was not as sensitive as that used here, a slight relaxation in the degree of localization would not have been detected. Note that the *D. melanogaster* version of this region was retained in all of the hybrid genes, and so its importance was not addressed in the experiments with these genes.

In addition to the sequence conservation, a large stereotypical secondary structure appears to have been conserved, and can be formed by the *bcd* mRNA 3′ noncoding regions of all species examined. Although the proposed structure is among the most energetically favorable, its assignment must remain tentative. Elucidation of RNA structure by comparative analysis requires a much larger number of examples, and should be supported by other types of evidence (Fox and Woese, 1975; Noller and Woese, 1981; Gutell *et al.* 1985; Pace *et al.* 1989). Nevertheless, the large size of the localization signal, the possibility of forming an evolutionarily conserved secondary structure, and the behavior of deletions mutants together support assignment of localization signal activity to such a structure. Moreover, the proposed structure suggests a model for organization of the localization signal that is consistent with all available data.

In this model, the secondary structure can be viewed as consisting of two general domains. One serves to form a particular recognition site for interactions with a binding protein (or proteins), and the other facilitates the correct folding of the recognition domain. The

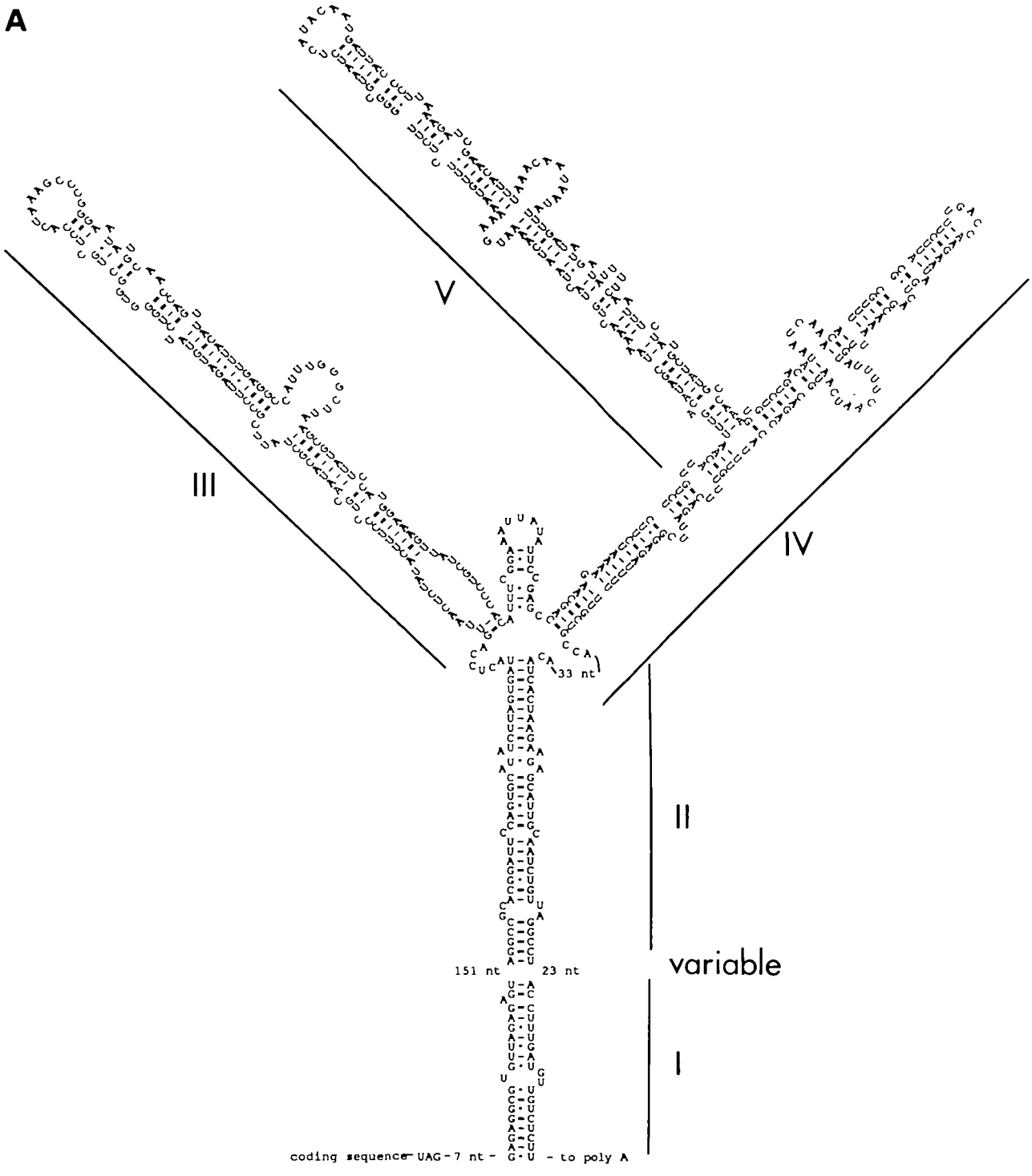
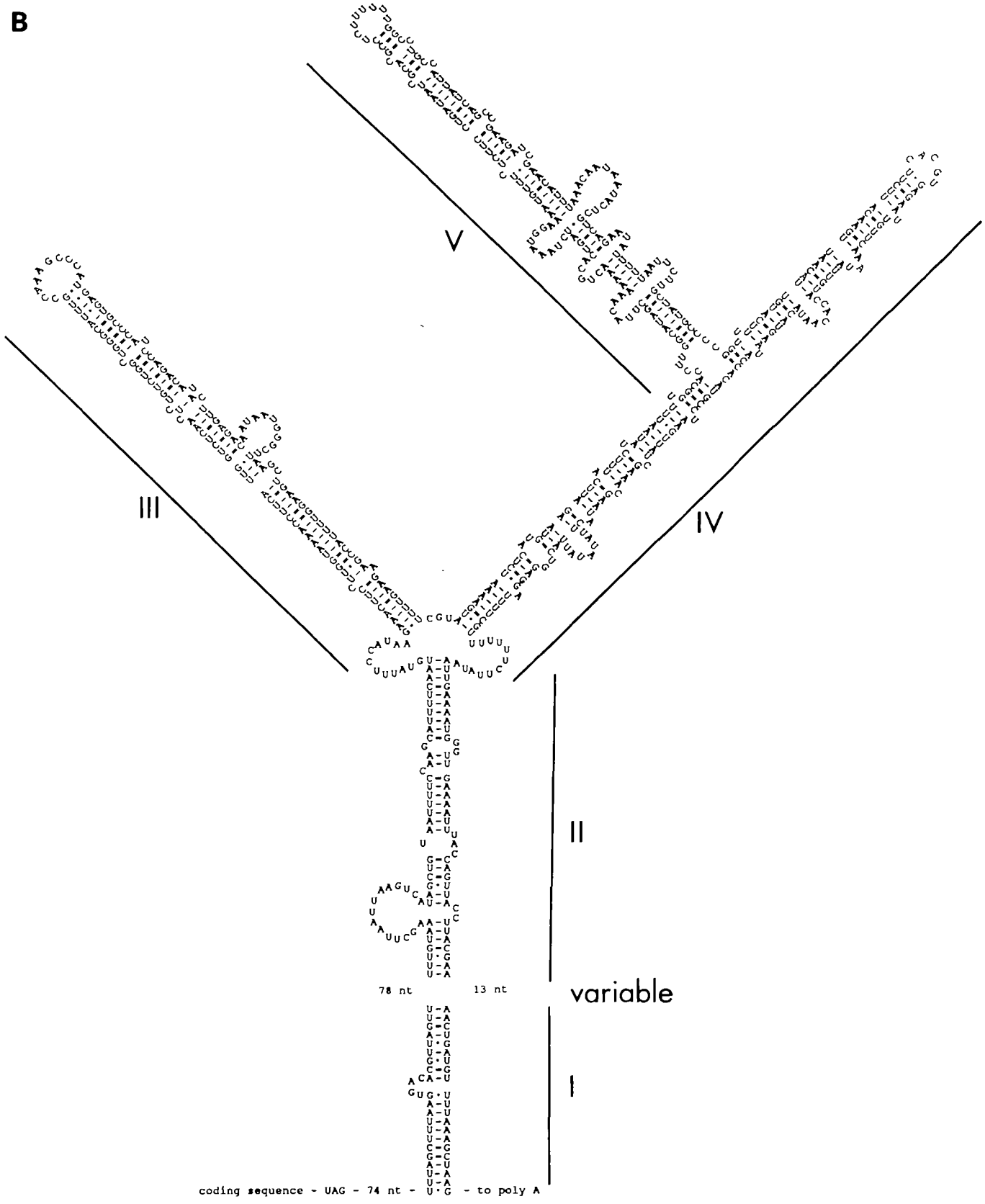


Fig. 3. Predicted stereotypic secondary structures for the *D. melanogaster* and *D. virilis* 3′ *bcd* regions. (A) The *D. melanogaster* structure. (B) The *D. virilis* structure. The structures were determined as described in Materials and methods. For each, the sequence abuts the coding region at the lower left and extends towards the 3′ end of the mRNA on the lower right. The five major stems are labeled (I–V), as is the variable region. To facilitate comparison among all species, the position of each stem is also indicated above each related group of sequences in Fig. 1. Stem III is the most notable element of the structure in the sense that it is absent only in structures so suboptimal that they are unlikely to exist. Curiously, the terminal and side loops of stem III are complementary, and this potential for base pairing is conserved in all species examined.


B
V
III
IV
II
variable
I
78 nt
13 nt
coding sequence - UAG - 74 nt - U·G - to poly A

recognition domain should be highly conserved (as specificity of binding is necessary) and therefore is likely to consist of some or all of the most highly conserved regions, namely stem-loop structures III, IV and V. The terminal and side loops of stem III are appealing candidates for binding sites as both contain nucleotides that are invariant among the seven species. Given the existence of such a binding site, the remainder of the structure might be conserved simply to prevent disruption of the correct folding of the appropriate stem-loop structure(s). For example, an isolated stem-loop containing a binding site might be relatively susceptible to the encroachments of competing secondary structures or to melting by the processive invasion of a single-stranded RNA-binding protein. In contrast, the same stem-loop embedded within a larger structure would be resistant to such attacks. The composition of these flanking structural domains would not need to be highly conserved, and stems I and II and the variable region (or some other folding of those regions) could be present to provide this function. Although this model is speculative, it does make several predictions about the localization signal that can be tested by the construction of specific mutations. Analysis of those mutations, together with a more extensive phylogenetic comparison, should provide a much clearer picture of how anterior localization of *bcd* mRNA occurs.

Ron Blackman, John Hunt, and Gail Simmons generously provided libraries and DNA samples. I thank Atsuko Nakanishi for performing the injections, Pat Miceli and Manfred Kittel for technical help, and Bruce Baker, Tom Goralski, Brian Oliver, Gary Struhl and Robin Wharton for their critical reviews of the manuscript. Part of this work was performed in Gary Struhl's lab at Columbia University, and I thank my colleagues there, Gary Struhl and Robin Wharton in particular, for their advice and encouragement. This work was supported by a grant from the NIH (RO1 GM42612) and in part by the Howard Hughes Medical Institute.

References

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M. and Nüsslein-Volhard, C. (1988). The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* **7**, 1749–1756.

Beverley, S. M. and Wilson, A. C. (1984). Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. *J. molec. Evol.* **21**, 1–13.

Beverley, S. M. and Wilson, A. C. (1985). Ancient origin for Hawaiian Drosophilinae inferred from protein comparisons. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4753–4757.

Cariou, M. L. (1987). Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*. *Genet. Res.* **50**, 181–185.

Corces, V., Pellicer, A., Axel, R. and Meselson, M. (1981). Integration, transcription, and control of a *Drosophila* heat shock gene in mouse cells. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7038–7042.

Crick, F. (1970). Diffusion in embryogenesis. *Nature* **225**, 420–422.

Dente, L., Cesareni, G. and Cortese, R. (1983). pEMBL: a new family of single stranded plasmids. *Nucl. Acids Res.* **11**, 1645–1655.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.

Driever, W. and Nüsslein-Volhard, C. (1988*a*). A gradient of *bicoid* protein in *Drosophila* embryos. *Cell* **54**, 83–93.

Driever, W. and Nüsslein-Volhard, C. (1988*b*). The *bicoid* protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* **54**, 95–104.

Driever, W. and Nüsslein-Volhard, C. (1989). The *bicoid* protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature* **337**, 138–143.

Driever, W., Thoma, G. and Nüsslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340**, 363–367.

Fischer, J. A. and Maniatis, T. (1986). Regulatory elements involved in *Drosophila Adh* gene expression are conserved in divergent species and separate elements mediate expression in different tissues. *EMBO J.* **5**, 1275–1289.

Fox, G. E. and Woese, C. R. (1975). 5S RNA secondary structure. *Nature* **256**, 505–507.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986). Structure of the segmentation gene *paired* and the *Drosophila* PRD gene set as part of a gene network. *Cell* **47**, 735–746.

Frohnhöfer, H. G., Lehmann, R. and Nüsslein-Volhard, C. (1986). Manipulating the anterior posterior pattern of the *Drosophila* embryo. *J. Embryol. exp. Morph.* **97**, 169–179.

Frohnhöfer, H. G. and Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the *Drosophila* embryo by the maternal gene *bicoid*. *Nature* **324**, 120–125.

Frohnhöfer, H. G. and Nüsslein-Volhard, C. (1987). Maternal genes required for the anterior localization of *bicoid* activity in the embryo of *Drosophila*. *Genes & Dev.* **1**, 880–890.

Gutell, R. R., Weiser, B., Woese, C. R. and Noller, H. F. (1985). Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucl. Acids Res. Molec. Biol.* **32**, 155–216.

Jaeger, J. A., Turner, D. H. and Zuker, M. (1990). Predicting optimal and suboptimal secondary structure for RNA. *Meth. Enz.*, (in press).

James, B. D., Olsen, G. J., Liu, J. and Pace, N. R. (1988). The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell* **52**, 19–26.

King, R. C. (1970). *Ovarian Development in Drosophila Melanogaster*. Academic Press, New York.

Klinger, M., Erdelyi, M., Szabad, J. and Nüsslein-Volhard, C. (1988). Function of *torso* in determining the terminal anlagen of the *Drosophila* embryo. *Nature* **335**, 275–277.

Lehmann, R. and Nüsslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in *Drosophila*. *Cell* **47**, 141–152.

Macdonald, P. M. and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature* **324**, 537–545.

Macdonald, P. M. and Struhl, G. (1988). *Cis*-acting sequences responsible for anterior localization of *bicoid* mRNA in *Drosophila* embryos. *Nature* **336**, 595–598.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984). A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* **308**, 428–433.

McKnight, S. L. and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. *Science* **217**, 316–324.

Mitsialis, S. A. and Kafatos, F. C. (1985). Regulatory elements controlling chorion gene expression are conserved between flies and moths. *Nature* **317**, 453–456.

Myers, R. M., Tilly, K. and Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. *Science* **232**, 613–618.

Noller, H. F. (1984). Structure of ribosomal RNA. *A. Rev. Biochem.* **53**, 119–162.

Noller, H. F. and Woese, C. R. (1981). Secondary structure of 16S ribosomal RNA. *Science* **212**, 403–411.

Nomura, M., Traub, P. and Bechmann, H. (1968). Hybrid 30S ribosomal particles reconstituted from components of different bacterial origins. *Nature* **219**, 793–799.

NÜSSLEIN-VOLHARD, C., FROHNHÖFER, H. G. AND LEHMANN, R. (1987). Determination of anteroposterior polarity in *Drosophila*. *Science* **238**, 1675–1681.

PACE, N. R., SMITH, D. K., OLSEN, G. J. AND JAMES, B. D. (1989). Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA – a review. *Gene* **82**, 65–75.

RUBIN, G. M. AND SPRADLING, A. C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348–353.

RUBIN, G. M. AND SPRADLING, A. C. (1983). Vectors for P element-mediated gene transfer in *Drosophila*. *Nucl. Acids Res.* **11**, 6341–6351.

STEPHENSON, E. C., CHAO, Y.-C. AND FACKENTHAL, J. D. (1988). Molecular analysis of the *swallow* gene of *Drosophila melanogaster*. *Genes & Dev.* **2**, 1655–1665.

STRECKER, T. R., HALSELL, S. R., FISHER, W. W. AND LIPSHITZ, H. D. (1989). Reciprocal effects of hyper- and hypoactivity mutations in the *Drosophila* pattern gene *torso*. *Science* **243**, 1062–1066.

STRUHL, G., STRUHL, K. AND MACDONALD, P. M. (1989). The gradient morphogen *bicoid* is a concentration-dependent transcriptional activator. *Cell* **57**, 1259–1273.

TAUTZ, C. (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres. *Nature* **332**, 281–284.

THEURKAUF, W. E., BAUM, H., BO, J. AND WENSINK, P. (1986). Tissue-specific and constitutive a-tubulin genes of *Drosophila melanogaster* code for structurally distinct proteins. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8477–8481.

TURNER, D. H., SUGIMOTO, N. AND FREIER, S. M. (1989). RNA structure prediction. *A. Rev. Biophys. Biophys. Chem.* **00**, 167–192.

WHARTON, R. P. AND STRUHL, G. (1989). Structure of the *Drosophila* BicaudalD protein and its role in localizing the posterior determinant nanos. *Cell* **59**, 881–892.

ZUKER, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48–52.

(*Accepted 16 May 1990*)