

Molecular Coevolution Between Developmental Genes In Insect Species

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester



by

Claire Fazakerley BA (Oxon)
Department of Genetics
University of Leicester

February 1995
revised January 1996

UMI Number: U078240

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U078240

Published by ProQuest LLC 2015. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346



for my parents

Acknowledgments

Firstly, I would like to thank Professor Gabby Dover for his patient and enthusiastic supervision of this project, and in particular for his long distance help and support during the revision of my thesis.

I would like to thank Martin Williamson, Dietard Tautz and Ralf Sommer for providing materials used in this work, and also Dietard again for helpful advice. I would also like to thank those people who taught me the techniques I have used - in particular, Ralf for showing me how to screen libraries, Juliet for demonstrating the isolation of RNA and Mike for teaching me to inject flies. Also Min, for keeping the lab running smoothly.

I would also like to thank those people who have made the past four years fun - including Caroline, Lynn and in particular Karen and Kate (who both know all about life in Sunny Leicester). Mike, Max and Lee have been the most tolerant of housemates. Neil, Lara and Matt, for pleasant weekends in Cambridge. My parents, and my sister Tori - who has had to put up with more than her fair share of well meant sisterly advice.

And lastly, I would like to thank Warren, who despite distance and illness has always managed to be there when I really needed him.

Abstract

Changes in the regulatory sequences of the genes involved in development are thought to be important in the evolution of morphology. However, molecular coevolution between functionally interacting genetic elements allows sequence divergence to be tolerated whilst the functional interaction is maintained. Molecular coevolution can lead to species-specificity in the sequence basis underlying molecular interactions.

The concentration-dependent activation of *hunchback* (*hb*) expression in the anterior half of the *Drosophila melanogaster* embryo by the gradient of bicoid (*bcd*) protein represents a primary step in the elaboration of pattern along the anterior-posterior axis, and this interaction is conserved in the housefly, *Musca domestica*.

In order to investigate the possibility that the molecular basis of this interaction may have coevolved, the *bcd* and *hb* genes have been partially sequenced from *M. domestica* and compared to those of *D. melanogaster*. Analysis of the putative *M. domestica* *hb* regulatory region identified three candidate *bcd* binding sites, with a consensus sequence of TTTAATCC, rather than the TCTAATCC of *D. melanogaster*. Comparison of the *bcd* sequences revealed 5 changes within the 60 amino acids of the homeodomain. Hence, it is possible that *M. domestica* *bcd* may have a subtly altered binding specificity, pointing towards the possibility that the coordinated changes in the binding site sequences have elicited compensatory changes in the *M. domestica* *bcd* homeodomain.

Preliminary analyses have been made of the functional significances of the observed differences. Although the functional significance of the observed differences in the *bcd* and *hb* genes is not fully understood, the possibility remains that the molecular nature of the interaction between *bcd* and *hb* has diverged between *M. domestica* and *D. melanogaster*.

Abbreviations

A	Adenine
<i>Antp</i>	<i>Antennapedia</i>
Arg	Arginine
ATP	Adenosine 5'-triphosphate
<i>bcd</i>	<i>bicoid</i>
bp, kb	Basepair; Kilobase
BSA	Bovine serum albumin
<i>btd</i>	<i>button head</i>
C	Cytosine
cDNA	Complementary deoxyribonucleic acid
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytosine 5'-triphosphate
ddATP	2'3'-dideoxyadenosine 5'-triphosphate
ddCTP	2'3'-dideoxycytosine 5'-triphosphate
ddGTP	2'3'-dideoxyguanosine 5'-triphosphate
ddTTP	2'3'-dideoxythymidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DIG	Digoxygenin
DNA	Deoxyribonucleic acid

dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	Ethylenediaminetetra-acetic acid
<i>ems</i>	<i>empty spiracles</i>
<i>en</i>	<i>engrailed</i>
<i>eve</i>	<i>even-skipped</i>
<i>exu</i>	<i>exuperantia</i>
<i>ftz</i>	<i>fushi tarazu</i>
G	Guanine
g, mg, µg, ng, pg	Grams; milli-, micro-, nano-, pico-.
HCL	Hydrochloric acid
<i>hb</i>	<i>hunchback</i>
HIV-1	Human immuno-deficiency virus -1
k_D	Dissociation rate constant
<i>kni</i>	<i>knirps</i>
<i>Kr</i>	<i>Krüppel</i>
l, ml, µl	Litre; milli-, micro-
M, mM, µM, nM	Molar; milli-, micro-, nano-
min.	Minute
mRNA	Messenger ribonucleic acid
<i>nos</i>	<i>nanos</i>
NRE	nanos response element

OD	Optical density
<i>otd</i>	<i>orthodenticle</i>
PBS	Phosphate Buffered Saline solution.
PCR	Polymerase Chain Reaction
pol	polymerase
RACE	Rapid amplification of cDNA ends
rDNA	Ribosomal deoxyribonucleic acid
RNA	Ribonucleic acid.
rpm	Revolutions per minute
<i>Scr</i>	<i>Sex combs reduced</i>
SDS	Sodium Dodecyl Sulphate
sec	Second.
SRI	Seminal Regulatory Interaction
SSC	Saline Sodium Citrate
<i>stu</i>	<i>staufen</i>
<i>sww</i>	<i>swallow</i>
T	Thymine
TBE	Tris-borate EDTA
<i>tl</i>	<i>tailless</i>
Tris	Tris-(hydroxymethyl)-methylamine[2-amino-(2-hydroxymethyl)-propan-1,3-diol]
<i>Ubx</i>	<i>Ultrabithorax</i>

UTR	untranslated region
UV	Ultra Violet
<i>w</i>	<i>white</i>

Contents

<i>Acknowledgements</i>	III
<i>Abstract</i>	IV
<i>Abbreviations</i>	V
Chapter 1 - Introduction	7
Molecular coevolution	9
Molecular drive	8
Molecular coevolution	10
Examples of molecular coevolution	11
rDNA - pol I transcription complex	11
the <i>per</i> gene	11
Developmental Genes	12
Developmental genetics	12
The Developmental Gene Hierarchy	13
Properties of the developmental gene system	13
Redundancy	14
Fast-evolving repeated regions	17
Molecular coevolution and developmental genes	19
Choice of genes and species	19
Why <i>bcd</i> and <i>hb</i> ?	20
Well characterised	20
Simple interaction	21
Homeodomain	21
Redundant	22
Why <i>M. domestica</i> ?	22
The interaction between bicoid and <i>hunchback</i>	23
bcd activates zygotic <i>hb</i> expression	23
maternal expression of <i>hb</i>	24
cooperative binding of <i>bcd</i> ?	25
synergistic interaction of <i>bcd</i> and <i>hb</i>	25
setting the <i>hb</i> expression domain	27
Chapter 2 - Materials & Methods	30
MATERIALS	30
consumables.	30
Oligonucleotides	30
Buffers and Solutions	31
Media	32
Bacterial strains	33
METHODS	33
General methods for nucleic acid handling	33

Phenol/chloroform extraction of DNA	33
Ethanol precipitation of DNA	34
Estimation of DNA concentration	34
Restriction endonuclease digests of DNA	34
Agarose gel electrophoresis	35
Of DNA	35
Of RNA	35
Isolation of DNA from agarose gels	36
Southern blotting of DNA	36
Northern blotting of RNA	37
labelling DNA	37
radio-labelling	37
digoxigenin labelling	38
DNA hybridisation and detection	38
With ³² P labelled probes	38
With DIG labelled probes	39
Subcloning procedures	40
Nested Deletions	40
Ligation	40
Transformation	41
Nucleic acid preparation	41
Isolation of Genomic DNA	41
Isolation of RNA	42
Isolation of Plasmid DNA	43
Small scale	43
Large scale	44
Single-stranded DNA	45
Isolation of Phage DNA	45
Plate lysate method	45
Liquid culture lysate method	46
Polymerase chain reaction	47
Standard	47
Colony and plaque screening	47
Library Screening	47
Preparation of the bacteria	47
Titration of the library	47
Plating the library	48
Making the filters	48
Positives, and rescreening	49
Colony screening	50
Sequencing and analysis.	50
Sequencing	50
Manual	50
Automated	51
Computer analysis	52
Protein methods	52
SDS-polyacrylamide gel electrophoresis	52
Protein Expression and Purification	53
Pilot experiment	53
Large scale preparation	54
<i>Drosophila</i> methods	55
Microinjection of embryos	55

Egg collection	55
dechoriation (manual)	55
Microinjection	56
<i>In situ</i> hybridisation of Embryos	57
Dechoriation (chemical)	57
Formaldehyde fixation	57
Pre-treatment	58
Hybridisation & washes	58
Staining and colour development	59
Mounting & photography	59
Chapter 3 - The <i>M. domestica</i> <i>bcd</i> gene	61
The <i>bcd</i> gradient	61
<i>bcd</i> localisation	62
<i>bcd</i> degradation	62
Reading the <i>bcd</i> gradient	63
<i>bcd</i> binding affinities	63
Synergistic interactions with other genes	63
Activation of gene expression	64
Conservation of gene expression	64
METHODS	65
Materials	65
Genomic DNA libraries.	65
<i>bcd</i> homeodomain fragment.	66
Methods	66
Library Screen	66
Restriction Mapping and Sub-cloning	66
Sequencing	67
DNA sequence analysis and alignments	68
Northern Analysis	69
RESULTS	69
Sequencing the <i>M. domestica</i> <i>bcd</i> gene	69
The second and third exons	70
The missing exons	70
Comparisons	70
Structure of the Gene	70
Alternate splicing	70
Intron size	71
Transcript length	72
Protein Conservation and Divergence	72
opa repeats	73
Homeodomain	73
PEST domain	74
Acidic region	75
DISCUSSION	75
Comparisons	76
Conservation and Divergence	77

Alternate splicing	77
PEST	78
opa repeats	78
Homeodomain	80
A changed specificity of the homeodomain-DNA interaction ?	81
Structural studies	81
Genetic studies	83
A changed specificity?	84
Chapter 4 - the <i>M. domestica hb</i> gene	86
Gene structure	86
Protein structure	87
<i>Hb expression</i>	87
Control of early zygotic expression	88
Control of maternal and later zygotic expression	90
Molecular Coevolution	90
METHODS	91
Materials	91
Genomic DNA libraries	91
<i>hb zinc finger probe</i>	91
Methods	92
Library Screens	92
Restriction mapping and sub-cloning	92
Sequencing	93
Subcloning	93
Shotgun cloning	94
Custom primers	94
DNA sequence analysis	95
Northern Analysis	95
RESULTS	95
Sequencing <i>M. domestica hb</i>	95
Coding sequence	96
The first zinc finger domain	96
The N-terminal region	97
Upstream sequence	97
P2 promoter	97
Sequence comparisons	98
Matches to <i>D. melanogaster</i> consensus bcd binding sequence	98
Candidate site spacing	101
<i>hb</i> binding sites	101
P1 promoter	102
Sequence comparisons	102
Maternal transcript length	103
DISCUSSION	103
Inter-species comparisons	103
The <i>hb</i> P1 promoter	104
The <i>hb</i> P2 promoter	105

Identification of <i>M. domestica</i> bcd binding sites	106
Homogenisation of variant bcd-binding sites	109
Further evidence supporting the candidate sites	109
Chapter 5 - Binding Assays	86
Introduction	85
METHODS	114
Materials	114
Binding sites	114
Binding buffer	115
Methods	115
Expression of <i>M. domestica</i> bcd homeodomain fusion protein	115
Expression vector construct	115
Gel Shift Assays	116
Probe DNA	116
Binding reactions	116
Electrophoresis	117
Autoradiography and quantitation	117
Relative binding affinity	117
RESULTS	117
Protein Expression	118
<i>M. domestica</i> bcd binds to the <i>M. domestica</i> site	118
Relative affinities to <i>M. domestica</i> and <i>Drosophila</i> sites	118
Purification of the crude extracts	119
DISCUSSION	120
<i>M. domestica</i> bcd binds to the candidate site sequence	120
Mapping bcd binding sites	120
Relative affinity to the two consensus sequences.	121
A Changed Specificity of bcd?	121
Higher affinity of <i>M. domestica</i> bcd	122
Contribution of non-homeodomain regions of bcd protein	123
An unchanged specificity?	124
Number of sites	125
Protein-protein interactions	126
Concentration of bcd	127
Chapter 6 - Transformations	128
The results so far	128
Functional Assays	129
In this Chapter	131
METHODS	132
Materials	132
<i>D. melanogaster</i> strains	132
Methods	132
Construction of <i>M. domestica</i> hb transformation fragment	132

P-element mediated transformation	133
Preparation of DNA for injection	133
Crosses	133
Southern analysis	134
<i>In situ</i> hybridisation	135
Probes	135
RESULTS	136
Transformation with <i>M. domestica hunchback</i>	136
The Construct	136
Injection and Genetic Analysis	137
Selection of transformed lines for analysis	138
<i>in situ</i> hybridisation	139
With the <i>D. melanogaster hb</i> probe	139
With the <i>M. domestica</i> probe	139
DISCUSSION	140
Transformation of <i>D. melanogaster</i>	140
Transformed <i>M. domestica hb</i> does not express	141
Low concentration of transcripts?	141
Very low level expression	141
Not the authentic P2 promoter	142
Transcript instability	143
Cis-acting factors	143
Affinity of the binding sites	144
The <i>bcd</i> binding sites	144
The <i>hb</i> binding sites	144
Chapter 7 - Discussion	147
Early development in insects	148
The original morphogen - <i>bcd</i> or <i>hb</i> ?	149
<i>bcd</i> and <i>hb</i> in <i>M. domestica</i> and <i>D. melanogaster</i>	152
<i>M. domestica hb</i>	152
<i>M. domestica bcd</i> .	153
A divergent interaction	154
Flexibility and Constraints	156
The <i>bcd-hb-Kr</i> interaction	157
Molecular coevolution and the evolution	159
References	158

Chapter 1

INTRODUCTION

In October, 1994, the journal *Science* marked the centenary of the founding of the field of developmental biology by Wilhelm Roux and colleagues, by conducting a survey of 100 leading developmental biologists. From this survey, they concluded that the evolution of development is the second¹ most important unanswered question in the field of developmental biology (*Science*, 1994).

The study of the evolution of development provides a unifying point for the integration of reductionist and mechanical studies of genes and cells, and historical and narrative accounts of life's phylogeny (Gould, 1992). Even before the concept of evolution was expressed, the parallels between development and the grouping and classification of organisms had been noted; and with the emergence of evolutionary theory, many focused on the relationship between the development of an organism and its evolutionary history (Gould, 1977). Now, as the specific genes and molecular genetic systems that control the development of organisms are unravelled, it becomes possible to examine the evolution of particular developmental processes at the level of the actual genes that control them.

The molecular and genetic basis of early pattern formation in *Drosophila melanogaster* has been particularly well studied. Comparisons with other species have revealed the conservation of some aspects of development, as well as the presence of some differences that may explain variations in early patterning events (reviewed Patel, 1994). However, such comparisons have typically focused on patterns of homologous gene expression in fairly widely diverged species, and hence relate primarily to the

¹The most important unanswered question was considered to be the molecular basis of morphogenesis (the formation of the body's specialised organs and tissues).

problem of how widely different body patterns are put together (Akam, 1994; Slack *et al.*, 1993), rather than through the examination of more closely related species, relating to the processes by which this variation is generated.

That the evolution of morphological change is a consequence of alterations in regulatory, rather than structural genetic elements was first suggested by Wilson *et al.*, (1974). This concept is supported by accumulating bodies of data from comparative studies of developmental gene expression (reviewed Carroll *et al.*, 1995; Scott *et al.*, 1994). So, for example, the distinct developmental functions acquired by the related *paired*, *gooseberry* and *gooseberry neuro* genes of *Drosophila melanogaster* since their duplication reflect differences in their expression patterns rather than their divergent protein sequences- a clear example of evolution through regulatory changes (Li and Noll, 1994). *Ultrabithorax* directs haltere formation in *Drosophila* and second wing formation in butterflies through directing the expression of different sets of downstream genes in these species (Warren *et al.*, 1994). Similarly, whilst the protein products of murine *Pax6* gene and its *D. melanogaster* homologue *eyeless* can function heterologously (Halder *et al.*, 1995) - and appear to act as master regulators of eye development, the distinct eye morphologies of mice and flies presumably result from divergence in the regulation of downstream genes during evolution (Zuker, C.S., 1994; Quiring, R. *et al.*, 1994; Halder *et al.*, 1995).

Given the contribution of regulatory changes to morphological evolution, it is necessary to examine the nature of the processes that affect the generation of these alterations in regulatory elements in order to understand the evolution of morphology. By examining the fine-grained differences in the way genes interact with one another which distinguish species-specific developmental programs, the genomic and evolutionary processes that are the proximate causes of the divergence of closely related developmental systems may be discerned. So, in order to understand the evolution of development it is necessary to compare the precise molecular basis of the interaction between pairs of molecules, and to examine the underlying mutational processes affecting this interaction, between species of manageable phylogenetic distance (Dover, 1992).

As will be discussed further below, development is regulated through local, combinatorial interactions between independent genetic elements that are often modular and redundant, evolving from a complex interplay between natural selection and genomic processes of mutation and turnover. Therefore, aspects of the evolution of ontogeny may be understood through the process of 'molecular coevolution' (Dover and Flavell, 1984): the ability of one molecule to compensate for a change in another molecule, whilst subtle evolutionary changes occur in developmental functions. Key compensatory changes can take place as a consequence of the tolerance and flexibility stemming from the genetic and functional redundancy of developmental programs.

In this thesis, the interaction between the early developmental genes *bicoid* (*bcd*) and *hunchback* (*hb*) is compared in *Musca domestica* and *Drosophila melanogaster*.

Molecular coevolution

Nuclear genomes are in a state of continual change through processes of genomic flux - unequal crossing over, slippage, transposition, gene conversion and RNA-mediated genetic exchange. Slippage results in the continual gain and loss of short direct repeats within an allele (Figure 1.1). Unequal crossing over can lead to the production of tandem repeats, and where repetitive DNA exists will serve to randomly increase or decrease the number of repeats by inter-allelic exchange (Figure 1.2) as chromosomes pair "out of register". Whilst gene conversion does not alter the total number of repeats, it may promote the spread of a variant through a population of repeats, both within and between alleles (Figure 1.3).

MOLECULAR DRIVE

The processes described above can all result in a continual gain and loss in the number of repeats within a region of repeated sequence or multigene family, and hence

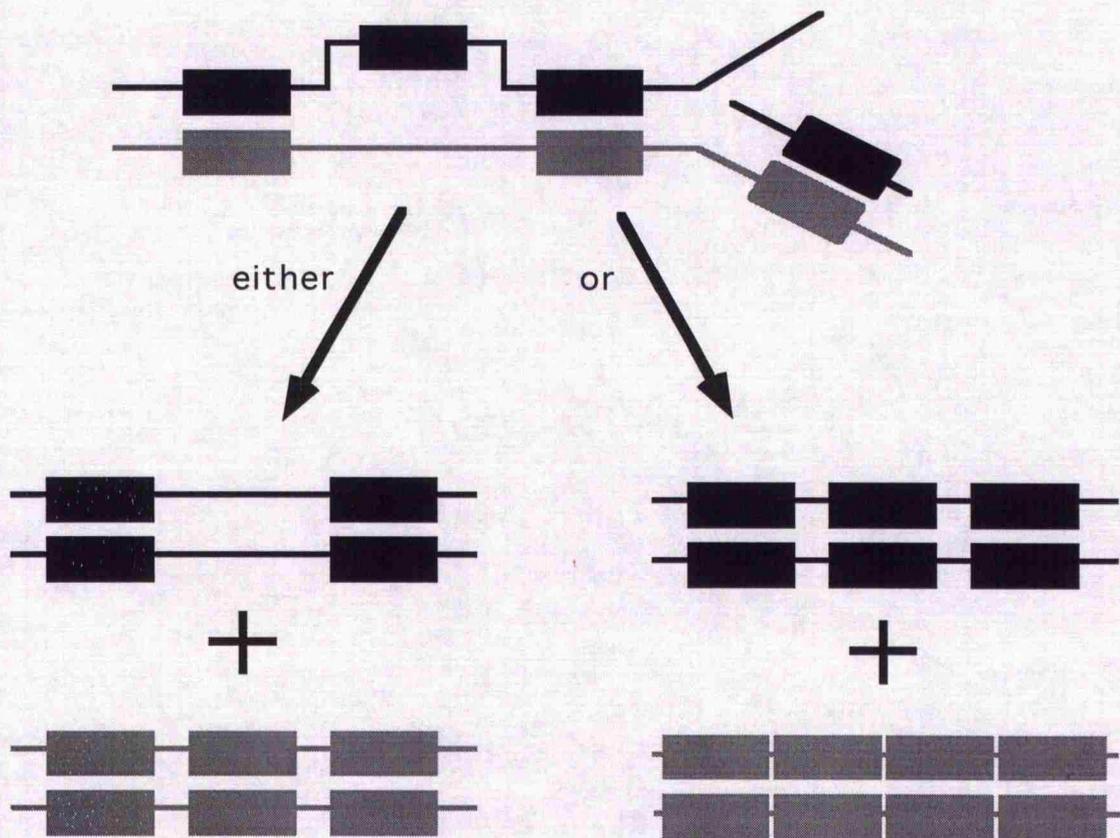


Figure 1.1 A model of slippage

in which, during replication, a repetitive motif on one strand slips and base pairs with another further down on the opposite strand. The resultant loop in the first strand can either be excised out or used as a template for synthesizing equivalent sequences on the second strand. By this means, continual gains and losses of repeats occur. (from Dover, 1989).

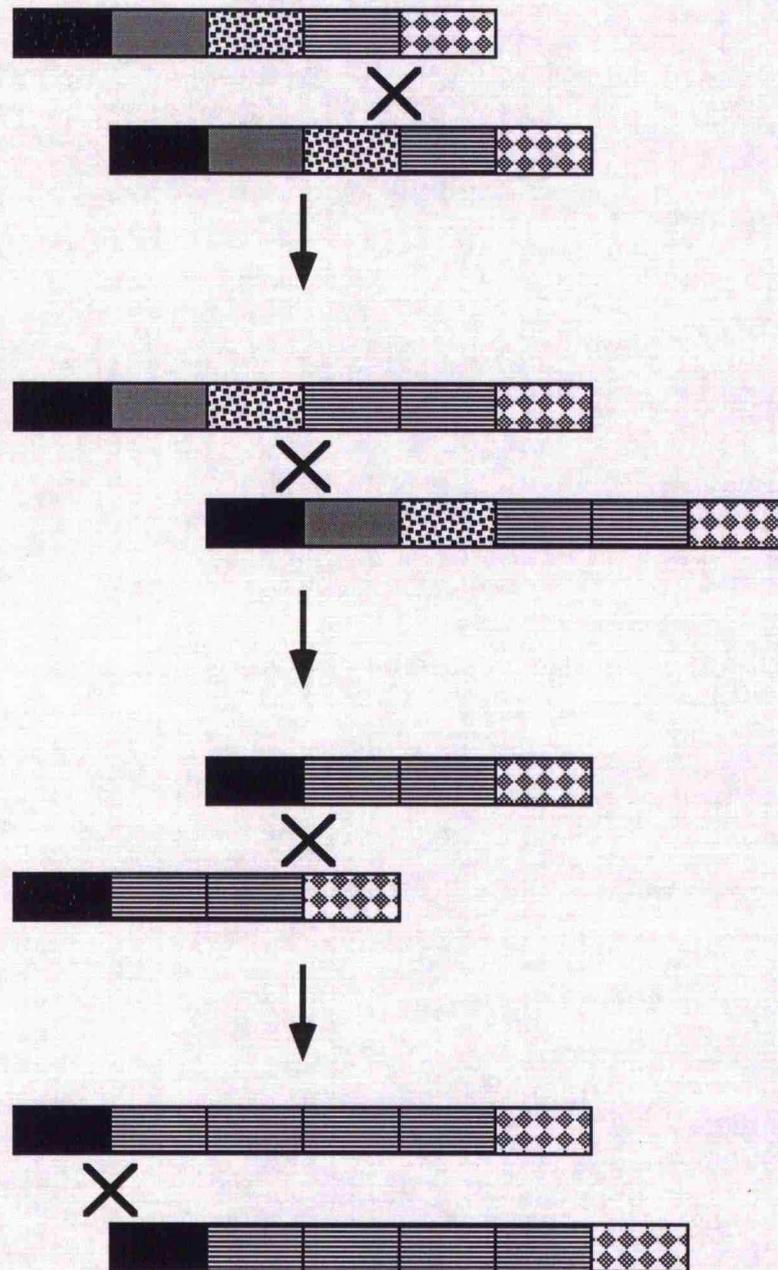


Figure 1.2. Unequal crossing over.

Differently patterned blocks represent variants of a repeated DNA sequence. Unequal crossing over between chromosomes can lead to the production of repeated segments of DNA. Once repeats exist, further unequal crossover events become more likely, as the chromosomes can then pair 'out of register'. This results in continual random fluctuations in the numbers of repeats, which will tend to homogenize all the repeats within an array. After Nei, 1987.

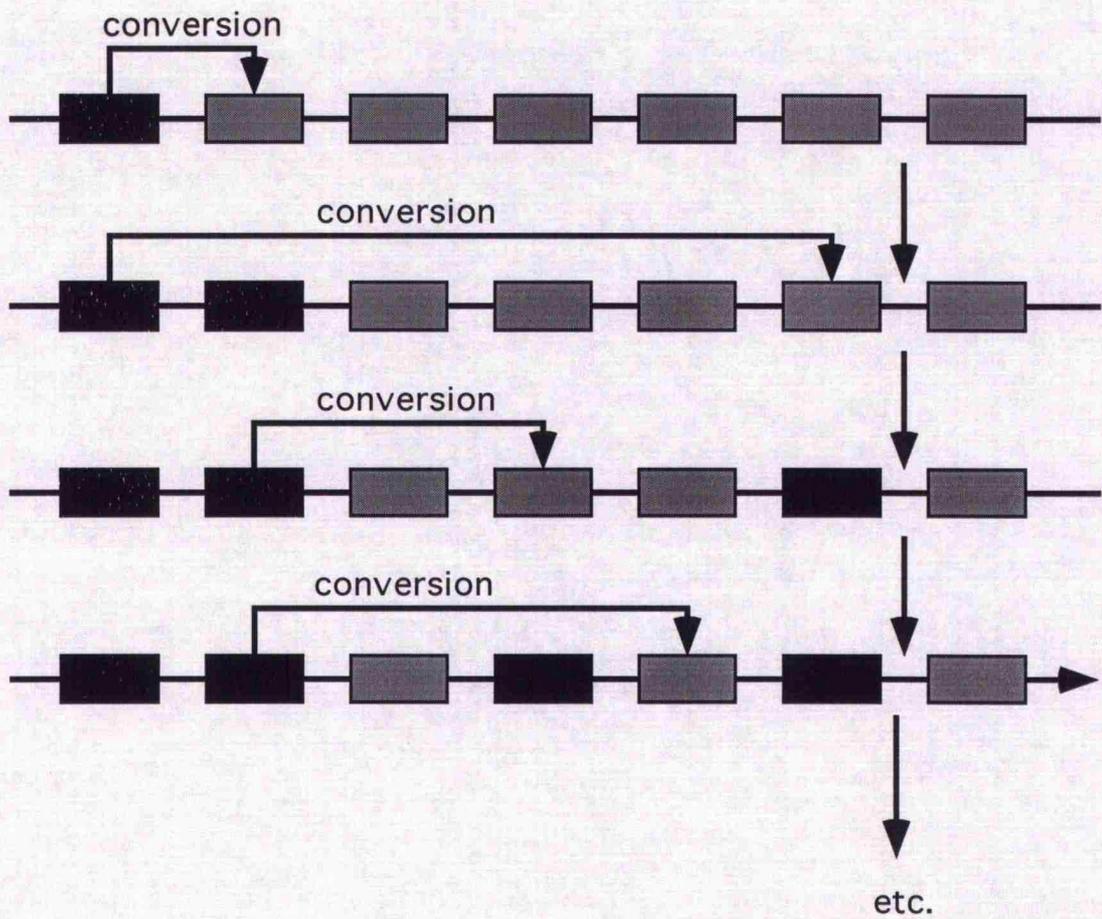


Figure 1.3. Gene conversion may promote the spread of a variant through a population of repeats of a DNA sequence.

can gradually promote the spread of a variant repeat through a sexual population. When mutations initially arise, most will be accidentally wiped out by the processes of genomic flux; however, for a small number, these same mechanisms will result in an increase in frequency by one or a few copies in the germ line of any individual. At the next generation, the mutant members would be distributed randomly among the progeny through the processes of meiosis and gametic fusion. As the spread of variants within a set of repeats (homogenisation) by the mechanisms of genomic flux is much slower than the spread of variants between individuals within a population (fixation) by sexual reassortment, all individuals within a population will tend to have the same proportion of old to new variant repeats at any given time (Dover, 1982, Ohta and Dover, 1984). The process by which this concerted evolution of repeated sequences occurs is termed molecular drive (Dover, 1982).

MOLECULAR COEVOLUTION

The fate of the mutant member gene as it increases in frequency through the population would depend on the nature of the mutation itself, and of the constraints imposed by other interacting molecules. Both genetic redundancy (a likely consequence of a large number of repeats) and functional redundancy (through overlapping gene functions) will mask the phenotypic effects of mutant repeats during the early stages of their spread. If the mutation has a critical effect on function, then it is likely that any population which has accumulated more than a tolerable number of such repeats will be eliminated by selection. If the effect is fairly small, then the gradual and cohesive spread of the variant repeat will create the conditions for the selection or drift of alleles of the genes coding for the proteins that are better able to interact with the changing spectrum of multiple target sequences - molecular coevolution (Dover and Flavell, 1984; illustrated in Figure 1.4).

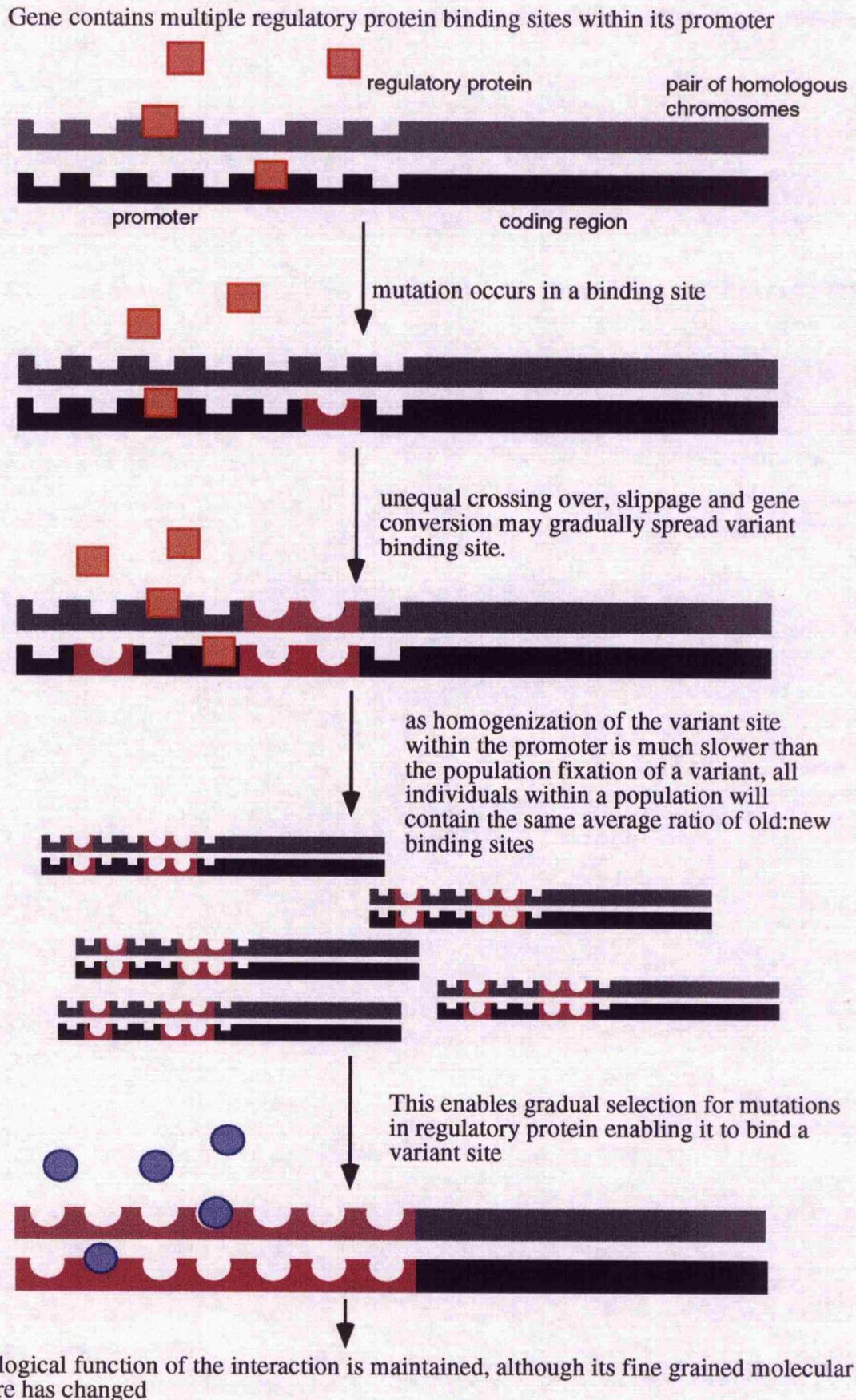


Figure 1.4. A diagram illustrating the molecular coevolution of a regulatory protein and its binding sites within the promoter of a gene.

EXAMPLES OF MOLECULAR COEVOLUTION

rDNA - pol I transcription complex

The classic example of molecular coevolution is that between the spacers of the rDNA family, and the interacting polymerase I complex of proteins.

The rDNA multigene family shows a pattern of concerted evolution, in which repetitive units share mutations that are specific for each species (reviewed by Federoff, 1979; Long and Dawid, 1980; Dover, 1982; Arnheim, 1983). The rDNA unit consists of three different genes coding for ribosomal RNA's, and intergenic spacers containing transcription promoters, enhancers and terminators. The interaction between the pol I complex and the rDNA promoters is species-specific, with the pol I complex from one species being unable to transcribe the rDNA of another (reviewed Grummt *et al.*, 1982; Reeder, 1984; Dover and Flavell, 1984; Arnheim, 1983; Moss *et al.*, 1985; Dover and Tautz, 1986). This suggests that the key polymerase I cofactors have coevolved with the slowly changing nature of the multiple promoters. The complex nature of these molecular interactions and how they differ between species has been examined (Bell *et al.*, 1989; Jantzen *et al.*, 1990).

the *per* gene

A case in which molecular coevolution may have been important in allowing the evolution of species-specific behavioural patterns within *Drosophila* is that of the *period* (*per*) clock gene. Interspecific comparisons of the *per* locus in *Drosophila* has provided some experimental evidence suggesting intragenic molecular coevolution between the repetitious threonine-glycine (TG) array found in the centre of the *per* gene and the flanking coding regions (Peixoto *et al.*, 1993). The number of TG repeats is associated with the number of amino acid replacements (but not silent site changes) in the more conserved sequences flanking the repeat regions. The high mutational pressure associated with slippage-like events in regions of high cryptic simplicity (Tautz *et al.*,

1986) such as the TG domain might drive compensatory mutations in other regions of the protein.

DEVELOPMENTAL GENES

The examples given above, and others described in the literature (reviewed Dover, 1992), demonstrate the general relevance of molecular coevolution to the evolutionary process. At one level, all organisms possess a similar range of functions, while at another level, the basic interactions which manage these functions are constantly changing - the tolerance in the system is such that interacting partners are able to continually alter and mutually change their composition. Molecular coevolution indicates that conservation and divergence at the sequence level do not necessarily signify functional and redundant activities respectively, but that organisms are capable of maintaining essential functions at the level of molecular interactions even though detailed aspects of the molecular interaction have changed.

The characterisation of the specific genes and molecular genetic networks that control early development in *Drosophila melanogaster* has revealed a redundantly specified developmental program consisting of a network of genes containing fast-evolving repetitive regions, themselves regulated through modular and redundant promoter elements. Therefore, morphology may be subject to molecular coevolutionary change. As discussed by Dover (1992), this sets the stage for a systematic study of molecular coevolution and evolutionary flexibility between selected pairs of interacting molecules in defined species.

Developmental genetics

The organism for which the genetic basis of the control of development is best understood is *Drosophila melanogaster*. The early development of *D. melanogaster* is controlled by a hierarchy or cascade of transcription factors, in which each level elaborates on the information specified by the level above (reviewed Akam, 1987;

Ingham, 1988; Lawrence, 1992). As *D. melanogaster* develops initially as a syncytial blastoderm, and the basic body plan is established before cellularisation occurs, the interactions between the different genes are mediated directly with many genes encoding transcription factors. Figure 1.5 summarises the establishment and elaboration of pattern along the anterior-posterior axis of the early embryo.

THE DEVELOPMENTAL GENE HIERARCHY

The anterior-posterior polarity of the embryo is determined by maternally transcribed factors, which establish morphogenetic gradients along the body axis. These gradients regulate the expression of the next layer of genes in the hierarchy, the gap genes, which are expressed in broad overlapping domains. In turn, the gap genes, (in conjunction with the maternal genes) regulate the expression of the pair-rule genes. The pair-rule genes are expressed in a series of seven stripes, and fall into two classes - the primary pair-rule genes, in which each stripe is individually regulated in a combinatorial fashion, and the secondary pair-rule genes, in which a single element drives the expression of all seven stripes in response to primary pair-rule gene expression. The segment polarity genes are expressed in a pattern of fourteen stripes, regulated by the pair-rule genes, and define the parasegmental boundaries - the basic repeated unit of the body pattern. Segment identity is defined by the expression pattern of the homeotic genes, in response to all of the above layers of the developmental gene hierarchy.

PROPERTIES OF THE DEVELOPMENTAL GENE SYSTEM

Development is regulated through a network of local, combinatorial interactions between independent genetic elements. As will be discussed below, many of these interactions are redundantly specified through the overlapping functions of various genes, themselves often modular and redundant, and which also contain fast-evolving

MATERNAL EFFECT GENES include *bicoid* and *nanos*. The bicoid protein forms a shallow gradient along the antero-posterior axis of the embryo, and directs expression of the gap genes.

GAP GENES such as *hunchback*, *Kruppel* and *knirps*. are expressed in broad overlapping domains. Mutations in gap genes result in the elimination of particular regions of the body, creating a gap in the antero-posterior polarity .

PAIR-RULE GENES such as *fushi tarazu*, *even-skipped* and *hairy* are expressed in a pattern of seven stripes, in response to the gap genes. Mutations in pair rule genes affect every alternate segment of the embryo.

Combinatorial action of the pair-rule genes drives the expression of **SEGMENT POLARITY GENES** such as *engrailed* and *wingless* in a pattern of fourteen stripes. Mutations in segment polarity genes affect each segment of the embryo.

HOMEOTIC GENES such as *Antennapedia*, *Ultrabithorax*, *Abdominal-A* and *Abdominal-B* are initially expressed in broad domains defined by the maternal and gap genes. These are modulated by interactions with the pair-rule and segment polarity genes to generate a series of unique parasegmental states. Mutations in homeotic genes result in the transformation of one segment type into another.

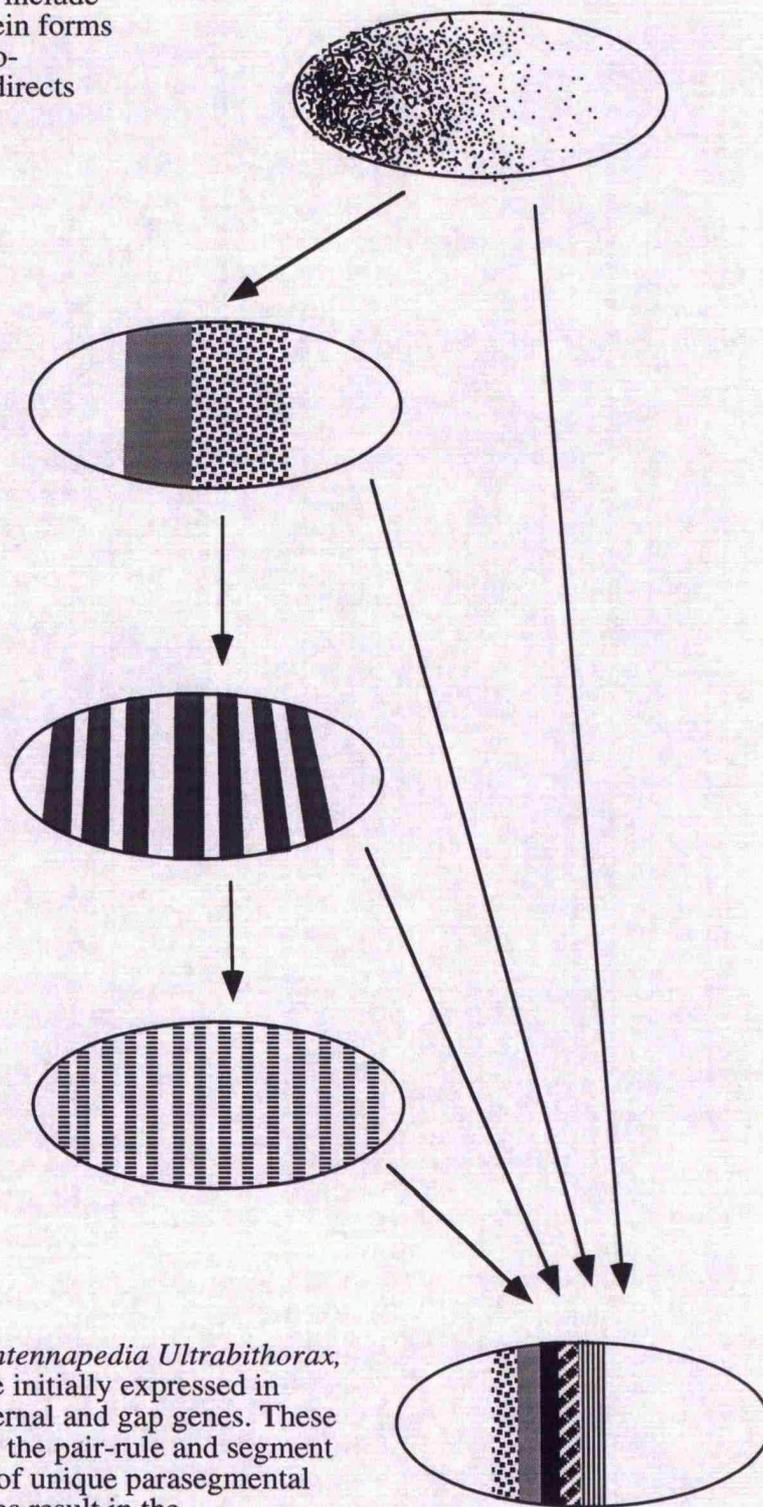


Figure 1.5. A simplified diagram illustrating the developmental gene hierarchy in *D. melanogaster*

repetitious regions likely to be subjected to the processes of genomic turnover (Tautz *et al.*, 1986).

Redundancy

Redundancy occurs where there is an overlap in the functions of genetic elements or regulatory pathways. This redundancy may be *genetic* - i.e. due to duplication of genetic elements, as in the case of the promoters of genes which contain multiple equivalent binding sites for regulatory factors. Redundancy may also be *functional*, where different, unrelated genetic elements or regulatory pathways fulfil the same function. The degree of redundancy that occurs may range from the ability of one element to fully compensate for the absence of a redundant partner, to a partial overlap with more limited compensatory abilities. A number of examples of redundancy within the early developmental system have been demonstrated. A selection of these are described below.

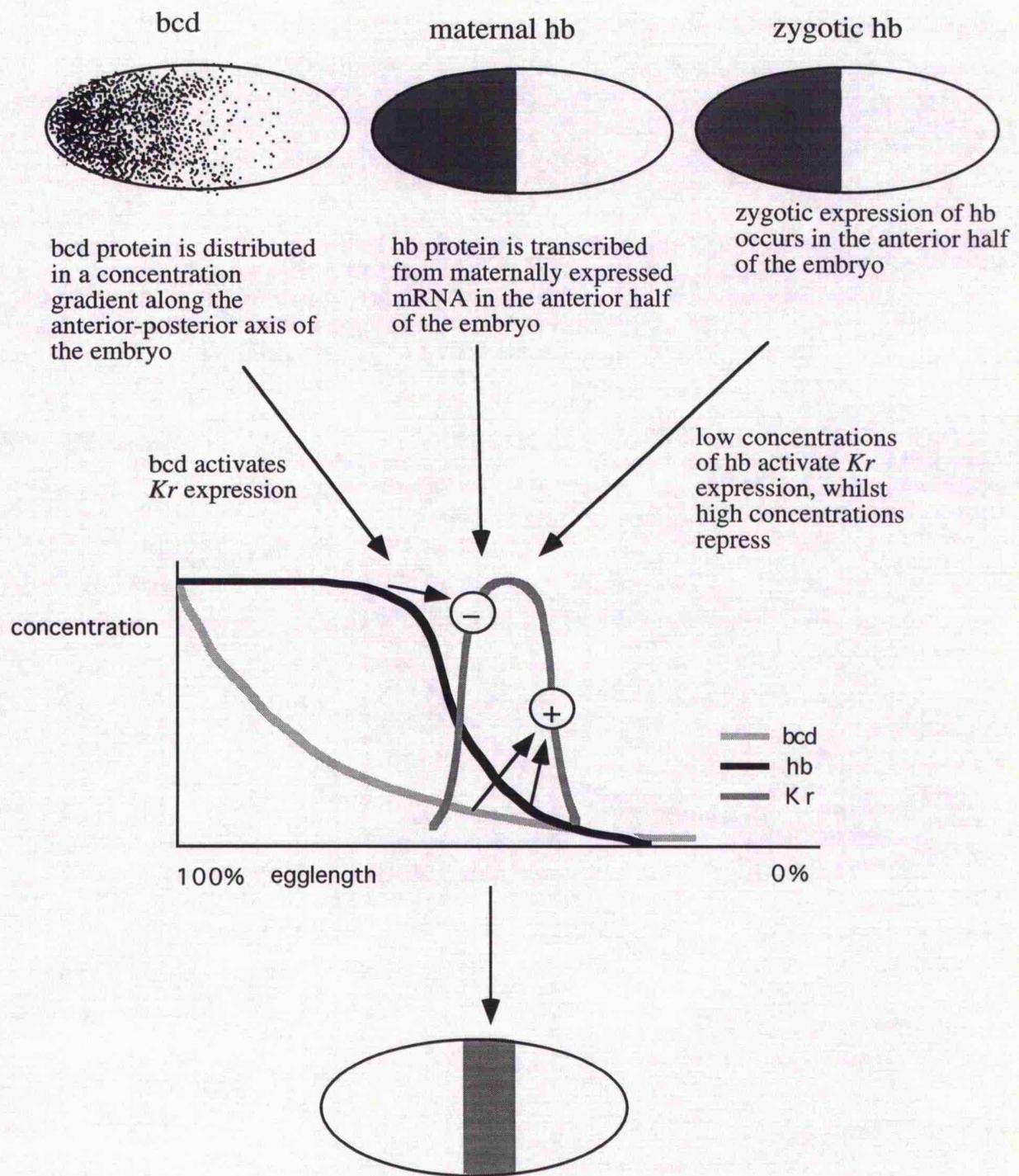
One classic example of redundancy is given by the regulation of the gap gene *hunchback (hb)*. *hb* protein is distributed in an anterior domain in the early *Drosophila* embryo. This distribution is redundantly specified - it occurs both through the posterior translational repression of universally distributed maternally transcribed *hb* mRNA by the posterior determinant nanos (Irish *et al.*, 1989), and through the activation of zygotic *hb* gene expression by the anterior determinant bicoid (*bcd*) in an anterior domain (Driever and Nüsslein-Volhard, 1989). These mechanisms are described more thoroughly later, but the important detail is that maternal *hb* expression is completely functionally redundant as the zygotic *bcd* dependent *hb* expression is able to fully compensate for the absence of the independently specified maternal expression (Lehmann and Nüsslein-Volhard, 1987; Hülskamp *et al.*, 1989; Irish *et al.*, 1989; Struhl, 1989).

Another example of the functional redundancy inherent within the developmental program of *D. melanogaster* is given by the regulation of expression of

the gap gene *Krüppel* (*Kr*). *Kr* expression is regulated by region specific activation and repression events. It is activated by low levels of the protein products of the maternal gene *bcd* and the gap gene *hb* (both of which are expressed in anterior domains), and it is repressed by high levels of hunchback (Hülskamp *et al.*, 1990; Hoch *et al.*, 1991; Hoch *et al.*, 1992; Gaul and Jäckle, 1987; see Figure 1.6). As *hb* itself is redundantly regulated, there are therefore three components involved in the activation of *Kr* expression - *bcd*, maternal *hb*, and zygotic *hb*. Each of these components alone is capable of directing the expression of *Kr* in a broad domain in the middle of the embryo (Hülskamp *et al.*, 1990; Struhl *et al.*, 1992). *Kr* expression is abolished only in embryos doubly mutant for both *bcd* and *hb*. As *bcd* and *hb* have clear functions apart from their regulation of *Kr* they are not interchangeable, since only a small part of their functions are overlapping. However, they are redundant with respect to their regulation of *Kr*. Thus their regulation of *Kr* expression provides a clear case of partial functional redundancy through overlapping gene function (Tautz, 1992; summarised in Figure 1.6).

The regulation of *Kr* expression also demonstrates genetic redundancy, in addition to the functional redundancy described above. The *Kr* promoter contains two separate enhancer elements for proper *Kr* regulation, each of which is capable of directing *Kr* expression in a central domain (Hoch *et al.*, 1990) and is therefore redundant. Furthermore, both of the enhancer elements contain multiple binding sites of varying affinity for *bcd* and *hb* (Hoch *et al.*, 1991).

Genetic redundancy has also been demonstrated for other developmental genes. For example, deletion analysis of the autoregulatory element (AE) of the pair-rule gene *fushi tarazu* (*ftz*) has defined multiple elements that are redundantly involved in enhancer activity (Schier and Gehring, 1993). The AE functions as a single autonomous unit, however up to 140 bp of any region within the AE can be deleted without affecting its qualitative function. This indicates that no single sequence element is strictly required for stripe expression, and points to extreme redundancy of qualitative information being present within the AE.



Each of the three components (bcd, maternal hb and zygotic hb) is capable of activating some *Kr* expression in a broad central domain - the activation of *Kr* expression therefore shows functional redundancy.

Figure 1.6.

A diagram illustrating the functional redundancy of the control of *Kr* expression.

Redundancy is also evident within the developmental systems of organisms other than *Drosophila*. For example, analysis of vulval development in *Caenorhabditis elegans* has revealed the presence of at least ten genes which produce mutant phenotypes only in combination with mutations in other genes (Ferguson and Horvitz, 1989). Similarly, the fact that targeted mutagenesis of genes potentially involved in pattern-formation in mice does not necessarily result in the expected strong phenotypes (for example Joyner *et al.*, 1991) may be attributed to the presence of redundancies.

It is commonly assumed that redundant elements should be quickly eliminated from the genome, and therefore will be rarely seen. Complete redundancies would not necessarily be evolutionarily stable - duplicated genes often lose function or adopt specialised tasks during evolution, unless there is gene conversion between them. Therefore, the evident redundancy within the developmental program begs the question "why is this apparent redundancy maintained?".

Redundant genetic elements will arise as a simple and inevitable consequence of the mechanisms of genetic flux described above. Once created, such redundant elements will persist within the genome simply because of the high degree of accuracy of eukaryotic DNA replication and repairs systems (Ohno, 1985). Furthermore, any gene conversion events between genetically redundant elements will act to prevent sequence divergence.

Functional redundancy may be maintained because development requires the transmission of information from one stage to the next (Tautz, 1992). At each of these steps, information may be lost - possibly as a consequence of genetic mutations or environmental perturbations. The selective advantage of possessing redundant specification of the developmental gene system is that it might insure against this loss of information, increasing the probability of successful completion of embryogenesis, and hence directly increasing the probability of survival of offspring. Even if the selective advantage conferred was very small, fixation could result fairly quickly as most such newly arising safe-guarding mechanisms would act as dominant alleles. Tautz (1992) extends this argument to suggest that the more highly redundant systems

“should be those that are evolutionarily more ancient, since the recruitment of different genes for specifying single decisions would be expected to take some evolutionary time”.

However, regardless of the reason for its maintenance, the demonstrable redundancy (both genetic and purely functional) of the regulation of ontogeny results in a tolerance that buffers the effects of mutational change, and as a consequence allows the possibility of molecular coevolution.

Fast-evolving repeated regions

Interspecific comparisons of developmental genes from various *Drosophila* species has shown that they are generally composed of a patchwork of conserved and diverged regions. Conserved patches, particularly within the coding regions, are often interspersed with regions of high cryptic simplicity - stretches of slippage-generated scrambled permutations of a number of short direct DNA motifs, as opposed to tandem arrays of single motifs (Tautz *et al.*, 1986). (Where the unit of slippage is out of phase with pre-existing motifs, this process results in the continual generation and combinatorial reshuffling of short and short lived motifs differing in sequence and length).

Regions of high cryptic simplicity are prone to action by slippage mechanisms, which occur at a significantly higher rate than point mutations, and bias the kinds of changes that occur (Tautz *et al.*, 1986; Levinson and Gutman, 1987). Such slippage-generated mutations, when occurring within a patch of high cryptic simplicity within a developmental gene, may well be functionally neutral. However, if they were not, then they could result in morphological changes - ones which would have arisen as a result of the action of internal genomic processes and not primarily as a result of adaptation to external selection pressures (Treier *et al.*, 1989).

Many developmental genes, including *engrailed* (Kassis *et al.*, 1986), *bicoid* (Seeger and Kaufman (1990) and *hunchback* (Treier *et al.*, 1989), contain fast-evolving regions of

'opa' repeats - scrambled permutations of the motif CAG usually encoding glutamines or histidines (see Treier *et al.*, 1989). As Dover (1989) has pointed out "although no direct functions have been ascribed to opa and some other strings of amino acids, nevertheless this category of sequence, in frequently varying between species in either copy number or composition or both, is no less a candidate for species differentiation than are the point mutation and single amino acid substitutions that are traditionally considered".

Multiple binding sites for a single regulatory factor within the regulatory region of a gene may be regarded as a small non-tandem repeat, and as such subject to genomic flux mechanisms. Whilst it is easier to see how mechanisms such as slippage and gene conversion can act to spread variants amongst tandem repeats (as for example in the case of the concerted evolution of the rDNA promoter), it may be harder to see how such processes could work on the small, non-tandemly repeated regulatory protein binding sites within the promoter of a developmental gene. However, there is evidence that suggests that this model is applicable.

Jones and Kafatos (1982) observed large numbers of mutations consisting of small deletions or duplications of direct DNA repeats involving both the deletion/reiteration of tandem repeats and the deletion of non-tandem repeats within the chorion multigene family of the silkworm *Antheraea polyphemus*. They have proposed a slippage-based model to explain the origin of this type of mutation, illustrated in Figure 1.7. This illustrates that slippage is a powerful mechanism for the generation of variation even within non-tandem arrays such as the multiple binding sites for a regulatory protein which might be found within the promoter of a developmental gene. Furthermore, misalignment could trap non-repetitive DNA between repeats, ultimately making it repetitive in turn. Therefore, slippage may well occur between repeats (such as reiterated regulatory protein binding sites) that are close to each other but not actually tandemly arrayed.

Comparisons of distribution of mutations within the late chorion locus multigene family of the silkworm *Bombyx mori* have revealed a patchwork distribution

of highly homologous domains separated by regions of lower homology to a reference gene (Burke and Eickbush, 1986), in which most variations are shared by multiple members of the gene family. This pattern is most readily explained by the occurrence of numerous small gene conversion events between these non-tandemly repeated genes (Eickbush and Burke, 1986) illustrating the potential of gene conversion for spreading variants amongst non-contiguous repeats.

Molecular coevolution and developmental genes

The examples described above illustrate the principle that developmental programs (in *D. melanogaster* at least) are mediated through redundantly specified networks of interactions between genes containing fast-evolving repetitious elements, themselves regulated through modular and redundant promoter elements. Such networks are, as discussed above, subject to processes of genomic turnover and are, to a certain degree, tolerant of mutational change. Therefore, in order to look at the genomic and evolutionary processes that are the proximate causes of the divergence of closely related developmental systems, it is necessary to examine the fine-grained differences in the way genes interact with each other that distinguish species ontogenies. This might be done by making interspecific comparisons of the molecular bases of interaction between equivalent pairs of genes (or their products) in relatively closely related species.

Choice of genes and species

In the work described in this thesis, the *bicoid* and *hunchback* genes of the housefly, *Musca domestica* were compared to those of *D. melanogaster*. Interactions between the segmentation genes in *D. melanogaster* probably constitute the best characterised cascade of transcription factors known for any developmental process in a higher eukaryote. The genetic basis of the regulation of development is currently best understood in *D. melanogaster*, making this the obvious base species with which to make

the comparison. The reasons for choosing *M. domestica*, and the *bicoid* - *hunchback* interaction are outlined below.

WHY BCD AND HB?

The interaction between *bcd* and *hb* is described more fully later (p. 14). The interaction between *bcd* and *hb* was chosen as the basis of this project for a number of reasons - it was thought to be relatively simple, it was the best characterised at the time, it is redundantly specified, and the *bcd* protein interacts with the *hb* promoter via its homeodomain.

Well characterised

The nature of the interactions between the genes controlling development in *D. melanogaster* was initially determined by examining the expression patterns of these genes in backgrounds mutant for one or more of the other genes. By early 1992, when this project was begun, only a few of these interactions had been characterised molecularly. Binding sites for a number of genes implicated in its regulation had been identified by footprinting within an element identified through promoter fusion experiments as responsible for expression of the second stripe of the pair-rule gene *even-skipped* (*eve*) (Small *et al.*, 1991). Similar studies had characterised the promoter elements of the gap gene, *Kr* (Hoch *et al.*, 1991), and various elements controlling expression of the individual stripes of the pair-rule gene *hairy* (*h*) (Riddihough and Ish-Horowicz, 1991). For other genes such as *ftz*, direct regulatory targets had been identified *in vivo*, but the actual binding sites had not yet been determined (Schier and Gehring, 1992).

Notwithstanding the above, the best characterised interaction was that between *bcd* and *hb*. The *bcd*-dependent regulatory element of the *hb* gene had been defined *in vivo*, and binding sites for the *bcd* protein identified *in vitro* (Driever and Nüsslein-Volhard, 1989). Multiple copies of *bcd* binding sites were capable of mediating *bcd*-

dependent expression of reporter genes when used as artificial promoters in heterologous systems, indicating that the interaction is mediated directly, and that the sites identified *in vitro* were capable of mediating the interaction observed *in vivo* (Driever and Nüsslein-Volhard, 1989, Driever *et al.*, 1989, Struhl *et al.*, 1989, Hanes and Brent, 1991).

Simple interaction

The interaction between *bcd* and *hb* also had the virtue of being simple. The early zygotic domain of *hb* expression is determined by *bcd* alone, whereas the regulation of many downstream genes is more complex. The regulation of *even-skipped* (*eve*) stripe 2, for example, involves four different proteins - *bcd*, *hb*, *giant*, and *tailless* (Small *et al.*, 1991). Furthermore, the *bcd* binding sites within the *hb* promoter are located fairly close to the start of the coding sequence, and are not widely distributed (Schröder *et al.*, 1988; Struhl *et al.*, 1989, Driever and Nüsslein-Volhard, 1989). Some developmental genes have regulatory elements distributed over up to 40 kb of DNA (for example, *Ultrabithorax* (*Ubx*); Irvine *et al.*, 1991, Müller and Bienz, 1991), making interspecies comparisons difficult.

Homeodomain

Bcd activates *hb* expression by binding to multiple sites within the *hb* promoter (Driever and Nüsslein-Volhard, 1989). It contains a homeodomain, which is the best characterised of the eukaryotic DNA binding domains. If genomic turnover processes have led to the spread of variant binding sites, then compensatory mutations might have been elicited from the DNA binding domain. As some structural (Otting *et al.*, 1990; Kissinger *et al.*, 1990; Qian *et al.*, 1989) and genetic (Hanes and Brent, 1989; Hanes and Brent, 1991) studies had been made of homeodomain DNA binding, it would be possible to make predictions of how any observed mutations in the homeodomain or the binding sites might affect the interaction between the two. Such predictions would

be harder to make with less well characterised DNA binding domains such as the zinc finger domains found in many of the gap genes.

Redundant

The interaction between *bcd* and *hb* shows both genetic and functional redundancy. The expression of *hb* in an anterior domain extending over 50% of egg length is specified both maternally and zygotically (described in more detail below), with the zygotic expression able to compensate fully for the maternal expression (Lehmann and Nüsslein-Volhard, 1987), giving functional redundancy. The interaction is also genetically redundant, with the *hb* promoter containing multiple binding sites for the *bcd* protein (Driever and Nüsslein-Volhard, 1989).

WHY *M. DOMESTICA*?

In examining the conservation or divergence of the molecular interactions involved in the early pattern process it is necessary firstly to identify homologous genes and then to assume that they have homologous functions. Sommer and Tautz (1991) have argued that the assumption of homologous function may be tested if the first step in the analysis is the comparison of the expression of homologous genes in a species that shows homologous embryonic development. Comparisons of the *bcd* gene have been made between *D. pseudoobscura* and *D. melanogaster* (Seeger and Kaufman, 1990), and of the *hb* gene between *D. virilis* and *D. melanogaster* (Treier *et al.*, 1989). The 40-60 million years that separate these species has resulted in the accumulation of changes in sequence and expression domains that reflect the potential flexibility of the individual components of the hierarchy of genetic interactions. However, both the *D. pseudoobscura bcd* gene and the *D. virilis hb* gene are able to fully substitute for the homologous genes in *D. melanogaster*. (Seeger and Kaufman, 1990; Luk *et al.*, 1994; Lukowitz *et al.*, 1994), so molecular coevolutionary change has not obviously occurred in these genes between these closely related species.

M. domestica and *D. melanogaster* are separated by 100 million years (Hennig, 1981). Despite this, their morphology and early embryology are very similar (Weismann, 1866). Sommer and Tautz (1991) have demonstrated that the primary expression domains of a number of genes of the early developmental hierarchy (including *bcd* and both the maternal and early zygotic components of *hb* expression) are conserved in *M. domestica*, which they argue indicates that the genes are functionally equivalent. There are, however, a number of interesting changes in the secondary expression domains. Therefore, *M. domestica* is sufficiently close in embryology to *D. melanogaster* that homologous genes can be identified and homologous function inferred, but potentially separated by sufficient evolutionary time for molecular coevolutionary change to have occurred.

The interaction between bicoid and *hunchback*

When this project began, the model of the interaction between *bcd* and *hb* could be summarised as follows: a concentration gradient of the maternal morphogen *bcd* determines the anterior-posterior axis, and directs the expression of *hb* in a concentration-dependent manner. *Hb* contains high affinity *bcd* binding sites, and so requires a low threshold level of *bcd* for activation, resulting in *hb* expression over an anterior domain extending over 50% of egg length. Recently published work has suggested that this model is over-simplified, and in particular, that the *hb* protein is itself required for *hb* gene expression (Simpson-Brose *et al.*, 1994). The current understanding of the interaction between *bcd* and *hb* is outlined and discussed, below.

BCD ACTIVATES ZYGOTIC *HB* EXPRESSION

Hb is expressed in a domain covering the anterior half of the embryo, from 100-50% egg length, under the control of *bcd*. The *bcd* protein forms an exponential concentration gradient along the anterior-posterior axis of the embryo. Above a certain

concentration threshold, bcd protein activates *hb* expression by binding to multiple sites within the *hb* promoter; with the number and affinity of the sites determining the extent of the domain of expression (Driever and Nüsslein-Volhard, 1989; Driever *et al.*, 1989; Struhl *et al.*, 1989).

The *hb* gene contains two promoters. The P1 (distal) promoter regulates maternal and later zygotic expression of *hb*, whilst the P2 (proximal) promoter regulates the bcd-dependent early zygotic expression of *hb* (Tautz *et al.*, 1987). *In vitro* footprinting has indicated that the P2 promoter contains a number of binding sites for the bcd protein (Driever and Nüsslein-Volhard, 1989). These are at least partially redundant - reporter genes that contain some but not all of the binding sites are capable of driving reporter gene expression in the proper domain (Driever and Nüsslein-Volhard, 1989; Driever *et al.*, 1989).

MATERNAL EXPRESSION OF *HB*

Maternally transcribed *hb* mRNA is distributed evenly throughout the egg (Tautz *et al.*, 1987; Tautz and Pfeifle, 1989). *Nanos* (*nos*) mRNA becomes localised to the posterior pole of the egg, and translation and diffusion result in the formation of a *nos* gradient from posterior to anterior pole, in a manner analogous to the formation of the bcd gradient (Wang and Lehmann, 1991). *Nos* represses the translation of *hb* mRNA by binding to a *nos* response element (NRE) within the 3' untranslated region (UTR) of the *hb* mRNA (Wharton and Struhl, 1991). This results in the maternally encoded *hb* only being translated in the anterior half of the embryo - the same expression domain as results from the zygotic activation of *hb* expression by bcd (Tautz, 1988).

The maternal expression of *hb* is redundant. Eggs from *hb*⁻ females (and therefore lacking maternally expressed *hb*) can be rescued by a single copy of paternally encoded *hb*. (Lehmann and Nüsslein-Volhard, 1987).

CO-OPERATIVE BINDING OF BCD?

Co-operative binding of bcd to the *hb* promoter has been suggested (Driever and Nüsslein-Volhard, 1989; Struhl *et al.*, 1989; Hanes and Brent, 1989) but has not yet been conclusively demonstrated.

Co-operative binding of bcd proteins to the multiple sites within the *hb* promoter was initially invoked as a mechanism to explain the transformation of the information provided by the smoothly graded distribution of bcd into the more sharply defined domain of *hb* expression (Driever and Nüsslein-Volhard, 1989; Struhl *et al.*, 1989). It is also suggested by experiments involving LexA-bcd fusion proteins. The DNA bound form of LexA-containing proteins is a dimer (Brent and Ptashne, 1981; Brent, 1982), however, the amino terminus of the LexA protein (that part contained within the fusion protein) does not dimerize efficiently, and so the fact that LexA-bcd recognises LexA operators suggests that the bcd moiety contributes to dimerization (Hanes and Brent, 1989). This ability to dimerize, taken in conjunction with the ability of multiple copies of fragments containing a single site to activate transcription, whilst single copies fail (Struhl *et al.*, 1989; Hanes and Brent, 1989, Driever *et al.*, 1989a), suggests the possibility of co-operative binding of bcd oligomers to DNA.

SYNERGISTIC INTERACTION OF BCD AND HB

The initial model of the bcd:*hb* interaction suggested that the expression of *hb* depended on bcd. Recently, work has been published suggesting that another factor(s) may also be involved (Hanes *et al.*, 1994; Simpson-Brose *et al.*, 1994), and that this other factor may in fact be the *hb* protein itself (Simpson-Brose *et al.*, 1994). A number of homeodomain proteins have been shown to bind DNA as either homo- or heterodimers.

Hanes *et al.*, (1994) have demonstrated that artificial promoters containing multiple high affinity bcd binding sites require different site spacings for maximal expression in *D. melanogaster* and in yeast. In *S. cerevisiae*, the preferred site spacing is

11 bp, and in *D. melanogaster* 25 bp. They suggest that this difference could be due to ancillary proteins that might facilitate co-operative interactions but impose different, species-specific spacing requirements. For example, in yeast, activation by *bcd* also requires the SNF2/SWI2 transcription factors (Laurent and Carlson, 1992), which are homologous to the product of the *D. melanogaster* gene, *brahma* (Tamkun *et al.*, 1992). These proteins might impose different site spacings in each organism (Hanes *et al.*, 1994).

Simpson-Brose *et al.* (1994) argue that the inability of reporter gene constructs to replicate the sharp borders of expression seen for *hb*, and the inability of artificial promoters made up of multiple high affinity *bcd* binding sites to drive expression over the full *hb* domain indicates that some component is missing. However, the small domain of expression is also seen when the authentic P2 promoter is introduced into *D. melanogaster*, and may be due to the absence of a P1-regulated stripe that overlaps the domain of *bcd*-dependent *hb* expression (Lukowitz *et al.*, 1994).

A synergistic interaction between *bcd* and *hb* has been demonstrated by transient co-transfection experiments (Small *et al.*, 1991). Simpson-Brose *et al.* (1994) use a system of transgenes in *hb*-flies (both maternal and zygotic components removed) to demonstrate a synergistic effect of *bcd* and *hb* on the expression of reporter genes with artificial promoters containing a series of *hb* binding sites in tandem with a series of *bcd* binding sites. However, as these experiments do not use the native *hb* promoter, they only demonstrate that a synergistic interaction between *bcd* and *hb* can occur, and not that it actually does in the particular case of the *hb* promoter.

Simpson-Brose *et al.* (1994) also examined the expression of *hb* mRNA in a series of embryos lacking maternal *hb*, but with varying doses of zygotic *hb*. In these embryos, an increasingly anteriorly restricted domain of *hb* expression is seen as the zygotic *hb* dose is reduced from two to nil. However, in concluding that this indicates the importance of the *hb* protein in promoting *hb* gene expression, they have neglected the varying contribution of the P1 regulated *hb* stripe that overlaps the posterior border of the P2 (*bcd*-dependent) *hb* expression. This stripe has been demonstrated to be due to

hb auto-regulation (Lukowitz *et al.*, 1994; Hülkamp *et al.*, 1994), and will therefore be present in embryos containing two doses of zygotic *hb*, and absent in those containing no *hb* at all - accounting at least partially for the apparent hb dependence of *hb* expression.

A hb binding site has been identified by *in vitro* footprinting within the cluster of bcd binding sites upstream of the *hb* promoter (Treisman and Desplan, 1989). However, it is only partially conserved in *D. virilis* (12/16 matches, Treier *et al.*, 1989), and yet the *D. virilis hb* gene is correctly regulated when transformed into *D. melanogaster* (Lukowitz *et al.*, 1994). Furthermore, a fragment of the *hb* promoter containing one high and three low affinity bcd binding sites is capable of directing reporter gene expression, whilst one containing two high affinity bcd binding sites and the *hb* binding site is not (Struhl *et al.*, 1989).

Therefore, whilst it has been demonstrated that bcd and hb can act together synergistically, and this may be important in the regulation of downstream genes such as *eve* (Small *et al.*, 1991), a definitive role for this synergism in the activation of early zygotic *hb* expression remains to be demonstrated.

SETTING THE HB EXPRESSION DOMAIN

The extent of the early zygotic domain of *hb* expression has been shown experimentally to depend on the number and affinity of the bcd binding sites within the *hb* promoter. Multimerisation of a DNA fragment from the *hb* promoter containing both a high and a low affinity bcd binding site was found to lead to progressively more posterior limits of expression (Struhl *et al.*, 1989); whilst lower affinity binding sites were found to direct expression over smaller domains, as expression is only activated in regions of high bcd concentration (Driever *et al.*, 1989).

If bcd binding is co-operative, then this may integrate both the number and affinity of the binding sites (Beachy *et al.*, 1994), and so a large number of low affinity binding sites could be functionally equivalent to a promoter containing a small number

of high affinity sites. If a synergistic interaction with *hb* is also required for *hb* expression (discussed above), then the number and affinity of *hb* binding sites might also be involved in setting the posterior border of *hb* expression. The domain of *hb* expression is also determined by the concentration of *bcd*. In embryos containing extra copies of the *bcd* gene, the expression domain of *hb* is extended posteriorly (Driever and Nüsslein-Volhard, 1988).

Therefore, there are at least three components setting the extent of the domain of *hb* expression - the concentration of *bcd*, the number of *bcd* binding sites, and the affinity of the *bcd* binding sites.

In order to investigate the possibility that the *bcd:hb* interaction may have been subject to molecular coevolution between *D. melanogaster* and *Musca domestica*, the *bcd* and *hb* genes have been partially sequenced from *M. domestica*, and compared to those of *D. melanogaster*. From the comparisons it is argued that it is possible that the *M. domestica* *bcd* homeodomain has a slightly changed specificity compared to *D. melanogaster* *bcd*. Examination of sequence upstream of the *M. domestica* *hb* coding DNA reveals no regions of conservation between *M. domestica* and *D. melanogaster*, and no exact matches are found to the *D. melanogaster* consensus *bcd* binding sequence either. Three candidate high affinity *bcd* binding sites are identified within the *M. domestica* *hb* promoter, defining a consensus sequence of TTTAATCC, as opposed to the *D. melanogaster* consensus, TCTAATCC, with each of the *Musca* sites showing the same C to T change. This points towards the possibility that changes in the sites' sequences have elicited compensatory changes in the *M. domestica* *bcd* homeodomain.

Gel retardation assays using *M. domestica* *bcd* homeodomain expressed in *E. coli* are have been used to investigate its relative affinity for the *M. domestica* and *D. melanogaster* consensus *bcd* binding sites *in vitro*, and the *M. domestica* *hb* gene was

transformed into *D. melanogaster* in order to examine the ability of the *D. melanogaster* bcd protein to regulate the expression of the *M. domestica hb* gene *in vivo*.

Chapter 2

GENERAL METHODS & MATERIALS

This chapter provides a description of the general materials and methods used. Details of the exact experimental conditions, and how these techniques were used, are given in the relevant results chapters.

MATERIALS

CONSUMABLES.

All chemicals, reagents and plastic ware used were standard and purchased from recognised suppliers of molecular biology reagents (Applied Biotechnologies Limited, Boehringer Mannheim Biochemica, Fisons, Gibco-BRL, New England Biolabs, Serva, Sigma, Pharmacia and Perkin-Elmer-Cetus) according to cost, availability and applicability.

OLIGONUCLEOTIDES

Hexadeoxyribonucleotides for random priming were supplied by Pharmacia. Oligonucleotides for DNA sequencing and polymerase chain reaction amplification of DNA were synthesised by J. Keyte (Department of Biochemistry, University of Nottingham) and by D. Langton (Department of Biochemistry, University of Leicester), and were ethanol precipitated and dissolved in TE before use. The sequences of the oligonucleotides used are given in Table 2.1

Name	Sequence (5' to 3')	Description
B16P	TTAATGGCAATATTAGG	<i>hb</i> , sequencing primer (Figure 4.5). Nucleotides 2621-2637 in Figure 4.7
B17.1P	AGAGGGAGGTAAAATCCG	<i>hb</i> , sequencing primer (Figure 4.5). Nucleotides 3869-3853 in Figure 4.7
B17.2P	TCCGTGGTTTACAGCGC	<i>hb</i> , sequencing primer (Figure 4.5). Nucleotides 4078-4062 in Figure 4.7
BCD1	AGAATTCACGACGCACCCGTAC	<i>bcd</i> , nucleotides 1113-1229 in Figure 3.7 (includes added 5' <i>Eco</i> RI site)
BCD4	GATCAAGCTTACGATTGAATTTGTGACGTCG	<i>bcd</i> , nucleotides 1292-1271 in Figure 3.7 (includes 3' added <i>Hind</i> III site)
E11	GTTTGGATATAGCTCC	<i>bcd</i> sequencing primer (Figure 3.2). Nucleotides 201-217 in Figure 3.7
E14	AATTCAGAACGGTTATC	Sequencing primer (Figure 3.2). Nucleotides 2413-2397 in Figure 3.7
HB2	GTCCAAAACCATGCGCG	<i>D. melanogaster hb</i> , 1st zinc finger domain, 3' primer. Aligned with <i>Musca</i> nucleotides 5517-5501 in Figure 4.7.
HB3	AGTGCAAGACCTGCGGC	<i>D. melanogaster hb</i> , 1st zinc finger domain, 5' primer. Aligned with <i>Musca</i> nucleotides 5171-5187 in Figure 4.7
MAL	GGTCGTCAGACTGTCGATGAAGCC	forward primer for sequencing inserts in the vector pMALc2.
MD2	GTGTTGCCCGCCATGAATC	<i>hb</i> , sequencing primer (Figure 4.5). Nucleotides 4843-4861 in Figure 4.7
Reverse	AACAGCTATGACCATG	sequencing or amplification of inserts in M13 or Bluescript vectors
TC2P	CTCCTATACATCCTTAG	<i>hb</i> , sequencing primer (Figure 4.5). Nucleotides 319-335 in Figure 4.7
Universal	GTA AACGACGGCCAGT	sequencing or amplification of inserts in M13 or Bluescript vectors

Table 2.1. Oligonucleotide sequences

BUFFERS AND SOLUTIONS

Column wash solution: 20 mM Tris.Cl, 200 mM NaCl, 1 mM EDTA.

Church buffer: 0.5 M Na₂HPO₄, pH 7.2, 1% BSA, 1 μM EDTA, 7% SDS.

Denaturing solution: 1.5 M NaCl, 0.5 M NaOH

DIG hybridisation buffer: 5x SSC, 1% (w/v) blocking reagent (Boehringer Mannheim Biochemica), 0.1% (w/v) N-lauroylsarcosine, 0.02% (w/v) SDS.

GFM buffer: 1.1 M glyoxal (Sigma), 78% deionised formamide, 0.06 x MOPS.

Hepes buffer: 0.1 M Hepes pH 6.9, 2 mM MgSO₄, 1 mM EGTA (stock solution 0.5 M, adjusted to pH 8 with NaOH).

Injection buffer: 1 x injection buffer is 5 mM KCl, 0.1 mM sodium phosphate, (pH 6.8).

In situ hybridisation solution: 50 % formamide, 5 x SSC, 50 μg/ml heparin, 0.1% Tween-20, 100 μg/ml sonicated and denatured salmon sperm DNA. Stored at -20 °C.

Lambda diluent: 10 mM Tris.Cl, 10 mM MgSO₄ (pH 7.5).

Maleic acid buffer: 0.1 M maleic acid, 0.15 M NaCl, pH 7.5.

MOPS buffer: 20 mM MOPS (3-[N-morpholino]-propanesulphonic acid), 5 mM sodium acetate, 0.1 mM EDTA, (pH 7.0).

Neutralising solution: 1.5 M NaCl, 1 M Tris.Cl (pH 8.0).

PBT: PBS + 0.1 % Tween-20.

Phosphate buffered saline solution (PBS): 130 mM NaCl, 10 mM sodium phosphate (pH 7.2).

PP: 4 % paraformaldehyde in PBS.

Protein sample buffer: 1 x protein sample buffer is 0.0625 M Tris.Cl, pH 6.75, containing 2% SDS, 5% mercaptoethanol, 10% glycerol and 0.01% bromophenol blue.

RNA sample buffer: 40% deionised formamide, 50% glycerol, 1 x MOPS pH 7.0, 0.025 % xylene cyanol, 0.025% bromophenol blue.

SM: 0.1 M NaCl, 50 mM Tris·Cl, 0.2% (w/v) MgSO₄·7H₂O, 0.1% (w/v) gelatin (pH 7.5).

SSC (20x): 3 M NaCl, 0.3 M Na₃citrate (pH 7.0).

TAE (50x): 2 M Tris-acetate, 0.05 M EDTA.

TBE (10x): 0.89 M Tris-borate, 2 mM EDTA (pH 8.3).

TE: 1 mM Tris·Cl, 1 mM EDTA (pH 8.0).

TSB: Luria broth (pH 6.1) containing 10% PEG-4000, 5% DMSO and 20 mM Mg²⁺ (10 mM MgCl₂ + 10 mM MgSO₄).

Washing solution: 100 mM NaCl, 50 mM MgCl₂, 100 mM Tris·Cl pH 9.5. Made fresh before use.

MEDIA

Luria broth

1 % (w/v) Bacto-tryptone (Difco), 0.5% (w/v) Bacto-yeast extract (Difco), 1% (w/v) NaCl.

rich broth + glucose & carbenicillin

1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl and 0.2 % (w/v) glucose, autoclaved. Sterile carbenicillin was then added to 100 µg/ml.

NZCYM

1% (w/v) NZ amine, 0.5% NaCl, 0.1% casamino acids, 0.5% (w/v) Bacto-yeast extract (Difco), 0.2% (w/v) MgSO₄·7H₂O (pH 7.5).

egg laying plates

35 g agar was mixed with 1 l dH₂O, and dissolved by autoclaving for 20 minutes. 25 g sucrose was dissolved in 500 mls of dH₂O, and microwaved at full power for 4 minutes. Then 500 mls of apple juice was mixed with the warm sucrose solution. The sucrose solution was then mixed with the molten agar, and 4 g of Nipagin dissolved in 20 mls of 95% ethanol added. Froth was removed from the surface using a paper towel, and the mixture poured into petri dishes. Once set, the egg laying plates were stored at 4 °C. The plates were allowed to warm to room temperature and any condensation removed with a paper towel before using.

BACTERIAL STRAINS

LE392 (Murray, 1977):- F⁻, *hsdR574*, (*r_K*⁻, *m_K*⁻), *supE44*, *supF58*, *lacY1* or Δ (*lacIZY*)₆, *galK2*, *galT22*, *metB1*, *trpR55*.

DH5 α F' (Hanahan, 1983):- F', Φ 80*dlacZ* Δ M15, *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*, (*r_K*⁻, *m_K*⁺), *supE44*, *relA1*, *deoR*, Δ (*lacZYA-argF*)U169.

METHODS

General methods for nucleic acid handling

PHENOL/CHLOROFORM EXTRACTION OF DNA

Phenol/chloroform extraction was used to purify DNA from proteins during the initial extraction process and as an additional purification step following some enzymatic treatments. An equal volume of phenol/chloroform (Phenol:Chloroform:Isoamyl alcohol in the ratio 25:24:1. Phenol with hydroxyquinone added to 0.1%, equilibrated with Tris-Cl pH 8.0) was added to the nucleic acid sample and the mixture emulsified. The mixture was centrifuged to separate the phases and

the upper aqueous layer transferred to another tube leaving behind the proteinacious interface. Further phenol/chloroform extraction's were performed until the interface was clean and the aqueous layer clear. An extraction with chloroform was then performed to remove traces of phenol, and then the DNA was ethanol precipitated.

ETHANOL PRECIPITATION OF DNA

Ethanol precipitation was used to de-salt, concentrate or recover DNA following manipulation. 1/10th volume of 2M sodium acetate (pH 7) and 2 volumes of ethanol or 1 volume of isopropanol were added to the DNA solution. The solution was mixed, then chilled to precipitate the DNA. The DNA was pelleted by centrifugation in a bench top microfuge for 10 minutes at high speed (12 000 rpm). The pellet was then rinsed in 70-80% ethanol, dried and dissolved in the required amount of distilled water

ESTIMATION OF DNA CONCENTRATION

DNA concentration was assayed by measurement of UV absorbance at a wavelength of 260 nm in a spectrophotometer (Ultrospec Plus, LKB), given that 1 A_{260} unit of double-stranded DNA = 50 $\mu\text{g}/\text{ml}$, and 1 A_{260} unit of single-stranded DNA = 33 $\mu\text{g}/\text{ml}$. For smaller amounts of DNA, such as individual restriction digests, concentrations were estimated by visual comparison of aliquots of the samples with DNA of known concentration after agarose gel electrophoresis and ethidium staining.

RESTRICTION ENDONUCLEASE DIGESTS OF DNA

Restriction enzyme digests were performed in the manufacturers supplied buffer at the temperature recommended. Incubations were typically for 1-5 hours with 1 unit of enzyme per microgram of DNA.

AGAROSE GEL ELECTROPHORESIS

Of DNA

Agarose gels in the concentration range 0.8 to 1.5% were typically used, depending on the size of the fragments to be resolved. Gels were run in TAE (40 mM Tris-acetate, 20 mM sodium acetate, 0.2 mM EDTA, pH 8.3). Ethidium bromide was added to the gels at a concentration of 0.5 mg/ml. Gels were either 6 cm or 10 cm long, depending on the separation required. Loading dye (0.1% bromophenol blue, 0.1% xylene cyanol, 30% glycerol) was added to the samples prior to loading. DNA samples were run alongside markers of known molecular weight (either λ cut with *Hin* dIII, range 2-23 kb; or Φ X174 cut with *Hae*III, 70 bp to 1.3 kb range) Current applied to the gel was adjusted to suit a particular run, monitoring of the run could be achieved by following the migration of the bromophenol blue, or visualisation of the DNA by UV fluorescence on a transilluminator. For high quality photography, DNA was visualised by UV fluorescence on a transilluminator, and photographed with a Polaroid MP-4 camera using Kodak negative film (T-max Professional 4052). Films were processed with Kodak LX24 developer, FX40 fixer and HX40 hardener. For much more rapid, lower quality gel documentation, a video imaging system (GDS2000, UVP International) was used.

Of RNA

10 cm long, 1 % agarose gels in 1 x MOPS buffer were used. The gels were run in 1 x MOPS buffer. 1 - 3 volumes of GFM buffer were added to each sample, and then the samples were heated at 55 °C for 20 minutes, before the addition of 0.1 volumes of RNA sample buffer, and then loaded on to the gel. RNA samples were run alongside size markers (Promega) of known molecular weight. The gels were run at around 100 V, until the first of the blue markers was leaving the end of the gel. For photography, the gel was stained in ethidium bromide (0.5 mg/ml in 1 x MOPS) and the RNA visualised

by UV fluorescence on a transilluminator, and photographed as described for DNA above.

ISOLATION OF DNA FROM AGAROSE GELS

To selectively obtain and purify DNA fragments of a certain size prior to ligation, random oligonucleotide labelling etc., the DNA was electrophoresed in an agarose gel until adequate separation was achieved. The gel was viewed under low power UV light, and the slice of gel containing the DNA cut out. The gel slice was then placed in a small basket of 3MM paper inserted into a tube, the bottom of which had been pierced. This tube was then placed inside a second tube, and spun for 1 minute in a microcentrifuge. The DNA was then recovered from the buffer, which collected at the bottom of the second tube, by ethanol precipitation.

SOUTHERN BLOTTING OF DNA (Southern, 1975)

The gel to be blotted was photographed with a ruler along side to record the positions of the DNA markers. Gels were shaken gently for 20 minutes in denaturing solution, rinsed in water, then shaken for 20 minutes in Neutralising solution. If large fragments of DNA (>10 kb) were to be transferred, the gel was depurinated (before denaturation) by gently shaking for 15 minutes in 0.25 M HCl. For capillary blots, gels were then placed on a wick of 3 MM blotting paper (Whatman International Ltd.) previously soaked in 20x SSC over a reservoir of buffer. All of the surface of the wick not covered by the gel was covered by clingfilm, making sure there were no gaps between gel and clingfilm. A prewetted membrane (Hybond-N+, Amersham International), cut to the size of the gel, was placed over it, and followed by two sheets of wetted 3MM and a stack of Kleenex paper towels. Weight was applied to the top of the stack via a glass plate to ensure even pressure onto the gel. Blots were left for 5 to 16 hours. As a rapid alternative, the denatured gel would be blotted under vacuum, using a BioRad vacuum blotter according to the manufacturers instructions. In this

case, transfer would take place in denaturing solution. After transfer, the membrane was rinsed in 2x SSC, and air dried. DNA was fixed by 1 minute exposure to UV light using a UV transilluminator.

NORTHERN BLOTTING OF RNA

The marker lane of the gel to be blotted was cut off, stained with ethidium and photographed as described under "agarose gel electrophoresis of RNA", whilst the remainder of the gel was kept unstained. The RNA was transferred from the unstained gel to a nylon membrane (Hybond-N, Amersham International) by capillary action as described above for Southern blotting of DNA.

LABELLING DNA

radio-labelling

by random primer extension.

200 ng of probe DNA (in a volume of 37.5 μ l) was boiled for 5 minutes, then quench on ice for 1 minute. Then, 10 μ l of oligonucleotide labelling buffer (90 OD units/ml random hexanucleotides in 250 mM Tris-HCl, pH 8.0, 25 mM MgCl₂, 10 mM DTT, 0.1 mM each dATP, dCTP, dGTP, dTTP), 0.5 units of Klenow and 2 μ l of [α -³²P]-dCTP (Amersham, 3000 Ci/mmol, 10 μ Ci/ μ l) were added, and mixed thoroughly by pipetting. The labelling reaction was allowed to proceed at 37 °C for 1 to 3 hours. Unincorporated nucleotides were removed by ethanol precipitation, and the probe was boiled for 5 minutes before use.

by end-labelling

Double-stranded DNA with an appropriate 5' overhang was end-labelled with ³²P α dCTP. 100 ng of DNA was added to 20 μ l of 0.2 mM dATP, 0.2 mM dGTP, 0.2 mM dTTP in 1 x Klenow buffer, and mixed thoroughly by pipetting. After the addition of 2 μ l [α -³²P]-dCTP and 1 unit of Klenow, the labelling reaction was allowed to proceed at

room temperature for 30 minutes. Unincorporated nucleotides were removed by ethanol precipitation, and the labelled DNA pellet resuspended in TE buffer and stored (if necessary) at -20°C for 4 - 6 weeks.

digoxigenin labelling

DNA was labelled with digoxigenin (DIG) using the reagents from the DIG labelling and detection kit (Boehringer-Mannheim), according to the manufacturers instructions.

1 µg of probe DNA was boiled for 10 minutes, and then quickly chilled in an ice/NaCl bath. The chilled DNA was mixed with 20 µl of 1 x hexanucleotide mix, 1 x labelling mix (10 x mix is 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP and 0.35 mM DIG-dUTP) containing 1 unit of Klenow. After brief microcentrifugation, the mix was left at 37 °C over night. The reaction was diluted to 50 µl with TE, and stopped by the addition of EDTA to 20 mM final concentration.

DNA HYBRIDISATION AND DETECTION

With ³²P labelled probes

The filters were prehybridised for 30 minutes at 65°C in 10 mls of Church buffer containing in a bottle rotating in a hybridisation oven (Hybaid). The buffer was then replaced with fresh Church buffer, to which freshly denatured ³²P labelled probe DNA had been added, and the filters hybridised at 65°C overnight.

The filters were washed twice in 2x SSC at room temperature for 10 minutes. Then a series of washes of increasing stringency were performed, with the filters being monitored with a Geiger counter between each wash. Each wash was performed at 65 °C for 20 minutes. Typically, the series would consist of washes in 1x SSC, 0.1% SDS; then 0.5x SSC, 0.1% SDS, then 0.3x SSC, 0.1% SDS, and finally 0.1x SSC, 0.1% SDS. After this, the filter should emit radiation at about background levels. The filters were placed

on a double layer of Whatman 3MM paper to absorb excess liquid (being careful not to let them dry out if the filters were to be re-probed), then sandwiched in SaranWrap, and placed in an autoradiography cassette containing intensifying screens, with a sheet of Fuji RX100 X-ray film, and exposed either at -80°C with an intensifying screen or at room temperature without. The length of exposure varied from 1 hour to 14 days, depended on the estimated signal strength and the band intensity required.

With DIG labelled probes

DNA hybridisation and detection with DIG labelled probes was performed using reagents from the DIG labelling and detection kit (Boehringer Mannheim), according to the manufacturers instructions.

The filters were prehybridised for at least 1 hour at 65 °C in 20 mls of DIG hybridisation buffer per 100 cm² of filter contained in a bottle rotating in a hybridisation oven (Hybaid). The solution was then replaced with 2.5 mls/cm² of DIG hybridisation buffer containing approximately 20 ng/ml of freshly denatured DIG-labelled probe DNA, and hybridised at 65 °C for at least 6 hours.

The filters were then washed twice, for 5 minutes each, in 2 x SSC, 0.1% SDS at room temperature, then twice, for 20 minutes each, in 0.1 x SSC, 0.1% SDS at 68 °C. For detection, the filters were first washed briefly in maleic acid buffer + 0.3 % (w/v) Tween-20, then incubated for 30 minutes with 1 ml/cm² of 1% blocking reagent (Boehringer Mannheim Biochemica) in maleic acid buffer. The filters were then incubated for 30 minutes with 150 mU/ml of antibody-conjugate (polyclonal anti-digoxigenin Fab-fragments, conjugated to alkaline phosphatase, Boehringer-Mannheim Biochemica) diluted in 1% blocking reagent in maleic acid buffer. Unbound antibody-conjugate was removed by washing the filters twice for 15 minutes in maleic acid buffer + 0.3% (w/v) Tween-20. The filters were equilibrated for 2 minutes in washing solution, then incubated with 0.1 mls/cm² of colour solution (45 µl NBT-solution, 35 µl X-phosphate (both from Boehringer Mannheim) in 10 mls washing solution) in the dark

until the desired bands were detected. The reaction was then stopped by washing the filters for 5 minutes in TE, and the results documented by photography of the filter.

SUBCLONING PROCEDURES

Nested Deletions

Nested deletions (Henikoff, 1984) were produced by the digestion of plasmid DNA with exonuclease III. Plasmid DNA, cut with restriction enzymes to produce one end with a protruding 5' overhang (the deleted end) and one protruding 3' overhang (the protected end) was digested with 50 u μg^{-1} exonuclease III at 37 °C for 1-20 minutes, in 66 mM Tris·Cl pH 8.0, 0.66 mM MgCl_2 . Samples were then digested with mung-bean nuclease (5 unit μg^{-1}) for 1 hour at 37 °C. Each sample was then loaded onto a 0.8% low melting temperature agarose for electrophoresis. Each lane contained a major band, and smears above and below. The major bands were excised from the gel; and after the addition of 2 volumes of TE, the gel slices were melted at 65 °C. The DNA was then used directly in for filling in with Klenow polymerase at 37 °C, followed by ligation overnight.

Ligation

Ligations were essentially done according to Sambrook *et al.* (1989). DNA ligation reactions were carried out in 50 mM Tris·Cl pH 7.6, 10 mM MgCl_2 , 5% PEG-6000, 1 mM ATP and 1 mM DTT, to which vector and insert DNA had been added in various molar ratios (typically 1:3 for cohesive ends). 1 unit T4 DNA ligase was then added, and the ligation allowed to proceed for 3 hours to overnight at room temperature (cohesive ends), or overnight at 16 °C (blunt ends).

If the vector had been digested with a single restriction enzyme, it was treated with calf intestinal alkaline phosphatase (CIAP) to prevent self-ligation. Phosphatase reactions were carried out in 50 mM Tris·Cl, pH 9.0, 1 mM MgCl_2 , 0.1 mM ZnCl_2 and 1 mM spermidine to which 0.01 u/pmol ends of CIAP was added, and were allowed to

proceed for 60 minutes at 37 °C. The reaction was stopped by the addition of EDTA to a final concentration of 2.5 mM. The CIAP was removed from the reaction by phenol/chloroform extraction, or the gel purification of the vector DNA prior use in ligation reactions.

Transformation

Transformation of *E. coli* was performed as described by Chung and Miller (1988).

Bacterial cells were grown to the early log phase ($OD_{600} = 0.3-0.6$) in LB broth, and the pelleted by centrifugation ($1000 \times g$ for 10 minutes at 4 °C). The cells were then resuspended in 0.1 volumes of transformation and storage buffer (TSB) at 4 °C and incubated on ice for approximately 10 minutes. 0.1 ml aliquots of the cells were pipetted into cold polypropylene tubes and mixed with 100 pg of plasmid DNA, and then returned to ice for 5-30 minutes, then plated on antibiotic-containing agar plates for the selection of transformants.

For cloning experiments into Bluescript, ligation mixes were transformed into *E. coli* DH5 α F' and plated onto media containing carbenicillin, for selection, and X-gal and IPTG for the identification of recombinant clones. An insert interrupts the vector's β -galactosidase gene, and so the chromogenic substrate X-gal cannot be converted to a blue product; white colonies should therefore be recombinant.

Nucleic acid preparation

ISOLATION OF GENOMIC DNA

Flies were ground to a fine powder under liquid nitrogen using a pestle and mortar. The powder was then transferred to a 15 ml falcon tube. 5 mls of homogenisation buffer added, and the powder and buffer mixed gently but quickly, by inversion of the tube. 250 μ l of 10% SDS was then added, and mixed in gently by

inversion. The resulting viscous solution was then incubated with Proteinase K at a final concentration of 0.1 mg/ml at 65 °C for 2 hours. The solution was then extracted with an equal volume of Tris-buffered phenol (pH 7.5), and the aqueous and organic layers separated by centrifugation at 1000 x g for 10 minutes at room temperature. The aqueous layer was then transferred to a clean tube using an inverted 25 ml glass pipette, with care being taken to avoid disturbing the interface. The extraction was repeated using an equal volume of chloroform, and after centrifugation, the aqueous layer was again transferred to a clean tube. The DNA was then ethanol precipitated, pelleted by centrifugation (1000 x g, 10 minutes). The pellet was washed in 70% ethanol, air-dried and resuspended in TE.

ISOLATION OF RNA

The method used for the extraction of RNA is based on that of Chomczynski and Sacchi (1987) with modification by Puissant and Houdebine (1990). All glassware was baked in an oven for 12 hours prior to use (bottle tops were autoclaved). Pasteur pipettes were siliconised (covered in a siliconising solution (2% dichloromethylsilane, BRL), left for 1 hour in a fume hood, then rinsed extensively, first in tap water, then distilled water) before baking. All solutions (except protein solutions, which were just autoclaved) were treated with DEPC, and made with DEPC-treated water (1 ml of DEPC (diethyl pyrocarbonate; Sigma) was added to every litre of H₂O, and allowed to stand overnight at room temperature, before autoclaving).

Flies were homogenised in 4 mls of Solution D (100g of guanidinium thiocyanate dissolved in 117.2 ml DEPC-treated water, 7.04 ml 0.75 M sodium citrate and 10.76 ml 10% sarcosyl; to which 0.36 ml/ 50mls β-mercaptoethanol has been added), and then 400 µl of 2 M sodium acetate, 4 ml water saturated phenol and 800 µl of chloroform-isoamyl alcohol (24:1) were added, and mixed by inversion. After 10 second vortex, followed by 15 minutes cooling on ice, the solution was transferred into siliconised 30 ml corex tubes then centrifuged at 10 000 rpm for 20 minutes at 4 °C in a Sorval RC3C centrifuge, using an SS34 rotor. The aqueous phase was removed, and filtered through

a Millipore filter into a siliconised 15 ml corex glass tube. 1 volume of isopropanol was then added, mixed in well, and the sealed tube placed a -20 °C overnight.

The RNA was then pelleted by centrifugation at 10 000 rpm for 20 minutes. The supernatant was removed, and the pellet resuspended in 4 ml 4 M LiCl. The RNA was then pelleted again by centrifugation at 10 000 rpm for 20 minutes, and this time resuspended in 2 mls of 10 mM Tris-Cl (pH 7.5), 1 mM EDTA, 0.5% SDS. 2 mls of chloroform was then added, mixed by vortex and the phases separated by centrifugation at 10 000 rpm for 10 minutes. The aqueous phase was collected into a siliconised 15 ml corex tube, 1 volume of isopropanol was then added, mixed in well, and the sealed tube placed at -20°C overnight.

The RNA was pelleted by centrifugation at 10 000 rpm for 20 minutes, and then resuspended in 1 ml 75% ethanol, and transferred to microcentrifuge tube. The RNA was re-pelleted in a microcentrifuge for 30 minutes at 4 °C, and the supernatant removed. The pellet was air-dried and then resuspended in approximately 500 µl of dH₂O, and stored at -70 °C.

ISOLATION OF PLASMID DNA

Small scale

1 ml of an overnight culture of cells containing the plasmid was transferred to a microcentrifuge tube, and the cells pelleted by centrifugation for 1 minute. The cells were resuspended in 200 µl of 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, and then lysed with 450 µl of 0.3 M NaOH, 1% SDS. 350 µl of cold 1.32 M potassium acetate (pH 4.8) was added, and mixed gently by inversion. The resulting mixture was then incubated on ice for 5 minutes. The white precipitate was pelleted by microcentrifugation at high speed for 10 minutes, and the supernatant decanted into a fresh tube. The cleared supernatant was then extracted with 600 µl of chloroform:isoamyl alcohol (24:1), mixed by vortex. The organic and aqueous layers were separated by centrifugation (1 minute at high speed) and the aqueous layer transferred to a new tube. The plasmid DNA was

precipitated by the addition of 1 ml of isopropanol, and pelleted by centrifugation. The pellet was washed with 70% ethanol, allowed to air-dry, and resuspended in 50 µl TE.

Large scale

600-1000 ml of an overnight culture of cells containing the plasmid of interest were pelleted by centrifugation at 14 000 x g for 10 minutes at 4 °C. The supernatant was discarded and the pellet resuspended in 30 mls of 50 mM Tris·HCl (pH 7.5), 10 mM EDTA and 100 µl RNase A. 30 mls of Cell Lysis solution (0.2M NaOH, 1% SDS) was then added, and mixed gently, but thoroughly, by inversion until the solution became clear and viscous. 30 mls of 1.32 M potassium acetate (pH 4.8) was then added and mixed immediately by inversion of the centrifuge bottle several times. The resulting white precipitate was pelleted by centrifugation (14 000 x g for 15 minutes at 4 °C), and the cleared supernatant decanted into a clean centrifuge bottle. 0.6 volumes of isopropanol were then added to the decanted supernatant, and mixed by inversion. After centrifugation at 14 000 x g for 15 minutes at 4 °C, the supernatant was discarded and the DNA pellet resuspended in 5 ml of TE buffer. 20 mls of Wizard Megapreps DNA purification resin (Promega) was then added to the DNA solution, and mixed by swirling. The tip of a Wizard Megacolumn (Promega) was inserted into vacuum source, and the DNA/ resin mix transferred to the column. A vacuum was applied to pull the DNA/resin mix into the column. The column was then rinsed with 50 mls of plasmid column wash solution (125 parts of 200 mM NaCl; 20 mM Tris·HCl, pH 7.5; 5 mM EDTA to 170 parts of 95% ethanol), using the vacuum to draw the wash solution through the column. The resin was then rinsed with 10 mls of 80% ethanol, and then dried by continuing to draw a vacuum for an additional 10 minutes. 3 mls of preheated (70 °C) TE buffer was then applied to the column, and after a 1 minute wait, the DNA eluted by spinning the column, in its reservoir, at 1300 x g for 5 minutes. The column was then discarded, and the plasmid DNA solution transferred to microcentrifuge tubes for storage at -20°C.

Single-stranded DNA

Single stranded DNA was isolated from Bluescript plasmids by taking 15 µl of an overnight culture of cells containing the plasmid, mixing with 25 µl of Luria broth in a microcentrifuge tube, and adding 1 µl (where the titre $\geq 1 \times 10^{10}$ phage/ml) of VCS-M13 helper phage (Stratagene). The phage were allowed to infect the cells for 15 minutes at room temperature, and then the infected culture was transferred to a test-tube containing 2 mls of Luria broth supplemented with carbenicillin and kanamycin. The culture was then incubated at 37 °C for 12 hours, with vigorous shaking. 1.5 mls of the culture was then transferred to a microcentrifuge tube, and centrifuged at high speed for 30 seconds. 1 ml of the supernatant was transferred to a fresh tube, and the phage precipitated by the addition of 150 µl of 2.5 M NaCl, 20% PEG-6000 followed by 15 minutes incubation on ice. The phage were then pelleted by 5 minutes microcentrifugation. All traces of the supernatant was removed using a Pasteur pipette with a fine, drawn-out tip. The phage pellet was then resuspended in 400 µl of 0.3 M sodium acetate, 0.1 mM EDTA. The solution was then extracted with an equal volume of phenol/chloroform (1:1 mixture), mixed by vortex for 1 minute, and the phases separated by microcentrifugation for 1 minute. The aqueous layer was transferred to a clean tube, and the DNA precipitated by incubation on ice for 20 minutes with 1 ml of ethanol. The DNA was pelleted by microcentrifugation for 15 minutes at high speed. After washing with 70% ethanol, the pellet was air-dried and resuspended in 30 µl TE buffer.

ISOLATION OF PHAGE DNA

Plate lysate method

100 µl aliquots of phage containing about 1×10^5 pfu were incubated at 37°C for 20 min. with equal volumes of LE392 bacteria (treated as described for library screening). The concentration of phage should be sufficient to result in confluent plaques. Each aliquot of phage and bacteria was then mixed with 3 mls of Luria Broth

top agarose (melted, and kept at 45°C), and overlaid onto a fresh Luria Agar plate. Top agarose was used rather than top agar, as agar contains inhibitors that interfere with restriction enzyme digest. The plates were inverted and incubated at 37°C overnight, or until the plaques became confluent. Each plate was then overlaid with 5 mls of λ diluent, and shaken gently for 1-2 hours to elute the phage. The buffer was then transferred to a 15 ml Sarstedt tube, and centrifuged at 2000 rpm of 10 minutes at 4°C to remove the debris. The supernatant was transferred to an Oakridge tube and incubated for 1 hr at 37°C with 1 μ l of 10 mg/ml RNase and 1 μ l of 1 mg/ml DNase 1. An equal volume of a solution containing 20% (w/v) PEG 8000, 2 M NaCl in λ diluent was added, mixed by vortexing briefly, and then the mixture was incubated on ice for 1 hr. The phage were then pelleted by centrifugation at 10 000 rpm at 4°C for 20 minutes, using a Sorvall centrifuge and SS34 rotor. The supernatant was discarded, and the phage resuspended in 0.5 mls TE, using a vortex mixer. The solution was transferred to an eppendorf tube, and incubated at 65°C for 5 minutes with 5 μ l of 10% SDS. Then 10 μ l of 5 M NaCl was added, and the solution phenol/chloroform extracted. The DNA was then precipitated by adding of an equal volume of ice cold isopropanol and standing at -70°C for 20 minutes. The DNA was recovered by spinning in a microcentrifuge for 15 minutes at 4°C. The DNA pellet was washed in 70% ethanol, and air dried, then resuspended in 50-100 μ l of TE.

Liquid culture lysate method

A single fresh phage plug was added to 10 mls of Luria broth supplemented with 0.2% maltose and 10 mM Mg_2SO_4 . To this was added 100 μ l of fresh LE392. The culture was incubated at 37°C with vigorous shaking for 5 hours. If lysis was not seen, a further 10 mls of Luria broth plus 0.2% maltose and 10 mM Mg_2SO_4 was added, and the incubation continued overnight, or until lysis occurred. Once the cultures had lysed, a drop of chloroform was added and the culture shaken briefly to kill the bacteria. Then it was transferred to an Oakridge tube, and from this point treated as described for the plate lysate method, scaling volumes proportionately.

Polymerase chain reaction

STANDARD

PCR reactions were carried out in 50 μ l volume of 10 mM Tris·Cl pH 8, 1.5 mM MgCl₂, 50 mM KCl, 0.01% w/v gelatin, 0.01% w/v Triton-X (Sigma) to which 0.2 mM of each dNTP, 2 units of Taq polymerase and of the primers were added. The mixture was overlaid with 40 μ l of mineral oil. After an initial 1 minute at 96 °C, 25 cycles of amplification were carried out using a step program (94°C, 40 sec; 50°C, 2 min.; 72°C, 3 min.), followed by a 15 minute final extension at 72°C. The PCR reactions were carried out using a DNA thermal cycler (Perkin-Elmer-Cetus).

Colony and plaque screening

LIBRARY SCREENING

This was basically done according to Sambrook *et al.* (1989)

Preparation of the bacteria

The *E. coli* strain LE392 was used for plating libraries. An overnight culture was grown in Luria Broth supplemented with 0.2% maltose. The cells were pelleted by centrifugation (1000 x g for 10 minutes). The supernatant was poured off and the cells resuspended in 0.25 volumes of 10 mM MgSO₄.

Titration of the library

The titre of phage in the library was calculated by making serial dilutions of the library in SM buffer. 100 μ l of each dilution, and a negative control consisting of SM buffer only, was mixed with 200 μ l of an overnight culture of LE392 and incubated in an

ependorf tube at 37°C for 20 minutes to allow the phage to infect the bacteria. The contents of each tube were then added to a small test tube containing 3 mls of melted NZCYM top agar at 45° C, and rapidly mixed, then overlaid onto NZCYM plates (prewarmed to 37°C). The plates were left at room temperature for five minutes to set, then inverted and incubated at 37° C overnight. The titre of the library was then calculated by counting the number of plaques formed for a given dilution.

Plating the library

The library was plated essentially as described for titration, except that it was plated on either 10x10 cm, or 23x23 cm, square plastic petri dishes (Nunc). Large square plates were used instead of ordinary petri dishes as it results in a much smaller number of filters, making the subsequent steps easier. The library was plated on NZCYM. An aliquot of the library was diluted, and 500 µl of this dilution mixed with an equal volume of a fresh bacterial culture was added to 30 mls of NZCYM top agar and overlaid onto each 23x23 cm plate (volumes scaled down accordingly for the 10x10 cm plates). The dilution was calculated to give a density of 150 pfu's/cm². This gives single plaques at high density. The number of plates was calculated to ensure that at least 3x the genome size was plated, in order to ensure that the screen was representative. After overlaying, the plates were left to harden at room temperature for 30 minutes, then inverted and incubated at 37°C for approximately 8 hours, until the plaques had formed.

Making the filters

The plates were cooled at 4°C for 30 minutes. This reduces the stickiness of the top agar, making it less likely to adhere to the filters when they are lifted. 20x20 cm² pieces of Hybond N+ were cut, and marked with three crosses in an asymmetric triangle using a laundry marker pen. The filters were laid carefully over the plates, with the marked side down, and left cooling for a further 10 minutes. The positions of

the crosses were marked on the underside of the plates, to allow subsequent re-orientation of filters with plates. The filters were carefully peeled back from the plates, using tweezers to lift back one corner first. They were then denatured by placing marked side down in denaturing solution for 3 minutes, neutralised in neutralising solution for 5 minutes, and then twice washed briefly in 2x SSC. The filters were sandwiched between paper towels to remove excess 2x SSC, then placed between fresh towels and fixed by baking at 80°C for 2 hours, and then hybridised with an appropriate probe.

Positives, and rescreening

The film was developed and all possible positive signals identified. False positives corresponding to bits of dirt on the filters were ignored. The location of all other potential positives was noted, and aligned with the library plates. Using a sterile spatula, a (roughly 5 mm x 5 mm) rectangle of the top agarose surrounding the location of each potential positive was removed and placed individually in an eppendorf tube containing 0.5 mls of SM buffer. The phage were left to elute at room temperature for about 1 hr. A 1:100 dilution was made of each eluted potential positive and 1, 10 and 90 µl of this dilution incubated with LE392 and overlaid onto NZCYM petri dishes as described for library titration. After plaque formation, the dilution of each potential positive that gave in the region of 200 plaques was selected, and filters lifted and hybridised as described for the initial library screen. At this stage a positive will give an above background signal, and it should be possible to clearly align the positive signal to a plaque on the plate. If this is the case, the plaque is removed as a plug using a Pasteur pipette, then expelled into 0.5 mls of SM buffer and eluted for 1 hr at room temperature, or overnight at 4°C. As phage will diffuse through the agar plate, it is necessary to further purify the positive phage by repeated plating, screening and re-picking until all the phage on a given plate are positives.

COLONY SCREENING

Recombinant plasmid clones were identified by colony hybridisation as described by Buluwela *et al.* (1988). Bacterial colonies were lifted from agar plates onto Hybond N+ (Amersham International) by placing dry filters onto plates to contact colonies, and peeling off the filter. The filters were then laid, colony side up, on Whatman 3MM paper soaked in 2 x SSC, 5% SDS, and left for 2 mins. The dish with the filters was then transferred to a microwave oven with a rotating turntable and treated for 2.5 minutes at full setting (650 watts) to lyse the cells, denature the DNA, and fix the DNA to the membrane. Filters were then subjected to standard prehybridisation for 30 mins. and hybridisation for 4 hours at 65 °C, followed by normal washes.

Sequencing and analysis.

SEQUENCING

Manual

Sequencing reactions were performed using the chain-termination method of Sanger *et al.* (1977), using single or double stranded template DNA.

0.5 -1.0 pmol of sequencing primer was annealed to 3-5 µg of single stranded DNA in 10 µl 1 x sequencing reaction buffer (40 mM Tris-HCl pH 7.5, 20 mM MgCl₂, 50 mM NaCl), by heating for 2 minutes at 65 °C, cooling slowly to 35 °C over 15 minutes then chilling on ice. For double stranded DNA, primer and template were heated to 97 °C for 5 minutes, in the absence of reaction buffer, which was added after cooling slowly to 35 °C over 15 minutes.

The ice cold annealed DNA was added to 1 µl of 0.1 M DTT, 2 µl of a 1:5 dilution of Labelling Mix (7.5 µM dGTP, 7.5 µM dCTP, 7.5 µM dTTP), 0.5 µl of [³⁵S] dATP, and 3.2 units of T7 DNA polymerase in a final volume of 16.5 µl, mixed, and incubated at room temperature for 2-5 min., to label. 3.5 µl of the labelling reaction was transferred to each of four tubes, each of which contained 2.5 µl of a single Termination Mixture (50 mM NaCl, 80 µM of each dNTP, plus 8 µM of either ddGTP, ddATP, ddTTP or

ddCTP). The termination reactions were incubated at 37 °C for 5 minutes, and then stopped with the addition of 4 µl of Stop Solution. Samples were heated to 75 °C for 2 minutes prior to loading 2-3 µl of each onto a sequencing gel.

Initial sequencing reactions were electrophoresed on 5% denaturing polyacrylamide sequencing gels using a Sequi-Gen (Bio-Rad) nucleic acid sequencing cell, for 2-5 hours. Typically, these sequencing reactions were then repeated, with a portion of each of the termination mixes replaced with extension mix (180 µM each dATP, dCTP, dTTP and dGTP, 50 mM NaCl), and run on a wedge gel for an extended gel run time, in order to obtain the maximum length of reliable sequence data.

Automated

Taq cycle sequencing termination PCRs were carried out using the reagents from the Taq DyeDeoxy Terminator Cycle Sequencing kit (Applied Biosystems Inc.), and according to the manufacturers instructions.

Sequencing reactions contained 500 ng of single stranded DNA and 0.8 µM primer, or 1 µg of double stranded DNA and 3.2 µM primer in 20 µl of reaction mixture, overlaid with a drop of mineral oil. The reaction mixture contained 1 µl each of the G, A, T and C DyeDeoxy Terminators and 4 units of AmpliTaq polymerase in 20 µl of 7.5 µM each dATP, dTTP, dCTP, 37.5 µM dITP, 80 mM Tris.Cl, 1 mM MgCl₂, 20 mM (NH₄)₂SO₄ pH 9.0. 25 PCR cycles of 96 °C (30 seconds), 50 °C (15 seconds), 60 °C (4 minutes), were performed on a Model 480 Perkin-Elementer-Cetus DNA thermal cycler, with 1 °C/s ramp rate between all temperatures.

Excess DyeDeoxy terminators were removed by phenol/chloroform extraction, followed by ethanol precipitation. The resulting pellet was dried under vacuum, and sent to the Protein and Nucleic Acid Chemistry Laboratory (University of Leicester) for electrophoresis on an Applied Biosystems Inc. Model 373A automated DNA sequencer.

COMPUTER ANALYSIS

Unless otherwise specified, DNA sequences were analysed using a Silicon Graphics Inc. 4D/480S mainframe computer system running IRIX, using the Genetics Computer Group Sequence Analysis Software Package version 7 developed at the University of Wisconsin (Genetics Computer Group, 1991).

Protein methods

SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS

Electrophoresis of protein extracts and purified proteins was performed using a Mighty Small II slab gel electrophoresis unit (Hoefer Scientific Instruments) under denaturing conditions using a discontinuous buffer system based on that of Laemmli (1970).

The resolving gel consisted of 10% acrylamide in 0.375 M Tris·Cl pH 8.8, 0.1% SDS, with ammonium persulfate and TEMED added to 0.07% v/v and 0.1% w/v respectively to initiate and catalyse the polymerisation reaction. The stacking gel consisted of 4% acrylamide in 0.125 M Tris·Cl pH 6.8, 0.1% SDS, with ammonium persulfate and TEMED added to 0.05% w/v and 0.05 % v/v respectively. The gels were run in 1 x Tris-glycine buffer (0.025 M Tris pH 8.3, 0.192 M glycine, 0.1% SDS).

An equal volume of 2 x protein sample buffer (1 x buffer is 62.5 mM Tris·HCl, pH 6.75 containing 2% SDS, 5% mercaptoethanol, 10% glycerol and 0.001% bromophenol blue) was added to each sample, and the sample boiled for 5 minutes before loading. Electrophoresis was carried out a 20 ma constant current, until the blue dye reached the end of the gel (approximately 1 hour). The gels were then stained in 0.125 % Coomassie Blue (Sigma), 50% methanol, 10% acetic acid with gentle shaking for 1-4 hours, and destained in 10% methanol, 10% acetic acid for several hours.

PROTEIN EXPRESSION AND PURIFICATION

Pilot experiment

80 mls of Rich Broth + glucose + carbenicillin was inoculated with 0.8 mls of a fresh overnight culture of DH5 α F' cells containing the expression plasmid. The culture was grown at 37 °C with shaking, until it reached A₆₀₀ of about 0.5. A 1 ml sample was taken, and microcentrifuged for 2 minutes. The supernatant was removed, and the pelleted cells resuspended in 50 μ l of protein sample buffer and stored at -20 °C (uninduced cells).

Protein expression was then induced by adding IPTG to a final concentration of 0.3 mM, and the culture incubated for a further 2 hours at 37 °C. After this incubation, a 1 ml sample of the induced cells was taken, microcentrifuged for 2 minutes, and the pellet resuspended in 100 μ l of protein sample buffer, then stored at -20 °C.

The remaining culture was split into two. The cells in one portion were pelleted by centrifugation at 4000 \times g for 10 minutes at 4 °C (Sorvall, SS34 rotor). The pelleted cells were resuspended in 5 mls of column buffer, and frozen overnight at -20 °C, then thawed slowly in cold water. Once thawed, the cells were placed in an ice water bath, and sonicated in pulses of 15 seconds (in this case, 7 \times 15 sec pulses) until the released protein reached a maximum as determined by the Bradford Assay.

The sonicated cells were centrifuged at 9000 \times g, at 4 °C for 20 minutes. The supernatant, which contains the soluble material, was decanted and saved on ice. This is the crude extract. The pellet (which contains the insoluble material) was resuspended in 5 mls of column buffer.

A sample of the fusion protein was purified from the crude extract by extraction with amylose resin. A 200 μ l aliquot of amylose resin (New England Biolabs) was placed in an eppendorf tube, and briefly centrifuged. The supernatant was removed, and the resin resuspended in 1.5 mls of column buffer. The resin was centrifuged, the supernatant discarded, and the resin resuspended in a further 1.5 mls of column buffer. This was repeated once more, and the resin resuspended in 200 μ l column buffer. 50 μ l

of the resuspended resin was mixed with 50 μ l of the crude extract and incubated on ice for 15 minutes, then microcentrifuged for 1 minute. The supernatant was then discarded, and the resin resuspended in 1 ml of column buffer. The resin was then pelleted again, by microcentrifugation for 1 minute, the supernatant discarded, and the resin resuspended in 50 μ l of protein sample buffer.

10 μ l of each of the samples of uninduced cells, induced cells, crude extract, insoluble material and purified protein were run on a 10% SDS-PAGE gel to confirm that this system would produce soluble, purifiable bcd fusion protein.

Large scale preparation

500 mls of Rich Broth + glucose was inoculated with 6 mls of a fresh overnight culture of pMALBCD1 in DH5 α F'. The culture was grown to A₆₀₀ of about 0.5, and then expression of the fusion protein was induced by addition of IPTG to a final concentration of 0.3 mM. The cells were pelleted by centrifugation at 4000 x g for 20 minutes (Sorvall, GS4 rotor). The supernatant was discarded, and the pelleted cells resuspended in 25 mls of column buffer. The resuspended cells were frozen overnight at -20 °C, then thawed slowly in an ice-water bath.

The thawed cells were kept on ice whilst they were sonicated in pulses of 15 seconds. Protein release was monitored using the Bradford Assay, and sonication continued until this had reached a maximum - which was found to be after 7 pulses. The sonicated cells were centrifuged at 9000 x g for 30 minutes (Sorvall SS34 rotor) to pellet the insoluble material. The supernatant (the crude extract) was decanted.

2 mls of amylose resin (New England Biolabs) was poured into a column (BioRad), and allowed to settle, then rinsed with 8 column volumes of column buffer. A 1:5 dilution of crude extract in column buffer was loaded onto the column, and then the column was rinsed with 10 column volumes of column buffer. The fusion protein was eluted with column buffer + 10 mM maltose. Fractions of 0.5 mls were collected, and assayed for protein using the Bradford Assay.

BRADFORD ASSAY FOR PROTEIN QUANTITATION

The Bradford assay was used to determine protein concentrations. A series of dilutions of a 10 mg/ml stock solution of BSA were generated (200, 100, 75, 50, 25 and 10 µg/ml), and 800 µl of each of these was mixed with 200 µl of Bradford reagent (BioRad), vortexed, and allowed to stand for 5 minutes. The absorbance at 595 nm was read against a blank of 0.8 ml water and 200 µl reagent. These values were used to plot a standard curve of absorbance against µg protein from which the protein content of a similarly constructed series of dilutions of the protein sample could be read off.

Drosophila methods

MICROINJECTION OF EMBRYOS

Egg collection

3-6 day old flies were placed in clean glass tubes, plugged at the top with a foam bung. A fine piece of net mesh was stretched over the bottom ends of the tubes, and fixed in place with elastic bands. The tubes were placed upright on egg laying plates. The flies lay their eggs onto the plates through the mesh, but as the mesh prevents the flies escaping they can easily be transferred to fresh plates. The first collection of the day was discarded, since the females may store their fertilised eggs for several hours before oviposition, and so these will be too advanced in development for use in transformation. Subsequently, collections were made at 30-60 minute intervals.

dechoriation (manual)

A square of double sided sticky tape (3MM) was fixed to a cover slip, and the cover slip mounted onto a microscope slide using a drop of water. The freshly laid eggs were scraped from the egg laying plates using a mounted needle, and dechorionated manually by rolling them gently on the sticky tape, under a dissecting

microscope. Each dechorionated egg was then transferred to the right hand edge of the sticky tape, and aligned with their posterior ends hanging over the edge of the tape. Eight minutes was spent dechorionating and aligning a batch of embryos on a single cover slip, these were then allowed to desiccate for a further 2 minutes, before covering the eggs with Voltalef oil (Grade 10S). In this way a batch of eggs with a graded degree of desiccation was produced. A graded distribution ensures that at least some eggs within each batch are correctly desiccated, whereas with chemical dechorionation batches of uniformly incorrectly desiccated embryos may be produced.

Microinjection

Microinjection was performed using an Olympus CK2 inverted microscope and Narishige micromanipulator. Glass microcapillary tubes (GC100TF-15, Clark Electromedical Instruments) were pulled into fine pointed needles using a Flaming Brown micropipette puller (Model P80/PC, Sutter Instruments Co.) by M. Hennessey. The needle was mounted in the micromanipulator, and an open bevelled edge produced by scraping the needle point along the edge of a glass cover slip fixed to a microscope slide.

The microscope slide (with cover slip of oil covered embryos) was mounted on to the stage of the microscope with the posterior ends of the embryos towards the needle, and the midlines of the embryos brought into clear focus. The micromanipulator was used to bring the needle into the field of view, and into the plane of clear focus. Once the needle and the eggs were aligned, the embryos vitelline membrane was pierced with the needle by moving the stage, and DNA injected into the posterior region of the embryo.

The needle was attached to a mounted glass syringe (Narishige) by an air-filled plastic tube. The syringe plunger is fixed in place with a screw mechanism, and by adjusting its position it is possible to produce a pressure sufficient to constantly dribble

DNA from the needle tip. When the needle is inserted into the embryo it is possible to see the DNA flowing from the needle tip into the egg, with the amount of DNA depending on the length of time the needle was inserted into the egg, and the degree of desiccation. Any embryos which had developed to cellular blastoderm stage or beyond were not injected, and were destroyed.

Once all the embryos had been injected, and the uninjected ones destroyed with a mounted needle, the embryo laden cover slip was detached from the microscope slide and placed on a petri dish of fly food, dyed red with food colouring. A fragment of capillary tube was placed on the cover slip, just in front of the row of injected embryos, and a few more drops of oil added. This ensures that the embryos are covered with sufficient oil to prevent desiccation. The embryos were then incubated at 18 °C. When they hatch, about 2 days later, they were gently scooped up using a mounted needle, and transferred to vials of fly medium, and incubated at 18 °C, until they eclose. The red colouring of the food in the petri dish is to enable larvae that wander off the coverslips to be easily located, as otherwise they are camouflaged by the food.

***IN SITU* HYBRIDISATION OF EMBRYOS**

Dechoriation (chemical)

Eggs were collected on fresh apple juice agar plates over a period of 4 hours. The eggs were carefully scraped from the plates using a mounted needle, and transferred to small wire baskets in a drop of water. They were then dechorionated for 5 minutes in 3% sodium hypochlorite solution, and then washed in 0.1 % Triton X-100 (Sigma) and rinsed in water.

Formaldehyde fixation

Dechorionated embryos were transferred to 4 mls of Hepes buffer in a glass vial, and 0.5 mls of 37 % formaldehyde solution (BDH) and 5 mls of heptane were added.

The vial was then shaken vigorously for 20 minutes, in order to maintain an effective suspension. The layers were then allowed to separate (the embryos were at the interface), and the lower phase was removed. 10 mls of methanol were added, and the vial shaken for 30 seconds, then left to stand. At this point the majority of embryos devitellinised and sank to the bottom of the vial, from where they were collected, and transferred to an eppendorf tube. The embryos were washed a couple of times in methanol, and if required, were stored in methanol for up to several weeks at 4 °C.

The embryos were rehydrated by passing them through an ethanol series (70, 50 and 30% ethanol), then incubated for 10 minutes in PBS, followed by re-fixation in PP for 20 minutes, before pre-treatment.

Pre-treatment

The following steps were performed in 1 ml volumes in eppendorf tubes, on a revolving wheel and avoiding potential RNase contamination (using DEPC treated solutions, as described for "isolation of RNA").

The embryos were washed 3 x 2 min. in PBT. They were then incubated for 3-5 min. in 50 µg/ml proteinase K in PBS. The precise time of incubation varies according to the batch of proteinase K, and was therefore optimised for each batch of enzyme used. The incubation time is critical as too little digestion results in low signal and high background, but too much results in the disintegration of the embryos in the subsequent steps. The proteinase K digestion was stopped by 2 x 2 min. washes in 2 mg/ml glycine in PBT. This was followed by 2 x 5 min. washes in PBT. The embryos were re-fixed in PP for 10 minutes, then washed 3 x 10 minutes in PBT.

Hybridisation & washes

The embryos were incubated for 20 minutes in a 1:1 mix of hybridisation solution and PBT, then for 20-60 minutes in hybridisation solution alone. The embryos

were then pre-hybridised for 1 hr at 45 °C. Most of the supernatant was removed, and 0.5 µg/ml of probe DNA (heat denatured in the presence of sonicated salmon sperm DNA) added. Hybridisation took place overnight at 45 °C in a water bath (without shaking).

After hybridisation, a series of 20 minute washes in decreasing proportions of hybridisation solution : PBT were performed - 4:1, 3:2, 2:3, then 1:4, followed by two 20 min. washes in PBT alone.

Staining and colour development

2 µl of the antibody-conjugate solution (DIG labelling and detection kit, Boehringer Mannheim) was preabsorbed against about 50 µl of fixed embryos in 400 µl of PBT for 4 hours at room temperature. The preabsorbed antibody was then further diluted to a final volume of 4 mls in PBT. The hybridised embryos were incubated for 1 hour on a revolving wheel at room temperature with 500 µl of the diluted, preabsorbed antibody.

The embryos were then washed four times, for 20 minutes each, in PBT. This was followed by three 5 minute washes in freshly made up washing solution + 1 mM Levamisole, 0.1 % Tween-20. 4.5 µl of NBT and 3.5 µl X-phosphate solution (both from the DIG labelling and detection kit, Boehringer Mannheim) were added to the final wash, and mixed thoroughly. The colour developed in 10-60 minutes, in the dark. Colour development was monitored by placing a sample of the embryos in a watch glass and examining them at intervals under a dissecting microscope. Colour development was stopped by washing the embryos several times in PBT.

Mounting & photography

After dehydrating the embryos through an ethanol series, they were mounted in Euparal. The embryos were photographed under Normarski DIC illumination using an

Olympus BH-2 NIC microscope, and the 10x objective lens. The exposure time was determined automatically by an Olympus AD system exposure control unit. The film used was PAN F (ASA 50, Ilford).

Chapter 3

THE *MUSCA DOMESTICA* BICOID GENE

As described in Chapter 1, *bicoid* (*bcd*) is a maternally expressed gene involved in the determination of anterior-posterior polarity along the length of the *Drosophila* embryo. Females mutant for the *bcd* gene produce embryos lacking head, thoracic and anterior abdominal segments but containing a transformation of the (anterior) terminal acron into a (posterior) telson instead (Frohnhofer and Nüsslein-Volhard, 1986).

THE BCD GRADIENT

Bcd mRNA is transcribed in the nurse cells of the maternal ovaries, and secreted into the anterior tip of the developing oocyte where it becomes tightly localised (Frigerio *et al.*, 1986; Berleth *et al.*, 1988; St. Johnston *et al.*, 1989). Translation of the mRNA begins soon after egg deposition, and diffusion of the translated *bcd* gene product away from its localised source, coupled with the rapid degradation of the protein throughout the embryo leads to the establishment of a stable exponential gradient of protein along the anterior-posterior axis (Driever and Nüsslein-Volhard, 1988a). The concentration dependent activation of downstream target genes allows the conversion of the *bcd* gradient into regions of distinct developmental outcomes. Thus, *bcd* can act as a morphogen (Turing, 1952), conveying positional information (Wolpert, 1969) through its concentration gradient.

***bcd* localisation**

Localisation of the *bcd* mRNA at the anterior tip of the developing oocyte (providing the source for the stable gradient) is mediated by *cis*-acting localisation signals contained within the 3' untranslated region of *bcd* mRNA (Macdonald and Struhl, 1988; Macdonald *et al.*, 1993; Ferrandon *et al.*, 1994). The localisation signal appears to be modular - a number of different elements have been identified within the *bcd* mRNA localisation signal (Macdonald *et al.*, 1993; Gottlieb, 1992). The protein products of a number of genes eg. *exuperantia* (*exu*), *swallow* (*sww*) and *staufer* (*stau*) are required for the correct determination of this process (St. Johnston *et al.*, 1989; Stephenson *et al.* 1988), and mutations in the *trans*-acting factors alter *bcd* mRNA distribution at different points during localisation (reviewed St. Johnston and Nüsslein-Volhard 1992).

The use of mRNA pre localisation to set up the *bcd* protein gradient appears to have been conserved amongst the higher dipterans - anteriorly localised *bcd* mRNA has been observed within the *D. melanogaster* and in species as distant as *Musca*, and members of the Calliphoridae (MacDonald, 1990; Seeger and Kaufman, 1990; Sommer and Tautz, 1991; Schröder and Sander, 1993).

***bcd* degradation**

The "sink" for the creation of a stable gradient of *bcd* is its rapid degradation throughout the embryo. *bcd* has a short half-life, probably due to its possession of a PEST sequence (Driever and Nüsslein-Volhard, 1988a). These are regions rich in proline, glutamic acid, serine and threonine, and flanked by clusters containing several positively charged amino acids. They are associated with proteins with short half lives and are thought to act as signals for rapid degradation (Rogers *et al.*, 1986; Rechsteiner, 1987). The PEST sequence has been conserved in *D. pseudoobscura* bicoid, reflecting the importance of the rate of degradation in the creation of a stable concentration gradient (Seeger and Kaufman, 1990).

READING THE BCD GRADIENT

bcd binding affinities

The concentration gradient of *bcd* instructs the pattern of expression of the gap genes, which form the next level down in the hierarchy of the developmental genes. Above a certain concentration threshold, the *bcd* protein will bind directly to sites within the promoter of target gap genes such as *hunchback* (*hb*) (Driever and Nüsslein-Volhard, 1988b; 1989a; Driever *et al.*, 1989a; Struhl *et al.*, 1989), *orthodenticle* (*otd*) (Finkelstein and Perrimon, 1990), *empty spiracles* (*ems*) and *button head* (*btd*) (Cohen and Jürgens, 1990), activating gene expression. The affinity of the interaction between *bcd* and its sites in any given gene determines the level of this concentration threshold - and hence the extent of the expression domain of that particular target gene. So, *bcd* has been shown to bind directly with high affinity to sites within the promoter of *hb* (Driever and Nüsslein-Volhard, 1988b; 1989; Driever *et al.*, 1989a; Struhl *et al.*, 1989) resulting in the activation of *hb* expression through the entire anterior half of the embryo, whereas gap genes expressed in more anterior domains (which have higher concentrations of bicoid), such as the head specific genes *otd* (Finkelstein and Perrimon, 1990), *ems* and *btd* (Cohen and Jürgens, 1990), are thought to require lower affinity sites (Driever *et al.*, 1989a; Liaw and Lengyel, 1993). Essentially, there is an inverse relationship between concentration and affinity - in regions of high *bcd* concentration, low affinity sites are sufficient for activation, whereas in regions of low *bcd* concentration, *bcd* dependant activation of gene expression requires high affinity binding sites.

Synergistic interactions with other genes

However, the affinity of *bcd* for binding sites is not the only factor determining the extent of the expression domain of some *bcd* dependant genes. *bcd* is responsible for activating expression of the gap gene *Krüppel* (*Kr*) (Hoch, Seifert and Jäckle, 1991),

and the second stripe of the pair rule gene *evenskipped* (*eve*) (Small *et al.*, 1991), which it does in conjunction with the hb protein. A synergistic interaction between *bcd* and hb appears to allow the activation of these genes in relatively posterior domains, despite containing relatively weak *bcd* binding sites (Small *et al.*, 1991). This synergistic interaction between *bcd* and hb may possibly, as discussed in Chapter 1, be important in allowing the expression of *hb* over its full domain (Simpson-Brose *et al.*, 1994).

ACTIVATION OF GENE EXPRESSION

bcd is a weak activator (Struhl *et al.*, 1989). Activation of gene expression by the bound *bcd* protein is at least in part due to an acidic domain within the carboxyl-terminal part of the protein (Driever *et al.*, 1989b). However truncations of the *bcd* protein that remove this region are still capable of some transcriptional activation in a number of assays (Driever *et al.*, 1989b; Struhl *et al.*, 1989). This indicates that the region immediately C-terminal to the homeodomain also has a degree of activating ability. This region is not acidic, but does have a high serine-threonine content which it has been suggested might be acidified through phosphorylation (Struhl *et al.*, 1989) - the *bcd* protein has been shown to be phosphorylated in early embryos (Driever and Nüsslein-Volhard, 1989a; Ronchi *et al.*, 1993). However, Ronchi *et al.* (1993) have suggested that phosphorylation (by the torso (*tor*) receptor mediated signal transduction cascade) leads to the down regulation of *bcd*, and so the mechanism for activation by *bcd* (as opposed to binding) is left unclear.

CONSERVATION OF GENE EXPRESSION

Bcd has not been found outside the Diptera. Transplantation and rescue experiments, in which poleplasm from various insect species was injected into embryos from mutant *D. melanogaster* lacking a functional *bcd* gene, have revealed a *bcd*-like activity in the blowflies *Lucilia* and *Calliphora*, as well as *M. domestica* and various *Drosophila* species (Schröder and Sander, 1993). However, these experiments showed

an increased qualitative failure of rescue with increasing taxonomic distance. This failure to rescue suggests that an evolutionary change may have occurred that has reduced the affinity of the donor species' bcd protein for the regulatory regions of bcd-dependent *D. melanogaster* target genes (Schröder and Sander, 1993).

However, the expression of bcd appears to be conserved in the housefly, *Musca domestica* (Sommer and Tautz, 1991), which diverged from *D. melanogaster* about 100 million years ago (Hennig, 1981). The conservation of bcd expression, in conjunction with the broad conservation of the downstream segmentation gene expression domains in *M. domestica* (Sommer and Tautz, 1991) implies that the function of bcd has also been conserved. Divergence of the protein coding regions of the bcd gene in *M. domestica* may therefore reflect lack of function, or possibly molecular co-evolution - as discussed in Chapter 1.

In this chapter, the sequence of a genomic fragment of DNA encompassing two thirds of the protein coding region of the *M. domestica* bcd gene is presented, and attempts to clone the remaining regions of bcd described. The *M. domestica* sequence is compared with those published for the *D. melanogaster* and *D. pseudoobscura* bcd genes, and the conservation and divergence of these are discussed.

METHODS

Materials

GENOMIC DNA LIBRARY.

A *M. domestica* genomic library constructed in λ EMBL3 was a gift from M. Williamson. A partial *Sau* 3A digest of *M. domestica* genomic DNA had been size-fractionated, and ligated into *Bam* HI-cut λ EMBL3 arms.

BCD HOMEODOMAIN FRAGMENT.

A PCR fragment containing part of the *M. domestica bcd* homeodomain, ligated into the M13mp18 vector, was a gift from R. Sommer. This fragment has been described in Sommer and Tautz (1990), and its location within the *M. domestica bcd* gene is illustrated in Figure 3.1.

Methods

LIBRARY SCREEN

The *M. domestica* genomic library was plated in the *E. coli* strain LE 392, screened at high stringency (0.1 x SSC, 0.1% SDS, 65 °C), and rescreened as described in Chapter 2 (General Methods). The library filters were probed with the *M. domestica bcd* homeodomain fragment. This was amplified by PCR using the M13 forward and reverse primers, and gel purified, then labelled with ³²P by random priming, as described in Chapter 2.

The number of library plaques it is necessary to screen in order to have a given chance (*N*) of a particular DNA sequence being present maybe calculated from the equation:

$$N = \frac{\ln(1 - P)}{\ln(1 - f)}$$

where *P* is the required probability and *f* is the average fractional proportion of the genome within each library phage. Therefore, if it assumed that the library phage contains an average insert of 15 kb, and the *M. domestica* genome is 5 x 10⁵ kb, then the number of phage to be screened in order to have a 99% chance of a given sequence being present is 1.5 x 10⁵.

In this screen, a total of 2 x 10⁵ plaques screened. 15 potential positives were identified from the primary screen, of which only one (λ bcd3.13) proved positive on rescreening. DNA was isolated from this phage using the liquid culture lysate method described in Chapter 2.

RESTRICTION MAPPING AND SUB-CLONING

The positive library clone, λ bcd3.13, was restriction mapped. Restriction enzyme digests were performed, and the DNA run on 0.8% agarose gels. These were blotted, and probed with the bcd probe used to screen the library. This allowed the calculation of a rough restriction map (Figure 3.1a), and the identification of a 3.2 kb *Eco* RI fragment that contained the homeodomain. The *Eco* RI fragment was isolated and

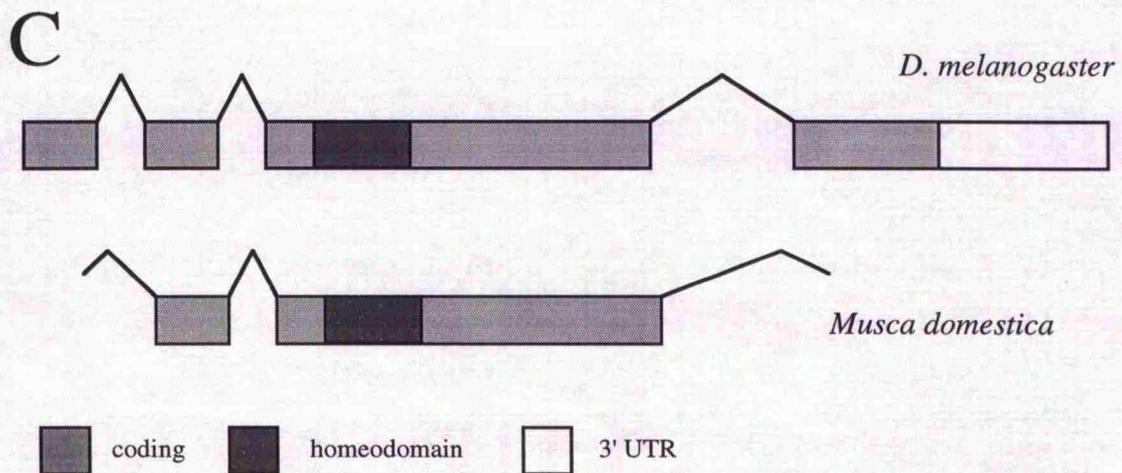
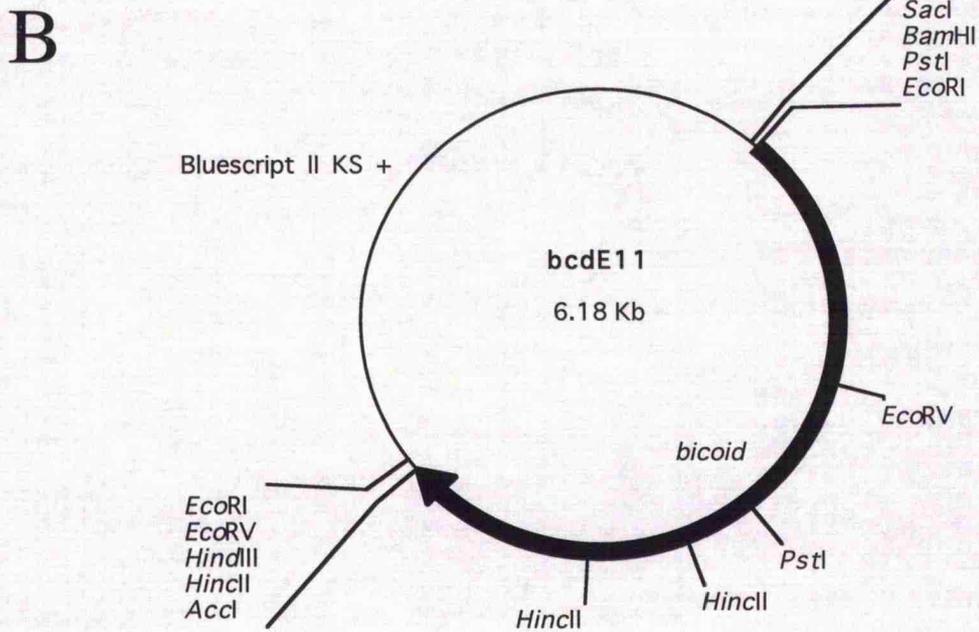
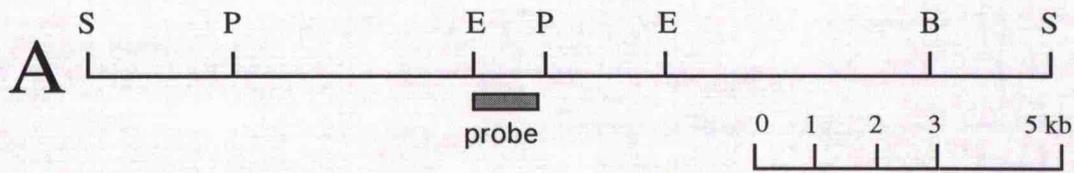


Figure 3.1 Sequencing the *Musca domestica bicoid* gene. **A.** The restriction enzyme map of the insert within λ bcd 3.13, indicating the region to which the *M. domestica bcd* homeodomain probe hybridised. Enzymes: S, *Sal* I; P, *Pst* I; E, *Eco*RI; B, *Bam*HI. **B.** Restriction map of plasmid Bcd E11. This contains the 3.2 kb *Eco*RI fragment of λ bcd 3.13 ligated into the *Eco*RI site within the polylinker of Bluescript II KS +. **C.** A schematic diagram of the structure of the *Drosophila melanogaster bcd* gene. The section of the *M. domestica bcd* gene contained within the 3.2 kb *Eco*RI fragment is illustrated below.

ligated into *Eco* RI cut Bluescript in both orientations, to give plasmids bcdE11 and bcdE14.(Figure 3.1b). Small scale preparations were made of the DNA of both of these plasmids, and they were further restriction mapped (Figure 3.1b). As *Eco* RV cuts within the homeodomain fragment used as a probe, it was easily possible to calculate the orientation of the sub-cloned *Eco* RI fragment, and within in it, the location of the homeodomain.

SEQUENCING

The ends of the 3.2 *Eco* RI fragment were obtained by sequencing single stranded DNA prepared from the plasmids bcdE11 and bcdE14, using the vector Universal primer. To obtain the complete sequence of the bcdE11 clone, a strategy of sub-cloning specific fragments of the insert was adopted. The insert was restriction mapped with a selection of enzymes, each of which also cut within the polylinker of the Bluescript II KS + vector. The insert was found to contain *Eco* RV, *Pst* I, *Hinc* II, *Sst* II, and *Acc* I restriction sites. These sites were in convenient locations to enable the construction of a bank of subclones spanning most of the insert, in portions small enough to be accurately sequenced. Sequencing was performed manually as described in Chapter 2.

The sparse regions (between the *Eco* RI and *Eco* RV sites, and between *Hinc* II and *Eco* RI) were covered by constructing nested deletions, as described in Chapter 2. All deletions used *Sac* I to produce deleted ends and *Sma* I to generate protected ends. Four time points were taken - after 1, 3, 5 and 7 minutes of Exo III digestion. Three colonies were selected for each time point. Restriction analysis of the resultant plasmids indicated that each minute of Exo III digestion deleted about 100 bp of DNA. Single stranded DNA from the nested deletion plasmids was used in cycle sequencing reactions run on an automated DNA sequencer as described in Chapter 2. A small number of plasmids failed to sequence. Restriction analysis suggested some enzyme sites were missing from the polylinker, suggesting that the "protected" end had been chewed back by the Exo III nuclease. This could have resulted in deletion of the primer site.

Two final gaps were filled by using custom primers (E11 and E14), in cycle sequencing reactions run on an automated DNA sequencer as described in Chapter 2. Their positions are indicated in Figure 3.2. Details of their sequences are given in Chapter 2.

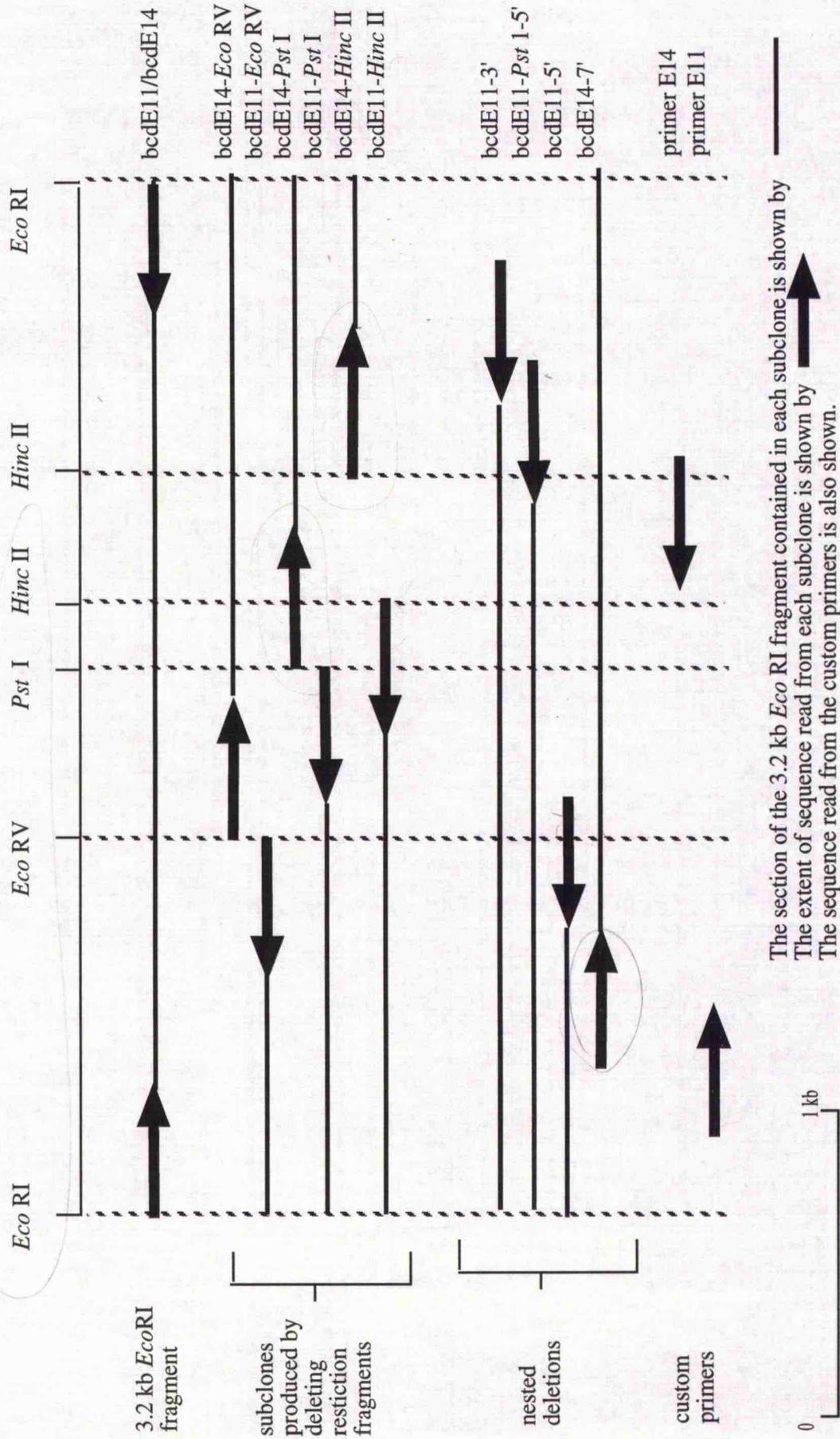
A summary of the subclones, and the extent of sequence data obtained from each, is given in Figure 3.2. Approximately 70% of the insert was read from two separate subclones, generally from opposite strands. Only in one case did the overlapping sequences not match. This was caused by a compression of 4 bases in one sequence that were clearly resolved when sequence from the opposite strand was read.

DNA SEQUENCE ANALYSIS AND ALIGNMENTS

Dotplots were performed at low stringency (at least 14 matches within a 21 nucleotide window) using the COMPARE and DOTPLOT programs of the UWGCG package. These show clearly that the *Eco* RI fragment contains the second and third exons of the *M. domestica bcd* gene. However, this fragment does not appear to contain the first and fourth exons (Figure 3.3).

The region of the *M. domestica* sequence containing the second and third exons (ie. nucleotide positions 1400-2400, Figure 3.6) was aligned with the corresponding section of the *Drosophila* sequences using the GAP program of the UWGCG package, and the default parameters. The GAP program uses the algorithm of Needleman and Wunsch (1970) to find the optimal alignment of two sequences, maximising matches and minimising gaps. The intron-exon boundaries were determined through homology to the *Drosophila melanogaster* and *Drosophila pseudoobscura bcd* gene. As the third exon is very diverged towards the C-terminal end, the position of exon 3-intron 4 boundary is not obvious from the sequence comparisons. It must occur between nucleotides 1975 (the end of last patch of homology) and 2099 (an in frame stop codon). There is a potential splice site at position 2038 (CGGTAAGC, compared to the consensus AGGTAAGT), which might be possible, and this has been taken as the boundary for all

Figure 3.2 Strategy for sequencing the 3.2 kb *Eco*RI fragment of the *Musca domestica bicoid* gene



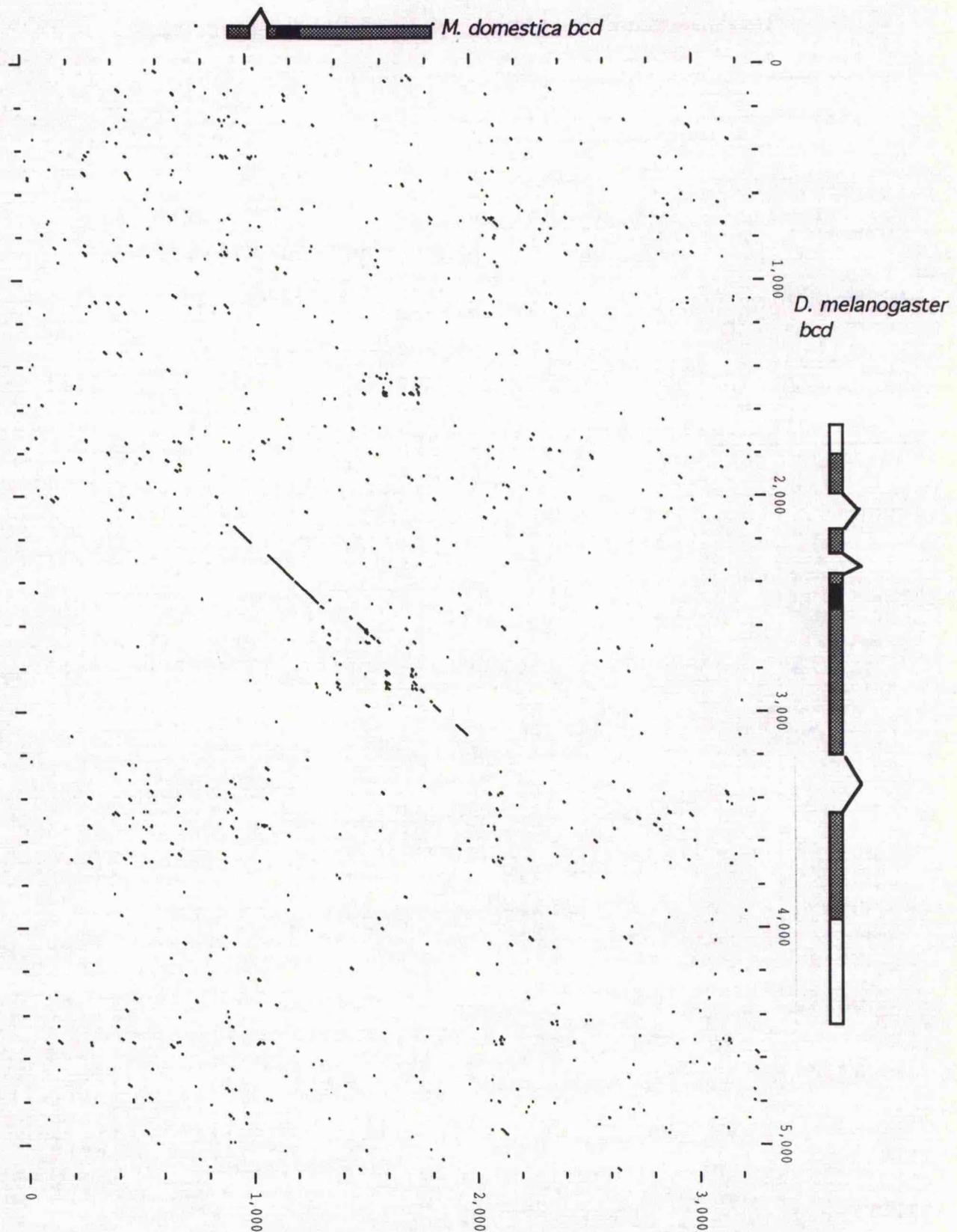


Fig 3.3. Dot matrix homology comparison of the *bcd* genomic region from *Drosophila melanogaster* and the sequenced region of the *Musca domestica bicoid* gene. The dot matrix comparison was generated using the COMPARE and DOTPLOT programs of the UWGCG DNA sequence analysis package. Fourteen matches or better over a 21 bp window (66%) are required to make a dot. The organization of the *D. melanogaster bcd* gene is indicated, as are the homologous regions of the *M. domestica bcd* gene (there are no regions with homology to the first or fourth exons of *bcd* within the region of the *M. domestica* gene that has been sequenced). Protein coding regions are shaded, the homeodomain is black.

subsequent analysis. Once the reading frame had been determined, the alignments were repeated using the translated amino acid sequences - this greatly improves the signal to noise ratio, and allowed for the alignment of the otherwise difficult C-terminal half of the third exon.

Multiple amino acid sequence alignments were performed using the CLUSTAL V package and the default settings. As this program tends to spread sequences out at the ends where one is shorter than the other and there are no strong regions homology, the alignment was adjusted by eye, taking into account those produced by the UWGCG GAP program.

The *M. domestica bcd* sequence was examined for PEST sequences using the PEST-FIND program (Rodgers *et al.*, 1986). This searches amino acid sequences for possible PEST sequences, and calculates their PEST score and hydrophobicity value.

NORTHERN ANALYSIS

Total RNA was isolated from adult female *M. domestica* flies. 10 µg of RNA was separated by electrophoresis and blotted onto nylon membranes, as described in Chapter 2. The northern blot was probed with the *M. domestica bcd* homeodomain probe described above, as described for Southern blots in Chapter 2.

RESULTS

Sequencing the *M. domestica bcd* gene

In *D. melanogaster*, the *bcd* gene is split into four exons, each containing distinct features. This is summarised in Figure 3.4.

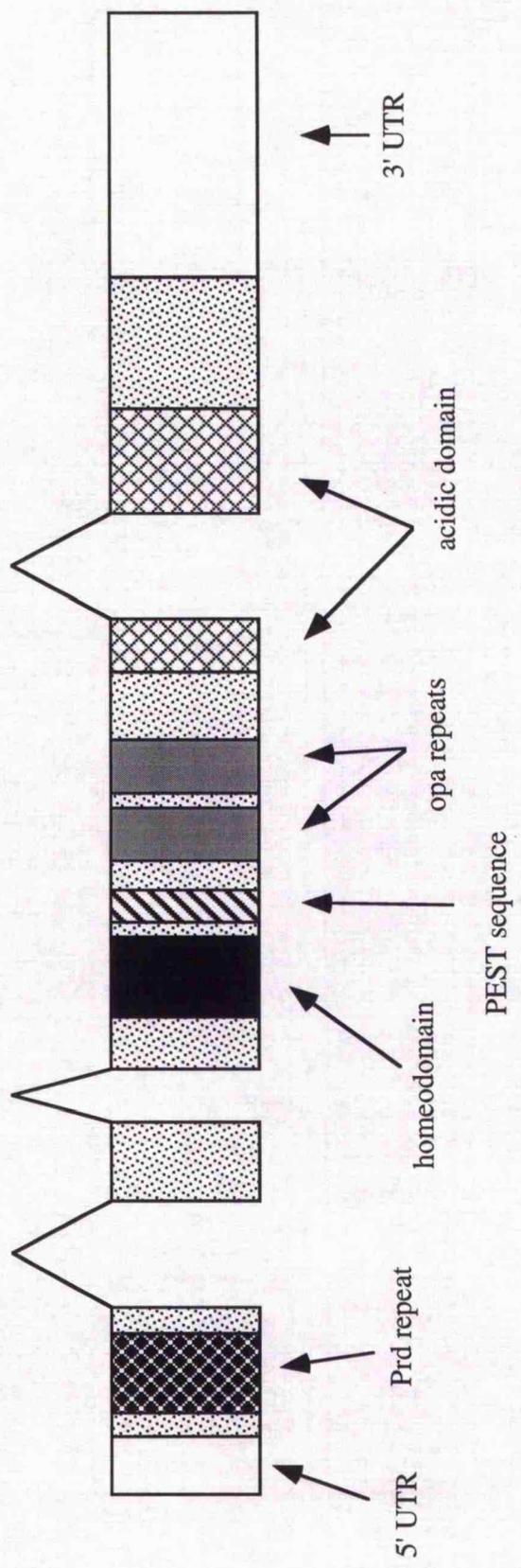


Figure 3.4. A schematic diagram of the structure of the *Drosophila melanogaster bicoid* gene

THE SECOND AND THIRD EXONS

A 3.2 kb *Eco*RI fragment of genomic DNA containing the second and third exons of the *M. domestica bcd* gene has been sequenced. The second and third exons were identified by homology to the *D. melanogaster* sequence, as described in the Methods section of this chapter. The complete sequence of the 3.2 kb *Eco* RI fragment of the *M. domestica bcd* gene is given in Figure 3.6.

THE MISSING EXONS

No regions with homology to either the first or fourth exons of *D. melanogaster bcd* were found within the 3.2 kb *Eco* RI fragment of the *M. domestica bcd* gene (Figure 3.3).

Comparisons

Figure 3.6 gives the nucleotide sequence of the 3.2 kb *Eco* RI fragment of the *M. domestica bcd* gene, and the protein translation. The alignment with the homologous region of the *D. melanogaster bcd* gene is also shown. A multiple alignment of the complete *D. melanogaster* (Berleth *et al.*, 1988) and *D. pseudoobscura* (Seeger and Kaufman, 1990) and partial *M. domestica bcd* protein sequences is given in Figure 3.7.

STRUCTURE OF THE GENE

Alternate splicing

In *Drosophila melanogaster*, two different *bcd* transcripts have been observed (Berleth *et al.*, 1988) - there is a major 2.6 kb transcript containing all four exons present in oocytes and cleavage stage embryos, and a weak 1.6 kb transcript (lacking the second and third exons) present throughout development (Figure 3.5a).

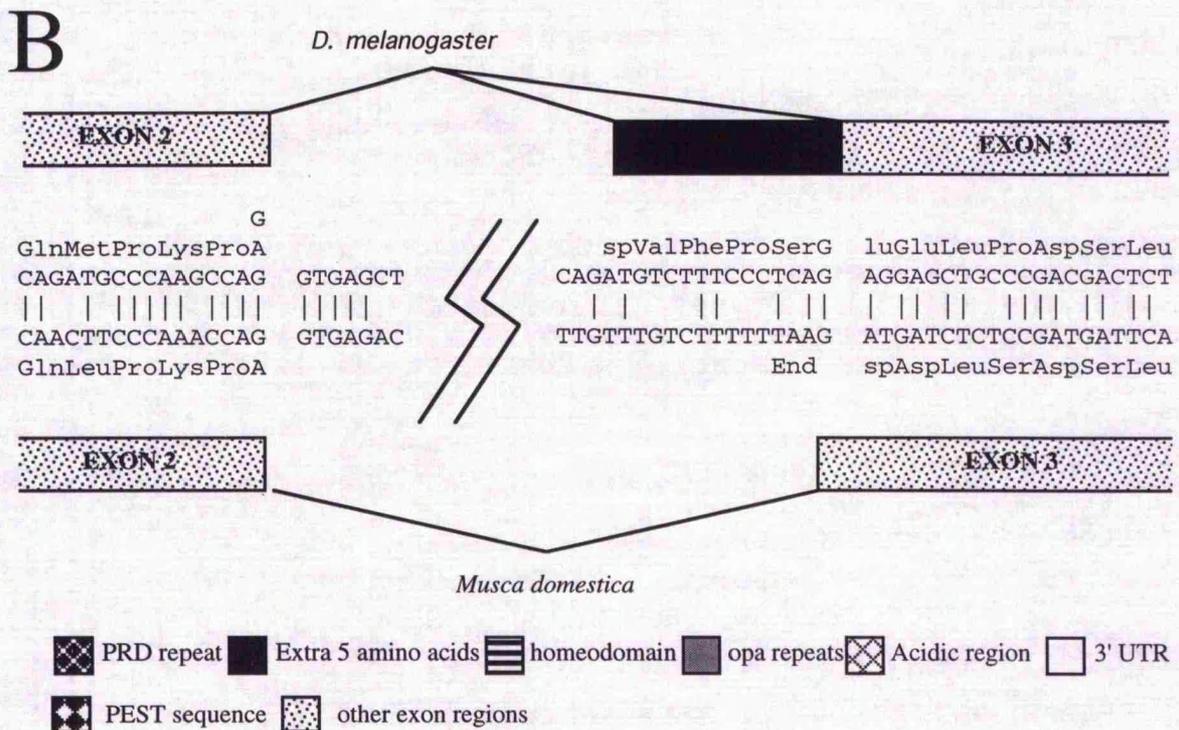
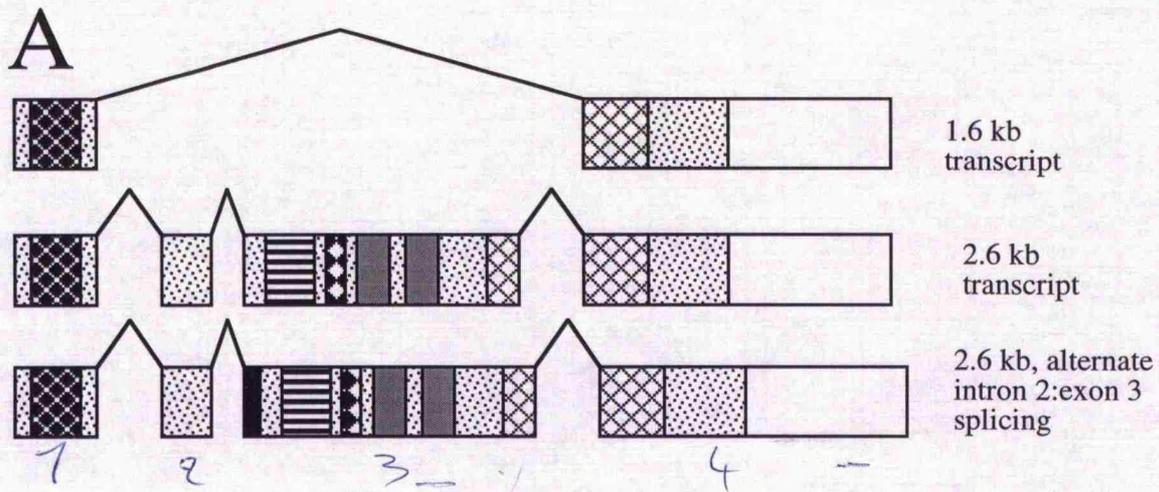


Figure 3.5. **A.** A schematic diagram illustrating the different transcripts of the *Drosophila melanogaster bicoid* gene. **B.** Comparison of the alternate splicing pathways between exons two and three.. Alignment of the intron 2 boundaries from *D. melanogaster* (top) and *Musca domestica* (bottom) with the potential splicing pathways in both species indicated. While the *D. melanogaster bcd* gene produces two protein products which differ by five amino acids due to the alternate splicing at this intron/exon junction, the *Musca bcd* gene can only produce a single protein which corresponds to the smaller of the two *D. melanogaster* gene products

Musca domestica bicoid DNA sequence

```

1          GAATTCAGTGTAGAAAGAAGTACATAACATTTGGAGAAAATC      40
41  GGTTCAGATAFGGCTATAGCTCCTATAATATAATATATTCATCCGATTTGTTGATTAAGTC  100
101 CCTCACATTCGTAATATGCATTCAAAAACAAAGATATTTGGTACCAGATATTTTATAAAC  160
161 CTCGAAATACCTCGGCCANATGTTATCAAGTTCGATTCAGTTTGGATATAGCTCCCAT  220
221 ATATATCCCTAATCCCATTAATTATAATTGTATAAAAAGTGAATATATTTTATTTT  280
281 TTGTTTAAACCAATAAATGTATACAAATTTTAAATGCATATAAAATTAACCGGGTAGCA  340
341 ACGAAGCCGGGTTTCAGGATATAAGATAGTTTATAGATACAGTTTCAAATTTAGAAATGTG  400
401 TAACGAAGATCGATACAAAATAAATTTGCCAAATTACATTCGAATCGGTTTGTATAT  460
461 ATATATATTTTCGGCCGTTTTCGTTTATTTGATGCGCCAAACACGCTATGTATGTATAT  520
521 ATGCATGTTTTTAACTTCGTATATTGGAAATTTTATTTGTTTCCCAATTCATCCCTCAA  580
581 AGGGAGATTTTGTGAGAGTGGGAGGAGTGGTCCATAGATTAGTTCCATATCTATGGT  640
641 TGGTAAAAATGGGCATATTTATTTCTTTAAAAATTTTACATTTGTTGTGAGTGATTTTC  700
701 CATCTAAGAACTCGATCAAGGATTTGTTGGATCATGTTTGGACTAAATCGTATAGTGTAT  760
761 CGTGCCGACAAGCCGACAAGACATCCCGACGGATATTTTAGATCGACTCCAATGTCCCTG  820
821 AAACGGACTTCGTCTAGTATGTACTTGGCATAACAAGTTATTTTAGCCAGAGACTTTAAT  880

          intron 1<<<<<LeuPheAspG1
881  TATATTCCTTCTCCAATTTTTAAATCTTTTCTTAATTTTATAATTCAGCTATTTGACGA  940
          --T--C--T--
          -----

uArgThrGlyAlaIleAsnTyrAsnTyrIleArgProTyrIleProAsnGlnLeuProLy
941  ACGTACTGGTGCACACTACAATAACACTATACGTCCATACATTCCTAACCAACTTCCCAA  1000
G--A--G--A--G--A-----C-----G--TC-G--C-----GA-G-----
-----Leu-----Met-----

sProA>>>>>intron 2
1001 ACCAGGTGAGACATTTTATATTTTCTTGGAAATTTCTTGTGCTTTTTTCACATTTGTTT  1060
G-----CTCAAAGCCAACAAGTCAGCCATCGTCTTATCAGATGTCTTTCCCTCAG
----G

          intron 2<<<<<spAspLeuSerAspSerLeuValSerCysArgProArgAr
1061  GTTGTGTTGTCTTTTTTAAGATGATCTCTCCGATTCAGTGGTTAGCTGTCGTCCACGACG  1120
          ---G--G--GC---C--T-----G-TGC-G--A-----T--
          luGlu---Pro-----MetArg-----

gThrArgThrThrPheThrSerAlaGlnIleAlaGluLeuGluGlnHisPheLeuGlnG1
1221  CACCCGTACAACATTTACCAGTCTCAGATTGCCGAATTGGAACAACATTTCTTCAAGG  1180
-----C--C--T-----CT---A--A--A--GC---G--G--C--T--G--G--
-----Ser-----

yArgTyrLeuThrSerSerArgLeuAlaGluLeuSerAlaLysLeuThrLeuGlyThrAl
1181  CCGATAFCTGACATCATCACGGCTGGCTGAACTTTCAGCGAAGCTAACTTTGGGTACAGC  1240
A-----C--C--G--CC-C--A--T--G--T--G-----A--G--CC---C-----
-----AlaPro-----Asp-----

aGlnValLysIleTrpPheLysAsnArgArgArgHisLysIleGlnSerAspGlnG1
1241  TCAGGTCAAGATTTGGTTCAAAAATCGCAGACGACGTCACAAAATTCATTCGGATCAACA  1300
C-----G-----A-----T--G--C--TC-G--T-----G--C-----G--
-----Hi-----

nLysGluPheSerCysAspGlyMetProLeuSerProSerLeuProThrThrIleLysSe
1301  GAAGGAATTTTCATGTGACGGTATGCCCTGTCCACATCACTTCCGACTACAATCAAATC  1360
C-----CCAG--C-AC--G--G-----T--C--G--GGGTATGAAACAGAGCGATGGCGA
s---AspGln---TyrGlu-----GlyMetLysGlnSerAspGlyAs

rGluProGlnGlySerAlaSerSerCysGlySerAsnAsnSerAsnArgPheAspIlePh
1361  GGAGCCCCAGGTTAGTCCAGCAGCTGTGGCAGTAACAATAGCAACCGGTTTCGACATCTT  1420
TCCC.....
pPro

```

1421 eLeuLeuPheValArgTrpThrProSerLeuGlnSerLeuSerIleAsnGlyAsnGlyGly 1480
 CCTCCTCTTCGTCAGGTGGACGCCAGCTTACAGTCGTTGAGTATAAACGGTAATGGCGG
G---A-TC-T--CT-G...---A--
-----Thr-----Leu.....-----

1481 ySerThrProAsnProLeuThrProSerProThrProThrProThrThrAsnLeuMe 1540
 CTCAACACCTAACCCATTGACACCATCACCGACGCCACCACCCACAACAATCTCAT
 AG-C--G--C---G-T-----T--G---C-----T-A--G-----TGCACACATGAC
 -Ala-----Ala-----Ser-----AlaHisMetTh

1541 tAspHisTyrSerGluProAlaPheAsnProTyrTyrTyrAsnAsnHisHisSerThrHi 1600
 GGATCATTACAGCGAGCCAGCGTTCAAATCCCCTACTATTACAATAATCATCATTCGACGCA
 ---G--C-----T-----CG-----CA--T-C--- GGAGG
 rGlu-----...Ser-----Ala-----AsnTyr-----GlyGly

1601 sHisHisHisHisGlnProProHisHisAlaThrLeuThrHisProTyrGlyCysSerAl 1660
 CCATCATCATCACCAGCCACCACATCATGCCACTCTCACTCATCCATATGGTTGTAGTGC
 ---CA---CGC---G-CAAT-G---CATGCACA-GCAGT---T-CC--AG-GG-GC-
 y---Asn---Ala---AlaAsnArg---MetHisMetGlnTyr---Ser---GlyGlyPr

1661 aGlyAlaThr.....GlyGlyGlnTyrTyrProProProProProProSe 1706
 AGGTGCTACC.....GGGGCCAGTATTATCCCCACCACCACCACCCTC
 ---AC--GGGTCGACCAATGTCAAT--C-----TC-TC-AG-AG-AG-AGGTC-ATAA
 o---ProGlySerThrAsnValAsn-----PhePheGlnGlnGlnGlnValHisAs

1707 rSerLeuGlnHisHisHis.....SerGlnHisGln 1737
 AAGCCTGCAGCACCATCAC.....AGCCAACATCA
 TCA--A-----A-TG---CACCAGGGCAACCACGTGCCCGCACCAGATGCAG--G--G--
 nHisGln---GlnLeu---HisGlnGlyAsnHisValProHisGlnMetGln---Gln--

1738 nGlnGlnTyrHisSerProHisProHisGlnPheGlnMetGluHisLysProHisAlaAl 1797
 ACAACAATATCACAGTCCGCATCCCCATCAATTTTCAGATGGAACATAAACCTCATGCCGC
 ---G--GGC---GCAG-A---ATA-----C---G-CT-CC-G--A--G-AAGCCAG---
 -----AlaGlnGlnGlnGlnTyr---His---AspPheGlnGln---GlnAlaSer--

1797 a.....ValIleLysGluAspPro.....AspTyrAsnPheAsnAsnProTyrTy 1842
 G.....GTCATCAAAGAAGATCCG.....GATTATAATTTCAACAACCCGTACTA
 CTGTGCGCTCC--GG-----G--C--A---GAGGCC--C--C-----G-T-----
 - LeuVal---aspGlu--- SerSer-----

1843 rMetArgMetProProThrAlaGlySerAsnProSerGlyValThrThrValGluProSe 1902
 TATGCGCATGCCACCACAGCGGGGAGTAATCCTTCGGGTGTGACCACAGTGGAGCCATC
 C-----A.....T-G-GAAT-TC-GGCG-CA-T-CATC-G-AT-C-CT-T-GCCCG
 -----SerGlyMet---GlyAlaThrAlaSerAlaSerAlaValAlaAr

1903 rSerAlaMetSerProAsnSerGluValTyrGluProLeuThrProLysAsnAspAspAs 1962
 TAGTGCAATGTCCAGAACTCAGAAGTCTATGAACCGTTGACACCGAAAAATGACGATAA
 AG-C--TGCC--G---GG---C--G-----C--G--A--A-----C--G-----A-G
 gGly--Ala-----Gly-----GluSe

1963 nSerSerLeuCysAsnGlyAla?????>>>>intron 3 2022
 TTCGAGTCTGTGCAATGGCGCCGGCGTAATGTCGATGTTGGTGATAATTTAGATGAAAC
 -C-----TGGCATCGGCAATCGGCGGACCTTGGCCATCGCCGTTGGCGAGACGGA
 rPro-----GlyIleGlyIleGlyGlyProCysAlaIleAlaValGlyGluThrGln

2023 CAAGGCCAAATTACGGGTAAGCGCTTTTCGTAATTTTATATATATAAATACTAAATTTTCAT 2082
 GGGCGCCGACGACATGACGACGGAACGAGCAAGAAGACGACGCTACAG
 uAlaAlaAspAspMetAspAspGlyThrSerLysLysThrThrLeuGln (intron 3)

2083 TTATATACTAGTAAAATAAATATATTAATTTCAACATTCCACTTTAATCCAGTACCCCA 2142
 2143 AAACCACATTGTACCAAAAACACCTCTCTAGGTTGTTTGGAAAATGTTGACTACAAGGAC 2202
 2203 ACGAAAAGTACTACATAGTACTATATTTGGTCAATTTGCATTTGGGTAAACATATAGTGT 2262
 2263 TGTATTGTTGTTGTTGTTGTTAGGGTACCCGAACATATGGAGTTGATTCGGAAATGTTG 2322
 2323 ACTATAACTGTACGAAAAATCAATAATAGTACTAAAATAGTACTAAAATTTGGTAAATGGG 2382

2383	GTAAAATGGATAGAGATAACCGTTCGAAATTTTGANANATCTTATTTAATCATATTTT	2442
2443	NAAATACCCNCCAACACTATGTGTGGGGGTATAATAAGTTTGATTGGGCCAAAAATTAT	2502
2503	GTAACACCTCGAAAAATTGGTTTCCCCCATAGAAAAGTATATATGGATCTTCCTATCTCG	2562
3563	CCGAGCTGAGTCGATGTATGTCCCGTATGTATGTCCGTTCGTCCGTCTGTCCCATGTTAA	2622
2623	TTTGTAATCACTCTACAGGTCGCAATTATGATCAGNCCATTCCTTTAGTCTTAGGTCAGA	2682
2683	TACGAATFGGTCAATATTGTCCAACCTGCTTCGGAACTCCCCATAAAGGTTCCAACAGT	2742
2743	TTGAATAATCACACTTCGAGGCTGGGGTATACATCTGTGGGACCCCTCCACCATTCATG	2802
2803	GAGGGTATATTAACPTTGTTCATCCGCTTGTAACCTCCGTAATATTAATAATATACCCTA	2862
2863	TAATGTATATATATATCTGGATCAGCGTAAATTCCTAAGTTCTACGAACTCTGGCGA	2922
2923	TGTCGGTATACCTGTTGTAATCACGCTATCTTTCGAATGAATGAGATAAAAAGGATAAAA	2982
2983	TTTTGCATACATGCATAGTTTATTAGTAGCCAGGCTAAGTTCGTAATGGATCATGTCAG	3042
3043	GCCACTTTTTCGTATAGCCCCATATAGGGTACCTCCCGATATTCGGTATTTTCATGAT	3102
3103	TTTAGCGACATTTATTTACCGGAATTTATTCAAAATTCATATTTGAAGTTCGTAATGGGTA	3162
3163	CTACAGCCTATGACAAAAATTTGTGAATGCAAGTCCACGCTCTCGTATAGGACCCATATA	3222
3223	ACGGTACCTCCCGATAGTTTGTAGTTTCATTTTCAGAATCAATCATTTGGTGAATTC	3279

Figure 3.6. The sequence of the 3.2 kb *EcoRI* fragment of the *Musca domestica bicoid* gene and its homology with *Drosophila melanogaster*. The *D. melanogaster* sequence is in italics, with dashes representing identities and dots representing gaps.

Alignment of bicoid protein sequences of *D. melanogaster*,
D. pseudoobscura and *Musca domestica*

```

MEL      MQQPPPDQNFYHHPLPHTHTHPHSHPHSHPHSHPHHQPQLQLPPQFRNPFDLLFDER
MUSCA    LFDER
PSEUDO   MQQPPPDQNFYHHPLPHTHTHP- PHPHPHPHPHPHHQPQLQLPPQFRNPFDLLFDER
                    *****

MEL      TGAINYNYIRPYLPNQMPKPDVFPSEELPDSLVMRPRRRTRTTFTSSQIAELEQHFLQGR
MUSCA    TGAINYNYIRPYLPNQLPKP-----DDLSDSLVSCRPRRTTTFTSAQIAELEQHFLQGR
PSEUDO   TGAINYNYIRPYLPNQMPKPDVFPSEELPDSLVMRPRRRTRTTFTSSQIAELEQHFLQGR
                    *****

MEL      YLTAPRLADLSAKLALGTAQVKIWFKNRRRRRHKIQSDQHKDQSYEGMPLSPGMKQS-DGD
MUSCA    YLTAPRLADLSAKLALGTAQVKIWFKNRRRRRHKIQSDQKQEFSCDGMPLSPSLPTTIKSE
PSEUDO   YLTAPRLADLSAKLALGTAQVKIWFKNRRRRRHKIQSDQHKDQSYDGMPLSPGLKTS-EGD
                    *****

MEL      P-----PSLQTLSLGG--GATPNALTPSPTPTPTAHMTE
MUSCA    PQGSASSCGSNNSNRFDIFLLFVRWTPSLQSLRINGGSGTPNPLTPSPTPTPTTNLMD
PSEUDO   P-----PSLQNLTLGG--GATPNALTPSPTPSATTAHLVE
                    *****

MEL      HYES-FNAYY--NYNGGHNAQANRHM-HMQYPSGGGPGPGSTNVNGGQFFQ--QQQV
MUSCA    HYSEPAFNPHYNNHHSTHHHHQPPHATLTHPYGCSAGAT---GGQYYPPP---
PSEUDO   HYGET-FNAYY--NYNHGHGQAQGRHVGHVHGQYSGAPGSQ---NGAQFFQTQQQQQL
                    *****

MEL      HNHQQQ-----LHHQ---GNHVPHQMQQQQQQAQQQQ-----YHHF
MUSCA    -----PPSSLQHHHSHQQQYHSPHPH-----QF
PSEUDO   HQQQQQQPPHHHQNHQQQQQLLHHQLPHTNHVPHQMQAQQQQQQQEQQQQQQLYHHF
                    *****

MEL      DFQKQASACRVLVKDEPEADYNFNSSYYMRSGMSG-----
MUSCA    QMEHKPHAA---VIKEDP--DYNFNPPYYMR-----MPPTAGSNPSPGVTT
PSEUDO   DFQKQTASACRV-VKDEPEADYNFNSSYYMRSGVGVAAAAAATAAPGTASSAVAA
                    *****

MEL      ATASAVARGAASPGEVYELTPKNDPESPLCGIGGPCAIAVGTEAADMDMGTS
MUSCA    VEFS-----SAMSFNSEVYELTPKNDNSSLN-GAG??
PSEUDO   AVSAAGEVVTALSAPGEVYELTPKNDPESPLC--GIGGPCATAVGDTDIADMDDGT
                    *****

MEL      KK--TTLQILEPLK-----GLDKSCDDGSSDDMSTGIRALAGTGNRGAFAKFGKPSPP
MUSCA    NKKTTTLQNLLEPLKSHTVVGLDKSCDDGSSDDMSTGMRVLSGRG---AFAKFGKPSAG
PSEUDO

MEL      QG--PQPPLGMGGVALGESNQYQCTMDTIMQAYNPHRNAAGNSQFAYCFN
MUSCA    QG--PQPPLGMGGVALGESNQYQCTMDTIMQAYNPHRNAAGNSQFAYCFN
PSEUDO

```

Figure 3.7. A multiple alignment of the amino acid sequences of the *Drosophila melanogaster*, *D. pseudoobscura* and *Musca domestica* bicoid proteins. The *D. melanogaster* sequence is from Berleth et al. (1988) and the *D. pseudoobscura* sequence from Seeger and Kaufmann (1990). Exon 2 is shown in bold type, the homeodomain is in red type, and PEST domains are green. The first and fourth exons are shown in italics - they have not yet been sequenced from *M. domestica* (see results section). The precise exon:intron boundary at the end of Exon 3 in *Musca* has not yet been determined (see methods). The alignment was made using the CLUSTAL V package as described in the methods section

One further alternative splice point occurs within the major 2.6 kb transcript, at the boundary between the second and third exons. This alternate splicing can result in the inclusion of an extra 5 amino acids at the 5' end of the third exon. Examination of this region in *M. domestica* (Figure 3.5b) indicates that the five amino acids that are unique to the larger protein product of *D. melanogaster* are not conserved, although the amino acids specified upstream of the splice donor in exon 2 and downstream of the second acceptor site for exon 3 are identical. Moreover, no possible splice acceptor sites are found within this intron in *M. domestica*, and there is an in frame stop codon immediately upstream of the second acceptor splice. Therefore, the alternate splicing at the beginning of the third exon seen in *D. melanogaster* is unlikely to occur in *M. domestica*. Similarly, *Drosophila pseudoobscura* can only produce a single protein corresponding to the smaller of the two *D. melanogaster bcd* protein products (Seeger and Kaufman, 1990).

Intron size

The size of the second intron is similar between *M. domestica* and *D. melanogaster*. (Figure 3.3), however, the first and third introns appear to have expanded in size in *M. domestica*. Introns 1 and 3 are each approximately 500 bp long in *Drosophila melanogaster*, and considerably shorter in *D. pseudoobscura* (Seeger and Kaufmann, 1990). More than 1 kb of DNA has been sequenced both upstream of exon 2 and downstream of exon 3, but the first and fourth exons do not appear to be located within this fragment of DNA; therefore, in *M. domestica* the first and third introns appear to be at least twice as large as those found in *D. melanogaster*. However, as the restriction map of λ bcd 3.13 has not been compared with a genomic map in southern blot experiments, it is possible that the insert is composed of a number of non-contiguous fragments of *M. domestica* genomic DNA. This may account for the absence of regions homologous to the first and fourth exons of *D. melanogaster bcd*; and the concomitant increase in the apparent size of the first and third introns .

Transcript length

An attempt was made to investigate the length of the *bcd* transcript in *M. domestica*, by using the *bcd* homeodomain PCR fragment to probe a Northern blot. No signal was obtained. As *in situ* hybridisation shows *bcd* RNA clearly present within the ovaries of adult *M. domestica* females (Schröder and Sander, 1993), the absence of a signal in the Northern blots obviously represents a technical failure in this instance.

PROTEIN CONSERVATION AND DIVERGENCE

The degree of conservation and divergence varies throughout the sequenced part of the coding region. The overall identity at the DNA level for exons 2 and 3 is 61.5% for *M. domestica*, compared to 78% for *D. pseudoobscura*. This becomes 58% and 83% for *M. domestica* and *D. pseudoobscura* respectively, at the amino acid level. The degree of amino acid divergence within exons 2 and 3 of *bcd* is far greater than that seen between *M. domestica* and *D. melanogaster* for either *armadillo*, which shows an overall amino acid conservation of 98% (Peifer and Weischaus, 1993), or the 5' exon of *Ultrabithorax* (72%; Wilde and Akam, 1987).

Many of the nucleotide changes are third base or synonymous codon substitutions. These have reached saturation level between *M. domestica* and *D. melanogaster* with approximately 70% of all silent sites showing changes. However, there is also a widely varying degree of amino acid substitution, insertion and deletion through the second and third exons (see Figure 3.7). Both the homeodomain and the whole of the second exon are highly conserved, each showing 92% identity at the amino acid level. This same region is identical at the amino acid level between the two *Drosophila* species. In contrast, the C-terminal half of the third exon is highly diverged, containing islands of homology separated by very widely diverged sequence.

opa repeats

The highly diverged C-terminal end of the third exon is dominated in *D. melanogaster* by 'opa' repeats (Figure 3.6) - long strings of poly-glutamine and other permutations of the (CAG)_n cyclic repeat (Wharton *et al.*, 1985), which are greatly expanded in *D. pseudoobscura* (Seeger and Kaufman, 1990), but are much reduced in *M. domestica*, instead being partially replaced with much shorter strings of histidines or prolines (Figure 3.7). This makes it difficult to align this portion of these proteins directly. A similar pattern of short stretches of conserved amino acids separated by clusters of highly diverged repeated sequences has been seen in other developmental genes, for example *engrailed* (Kassis *et al.*, 1986) and *hunchback* (Treier *et al.*, 1989), in comparisons between *Drosophila* species.

Homeodomain

The probe used to screen the library was a PCR derived fragment of the *M. domestica bcd* homeodomain (Sommer, 1992; Sommer and Tautz, 1991). This probe fragment includes 44 of the 60 amino acids of the homeodomain, and it shows six amino acid differences to *Drosophila*. The sequence of the genomic clone differs at two points from that of the probe. There is no T to C change at position 1574, and the nucleotide at position 1621 is a C, not a T (nucleotide positions from Figure 3.6). This means that one amino acid change seen by Sommer and Tautz (1991) is not present in this genomic sequence (Phe in *D. melanogaster* to Leu in *M. domestica*, amino acid 5 of the homeobox). No further differences to the *D. melanogaster* amino acid sequence were found in the regions of the *M. domestica bcd* homeodomain not included within the probe fragment.

In total, the *bcd* homeodomain shows a total of 5 changes out of 60 amino acids in *M. domestica*, a divergence of 8.3%, (compared to its absolute conservation between the two *Drosophila* species). This is high compared to the lower degree of divergence seen between the homeodomains of the *Antennapedia* class of homeotic genes.

Homeodomains from these genes have been sequenced in a number of different insect species much farther separated than *Drosophila* and *M. domestica*, including *Schistocerca gregaria*, *Tribolium castaneum*, *Manduca sexta* and *Bombyx mori* (Walldorf *et al.*, 1989; Tear *et al.*, 1990; Nagy *et al.*, 1991; Eggleston *et al.*, 1992; Ueno *et al.*, 1992; Stuart *et al.*, 1993), and typically show near identity (>58/60 amino acids conserved).

At the amino acid level, the degree of conservation within the homeodomain is no greater than that seen for exon 2. Both regions are absolutely conserved between the two *Drosophila* species, and 92% identical between *M. domestica* and *D. melanogaster*. Again, at the nucleotide level, the same degree of sequence identity (72%) is seen between *M. domestica* and *D. melanogaster* in both exon 2 and the homeodomain. However, in *D. pseudoobscura*, the homeodomain is less conserved than exon 2 (81% nucleotide identity to *D. melanogaster* for the homeodomain, compared to 89.5% for exon 2) - the homeodomain shows a higher percentage of synonymous codon and silent third base pair changes.

PEST domain

In *D. melanogaster*, *bcd* contains a possible PEST domain C-terminal to the homeodomain (amino acids 180-217, Driever and Nüsslein-Volhard, 1988a; see Figure 3.4). PEST regions are rich in proline, glutamic acid, serine and threonine and are generally flanked by clusters containing several positively charged amino acids. They are associated with proteins with short half lives, and are thought to act as signals for rapid protein degradation (Rodgers *et al.*, 1986). As *bcd* does in fact have a short half life - estimated at less than $\frac{1}{2}$ hr (Driever and Nüsslein-Volhard, 1988a) - the presence of a PEST region is not surprising. The rate of proteolytic degradation is an important variable in establishing and determining the slope of a stable gradient, and so it would be expected that the PEST region is well conserved in *D. pseudoobscura* (Seeger and Kaufman, 1990).

The *M. domestica bcd* coding sequence was analysed for the presence of possible PEST sequences using the PEST-FIND algorithm (Rodgers *et al.*, 1986). In *M. domestica*, the PEST domain seen in *D. melanogaster* is largely conserved, running from positions 1435-1547 (Figure 3.6). *M. domestica bcd* also contains a second possible PEST sequence (positions 1847-1951, Figure 3.6), not seen in *D. melanogaster*.

Acidic region

An acidic domain is found near the carboxyl-terminal end of the *D. melanogaster bcd* protein (amino acids 347-414, Driever *et al.*, 1989b; see also Figure 3.4). Initially, Driever and Nüsslein-Volhard suggested (1989) that this region might be required for transcriptional activation by the *bcd* protein, however the situation now looks more complicated. Truncations of the *bcd* protein that do not contain this region are capable of activation in a range of assays (Driever *et al.*, 1989a, b; Struhl., 1989). However, the inclusion of the acidic domain tends to allow more efficient transcriptional activation.

The acidic region spans the third intron, and so it is not possible to compare it fully with the *M. domestica* sequence. There is a well conserved stretch of 21 amino acids at the beginning of this acidic domain, but after that it becomes difficult to make alignments, and the exon 4 sequence is missing. The acidic domain as a whole is well conserved in *D. pseudoobscura*, (particularly with respect to its acidic nature) which it has been suggested may reflect functional importance (Seeger and Kaufman, 1990). This is reinforced by the conservation of the initial 21 amino acid stretch in *M. domestica*, which is sufficiently long diverged from *Drosophila* that any non-functional (non-constrained) sequence might be expected to have diverged.

DISCUSSION

The aim of this chapter has been to clone and characterise the *M. domestica bcd* gene, in order to investigate the possibility that its interaction with the *hb* promoter may have been subject to the processes of molecular coevolution (see Chapter 1).

The sequencing of a fragment of DNA containing the 2nd and 3rd exons of the *M. domestica bcd* gene has been described. Regions homologous to either the first or the fourth exons were not identified within the sequenced DNA. The purpose of this project was to examine the possibility of coevolution between *bcd* and *hb*. Whilst it is possible that other regions may contribute, the interaction between *bcd* and the *hb* promoter DNA is mediated by the *bcd* homeodomain. Therefore sequencing the homeodomain-containing region was of the highest priority. As the first and fourth exons have yet to be implicated in DNA binding by *bcd*, identification of these was of lower priority (whilst desirable for the sake of completeness) and in the end not pursued due to lack of time. One suitable avenue to be explored in pursuit of the missing exons would include constructing and screening a cDNA library. Another alternative would be to use selected fragments of the genomic clone in northern blots, in order to identify those which hybridise to mRNAs of the same size as *bcd* mRNA transcripts. A third possible approach would be to use RACE (rapid amplification of cDNA ends) PCR, as, for example, used by Peifer and Weischaus (1993) in sequencing the *M. domestica* homologue of the *D. melanogaster* segment polarity gene *armadillo*.

Comparisons

Typically, interspecies comparisons are used to identify conserved and diverged regions of homologous proteins. It is usually assumed that evolutionary conservation reflects functional constraints, and that sequence divergence reflects a lack of these. Alternatively, sequence divergence may reflect a functional divergence of the proteins. However, as discussed in Chapter 1, if an interacting pair of molecules have co-evolved, it is possible for a degree of sequence divergence to take place without disturbance of function. Hence, in order to examine the possibility that *bcd* and *hb* have co-evolved, it is necessary to look at the pattern of conservation and divergence of the protein.

The expression patterns of a number of early developmental genes are conserved between *M. domestica* and *D. melanogaster* (Sommer and Tautz, 1991), suggesting that their logical function within the developmental gene hierarchy has been maintained.

Even if the precise molecular basis is not conserved, the outcome of the interaction between *bcd* and *hb* appears to be the same in both *D. melanogaster* and *M. domestica* - that is, it results in the zygotic expression of the *hb* gene over 55% of the egg length. If the function of the protein is conserved, sequence divergence in specific regions of the *bcd* protein can be ascribed either to molecular coevolution, or to lack of functional constraint.

Does either the degree of divergence within *bcd* between *Drosophila* and *M. domestica*, or the nature of the changes involved, suggest that they are not neutral?

Conservation and Divergence

ALTERNATE SPLICING

The alternate splicing event resulting in the inclusion of the 5 extra amino acids in the 2.6 kb *bcd* transcript in *D. melanogaster* occurs just N-terminal to the homeodomain (see Figure 3.5b). Similar alternative splicing events have also been seen upstream of the homeobox in other developmental genes including *Ubx* (O'Connor *et al.*, 1988), *Antp* (Birmingham and Scott, 1988; Stroehrer *et al.*, 1988) and *labial* (Mlodzik *et al.*, 1988). The significance of these events is not clear, but the proximity to the homeobox has led a number of people to speculate that it may influence DNA binding specificity or potential interactions with other transcriptional regulatory proteins (above, and Seeger and Kaufman, 1990). However, as only a single form of transcript is sufficient for the *bcd* activity in both *D. pseudoobscura* and *M. domestica*, and as the single *D. pseudoobscura* form is capable of proper *hb* regulation in *D. melanogaster* (Seeger and Kaufman, 1990), then if the longer form does have a function in *D. melanogaster*, it might not be related directly to *hb* regulation.

PEST

In contrast to *D. melanogaster*, which contains a single PEST region (Driever and Nüsslein-Volhard, 1989), *M. domestica* contains two. How the presence of a second PEST sequence might affect the rate at which bcd is degraded is not entirely clear. As PEST regions are associated with rapid protein degradation (Rodgers *et al.*, 1986), it is possible that the presence of a second PEST domain would increase the rate of degradation. However, whilst there is a statistically significant correlation between the presence of a PEST sequence within a protein and it's displaying a short half life, the mechanism by which PEST sequences confer rapid protein degradation is not known. Therefore, it is not possible to make predictions about how the presence of a second region might affect rates of degradation.

If the presence of a second PEST domain did affect the stability of the *M. domestica* bcd protein, either by increasing or decreasing its stability, this would have important consequences. The slope of the bcd gradient along the length of the developing embryo is determined by three factors - the concentration of bcd at the anterior pole, the rate of diffusion of the protein, and the rate at which it is degraded (Driever and Nüsslein-Volhard, 1988a). If none of the other factors are changed, an increase in stability would result in a longer, shallower gradient, and conversely, a decrease in stability would give rise to a shorter, steeper one. Unless compensated for (by a change in bcd binding affinity, or the rate of bcd diffusion, for example), this could be expected to change the expression domains of all downstream genes. It would, therefore, be important to investigate the stability of the *M. domestica* bcd protein.

OPA REPEATS

As well as in *bcd*, opa repeats have been found in the genes encoding many other transcription factors involved in *D. melanogaster* development, including *engrailed* (*en*), *hb*, *Antennapedia* (*Antp*), *Deformed* (*Dfd*), and *fushi tarazu* (*ftz*); and also in the gene encoding the transmembrane receptor protein *Notch*. Variation in opa repeat length has

has been seen in *en* (Kassis *et al.*, 1986) and *hb* (Treier *et al.*, 1989) between *D. virilis* and *D. melanogaster*, as well as in *bcd* between *D. melanogaster* and *D. pseudoobscura* (Seeger and Kaufman, 1990), and now in *M. domestica*.

The opa repeats found within the *bcd* gene have expanded in *D. pseudoobscura* relative to *D. melanogaster*. In *M. domestica*, the opa repeats have contracted markedly, but have been partially replaced by shorter strings of prolines and histidines. This variation in opa repeat length may simply be due to the inherent mutational properties of simple sequence DNA subject to slippage-like processes. These occur at a significantly higher rate than point mutations (Levinson and Gutman, 1987; Tautz *et al.*, 1986), and so result in the fairly rapid generation of variation in copy-number of short repetitive motifs. Slippage-like processes implicate an inherent bias for the types of changes which may occur. If there is a certain prevalence of particular sequence motifs in a given region, it is likely that these will be the ones mainly involved in further rounds of the slippage process (Tautz *et al.*, 1986). Therefore, if slippage is acting on opa repeats, it would be expected to result in the addition or deletion of permutations of a (CAG)_n motif - which is what is observed in *bcd* between the two *Drosophila* species and *M. domestica*. So, the inherent instability of the opa repeat regions, due to the propensity of slippage to act on simple sequence, is reflected in the variation in the extent and precise location of the opa repeats.

Whilst the opa repeats seen in the *D. melanogaster* and *D. pseudoobscura* *bcd* genes are much reduced in *M. domestica*, they are partially replaced by short strings of prolines and histidines. Replacement of one triplet array by another is commonly observed (Dover, 1993). It is interesting that the opa (i.e. poly-glutamine) repeats should have been replaced by poly-proline, as it has been suggested that homopolymeric stretches of both glutamines and prolines can act as activators of transcription (Mitchell and Tjian, 1989).

Gerber *et al.* (1994) have found that, in cell transfection assays, strings of prolines or glutamines were able to activate transcription when fused to the DNA binding domain of the GAL4 protein, with the highest transcriptional activity observed for

constructs containing a (shorter) 10- oligomer of proline, or a (longer) 10-40 oligomer of glutamine. So a functional explanation for the initial presence of the opa repeats within many transcription factors, including *bcd*, could be that they are involved in transcriptional activation. However, whilst they may have a function in transcriptional activation, slippage results in rapid divergence.

If the opa repeats are involved in transcriptional activation, then (unless the poly-proline/histidines serve a similar function) it is possible that the drastic reduction in their extent within *M. domestica* *bcd* could have a morphological effect. If this is the case, then any morphological change would have not have arisen as a direct result of adaptation to selective pressures from the outside, but could represent a case of molecular drive contributing to morphological evolution (Treier *et al.*, 1989).

HOMEODOMAIN

The homeodomain and exon 2 constitute the most highly conserved regions of the *bcd* protein. The homeodomain does, however, contain 5/60 amino-acid differences to the *D. melanogaster* and *D. pseudoobscura* homeodomains (92% conservation at the amino acid level). This is a relatively high level of divergence compared to that seen for other homeobox containing genes, where there may be complete conservation of the homeodomain sequence even between *D. melanogaster* and humans. However, the most highly conserved homeodomains are generally found in the Hox group of homeotic genes, and this may reflect the different developmental constraints faced by this later acting, tightly co-ordinated cluster of genes (Akam *et al.*, 1994; Dawes *et al.*, 1994).

Surprisingly, the degree of conservation within the homeodomain is no greater than that seen for exon 2, with both showing the same degree of nucleotide and amino acid identity (75%, and 92% respectively) between *M. domestica* and *D. melanogaster*. Relatively fast divergence of the *bcd* homeodomain might reflect a lesser degree of functional constraint, molecular coevolution, or a divergent functional role.

A changed specificity of the homeodomain-DNA interaction ?

The interaction between *bcd* and *hb* is mediated by the homeodomain of *bcd*, which binds directly to high affinity sites within the *hb* promoter (see the introduction to Chapter 3). Given that the *M. domestica bcd* homeodomain is not identical to those of both *D. melanogaster* and *D. pseudoobscura*, it is interesting to ask if any of the changes observed might result in a changed binding specificity. Structural studies have been carried out for a number of different homeodomain-DNA complexes, although not for *bcd* itself. Genetic analyses have also provided information about homeodomain-DNA binding for a number of proteins, this time including *bcd*. On the basis of the models of the general homeodomain interaction resulting from these analyses, and taking into account the specific *bcd* information, it may be possible to draw inferences as to whether the changes seen in the *M. domestica bcd* homeodomain could affect its specificity.

STRUCTURAL STUDIES

The structure of the homeodomain-DNA complex has been resolved by nuclear magnetic resonance (NMR) for Antennapedia (Otting *et al.*, 1990; Billeter *et al.*, 1993), and by X-ray crystallography for engrailed (Kissinger *et al.*, 1990) and the more divergent MAT α 2 (Wolberger *et al.*, 1991) and Oct-1 (Klemm *et al.*, 1994) homeodomains. The overall arrangement of the homeodomain-DNA complexes is quite similar, despite the fact that the homeodomains belong to only distantly related homeobox families. This suggests that the general model of homeodomain :DNA interaction derived from these studies may be quite widely applicable.

The homeodomain consists of three helices. Helices 1 and 2 are arranged in an antiparallel alignment above the DNA, spanning the major groove at more or less right angles to the local direction of the DNA backbones. The recognition helix is suspended

at right angles below helices 1 and 2, aligned within the major groove. Most of the specific contacts occur between the recognition helix and the major groove of the DNA. The flexible N-terminal arm also makes specific contacts with the DNA, binding to the minor groove. The loop between helices 1 and 2, and the first few bases of helix 2 contact the DNA backbone. A schematic diagram of a homeodomain-DNA complex is given in Figure 3.8a, and Figure 3.8b summarises the contacts between protein and DNA on a residue by residue basis. These are then related to the changes in the homeodomain.

Of the amino acids that are different between *M. domestica* and *D. melanogaster* bcd, only one, amino acid 28, is in a position to contact DNA directly. In Antp, the side chains of Arg-28 contact the phosphate backbone of the DNA (Otting *et al.*, 1990; Billeter *et al.*, 1993). Helix swapping experiments between ftz and Sex combs reduced (Scr) also demonstrated the requirement for Arg-28 for efficient recognition of target sites (Furukubo-Tokunaga *et al.*, 1992). However, with both engrailed and MAT α 2 (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991), the arginine at position 28 is replaced by other amino acids which fail to contact the DNA backbone. In *D. melanogaster*, this position is occupied by alanine, which is unlikely to contact DNA as its side chain is very small. As it is replaced in *M. domestica* by the similarly small serine, a change in binding specificity seems unlikely to result.

From the structural data, it seems unlikely that the changes seen will *directly* affect the specificity of the bcd homeodomain - as there are no changes in the amino acids at those positions that are known to make direct contact with the DNA in other (non-bcd) homeodomains. However, it is possible that they could result in an alteration of the orientation of the helices with respect to one another and the major groove of the DNA, resulting indirectly in a changed binding specificity by changing the relative importance of different contacts between *D. melanogaster* and *M. domestica*. Furthermore, the bcd homeodomain may have a different mode of binding DNA - no structural studies have yet been made of the bcd homeodomain, and given the general degree of divergence of its homeodomain from those of the Antp class (approximately 40%), it may not be realistic to expect the details of the interactions to be applicable.

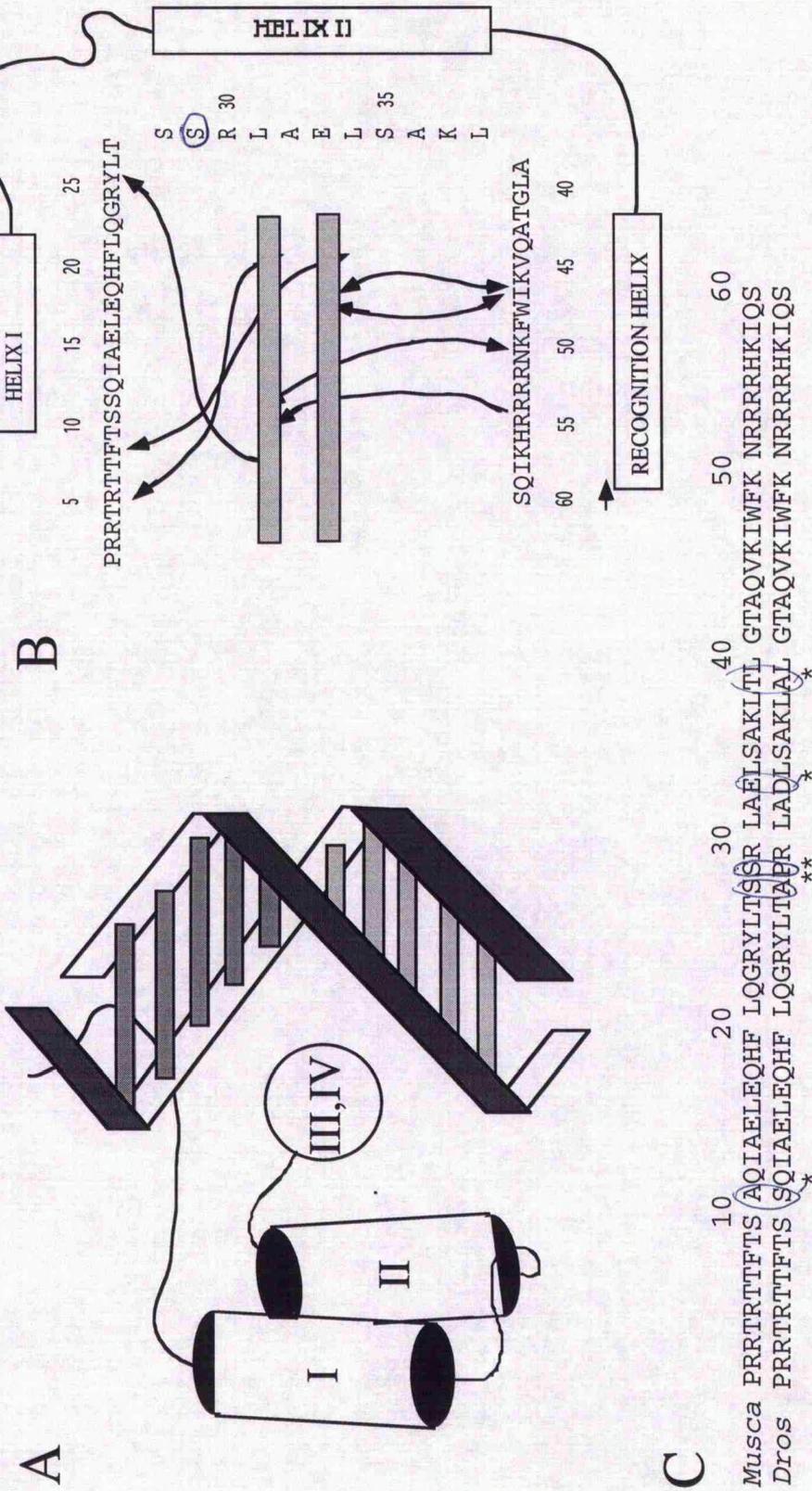


Figure 3.8. A - Schematic diagram of the general mode of interaction of a homeodomain with DNA (from Gehring et al., 1994). B. Summary of the intermolecular contacts between a homeodomain and its binding site (based on Otting et al., 1990). The amino acid sequence given is that for *Musca* bicoid. C. Alignment of the amino acid sequences of the *Musca* domestica and *Drosophila melanogaster* bicoid homeodomains, using the one letter protein code. The *D. melanogaster* sequence is from Berleth et al., (1988).

Indeed, the structural studies (Otting *et al.*, 1990; Kissinger *et al.*, 1990; Qian *et al.*, 1989) suggest that homeodomains contain a helix-turn-helix structural motif remarkable similar to that of the prokaryotic helix-turn-helix repressor proteins (Pabo and Sauer, 1984) which show a different mode of binding, with the N- rather than the C- terminal of the recognition helix inserted into the major groove of the DNA (Kissinger *et al.*, 1990). This suggests that it is not impossible that the precise way in which the homeodomain interacts with DNA varies from gene to gene. Although the general global fold of the homeodomains so far examined, and their modes of docking in major groove, are very similar (reviewed Gehring *et al.*, 1994) small differences in orientation could result in different contacts with the DNA. Therefore, the lack of correspondence of the amino acid positions at which changes have occurred between *M. domestica* and *D. melanogaster* bcd homeodomains to the positions implicated by structural studies of non-bcd homeodomains in *D. melanogaster* in determining specificity does not necessarily reflect a conserved bcd specificity.

GENETIC STUDIES

Homeodomain-DNA interactions have also been analysed genetically. Most genetic studies of the bcd homeodomain have concentrated on the interaction of the amino acid at position 50 of the homeodomain (position 9 of the recognition helix) which contacts the two nucleotides three prime to the TAAT core of the bcd binding site (consensus sequence TCTAATCCC). Changing this between lysine and glutamine is associated with a change in specificity from TAATCC to TAATGG (Hanes and Brent, 1989, 1991). This has also been observed for other homeodomain proteins (Percival-Smith *et al.*, 1990; Treisman *et al.*, 1989; Schier and Gehring, 1992), and is in agreement with the structural data. The amino acid at position 50 is unchanged in *M. domestica*, and so the specificity of bcd for the nucleotides 3' to the TAAT core of the binding site may be expected to be unchanged.

Other genetic results for bcd are generally in agreement with the model for homeodomain-DNA interaction suggested by the structural data (Hanes and Brent,

1989), with the carboxy-terminal rather than the amino-terminal end of the recognition helix being inserted into the major groove. However, there are some differences - whereas, in the structural analyses, residue 50 of the engrailed homeodomain contacts bp 8 (TGTAATTAC [Kissinger *et al.*, 1990]), and in Antp it contacts both bp 7 and 8 (TCTAATGGC [Otting *et al.*, 1988]), the genetic analysis of Hanes *et al.* (1994) indicates that, in *bcd*, residue 50 contacts bp 7 (TCTAATCCC), but its ability to discriminate is influenced by the identity of bps 8 and 9. It is possible that these discrepancies in the contacts made by the different homeodomains are artifactual - a result of the unusual conditions used in X-ray crystallography and NMR, but alternatively, it may reflect a difference in the positioning of Gln-50 within the major groove, as a result of the amino acids that are not conserved among the homeodomains of *bcd*, engrailed and Antennapedia (Hanes *et al.*, 1994). If these differences are not artefacts, then this provides evidence that the modes of binding between different homeodomains may be different, causing a subtle change in the nature of the interactions with DNA.

A changed specificity?

The structural studies and genetic analyses probably rule out any *major* changes in *bcd* binding specificity between *D. melanogaster* and *M. domestica*. From the structural and genetic experiments performed to date, it is not possible to argue that any of the changes seen in the *bcd* homeodomain will directly result in an altered binding specificity. However, there are no structural data for *bcd*, and there is genetic evidence that the precise positioning of at least one DNA contacting residue may be different between *bcd*, *en* and *Antp*. Therefore the possibility cannot be ruled out that small differences in the way *bcd* binds DNA compared to other studied homeodomains result in positions coming into play that are changed between *M. domestica* and *D. melanogaster*. The fact that none of the changes between the *M. domestica* and *D. melanogaster* *bcd* homeodomains are found in positions that directly contact DNA in structural studies, or have been shown to be important in determining specificity in genetic studies, need not absolutely mean that there are no (subtle) changes in the

binding specificity between the two species. Furthermore, as *bcd* binds to multiple sites, it is possible that small differences in the affinity of binding to each site would be summed to give larger overall differences in binding to multiple sites. It is also important to remember that changes in binding specificity could be moderated *in vivo* by protein-protein interactions - and that *bcd* and *hb* may possibly act synergistically in the activation of zygotic *hb* expression. These possibilities will be discussed further in Chapter 5.

Chapter 4

THE *MUSCA DOMESTICA* HUNCHBACK GENE

As described in Chapter 1, the segmentation gene *hunchback* (*hb*) is a member of the 'gap' class of genes controlling early development in *Drosophila*, and as such is involved in the transmission of maternal positional information to the next level down within the zygotic developmental gene hierarchy, that of the primary pair rule genes. In addition to its gap gene role, *hb* has a maternal function, and also shows later zygotic expression.

GENE STRUCTURE

The *Drosophila melanogaster* *hb* gene has been shown to contain two promoters (P1 and P2), leading to the production of two different transcripts (Tautz *et al.*, 1987; Figure 4.1). The 2.9 kb transcript is produced from the proximal P2 promoter in response to zygotic activation by the maternal gene *bicoid* (Schröder *et al.*, 1988; Driever and Nüsslein-Volhard, 1989a; Struhl *et al.*, 1989) and is necessary for gap gene function, whilst the distal P1 promoter is involved in maternal and later zygotic expression (Tautz *et al.*, 1987; Bender *et al.*, 1988; Margolis *et al.*, 1994; Lukowitz *et al.*, 1994) and results in a 3.2 kb transcript. Both transcripts contain distinct first exons, which splice to a common second exon containing the protein coding region (Tautz *et al.*, 1987; Bender *et al.*, 1988).

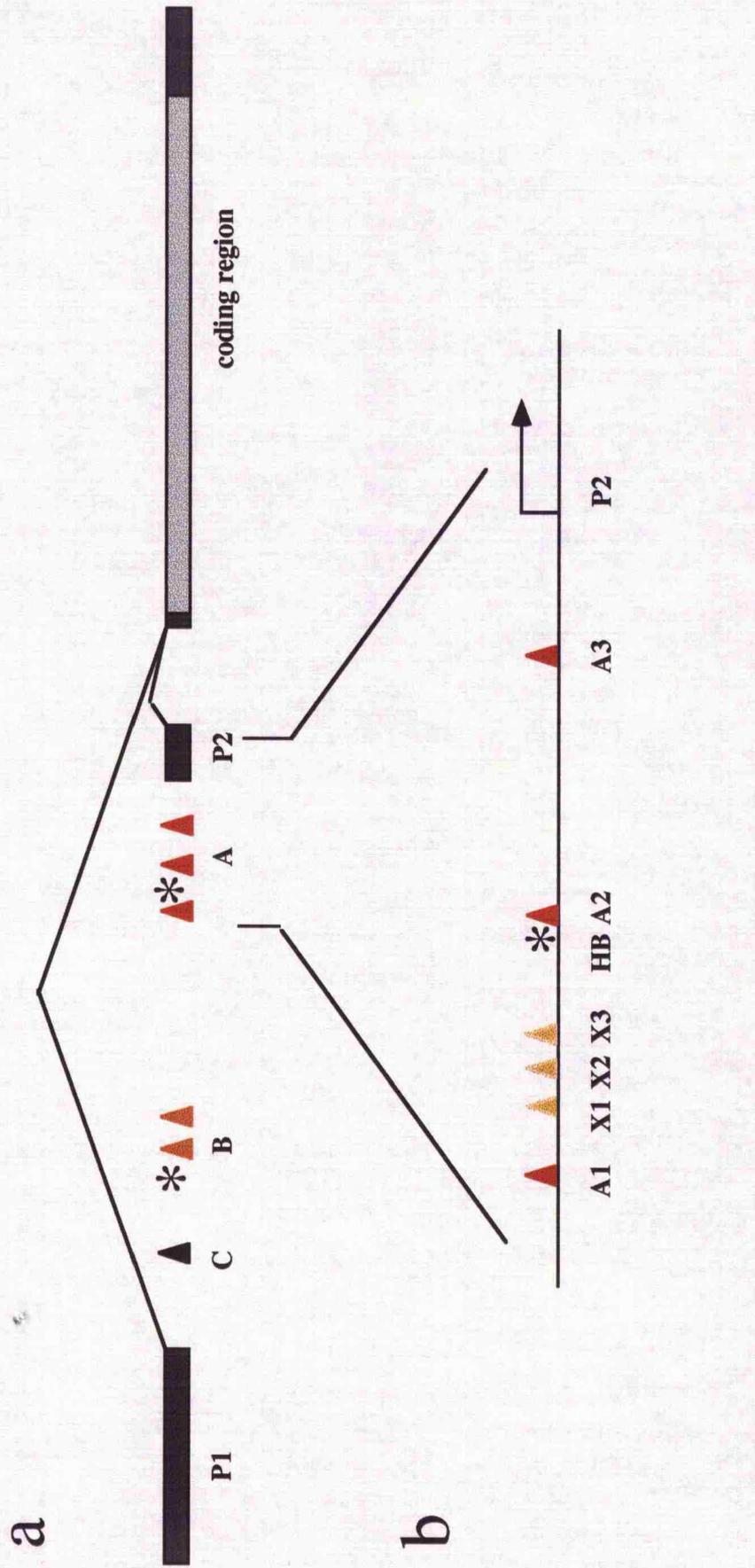


Figure 4.1

a, A schematic diagram of the structure of the *Drosophila melanogaster hunchback* gene, illustrating the relative locations of the two promoters, and the bicoid and hunchback binding sites (Driever and Nusslein-Volhard, 1989; Driever et al., 1989; Triesman and Despaln, 1989).
b, the A site region in more detail, showing the low affinity sites identified by Driever et al., (1989). P1 - maternal and later zygotic promoter. P2 - (bcd dependent) early zygotic promoter. The triangles represent the bcd binding sites, with the sets of high affinity A sites shown in red, and the lower affinity B sites in orange. The low affinity X sites are shown in yellow. Asterisks indicate hunchback binding sites..

PROTEIN STRUCTURE

The *D. melanogaster* hb protein (illustrated in Figure 4.4) contains two DNA binding zinc finger domains, composed of four and two his-cys zinc fingers respectively (Tautz *et al.*, 1987). It also contains a region designated "box A" by Tautz *et al.* (1987) that shows homology to another zinc finger containing gap gene, *Krüppel*, and to the *pol* gene of the retrovirus HIV-1, but whose function is so far unknown. Comparison with the sequence of *hb* in *D. virilis* (Treier *et al.*, 1989) has shown that the two zinc finger regions are highly conserved, with the first zinc finger domain showing only a single amino acid difference between the two species. However, the area in-between the two zinc finger domains is relatively highly diverged, correlated with a high degree of cryptic simplicity in the sequence, which would, as discussed in Chapter 1, lead to a frequent occurrence of slippage events (Tautz *et al.*, 1986).

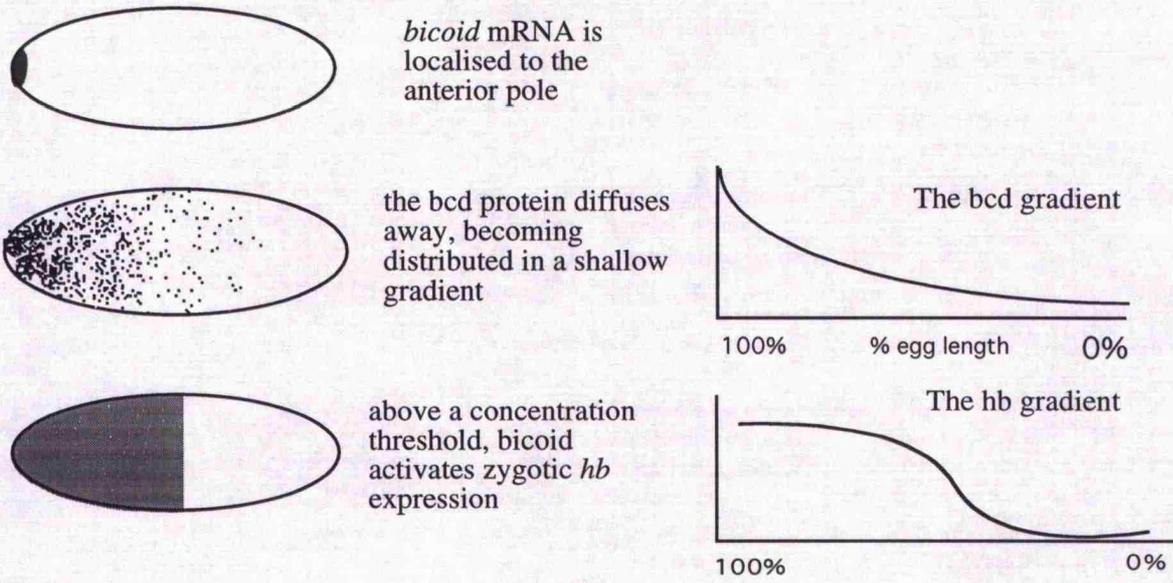
HB EXPRESSION

There are three separately regulated expression patterns of *hb* in the early embryo (Tautz, 1988) - maternal, early zygotic and later zygotic.

Maternal *hb* mRNA is transcribed during oogenesis, and is homogeneously distributed within the egg (Bender *et al.*, 1988; Tautz *et al.*, 1987; Tautz and Pfeifle, 1989). It is translationally inhibited by posterior determinant *nanos* in the posterior half of the embryo (Wang and Lehmann, 1991), resulting in a domain of maternal *hb* protein expression in the anterior half of the embryo (Tautz, 1988; Figure 4.2). Maternal expression of *hb* has been conserved in *D. virilis* (Treier *et al.*, 1989) and in *M. domestica* (Sommer and Tautz, 1991).

Early zygotic expression of *hb* occurs in a domain extending through the anterior half of the egg in response to activation above a certain concentration gradient by the product of the maternal gene, *bcd* (Tautz, 1988; Schröder *et al.*, 1988; Driever and Nüsslein-Volhard, 1989; Struhl *et al.*, 1989). This early zygotic *hb* expression domain is the same as that resulting from maternal expression (Figure 4.2). The early zygotic

A. Zygotic control of *hunchback* expression



B. Repression of maternal *hb* mRNA translation by *nanos*

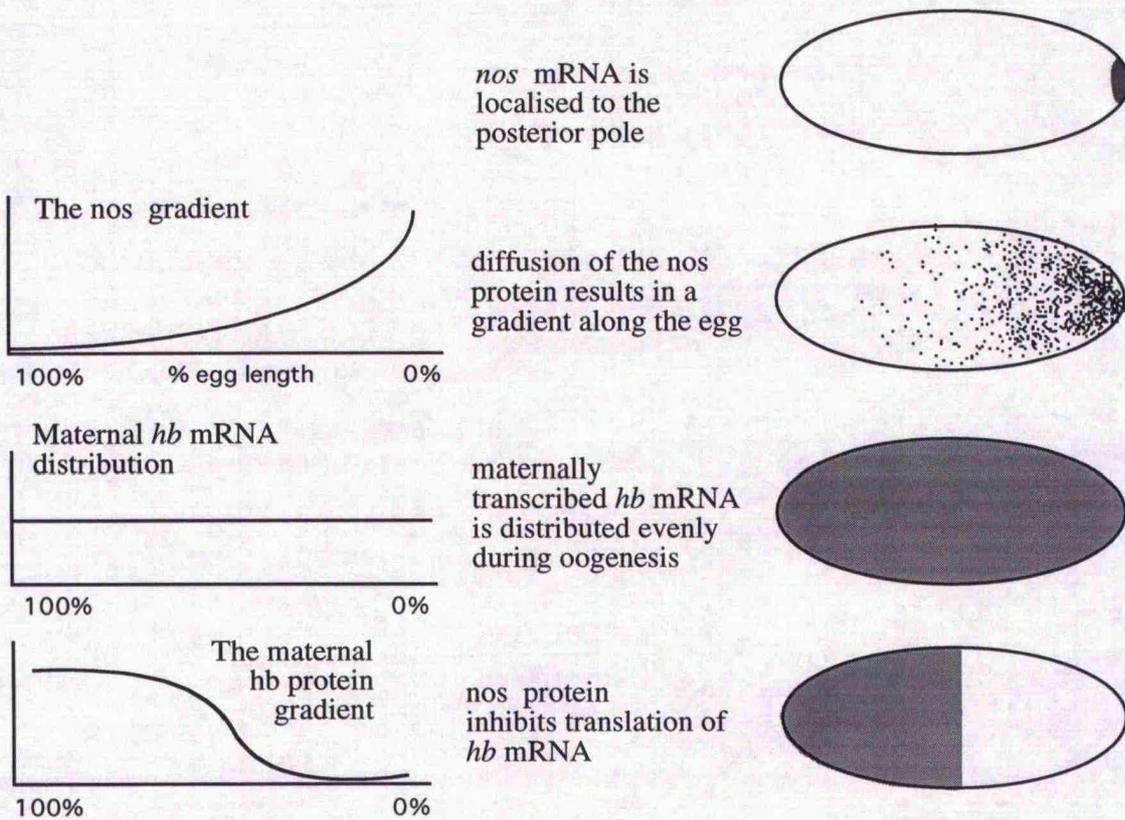


Figure 4.2. The redundant specification of the early zygotic domain of *hb* expression in *D. melanogaster*

expression of *hb* has been conserved in *D. virilis* (Treier *et al.*, 1989,) and in *M. domestica* (Sommer and Tautz, 1991), occurring in each case over about 55% of egg length.

The later zygotic expression of *hb* occurs at late blastoderm stage, in several domains, the most prominent of which are two stripes (one central, and one in the posterior region of the embryo), with a variable and weak domain of expression in the anterior region (Tautz, 1988; see also Figure 6.3, Chapter 6). The expression of these stripes is broadly conserved in *D. virilis* (Treier *et al.*, 1989; Lukowitz *et al.*, 1994) and *M. domestica* (Sommer and Tautz, 1991), and in *M. domestica*, is followed during gastrulation by a novel expression pattern in 13 irregular stripes (Sommer and Tautz, 1991).

Control of early zygotic expression

Initially, the phenotype of mutations in the *hb* gene (deletion of gnathal and thoracic segments - Lehmann and Nüsslein-Volhard, 1987), and the expression pattern of *hb* in *bcd* mutant embryos (Tautz, 1988) suggested *hb* as a possible target for regulation by the maternal gene, *bcd*. Subsequently, immunoprecipitation assays and footprinting experiments allowed for the identification of sites within the zygotic *hb* promoter that bind *bcd* *in vitro* (Driever and Nüsslein-Volhard, 1989). Three *bcd* binding DNA fragments (A, B & C) were identified upstream of the P2 promoter (Figure 4.1). Footprinting experiments showed that these fragments contained 3, 2 and 1 *bcd* binding sites respectively, defining a consensus sequence of TCTAATCCC, with the three "A" fragment sites (which show the closest match to the consensus) having the highest affinity. A further three minor, low affinity, *bcd* binding sites were identified by footprinting within the A fragment (Figure 4.1 - X1-3).

Subsequently, transient expression assays showed that the "A" fragment is necessary and sufficient to activate zygotic *hb* expression (Driever and Nüsslein-Volhard, 1989) and transformation experiments indicated that it is capable of mediating

expression of reporter gene constructs in the authentic zygotic *hb* spatial domain (Schröder *et al.*, 1988; Driever *et al.*, 1989a; Struhl *et al.*, 1989). These results, and experiments involving the activation of reporter plasmids by *bcd* in a (yeast) heterologous system (Hanes and Brent, 1991) suggest that the high affinity sites identified *in vitro* are important in mediating the activation of *hb* by *bcd* *in vivo*. However, whether the "X" sites are important in the regulation of *hb* expression *in vivo* is not clear. If *bcd* binding is independent (i.e. not co-operative), then these low affinity sites would be occupied only within a restricted domain - and yet the level of *hb* expression is constant throughout its anterior domain. However, if co-operative interactions were involved in *bcd*-binding, then these sites might play a role. It would be interesting to investigate their role by examining the expression of reporter genes in which the X sites had been replaced with random sequence.

As discussed in Chapter 1, recent experiments by Simpson-Brose *et al.* (1994) demonstrate a synergistic interaction between *bcd* and *hb*, and suggest that the synergism is possibly required for the full activation of early zygotic *hb* expression. This is supported by work showing that *hb* binds *in vitro* to a number of sites within its own promoter, adjacent to the A and B sets of *bcd* binding sites upstream of the *hb* promoter. (Treisman and Desplan, 1989; Figure 4.1).

Examination of the upstream sequence of the *D. virilis hb* gene has revealed that the three "A" *bcd* binding sites are fairly well conserved (Treier *et al.*, 1989) - sites A2 and A3 are identical, whilst site A1 has the sequence TCTAATCT in *D. virilis* as opposed to CGTAATCC in *D. melanogaster*. Furthermore, their spacing is conserved - despite the apparent occurrence of a number of insertion/deletion events in this region (Treier *et al.*, 1989), including one which removes two of the X sites. Transformation experiments indicate that this *D. virilis bcd* dependent element is fully functional in *D. melanogaster* (Lukowitz *et al.*, 1994), despite the minor changes in the A1 site sequence and the absence of two of the low affinity X sites.

Control of maternal and later zygotic expression

Genetic analyses of the *hb* P1 promoter using reporter gene constructs have identified regions that appear to be important in regulating maternal and later zygotic expression (Lukowitz *et al.*, 1994; Margolis *et al.*, 1994; 1995). Within these regions, there are a number of blocks of sequence conserved between *Drosophila melanogaster* and *D. virilis*, which may represent conserved binding sites for regulatory proteins (Treier *et al.*, 1989; Lukowitz *et al.*, 1994).

Later zygotic *hb* expression is under the control of the terminal system (Brönner and Jäckle, 1991; Casanova, 1990; Margolis *et al.*, 1995; Nüsslein-Volhard *et al.*, 1987; Tautz, 1988) and *Krüppel* and *hb* itself (Hülskamp, 1991). Examination of the region of DNA that mediates these expression aspects has identified binding sites for the products of the terminal system genes *tailless (tll)* and *huckebein (hkb)*, and for the Kr and hb proteins (Lukowitz *et al.*, 1994, Margolis *et al.*, 1995).

MOLECULAR COEVOLUTION

As discussed in Chapter 1, changes in the regulation of proteins are considered to be as important for the course of evolution as changes in the regulatory proteins themselves. Divergence in a regulatory region might reflect lack of function, adaptation or molecular co-evolution. In the last possibility, key compensatory changes can take effect as a consequence of the tolerance and flexibility stemming from genetic and functional redundancy.

The early zygotic expression of *hb* is genetically redundant in terms of having multiple binding sites for regulatory factors, and functionally redundant because of the additional maternal expression. The early expression patterns of both *bcd* and *hb* are conserved in *M. domestica* (Sommer and Tautz, 1991), which is consistent with the assumption that a regulatory interaction between *bcd* and *hb* is present in this species also. If this is so, then changes in the binding sites could represent molecular co-

evolution, rather than divergence. In this context it is important to analyse the number and composition of bcd binding sites in *M. domestica*.

In an attempt to identify the bcd dependant regulatory elements within the *M. domestica hb* gene, an extensive region of upstream DNA, together with the 5' portion of the *hb* coding region, has been sequenced. This has been compared to the sequences published for *D. melanogaster* (Tautz, 1987) and *D. virilis hb* (Treier *et al.*, 1989) genes, and putative regulatory regions have been identified.

METHODS

Materials

GENOMIC DNA LIBRARIES

The *M. domestica* EMBL 3 genomic DNA library was described in Chapter 3.

A *M. domestica* genomic DNA was a gift from Ralf Sommer and Dietard Tautz. The library was constructed through the ligation of size-fractionated, partial *Sau* 3A digested *M. domestica* genomic DNA into *Bam* HI-cut Lambda FIX arms.

HB ZINC FINGER PROBE

A PCR fragment containing the first zinc finger domain of the *M. domestica hb* gene (Figure 4.3), ligated into the M13mp18 vector was a gift from Ralf Sommer. This fragment has been described in Sommer and Tautz (1991).

Methods

LIBRARY SCREENS

The Lambda FIX library was found to be contaminated with bacteria. This contamination was removed by extracting a 1 ml aliquot of the library with 20 μ l of chloroform, on a revolving wheel for 6 hr's, followed by centrifugation for 30 seconds in a bench top micro-centrifuge.

Both libraries were plated and screened as described in Chapter 2. The *M. domestica hb* zinc finger fragment described above was used as a probe. The final washes were in 0.1 x SSC, 0.1% SDS at 65 °C.

A total of 1.5×10^5 plaques from the Lambda FIX library and 2×10^5 plaques from the EMBL 3 library were screened. As discussed on page 66, if it assumed that the library phage have an average insert size of 1 kb, and the *M. domestica* genome is 5×10^5 kb, then it is necessary to screen 1.5×10^5 phage per library in order to have a 99% chance of a given sequence being present in that library screen.

Six potential positives were identified from initial screen of the Lambda FIX library, none of which proved positive on the rescreen. However, of eight potential positives were identified on the initial screen of the EMBL 3 library, one, λ hb 3.15, proved positive on rescreening. DNA was isolated from this phage using the plate lysate method (Chapter 2)

RESTRICTION MAPPING AND SUB-CLONING

The positive *hb* library clone, λ hb 3.15, was restriction mapped. Restriction enzyme digests were performed, and the DNA run on 0.8% agarose gels. These were blotted, and probed with the *hb* probe used to screen the library. This allowed the calculation of a rough restriction map (Figure 4.3), and the identification of a 5.9 kb *Pst* I fragment that contained the first zinc finger domain of the *hb* gene. The *Pst* I fragment was isolated and ligated into *Pst* I- cut Bluescript to give the plasmid hbP1

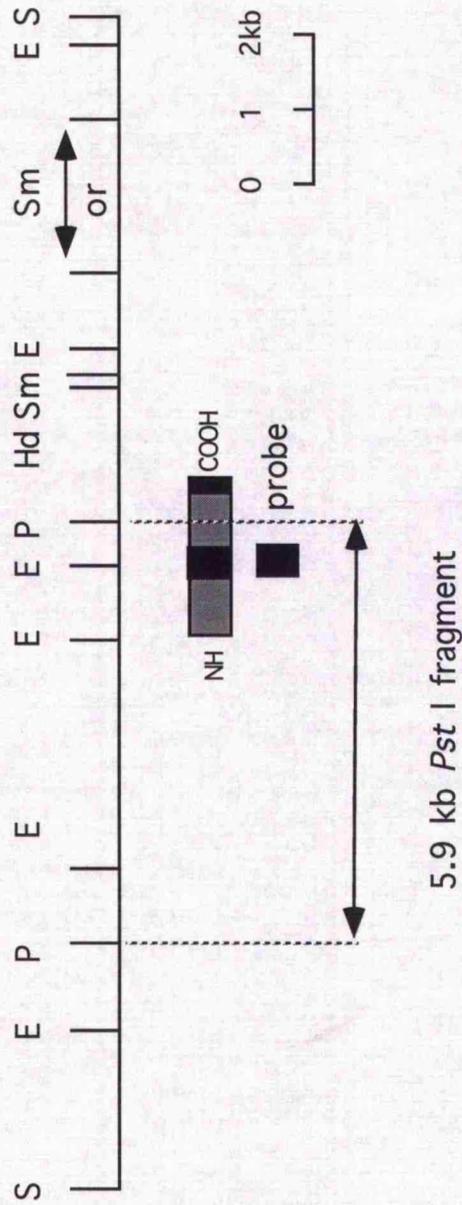


Figure 4.3

Restriction map of the insert of the library phage (λ hb 3.15) containing the *Musca domestica* *hunchback* gene. The 5.9 kb *Pst* I fragment that has been sequenced is indicated. The *hunchback* coding region is represented as a grey box, and the zinc finger domains within it as black boxes. The location of the *M. domestica* *hb* zinc finger probe is shown. Enzymes: S, *Sal* I; E, *Eco* RI; P, *Pst* I; Hd, *Hind* III; Sm, *Sma* I.

(Figure 4.4). This plasmid was then further restriction mapped. As both *Eco* RI and *Hinc* II were known to cut within the probe, it was possible from this to determine that the probe region was at the 5' end of the 5.9 kb *Pst* I fragment. Therefore, the insert either contained mostly coding or mostly upstream DNA

SEQUENCING

Initially, sequence was obtained from both ends of the 5.9 kb *Pst* I insert by using the vector universal and reverse primers and double stranded DNA. The sequence obtained from the reverse primer was identical to *M. domestica hb* coding sequence (Sommer, 1992), indicating that plasmid hbP1 contained 1.5 kb of coding and approximately 4.3 kb of upstream DNA.

A number of strategies were pursued in order to obtain the complete sequence of the hbP1 insert - subcloning of specific insert fragments, and random 'shotgun' cloning of restriction fragments; with custom primers being used in order to complete the sequence. A schematic diagram of the various sub clones and primer sites, together with the extent of the sequence obtained from each is given in Figure 4.5. Approximately 80% of the sequence was read from at least two separate subclones or primer sites.

Subcloning

Plasmid hbP1 was digested with *Eco* RI, and the four resulting bands isolated. The 500 bp, 900 bp, and 3.3 kb bands were ligated into *Eco* RI cut Bluescript in both orientations. The remaining band (3.9 kb) contains the vector DNA as well as 1 kb of insert, and so was simply ligated at low concentration to allow it to re-circularise. Further subclones were obtained by ligating the 0.9 kb *Eco* RI - *Eco* RV, and the 1.2 *Hinc* II - *Eco* RI fragments into similarly cut Bluescript. Sequence was obtained from either end of the insert DNA using the vector universal and reverse primers in manual sequencing reaction as described in Chapter 2.

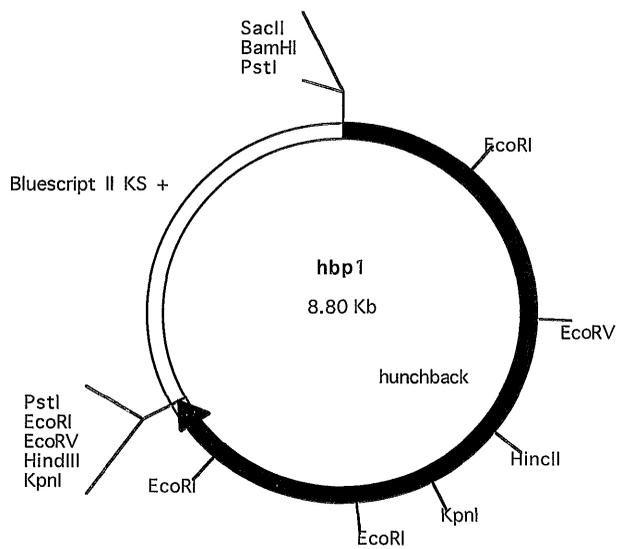


Figure 4.4.

Restriction map of the plasmid hbp1. This contains the 5.9 kb *Pst* I fragment of the *Musca domestica* *hunchback* gene ligated into Bluescript II KS +.

Shotgun cloning

Insert DNA was isolated from hbP1 and digested with either *Taq* I or *Sau* 3A. The digested DNA was phenol:chloroform extracted, then ligated into *Cla* I (*Taq* I digest) or *Bam* HI (*Sau* 3A) cut vector, and transformed into *E. coli* (DH5 α F'). Colonies containing plasmids with inserts were identified by blue/white colour selection. Twenty colonies were chosen at random for each ligation, and used to prepare single stranded DNA for sequencing. Sequencing reactions were performed manually as described in Chapter 2, using the vector's universal primer, and run on ordinary gels. A total of 15 different clones were obtained using this strategy. As it was possible for multiple fragments of the digested DNA to ligate into each plasmid, only sequences between pairs of *Sau* 3A or *Taq* I sites were considered continuous. It should also be pointed out that much of the sequence obtained from individual shotgun cloned fragments overlaps with independently derived sequence of known relative location (see Figure 4.5), and that no discrepancies were observed. The location and extent of the sequence obtained from each shotgun clone is shown in Figure 4.5.

Custom primers

Custom primers were used to fill in the remaining gaps. The relative locations, and extent of sequence obtained from each primer is given in Figure 4.3. The sequences of the primers are given in Chapter 2. The region between the *Hinc* II and *Kpn* I sites (roughly positions 3300-3600 bps, Figure 4.7) showed strong secondary structure, resulting in short runs of ambiguous sequence data when used in ordinary manual sequencing reactions. This region was sequenced by using the custom primers B17 and B17.2 in Cycle Sequencing reactions run on an automatic DNA sequencer, as described in Chapter 2. The increased temperature at which the cycle sequencing reactions were performed allowed clear data to be obtained.

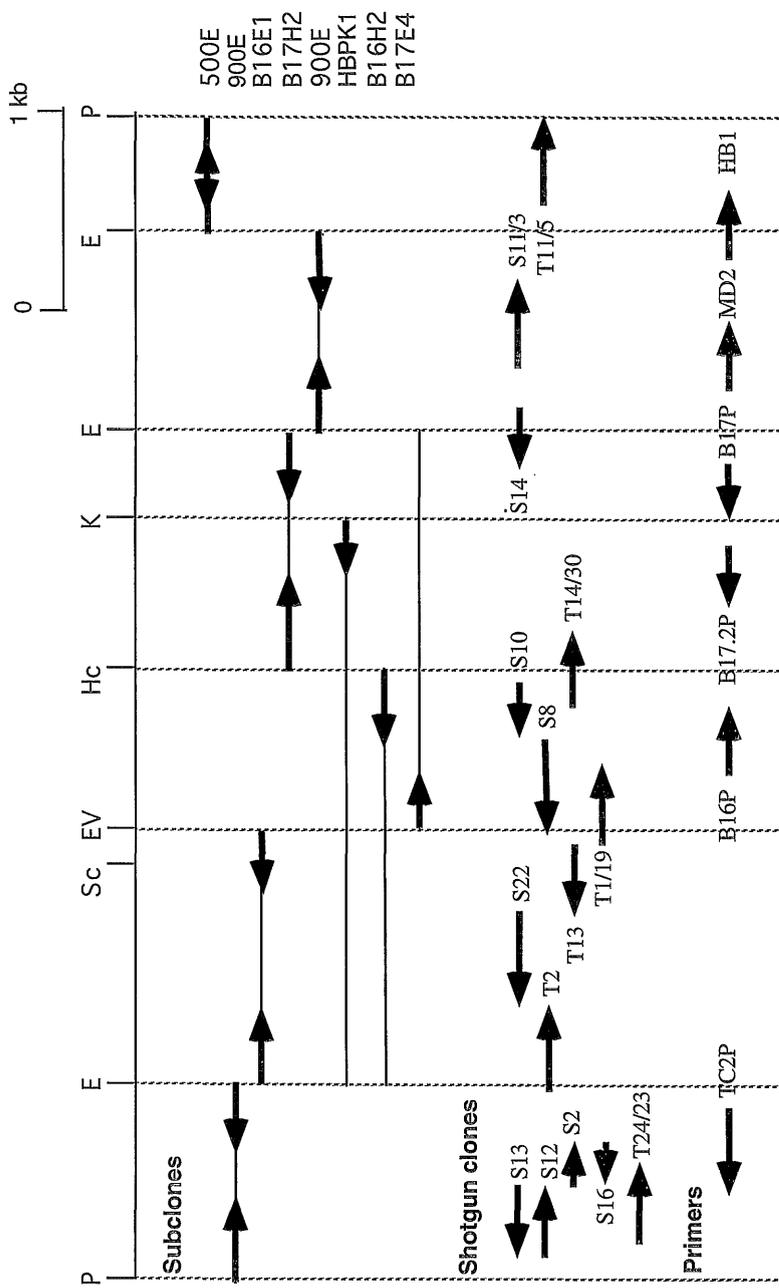


Figure 4.5 A schematic diagram illustrating the strategy used to sequence the 5.9 kb *Pst* I fragment of the *Musca domestica hunchback* gene. The thick black arrows indicate the extent of sequence obtained from each clone or primer. The thin black line shows the region of the *hb* gene within each subclone. P, *Pst* I; E, *Eco* RI; Ev, *Eco* RV; Hc, *Hinc* II; S, *Sac* II; K, *Kpn* I

DNA SEQUENCE ANALYSIS

Dotplot comparisons were performed using the COMPARE and DOTPLOT programs of the UWGCG package. The window size was 21 with a stringency of 14 matches (66% identity).

The *M. domestica hb* upstream sequence was searched for matches to the *Drosophila melanogaster bcd* binding site consensus sequence using the FIND program of the UWGCG package, allowing 0,1 or 2 mismatches.

Multiple alignments of the amino acid sequences of the *hb* proteins of *D. melanogaster*, *D. virilis* and *M. domestica* were generated using the CLUSTAL V package and the default parameters.

NORTHERN ANALYSIS

Total RNA was isolated from adult female *M. domestica* flies. 10 µg of RNA was separated by electrophoresis and blotted onto nylon membranes, as described in Chapter 2. The northern blot was probed with the *M. domestica hb* zinc finger probe described above, as described for Southern blots in Chapter 2.

RESULTS

Sequencing *M. domestica hb*

Two *M. domestica* genomic DNA libraries were screened with a probe to the first zinc finger region of the *M. domestica hb* gene, and a single positive phage isolated. From this, a 5.9 kb *Pst* I fragment has been sequenced. This fragment contains 1.5 kb of *hb* coding DNA, and 4.3 kb of upstream DNA. As the restriction map of the λ clone insert has not yet been compared with the genomic restriction map in Southern blot experiments, the integrity of the insert is not known. The following analysis assumes

that the upstream sequence is contiguous - that the sequence presented does represent the authentic *M. domestica hb* upstream sequence.

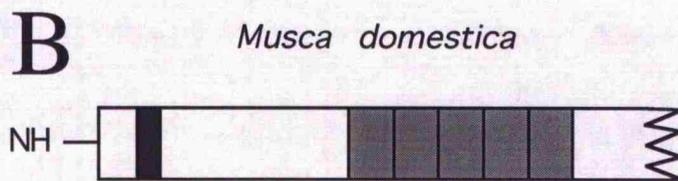
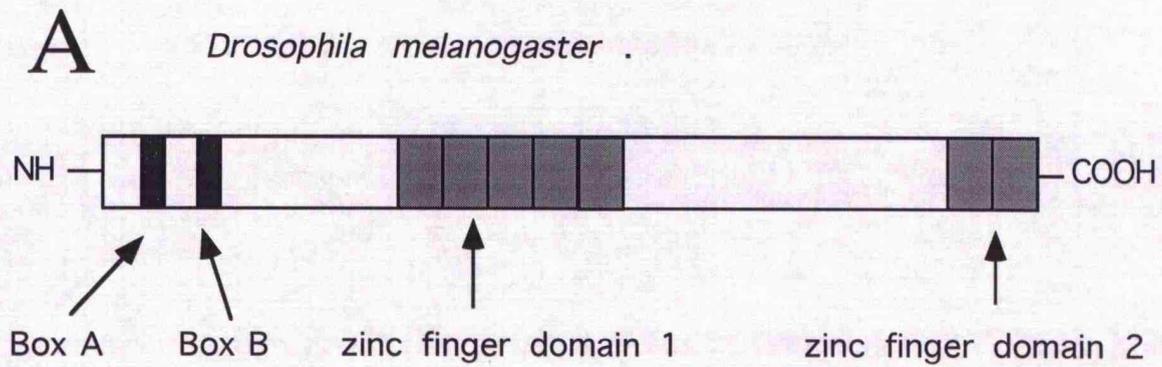
Coding sequence

The *Drosophila melanogaster* hb protein contains two sets of zinc fingers, the first lying roughly in the middle of the protein, the second at the C-terminal end. At the N-terminal end, regions of homology between hb and Krüppel have been noted; these are designated Box A and Box B (Tautz *et al.*, 1987). Figure 4.6 illustrates the structure of the *Drosophila melanogaster* hb protein.

The portion of the coding region of the *M. domestica hb* gene that has been sequenced in this study includes boxes A and B, and the whole of the first zinc finger domain, but does not extend as far as the second zinc finger domain. The region of the *hb* gene sequenced is illustrated in Figure 4.3. The complete sequence of the 5.9 kb *Pst* I fragment of the *M. domestica hb* gene is given in Figure 4.7, and a multiple alignment of the derived amino acid sequence with those of *D. melanogaster* (Tautz *et al.*, 1987) and *D. virilis* (Treier *et al.*, 1989) hb is given in Figure 4.8.

THE FIRST ZINC FINGER DOMAIN

The probe used to screen the library was a degenerate PCR derived fragment of the first zinc finger domain of the *M. domestica hb* gene (Sommer and Tautz, 1991, Sommer 1992, Sommer *et al.*, 1992). Before using the *hunchback* fragment as a probe in the library screen, the sequence was checked, and a number of differences to the previously reported sequence (Sommer, 1992) were noted. The sequence of the probe fragment also differs at a number of points to that of the library clone. These differences are summarised in Figure 4.9.



C

HIV-1 <i>pol</i>	C W W	A G I K Q E	F G I P Y N
		* * * * *	
<i>D. melanogaster Kr</i>	A A L	A G I K Q E	D V . H L D
		* * * * *	* * *
<i>D. melanogaster hb</i>	M F A	A N I K Q E	P G H H L D
	* * *	* * * * *	* *
<i>D. virilis hb</i>	M F A	A N I K Q E	P L S H H H
	* * *	* * * * *	*
<i>M. domestica hb</i>	M F A	A N I K Q E	P G T I N P

Figure 4.6. **A.** A schematic diagram of the *Drosophila melanogaster hunchback* protein, showing the relative positions of the two zinc finger domains, and Boxes A and B. **B.** A schematic diagram showing the section of the *Musca domestica hunchback* protein sequenced in this study. No region with homology to Box B was identified. **C.** Alignment of the Box A amino acid sequences. The alignment of the *D. melanogaster* and HIV-1 sequences was obtained from Tautz *et al.*, (1987). The *D. virilis hb* sequence was obtained from Treier *et al.*, (1989). Asterisks indicate pairs of conserved amino acids. A box has been drawn around the amino acids defined as constituting Box A in Tautz *et al.*, (1987).

The sequence of the 5.9 kb *Pst* I fragment of the *M. domestica hb* gene

1	CTGCAGATAAAAGACTGACCTGTGGAAAAATTTAGCTTTTCTATCTTAATTCATTTTAAAC	60
61	TGTAGCATGATTAACAACGACGGACAGAAGGACGGACAGGCTCAAATCGTCTTAGAATTT	120
121	TACGCTGATCAAGGGTCTAAAATTTATTTTCGACAAAGTTACATACTGAATGACGAATT	180
181	TAATATACCTCCATATCTATGTATGGTGAAGCCAATCCAAAATCAGGAAATTTTCAAAA	240
241	TTTCATCCAATCTACCCGGTACTGATAGTCTGCCACAATATATTAACAATCCCAAGTTT	300
301	TTAGCAAAAAAATGTGTCTCCTATACATCCTTAGGGCTTGCAATCTAATGTCATGTTT	360
361	TTTCCTTTAGTAAATCTATTTGGAAATAATTTTATTTTATTTTAAATGAGAAATTT	420
421	TCTATTTTCCCTAATCTTTTCTTTGAAAATCTATTTTGAACCTTTTAAAGGCAATAGGG	480
481	TCCTCTCAATAAAGTCCGTGATCTGACAGGCTTTCGAAAAATGTTGCATAAATTAATTTTC	540
541	TAGCCTTGCGTATTTTAAATAAAATCAAATGTTGGCACTGATGTCACAAAAA	600
601	AAATTAATTTTAAACACATATGGCTCTATTTCTAATAGCAGATCTAATATTTCTTTAAAAAG	660
661	TTTACTTATTTTCAGTATTTTAAATAGAGTAAAAACATTTTAAAAATAATTCAAATTTCCA	720
721	CCTTTGACGTCTTTTAAATTTTCTTTTAAATCTAATTTGAATTTTCTTTCAATT	780
781	CCTTTAAAAATGAATAGAAAAATAGTGTCTTCTAATAACTGCCTTTCAAAAGTCTCGTTGTA	840
841	GAATTTGTTGAGTTTATTTCTCTTTTCTTTTGGTTAAAAATCTTTTAACTCA	900
901	TATTTCCAGCAATCTTCCGAAAAATCCATGTTTGGCTCTACTTAAAAAATACTAAAA	960
961	AAACCATAGATTTTAAATTTTAAATCCCTAAAAAATCTCAGCTGTGAATTCCTTTACAAA	1020
1021	AAATTTCCAAAAATTTGATATAATTTTATCTCTGATAAAAAAACAACAAATTTGATAGTA	1080
1081	TACCAGAAACCCCAAAATATAGTTGAATATATATTTCCGAAAAATGCCAATTTGATGTC	1140
1141	CTTCTGAAGTTATCAGGAATCTCAGTTTTCATATACCAACCAATGTATAAAAAATAATA	1200
1201	GTTACTAAAAATATATGTTCTTTTAAAAAAGACAACGAAAGACGTCCTTAAAAATTTAT	1260
1261	CCAAAAAAGTATTTATTTAAGTTAAAAATTAATAAAAAATATTGAACAAAAA	1320
1321	TGAAAAAATATTTTTCAGTCTCCACTCATGGTAAATCAACGAGGTAAGTCAAATTCAAA	1380
1381	AAAAATATCCATCTCTGATTTTCTATAAATTTACTTTGGCAATTCACACACAGAAATAGG	1440
1441	TTCCCATCTCTGGATTTGATATATACATGGAATATAAATATAATGAATATGAATTTAAT	1500
1501	TTACATGTGTGAATAAATTTTGAACAACAAATTAACATATGATTTGATGACTTTACCAT	1560
1561	TCATTTATTCATACCAATGGTTTTCAGTTTACGCTTTCATATAGAGATTTTAAATTTTAA	1620
1621	GATTTATCAATAAAATGTCATACGATTTTGATTCGCATTTTCAAAAATATCCTTAAATAT	1680
1681	GTAAATAAGAAAGCTATAAGTTATTTAAAAATCACTCGAAATTTTTCATATTTATGTCCT	1740
1741	TTTAGTCCCTTTTAAAAATCTGTGTTCAAGTCGATTTTAAATTTTTFACATTTTGTAAAA	1800
1801	TATGAAAAATAAAATCCCTGATTTCTAGGATTTTAAATAAGGAAAAATAATTTGAAAT	1860
1861	ATTTTCAAAATAAAAAACTGAAATAAAAAATAATATGAAATAACATCAGATCCTACCAC	1920
1921	AAAAAATAAGTAAACATAATAAGAGCTACATCATTAAGTAAATAATTTGAAAGACTGA	1980
1981	AAATATTTTTCAGAAATATCAGCTCGGATTTCTTCTTGAATAATTTAAAAAATAAATCA	2040
2041	GTTTTAAAGCTCAGAGTTATAAAAAATAAAAGTGTATAAAAAATAATAAAATTTTAA	2100
2101	TTGTCCTTTTGGAAACATCTCGAGTCCGCGGAGTCTTTTCTTATAACCCCTATAATAGT	2160
2161	GTCTCTTTATTTTATTTTGAATTTTCTTCTTGAATCGTTTGTAAAAATAATTTCAATTTT	2220
2221	TAAAAATTTGATATCTTTTGTAAATCTATTTATTTTATTTTCCCTAAACATGGATTTGT	2280
2281	GAAGCCTCTGGATAGGATTTAATTTTCTTCTTGTGATATCTTACTAAATATTTGAAATTCGAA	2340
2341	AATTTTAAAAAATTTAATTTTCTTCTTGTGATATTTGTGCAATCAAGTGCCTTTCAAACTC	2400
2401	TATTTCTGAGTTGTTATGTTTCTTTTCTTTTCTTTTTCGAAAAATCACATCCACCGGCTGCC	2460
2461	AAAAATATTTCCGTCACCACTCTTTTCTTTTCTTTTATTTCTTTATGGGACCCCATGCCATTTTC	2520
2521	GGCATATCTTTTCTTCAATCACCTTCTCACATACCGCATGCACTTGTCTACTTGGCCA	2580
2581	CACACGTCGTGTAATTTATAGCGAAAAATTTTTCAGCTTAAATGGCAATAATAGGCTA	2640
2641	AATCTCGGGCATTATCCCGATCCCTTTTCTTTTACAAAAGTCAATTAATCCATTTCTTAAT	2700
2701	TCCGTTTCAATFAAATCCCGGAGGCGAGTGTGTACATACATACGCTACACACGCTTCGAAGT	2760
2761	TAAGTAAGGTTCAAGTTGTGTGTGCGGTTAAGTAAGCGATTAAATATATATGCTGTAACG	2820
2821	CTAAATCGCAAGATTAGCGCTTAAACGAAAAAAGAAAAATATCTGCATACGCGCTTAG	2880
2881	CTAAGGCAAAAAGGACTTATGTTTAAATCCATTTCAAAAAATCAACAAGGATTTGATCC	2940
2941	AGACTAACACGCGACGACACTTACGTACCTGACCTACCTGGTTGGAAGGAGTCCGGAGAG	3000
3001	GATGCTGGCGATGGCGGGGGGATTTGGTTGGTTGAAAAATTTCCGCTTTGTGTGTGTGTG	3060
3061	TCTGAGAGCCAGAGTAGATTTAGAAGAGCCCTACTTTTACCTGTGTGTGTGCGAGTGCCTG	3120
3121	TCACACACACATAAATGTCGACTTTTCCATCTAATCTCGTGTGTGTGAGGCTGTGCGTG	3180
3181	TTTAGATGCGTAAATGAGGCTGCATGTGTATTTCCACATCCTTTTGTTCGCTTGT	3240
3241	TGTTGCTTAGAACCAATCACCCATCTAACCCACTCTACTTCTCTTGTCTGTCTGTGTGTTT	3300
3301	TCT	3360
3361	TAAATATCTGAGGTTTCTTGGTTTGGAGCATCAGTTGCATATCAGCATCGAAACGTTGGA	3420
3421	AACATAAACACAGCTCCCATCTCTTCTTTTACTTTTCGATTTTCGAAACACAAAAATTTTCG	3480
3481	AAGAAGAGATATATAGTGTGTAACAGCAAGATTTTCCAAAAATCGAGTTTAAAAA	3540
3541	ATGGTNGAGANAGAAANATAAGNGGATGGATCGGGGGGAAATATAACTACCAAGAGTT	3600
3601	TTTATTAAGGGGAGAAACAATTTTCTTTTATTTTCTTCAATCAATTTTTCACAAA	3660
3661	AAATTAACAAAAAGTGGCAACGAAAAATGTTTAAACAAAACCGCGAGAAATTTGGATTC	3720
3721	CACTTTCCCTTTTGGATTTCTTTGTTAAGGTACCTATCGCCACCACCACCATTTGGA	3780

3781 TTACAAGTGTAAAAAATACAACAACAAAAATCGCCACAAAAATATCCTCCTCAAAA 3840
3841 AAAGGAGGAAAACGGATTTACCTCCCTCTCCCTCTCTCTGTGGAAAAAATAAAATCACAA 3900
3901 AAAAAATTTTTTTTTTGTTCFAAGAGAGTCCCTTCTGTGTTGTTGCTGTCCTTCCCT 3960
3961 GTCAAAGCCAGTAACAATAACCAACCAACAATTTGTCCCTTTGGACTCCCCTTTCCC 4020
4021 CGCCATGGATTTTCGCTGTAACCGGTAATTTTCGGCGCTTGGCGCTGTAAACCACGGAAA 4080
4081 AAATTAATTTTTTTCTTTTGTATTTTTTTGCGTGGCCTTTTTTTTGTGTTTGTGTTG 4140
4141 CCTCTAAATTTTTATAAATTTTATTTTATTAATTTTGTAAATTTTTTCTTATTTTTTTTT 4200
4201 TTTAGATCTCCTCCCCCTTTTFTTGGGAAGCTGGAACATAGACAAGGCATAGAGTGAA 4260

MetGlnAsnTrpAsp 5
4261 AAATCTTTCAAAAAGAAAAGGATTTTCGAATFCCCAAAAAATCATGCAAAATTTGGGAC 4320
4684-----G--C-----G
-----Glu

ThrAlaThrAsnAsnAsnSerAsnAsnAsnLysSerProAlaLeuThrMetGlnGlnThr 25
4321 ACTGCCACCAATAACAACAGCAACAACAACAATCACCAGCCCTCACAATGCAACAAAACA 4380
--GA-AG--
---ThrAla

AlaAlaAsnSerAsnAsnPheLeuGluHisAsnThrTrpTyrAsnGlnMetPheAlaAla 45
4381 GCCGCAATAGCAACAATTTCTGGAGCACACACATGGTACAATCAAATGTTGCGGCC 4440
.....-CG-C---C-A-GA-C-----G-C-----CAGC-----C--G--A
.....ThrThr---TyrGluGln-----Ala-----Ser-----

AsnIleLysGlnGluProGlyThrIleAsnProHisHisGlnProProGlnGlnHisSer 65
4441 AACATCAACAGGAAACCCGCACATAAACCCACATCACCAACCCCAACAGCAGCAGC 4500
--T-----G--A--T.....-AT--T-TCG--G-G
-----HisHisLeuAspGly

SerMetMetAlaSerGlnProGlnHisSerProLeuThrHisSerAlaAsnTyrLeuGlu 85
4501 AGCATGATGGCCTCCAGCCTCAACATTTCTCCCTGACCCACAGTGCCCAATTTATTTGGAG 4560
-AT-GCG-----AG-AGT--G-GC--A--G--A-TC---.TCGA-----C-CC---A
AsnSerVal-----Ser---ArgGln-----IlePro...SerThr---His-----

AsnTyrLeuLysGlnHisAlaHisAsnGlyGlyGlyAlaHisHisLeuGlnPheSerAsp 105
4561 AACTATCTCAAGCAACATGCCACAATGGCGGTGGTCCCATCATTTGCAAGTCTCGGAT 4620
C-G-TC-----G--GCAG--GC-G.....T--GCA---ACC-AT----
GlnPhe-----GlnGlnGlnGln.....LeuGlnGlnGlnProMet---

AsnSerGlyAlaMetThrProSerProAsnThrAsnValGlyArgValLysAspPheGly 125
4621 AACAGCGGTGCCATGACTCCCTCACCAATACCAATGTTGGGAGGTCAGGATTTTGGT 4680
-C-CTGT-C-----C--A-----GCCAA--C-A-CAA-ACAG-CT-C-GCA-...
ThrLeuCys-----SerGln---AspGlnAsnSerLeuGlnHis...

PheGluSerAsnThrSerSerAlaLeuAsnProSerLeuGlnGlnHisGlnLeuTyrGln 145
4681 TTTGAGTCGAACACATCGTCTGCCCTGAATCCAGTCTGCAACAACATCAGCTGTATCAG 4740
-AG--TGCT---.....T--C-G-AACAGT---TG--G--A---A---C---
TyrAspAlaAsn.....GlnGlnGln---Leu---Gln---Gln-----

GlnHisPheGlnGlnAlaGlnGlnAlaAlaAlaGlnAsnGlnValSerSerHisLeuPro 165
4741 CAGCATTTCCAACAGGCCAACAGCAGCGGCCAAAATCAGGTTAGCTCTCACTTGCCC 4800
-----GGCA-----G---.....C---TCACCATCAC--TC--ATG
-----Ala-----HisHisHisHisHis-----Met

LeuGlyPheAsnProLeuThrProProGlyLeuProAsnAlaValLeuProAlaMetAsn 185
4801 CTGGGTTTAAATCCCTGACGCCACCTGGTCTGCCCAATGCCGTTGTCGCCCATGAAT 4860
GGT--A--C-----G.....C-----C-G
Gly-----Pro---Gln

HisTyrTyrSerGlnGlnGlnGlnGlnGlnGlnArgGlnLeuGlnGlnThrProSer 205
4861 CACTACTCAGCCAACAGCAGCAGCAACAACAACAGCTCAGCTACAACAACCCCTCG 4920
----T---TG--GGCA-T-T--GA-CCAGT-CG--G...-CCACG-CCAC-T-TG---C
---Phe---GlyGlyAsnLeuArgProSerPro----ProThrProThrSerAla---

4921 ProSerAlaCysLeuSerAspAlaHisGluLysSerSerAlaLeuThrProArgHisThr 225
 CCCTCAGCCTGCCTCAGTGATGCCCATGAAAAATCATCTGCCCTCACACCACGCCATACC 4980
 A-AATT--GCC-G-TGCAGT----AC--GC-GCAGCGAGAAGT-GCAGG--...-TA--A
 ThrIle---ProValAlaVal---ThrGluSer---GluLysLeuGlnAla...Leu---

4981 ProProMetAspIleThrProProLysSerProLysThrThrValGlnAlaMetAspHis 245
 CCTCCCATGGACATAACCCACCCAAATCGCCCAAGACCCTGTACAAGCCATGGACCAC 5040
 --A-----TG-C--A--G--T--G-----GGCC-AGT-GAGT--GT-G.....A-T
 -----Val-----AlaLysSerSerGlnSer.....Asn

5041 GlnGlnHisProGluAspGlnAspLeuIleSerAsnSerSerGluAspLeuLysTyrIle 265
 CAACAACACCCAGAAGACCAGGATTTGATTTCCAATTCAGTGAAGATCTCAATACATT 5100
 ATTG-G-CGGAGA-G--G--C---CA---G--G--C-----C--G--CA-G--G-----G
 IleGluProGluLysGluHis---GlnMet-----Met-----Met

5101 AlaGluSerGluAspAspGlu...SerIleArgMetProIleTyrAsnSerHisGlyLys 284
 GCCGAATCCGAGGATGATGAG...TCCATACGCATGCCCATCTACAATTCGCATGGCAAA 5157
 -----G-----C-----TACCAA---C--G-----C-----G
 -----AspThrAsn-----

5158 MetLysAsnHisLysCysLysSerCysGlyMetValGlyIleThrLysMetGlyPheTrp 304
 ATGAAGAATCACAATGCAAAATCCTGTGGCATGGTGGGCATCACCAAGATGGGTTCCTGG 5217
 -----CT-----G-----GA-----C---G-----C-----G--AC-----
 -----Tyr-----Thr-----Val---Ala-----ValAsp-----

5218 GluHisAlaArgThrHisMetLysProGluLysIleLeuGlnCysProLysCysProPhe 324
 GAACATGCCAGGACCCACATGAAACCCGAAAAGATCCTGCAATGTCCCAAGTGTCCCTTT 5277
 -CG--CA--C-C-----A--C-----G--C--G-----C--G--C
 Ala---Thr-----Asn-----

5278 ValThrGluLeuLysHisHisLeuGluTyrHisIleArgLysHisLysAsnLeuLysPro 344
 GTCACCGAATCAAAATCATCTGGAGTATCACAATACGCAACACAAGAATCTCAAGCCA 5337
 -----GT-----G--C--CT-----C--T--C--G--G-----C-AA-----C
 -----Phe-----Gln-----

5338 PheGlnCysAspLysCysSerTyrSerCysValAsnLysSerMetLeuAsnSerHisArg 364
 TTCAGTGTGACAAAATGCTCCTACAGCTGTGTCAACAAGTCCATGTTGAATTCGCATCGC 5397
 -----C-----AG-----CG-----A-----C-A--C-----C---
 -----Thr-----

5398 LysSerHisSerSerValTyrGlnTyrLeuSerAlaAspCysAspTyrAlaSerLysTyr 384
 AAATCCCCTCCTCGGTGTATCAGTACCTCTCTGCCGACTGTGACTATGCCTCCAAGTAC 5457
 --G--G--AGT--T-----GT--G--G--T-----T--C--A-----T
 -----ArgCys-----Thr-----

5458 CysHisSerPheLysLeuHisLeuArgLysTyrGluHisLysProGlyMetValLeuAsp 404
 TGCCACTCGTTCAAATGCAATTTGCGTAAATATGAACACAAACCCGGCATGGTATTGGAT 5517
 -AGC-----G-----C-----G--GT-----G-----T-----C
 -----Gly-----

5518 GluGluGlyIleProAsnProSerValValIleAspValTyrGlyThrArgArgGlyPro 424
 GAGGAGGTATTTCCCAATCCCTCGGTGGTCATGTATGTCTATGGCACCCGGCAGGACCC 5577
 -----T--C--CC--G-----T-----C-----T--C-----G--T--T--T--G
 ---Asp---Thr-----Leu-----

5578 LysAsnLysSer..... 428
 AAGAACAAGTCG..... 5589
 ---G---AATGGTGGACCGATTGCCAGTGGAGGAAGTGGCAGCGGCAGCCGGAAGTCA
 ---Ser---AsnGlyGlyProIleAlaSerGlyGlySerGlySerArgLysSer

5590AlaAlaAsnAlaAlaLeuLysLysAlaCysSerAspLeuLysIleProProThr 446
GCAGCCAACTGCTGCTTTGAAAAAGCCCTGTAGTGATTTGAAAAATCCTCCCACT 5643
 AATGTT-----TGTC-----C-GCA-C--C--CAA-C-CAGCCAGCTC-GCCAGTCG-----A
 AsnVal-----Val---ProGlnGlnGlnGlnSerGlnProArgGlnProValAla---

```

5644 SerGlnLeuSerAlaAlaLeuGlnGlyPheProLeuGlnGlnGlnGlnGlnProGlnGln 466
TCTCAGTTGTTCGGCTGCCCTGCAAGGTTTCCTCTCCAGCAGCAACAGCAGCCGCAGCAG 5703
-----C--AGT--C-----A-----GGTT--AGGCAACTCCG-T-CT-C-
-----Val---GlyAsnSerAlaProPro

.....ProAlaSerProAlaLysSerSerSerSerVal... 477
5704 .....CCTGCTTCTCCGGCCAAAAGTTCCAGCTCTGTG... 5736
GCGGCATCTCCAGTGCTCCCGCTG--C--C-----T-----G---GTGGC-AG----GAA
AlaAlaSerProValLeuProLeu-----ValAla-----Glu

.....AlaSerGluLeuProAlaLeuThr..... 485
5737 .....GCCTCGGAATTGCCTGCCTTGACC..... 5760
CAGACGCCCAGCTTGCCAGTCCA--AATCTTC-----C-TC--G--AGCCTTCTGCAG
GlnThrProSerLeuProSerPro--AsnLeu-----Pro---AlaSerLeuLeuGln

.....LeuAsnMetSerLeuGlnGlnAsnLeuAla 495
5761 .....CTTAACATGTCATTGCAACAAAACCTGGCC 5790
CAGAACCACAACATGGCCTTCTTCCCCTACTGG--C-CAATC-C--GATGCTGGCC--
GlnAsnArgAsnMetAlaPhePheProTyrTrp--LeuAsn-----MetLeuAla---

GlnGlnGlnGln.....GlnGln 501
5791 CAACAGCAGCAG.....CAACAA 5808
-----A---GCCGCTGTCTTGGCCCAATGTGCGCCAAGAATGCGAGAGCAACTG--G---
-----AlaAlaValleuAlaGlnLeuSerProArgMetArgGluGlnLeu-----

GlnGlnSerProGlyAlaGlnSerHisSerSerGlnGlnGlnIleAsnAsnLeuLeuPro 521
5809 CAGCAGTCACCTGGAGCCCAAGCCACAGCAGCCAACAACAATCAATAATCTCCTACCT 5868
---A-CCAG-AGCAGAG-G-C-AT--GGAGGAGG-G--GG-CGATG-GT-CGAG-GTAAG
---AsnGlnGlnGlnSerAspAsnGlnGluGluGlu--AspAspGluTyrGluArgLys

ProLeuAsnSerLeuLeuGln 528
5869 CCTTTGGCCTCTTTGCTGCAG 5889
T-AG---A---GCCA--G-T 6313
SerValAsp---AlaMetAsp

```

Figure 4.7. Sequence of the 5.9 kb *Pst* I fragment of the *Musca domestica hunchback* gene, and its homology with *Drosophila melanogaster*. The *D. melanogaster* sequence is numbered as in Tautz et al. (1987). The *D. melanogaster* sequence is in italics, with dashes representing identities and dots representing gaps. The three candidate high affinity bicoid binding sites are underlined.

Alignment of the hunchback protein sequences of *D. melanogaster*,
D. pseudoobscura and *M. domestica*

```

D. melanogaster MQNWET-TA-----TTNYEQHNAWYNSMFAANIKQEP-----
D. virilis      MPNWETSTA-----APSYEQHNAWYSSMFAANIKQEPLSHHHHHHG
M. domestica    MQNWDTATINNNSNNKSPALTMQQTAANSNNFLEHNTWYNQMFANIKQEPGTINPHHQ
* * * * *
. . . * * * * *

D. melanogaster ---GHLDGNSVASSPRQSP---IPSTNHLEQFLKQQQQQLQQQ--PMDTLCA--MTP
D. virilis      QQHHDNHSNSNSGASSPRQSPSPPIPPSTQLEQYLKQQQQQQQQQQPMDTLCAAAMTP
M. domestica    PQQH---SSMMAQQPHSPL---THSANYLENYLKQHAHNGGGAHHLQFSDNSGAMTP
. . * * * * * . . * * * * * . . * * *

D. melanogaster SPSQNDQNSLQHYD-----ANLQQQLLQQQQYQHFQAAQQQHHHH---HL-MGGFNP
D. virilis      SPSNNDQNSLQHFD-----ATLQQQLLQQQQYQHFQAAQHQQQHH---HLALGGFNP
M. domestica    SP-NTNVGRVKDFGFESNTSSALNPSLQHQHLYQQHFQQAQQAQAAQNVSSHLPLG-FNP
* * . . . . . * * * * * * * * * * * * * * * * * * * * * *

D. melanogaster LTPPGLPNP---MQHFYGGNLRPSPQPTPTSASTIAPVAVATGSSEKQLQALTP---PM
D. virilis      LTPPGLPNP---MQHYGGNMRPSPQPTPTAAPTAVAAAQITG--DKLQALTP---PM
M. domestica    LTPPGLPNAVLPAMNHYSQQQQQQQRLQQTPTS--PSACLSAHEKSSALTPRHTPPM
***** . * * * . . * . . . * . . * * * * *

D. melanogaster DVTPPKSPAKSQS--NIEPEKEHDQMSNSEDMDKYMASEDDDTNIRMPIYNSHGKMKN
D. virilis      DVTPPKSPAKSQS--SAEPEKEHDLMSNSEDMDKYMASEDDDSNIRMPIYNSHGKMKN
M. domestica    DVTPPKSPKTTVQAMDHQHPEDQLISNSSEDLKYIAESEDES-IRMPIYNSHGKMKN
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

D. melanogaster YKCKTCGVVAITKVDFAWTRTHMKPKDKILQCPKCFVTEFKHHLEYHIRKHKNKQPFQC
D. virilis      YKCKTCGVVAITKVDFAWTRTHMKPKDKILQCAKCFVTEFKHHLEYHIRKHKNKQPFQC
M. domestica    HKCKSCGMVGIKMGFWEHARTHMKPEKILQCPKCFVTELKHHLEYHIRKHKNLKPFQC
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

D. melanogaster DKCSYTCVNKSMNLNHRKSHSSVYQYRCADCDYATKYCHSFKLHLRKYGHKPGMVLDEDE
D. virilis      DKCSYTCVNKSMNLNHRKSHSSVYQYRCADCDYATKYCHSFKLHLRKYGHKPGMVLDEDE
M. domestica    DKCSYSCVNKSMNLNHRKSHSSVYQYLSADCDYASKYCHSFKLHLRKYEHKPGMVLDEEG
***** * * * * * * * * * * * * * * * * * * * * * * * * * * *

D. melanogaster TPNPVLIDVYGTRRGPKSKNGGPIASGGSGSRSKNVAAPQQQSQPAQPVATSQ
D. virilis      TPNPVLIDVYGTRRGPKSKSFG-SGSSCSSTSKRSNASAAAAQQQQ---QPVATSQ
M. domestica    IPNPVVIDVYGTRRGPKNKS-----AANAALKKACSDLKIPPTSQ
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

D. melanogaster SAALQGFPL-----VQNSAPPAASVPLPL-PASPAKSVASVEQTPSLPSPANLLPPL
D. virilis      SAALQGFPPMPAAAAGTAAGAAGTAAPAAVAVPSPPAKSVASVEQAPSLPSA--LLPPL
M. domestica    SAALQGFPLQQQQ-----PQQPASPAKSSSSV-----ASELPAL
***** . * * * * * * * * * * * * * * * * * * * * * * * * * * *

D. melanogaster ASLLQQNRNMAFFPYWNLNLQMLAAQQQAQVLAQLSPRMREQLQQQNQQQ-----
D. virilis      ASLLQQNRNMAFFPYWNLNLQVLAQQQAQVLAQLSPRMADQLQQQQQQHQQQQQQQQ
M. domestica    T-----LNMSLQQNLAQQQQ-----QQQQSPG-----
. . * * * * * * * * * * * * * * * * * * * * * * * * * * *

D. melanogaster -----SDNQEEBQDDEYERKSVDSAMDLSQGTVPKEDQEQQQPQQPL---
D. virilis      QQQQQQLPAHSENEDEEEEEHEDDFERKSVDSAMDLSQGTVPKEEFQQQQQQQLPHS
M. domestica    -----AQSHSSQQQINNLLPPLASLLQ (end of M. domestica sequence)
. . . . . * * * * *

D. melanogaster ---AMNLKVEEETPLMSSSNASRRKGRVCLKDITLLQLRSEAMTSPEQ-LKVPSTPMPPTA
D. virilis      NHMAINLKLKDEDTPLISSSSASRRKGRVCLKDITLLQLKSAAMSSPEQQLKLPASVLPPTA
M. domestica

```

```

D. melanogaster SSPIAGR---KPMPEEHCSGTSSADESMETAHVPQANTISASSTASSSGNSSNASSNSNGN
D. virilis      SSPIAGSSANKQLADDDPCSGASSADESMETGRVPQVNISSASSTASSSGNSSNASSSTSNP
M. domestica

D. melanogaster SSSNS-SSNGTTSVAVAAPPSGTPAAAGAIYECKYCDIFFKDAVLYTIHMGYHSCDDVFKC
D. virilis      TAAATVATSGTVSSSSSSSTTTSSAPAIYECKYCDIYFKDAVLYTIHMGYHSCDDVFKC
M. domestica

D. melanogaster NMCGEKCDGPGVGLFVHMARNAHS
D. virilis      NMCGEKCDGPGVGLFVHMARNAHS
M. domestica

```

Figure 4.8.

Multiple amino acid alignment of *hunchback* from *Drosophila melanogaster* (Tautz et al., 1987), *D. virilis* (Treier et al., 1989) and *M. domestica* (partial sequence). The alignment was made using the CLUSTAL-V package, as described in Chapter 2. The zinc finger domains are shown in red, Box A in green and Box B in blue. Dashes indicate gaps introduced in the alignment, stars indicate amino acids conserved between all three species, and dots semi-conserved amino acids.

	Library clone sequence	GGCATCACCAAGATGGGTTTCTGG	5217
	new probe sequence	-C-----G---AC-----	
	original probe sequence	-C-----G--A---C-----	
5218	GAACATGCCAGGACCCACATGAAACCCGAAAAGATCCTGCAATGTCCCAAGTGTCCCTTT		5277

5278	GTCACCGAACTCAAACATCATCTGGAGTATCACATACGCAAACACAAGAATCTCAAGCCA		5337
	-----G-----		
5338	TTCCAGTGTGACAAAATGCTCCTACAGCTGTGTCAACAAGTCCATGTTGAATTCGCATCGC		5397
	-----G--T-----		
	-----G--T-----		
5498	AAATCCCACTCCTCGGTGTATCAGTACCTCTCTGCCGACTGTGACTATGCCTCCAAGTAC		5457
	-----T-G--G-----A-----		
	-----T-G--G--G-----A-----		
5458	TGCCACTCGTTCAAACATGCATTTGCGTAAATATGAACACAAACCCGGCATGGTATTGGAT		5517
	-----C-----G-----A---GC---C		
	-----C-----G--T--T-----G-----		
5518	GAGGAGGG		
	-----T--		

Figure 4.9. The sequence of the fragment of the *M. domestica hb* gene used as a probe to screen *M. domestica* genomic libraries. When the probe fragment was resequenced (new probe sequence), a number of differences were noted with the sequence as originally reported by Sommer (1992, original probe sequence). When these sequences are compared with the sequence obtained from the library, a number of further differences are noted. The library clone *hb* sequence is numbered as in Figure 4.7.

THE N-TERMINAL REGION

The multiple alignment of hb amino acid sequences reveals a few conserved features of the hb protein N-terminal to the first zinc finger domain. Notably, the first four amino acids encoded by hb are conserved between *M. domestica* and *D. melanogaster*. This suggests that the authentic start site of translation has been identified (position 4306 in Figure 4.7).

Two short regions of similarity between *hb* and the zinc finger containing gap gene *Krüppel* (also sharing similar locations within the proteins) have been designated Box A and Box B (Tautz *et al.*, 1987; Figure 4.6a). Box A also shows homology to the endonuclease region of the HIV-1 retrovirus *pol* gene (Tautz *et al.*, 1987; see also Figure 4.6c). Box A is highly conserved in *M. domestica* (Figure 4.6c).

In contrast, Box B (shown in green type in Figure 4.8) is only loosely conserved in *D. virilis* (Treier *et al.*, 1989), with the presence of a number of small insertions interrupting the amino acid sequence. It has completely diverged in *M. domestica* (Figure 4.8). No function has yet been ascribed to either box. However the conservation of box A in *M. domestica* would tend to suggest that it does have a significant function, whilst the divergence of box B might suggest a lack of function and a fortuitous homology to *Krüppel*.

Upstream sequence

P2 PROMOTER

As described in the introduction to this chapter, the bcd dependent activation of *hb* expression in *D. melanogaster* is mediated by the (proximal) P2 promoter. In *D. virilis*, the sequence immediately upstream of the P2 transcript is fairly well conserved over a region of 260 bp (Treier *et al.*, 1989), which includes the three high affinity bcd binding sites defined in *D. melanogaster* (Driever and Nüsslein-Volhard, 1989). Additional short stretches of high homology between these two *Drosophila* sequences

have also been seen further upstream of the P2 promoter (Treier *et al.*, 1987), while the sequences flanking these homologous regions seem to have diverged completely.

Sequence comparisons

Initially, the *M. domestica hb* sequence was compared with the *D. melanogaster* sequence by using dot plot analysis, as described in the methods section of this chapter. This did not reveal any significant regions of homology between the upstream regions of the two species (Figure 4.10). Similar comparisons of the regulatory regions of developmental genes between *Drosophila* species have typically revealed islands of high conservation of 40 bp or more (for example, Hooper *et al.*, 1992; Kassis *et al.*, 1986, 1989; Langeland and Carroll, 1993; Wilde and Akam, 1987; Treier *et al.*, 1989; Lukowitz *et al.*, 1994) despite a divergence time of around 60 MYA (Beverley and Wilson, 1984), thought from molecular clock calculations to be sufficient to allow unconstrained nucleotides to have diverged completely (Perler *et al.*, 1980).

Matches to *D. melanogaster* consensus bcd binding sequence

As it was not possible to identify the authentic bcd binding sites within the *M. domestica hb* gene on the basis of alignment and dot-plot comparison to the *D. melanogaster* regulatory region, attempts were made to identify possible candidate sites by searching for matches to the *Drosophila* consensus bcd binding sequence.

No exact matches to the *Drosophila* consensus bcd binding site are found within the sequenced region of the *M. domestica hb* gene. This raises the possibility that the binding sequence has changed.

The search was repeated allowing a single deviation from the *Drosophila* bcd binding sequence. This identifies two sites, at nucleotide position 286 (TCAAATCCC) and position 3151 (TCTAATCT) (Table 4.1). Extending the search to include sites with

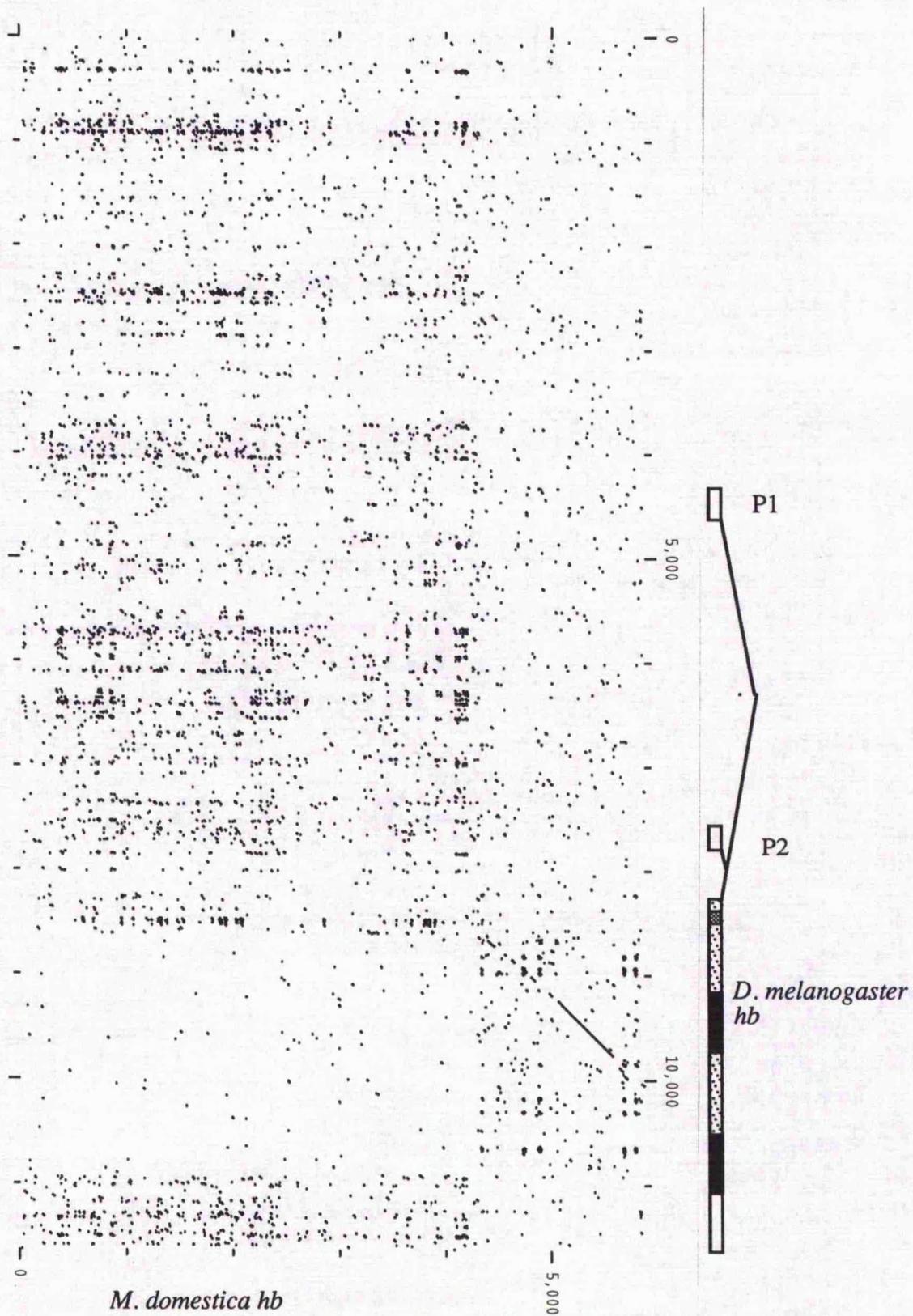


Figure 4.10 Dot matrix homology comparison of the sequences of the *Drosophila melanogaster hunchback* gene (sequence from Tautz et al., 1987; and Lukowitz et al., 1994) and the 5.9 kb fragment of the *Musca domestica hunchback* gene. The dot matrix comparison was generated using the COMPARE and DOTPLOT programs of the UWGCG DNA sequence analysis package. Fourteen matches or more over a 21 bp window (66%) are required to make a dot. The organization of the *D. melanogaster hb* gene is indicated. P1 is the promoter involved in maternal and later zygotic expression, the P2 promoter drives *bcd* dependent expression of *hb*. Non-coding regions of the *hb* transcript are represented by white boxes, the zinc fingers are black, Box A is shaded and all other coding regions are stippled.

two mismatches to the *Drosophila* consensus sequence identifies a further 18 sites within the *M. domestica hb* upstream region (Table 4.1).

TABLE 4.1. Close matches to the *Drosophila melanogaster* consensus bcd binding sequence within the *M. domestica hb* upstream sequence.

Position of first base of motif ^{***}	Sequence ^{**}	number of mismatches	contains a TAAT core?	binding affinity
286	T C AAATCCC	1	-	
346	TCTAAT G T C	2	+	
629	TCTAAT A G C	2	+	
980	T T AAATCCC	2	-	
2143	T A TAA C CCC	2	-	
2536	T C AAAT C A C	2	-	
2683	T T TAAT C CA	2	+	candidate high + A1
2709	T T AAATCCC	2	-	
2797	*T T TAAT C GC AAATTA G CG	2	+	candidate high + A8
2904	T T TAAT C CA	2	+	candidate high + A4
2929	*T C AAAT C CT	2	-	
3019	* A CCAATCCC	2	-	
3075	*TCTAAT C TA	2	+	medium?
3151	TCTAAT C TC	1	+	medium?
3178	*TCTAA A CAC	2	-	
3263	TCTA A CCAC	2	-	
3578	*T A TAT T CCC	2	-	
3777	*T G TAAT C CA	2	+	
4207	TCT C CTCCC	2	-	
4287	T C GAAT T CC	2	-	

All sequences are oriented 5'-3' unless indicated by *. **Mismatches to the consensus sequence are shown in bold type. ***The positions refer to Figure 4.7.

The *D. melanogaster* consensus bcd binding sequence contains the motif "TAAT" at its centre. The TAAT core is a critical feature of many homeodomain binding sites (Treisman *et al.*, 1992), including that of bcd (Driever and Nüsslein-Volhard, 1989), and structural studies have shown that amino acids highly conserved among all homeodomain proteins contact nucleotides within this core (Kissinger *et al.*, 1990; Otting *et al.*, 1990). The TAAT core appears to be essential for DNA binding activity, with the nucleotides flanking the core directing binding specificity (Catron *et al.*, 1993; Ekker *et al.*, 1991,1992; Wilson *et al.*, 1993). Bcd binds to sites lacking an intact TAAT core with low affinity (Driever *et al.*, 1989a). Therefore, even if the specificity of bcd binding has changed, it is unlikely that the TAAT core sequence would not be present in a high affinity bcd binding site.

The elimination of all sequences lacking a TAAT core from the list of candidates leaves 8 possible bcd binding sites within the *M. domestica hb* upstream sequence. Three of these sites contain a change (TT TAAT rather than the *Drosophila* consensus TC TAAT) 5' to the TAAT core, whereas the rest only contain changes 3' to the core sequence (Table 4.1).

The affinity of *D. melanogaster* bcd for a variety of mutant binding sites has been assayed through co-transfection experiments (Hanes and Brent, 1991). The C immediately 3' to the TAAT core was found to be critical - changing this to A or T abolished activation, suggesting sites 346 and 629 (Table 4.1) are unlikely to represent high affinity sites.

Remarkably, of the remaining six possible bcd binding sites, five are found in a fairly tight cluster, covering about 500 bp, and falling approximately 1-1.5 kb upstream of the start site of translation (Fig 4.7). They fall into two classes - TTTAATCC and TCTAATCT. Hanes and Brent (1991) and Driever and Nüsslein-Volhard (1989) have both shown that *D. melanogaster* bcd is capable of binding to variant sites which correspond to the *M. domestica* bcd binding site sequence. The TCTAATCT class has the lower affinity of the two, whilst both are of lower affinity for *D. melanogaster* bcd than

the *D. melanogaster* consensus binding sequence (Hanes and Brent, 1991). Therefore, the three TTTAATCC class sites appear to be the more likely candidates to be high affinity bcd binding sites within the sequenced region of the *M. domestica hb* gene.

In *D. melanogaster*, all three of the high affinity bcd-binding sites are oriented in the same direction. Of the three candidate bcd binding sites identified within the *M. domestica hb* upstream sequence, two are in the same orientation as the sites within the *D. melanogaster hb* promoter, whilst the central site is in a reversed orientation. The altered orientation of the central site need not necessarily affect its ability to function. Within the *Kr* promoter, the bcd binding sites are found in a mixture of both orientations (Hoch *et al.*, 1991). Similarly, reporter gene constructs containing two bcd binding sites in each orientation drive expression over a domain comparable to that of the native *hb* gene (Driever *et al.*, 1989a).

Candidate site spacing

One independent line of evidence in favour of the authenticity of the candidate bcd binding sites comes from their spacing. In *D. melanogaster*, the bcd binding sites are spaced 107 and 108 base pairs apart respectively (Driever and Nüsslein-Volhard, 1989). This spacing is more or less conserved in *D. virilis* (94 and 95 bps; Treier *et al.*, 1989) despite the apparent occurrence of a number of insertion and deletion events between the sites. It has been suggested that the spacing may be important for co-operative binding (Driever and Nüsslein-Volhard, 1989). In *M. domestica*, the candidate bcd binding sites are found 115 and 106 bp apart, an approximately conserved spacing.

hb binding sites

A synergistic interaction between bcd and hb has been demonstrated (Small *et al.*, 1992; Simpson-Brose *et al.*, 1994) and as discussed in Chapter 1, it has recently been suggested (though not conclusively demonstrated) that this synergistic interaction may be required for the expression of *hb* (Simpson-Brose *et al.*, 1994). Hb binding sites have

been identified within the *hb* promoter, adjacent to known bcd binding sites (Treisman and Desplan, 1989). Therefore, it is interesting to ask if there are any matches to the hb consensus binding sequence within the *M. domestica hb* upstream sequence and, in particular, if they are located adjacent to the candidate bcd binding sites. The hb consensus binding site sequence has been variously defined as $C/G^A/CATAAAAA$ (Stanojevic *et al.*, 1989) or ACNCAAAAAANTA (Treisman and Desplan, 1989). There are no exact matches to either of these sequences within the sequenced region upstream of the *M. domestica hb* coding DNA. Searching for sites with a single mismatch to these sequences identifies 23 sites, mostly falling in the AT-rich sequences flanking the candidate bcd binding sites.

P1 PROMOTER

In *D. melanogaster*, the P1 promoter directs maternal and later zygotic expression of *hb*. The P1 transcript contains a short open reading frame, which may enable translational control (Treier *et al.*, 1989). The nucleotides around the start codon ATG and the first few amino acids are conserved in *D. virilis* (Treier *et al.*, 1989).

Functional dissection of the region upstream of the P1 promoter has implicated a number of regions in the regulation of maternal and later zygotic *hb* expression (Margolis *et al.*, 1994). A number of conserved regions have been identified upstream of the P1 promoter between *D. melanogaster* and *D. virilis* (Treier *et al.*, 1989, Lukowitz *et al.*, 1994). Genetic analysis suggests that the later zygotic expression of *hb* is regulated by *Krüppel*, *hb* itself (Hülkamp, 1991), and the terminal system genes (Tautz, 1988), and some of the conserved regions upstream of the P1 promoter contain hb, Kr and tll consensus binding sites (Lukowitz *et al.*, 1994).

Sequence comparisons

Dotplot comparisons at low stringency (66% identity in a window of 21 nucleotides) did not reveal any detectable homology between the *D. melanogaster* P1

promoter, open reading frame and associated upstream conserved regions, and the *M. domestica* sequence (Figure 4.10). In *D. melanogaster*, the P1 promoter is located 4.5 kb upstream of the start site of translation and, in *D. virilis*, it is 5.5 kb upstream. In *M. domestica*, only 4.3 kb of DNA has been sequenced upstream of the candidate bcd binding sites.

Maternal transcript length

Northern analysis was performed on total RNA extracted from adult female *M. domestica*, using the *M. domestica hb* zinc finger probe. This indicates that the maternally expressed *hb* transcript is 5.1 kb in size (Figure 4.11a), a large increase when compared to the 3.2 kb long maternally derived (P1) transcript found in *Drosophila melanogaster*.

DISCUSSION

Inter-species comparisons

Inter-specific comparisons have previously been used to identify regions important for the regulation of a number of genes. If it is assumed that the regulation of a gene is conserved between the species compared, then evolutionarily conserved sequences in the promoter region can help to locate elements required for that gene's correct expression. Ideally, such comparisons are made between species that have evolved independently long enough to allow nearly complete divergence of unconstrained sequences, but not for so long as to allow too much divergence of more slowly evolving sequences. In this way, a number of conserved elements postulated to be involved in regulation have been observed in the promoter and upstream regions of developmental genes such as *engrailed* (Kassis *et al.*, 1986, 1989), *hairy* (Langeland and Carroll, 1993), *Ultrabithorax* (Wilde and Akam, 1987), *Antennapedia* (Hooper *et al.*, 1992)

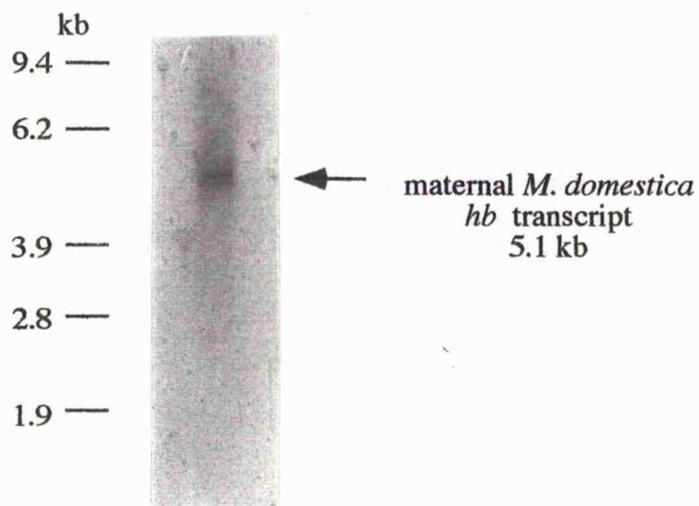


Figure 4.11.

A northern blot showing the maternal *hb* transcript in *M. domestica*. In *D. melanogaster*, the maternal transcript is 3.2 kb (Tautz et al., 1987), rather than the 5.1 kb seen here for *M. domestica*. The lane contains 10 μ g of total RNA extracted from adult female *M. domestica*. The *M. domestica* *hb* zinc finger fragment was used as the probe.

and *hb* (Treier *et al.*, 1989; Lukowitz *et al.*, 1994) in comparisons made between various *Drosophila* species.

A number of studies have demonstrated the validity of the assumption that conserved regions within the promoters represent regulatory regions. For example, a conserved region in the first intron of *engrailed*, not previously known to be important in its regulation (Kassis *et al.*, 1989) was subsequently found to be involved in control of *en* expression (Kassis, 1990). Similarly, regions 5' to the exons in *fushi tarazu* that have been shown in transformation experiments to be essential for the correct functioning of the gene have been shown to be conserved between a number of *Drosophila* species (Maier *et al.*, 1990, 1993). Interspecific comparisons can therefore be powerful ways of identifying putative regulatory elements. However, non-conserved regions may not necessarily be non-functional but may reflect important, divergent, species-specific functions. Or, they could have been subject to molecular co-evolution (see Chapter 1).

The *hb* P1 promoter

The P1 promoter drives maternal and later zygotic expression of *hb* in *D. melanogaster*. No homology to the unique first exon (see Figure 4.1) found within the transcript from the *D. melanogaster* P1 promoter was detected in *M. domestica*. Nor were any regions observed that correspond to the elements upstream of the P1 promoter that are conserved between *D. melanogaster* and *D. virilis*. It is entirely possible that the P1 promoter is not contained within the sequenced 5.9 kb fragment, but is located further upstream, given that the P1 promoter is located some 4.5 kb upstream of the start site of translation in *D. melanogaster*, and in *D. virilis* it is 5.5 kb upstream, but only 4.3 kb of upstream DNA is contained within the sequenced *M. domestica* fragment.

Given the degree of divergence of the P1 promoter-specific short open reading frame between *D. melanogaster* and *D. virilis* (Treier *et al.*, 1989), it is unlikely that this open reading frame could be identified on the basis of sequence homology. S1 nuclease

mapping experiments would allow the identification of the start site of transcription, allowing the approximate location of the P1 promoter to be determined.

The northern blot analysis indicates that the *M. domestica* maternally derived *hb* transcript is 5.1 kb long - a significant increase in comparison to the 3.2 kb of the *D. melanogaster* P1 promoter derived transcript (and 2.8 kb *D. melanogaster* P2 transcript). Sommer (1992) has amplified a fragment spanning both zinc finger domains of the *M. domestica* *hb* gene from genomic DNA by PCR, and subsequently obtained the sequence of this fragment. From this data, it appears that the size of the coding region from the beginning of zinc finger domain 1 to the end of zinc finger domain 2 (which corresponds to the C-terminal of the protein, Figure 4.6) is approximately conserved between *M. domestica* and *D. melanogaster*. The sequence data presented in this chapter (which overlaps with that of Sommer) indicates that the size of the region of the *hb* protein N-terminal to the first zinc finger domain (see Figure 4.7) is also approximately conserved between the two species - therefore the large increase in transcript length cannot be accounted for by an increase in the size of the *hb* coding sequence. It is possible that the increase may be accounted for by an increase in the size of the 3' or 5' untranslated regions, including the untranslated first exon (see Figure 4.1). Examination of cDNA library clones is one way in which this might be investigated.

The *hb* P2 promoter

The P2 promoter drives *bcd* dependent early zygotic expression of *hb* over an anterior domain of the *D. melanogaster* embryo (Tautz *et al.*, 1987). In this work, a 5.9 kb *Pst* I fragment of DNA containing to the 5' half of the coding region of the *M. domestica* *hb* gene and 4.3 kb of upstream DNA has been sequenced, and this sequence compared to those of the *D. melanogaster* and *D. virilis* *hb* genes, in an attempt to identify the *bcd* dependent regulatory elements of the *M. domestica* *hb* promoter.

IDENTIFICATION OF *M. DOMESTICA* BCD BINDING SITES

It was not possible to identify any significant regions of homology between the sequenced fragment of the *M. domestica hb* gene and the upstream regions of *D. melanogaster hb* through dot plot comparisons. This might at first glance seem surprising - given that comparisons of the regulatory region of a number of developmental genes between *Drosophila* species have revealed islands of highly homologous sequences despite a divergence time of around 60 MYA (Beverley and Wilson, 1984), thought from molecular clock calculations to be sufficient to allow unconstrained nucleotides to have diverged completely (Perler *et al.*, 1980). However, in this respect it is interesting to note that Malicki *et al* (1992) found that the a *cis*-regulatory element of the human HOX4B gene, whilst sharing functional similarities with the autoregulatory element of its *D. melanogaster* cognate *Dfd*, shares no extensive regions of primary sequence similarity. They conclude that this does not necessarily imply a lack of conserved functional elements, citing the example of the *Drosophila* and human *Adh* upstream control elements (Malicki *et al.*, 1992). These elements conserve a functionally related region that has little primary sequence similarity but still contains overlapping binding sites for the transcription factors AEF-1 and C/EBP (Falb and Maniatis, 1992). Similarly, most aspects of the expression of the *esterase-6* gene are shared between *D. melanogaster* and *D. pseudoobscura*, despite the promoter sequences being so different as to be unalignable (Oakeshott *et al.*, 1995).

There are a number of possible explanations for the absence of detectable homologous regions. One possibility is that the *M. domestica hb* sequence is not contiguous - that the upstream sequence is an unrelated stretch of DNA. As discussed above, as the restriction map of the λ clone insert has not been compared with the genomic map, this possibility cannot be ruled out. However, it would be a remarkable coincidence if the postulated site of discontinuity was located immediately upstream of the start site of translation. Therefore an alternate explanation would be required to account for the fact that the lack of homology begins immediately upstream of the start site of translation.

Another explanation for the degree of divergence seen in the upstream regions would be that they are not functionally homologous - if there was no bcd-dependent regulation of *hb* expression in *M. domestica* then no homology in the upstream DNA sequence would be expected. As mentioned in Chapter 1, given the homology of both embryology and early gene expression between *D. melanogaster* and *M. domestica*, it seems reasonable in the first instance to assume functional homology. However, as functional homology has not been proved, in the absence of conserved regions of upstream sequence, the possibility of very divergent regulation must be taken in to account. This theory is considered further in Chapter 7, where the evolution of the bcd-*hb* interaction is discussed.

However, in considering the apparent lack of homology upstream of the *M. domestica hb* gene, the applicability of molecular-clock assumptions must also be questioned. In the comparisons of the upstream regions of various developmental genes in various *Drosophila* species, the observed pattern is typically one of short islands of high homology in seas of completely diverged sequence. It is possible that this pattern occurs as a result of strong functional constraint on the conserved regions whilst the rest of the sequence is free to accumulate mutation. However, other internal genomic processes may also contribute to this conservation/divergence pattern (Dover and Tautz, 1986). Examination of the upstream sequence of the *hb* gene reveals cryptic simplicity (Treier *et al.*, 1989). Therefore slippage-generated mutations acting to scramble sequence might be more important than point mutations in generating divergent sequences - measured rates of slippage are at least two orders of magnitude faster than point substitutions (Tautz *et al.*, 1986; Levinson and Gutman, 1987). If there are a number of short functionally important regions embedded within sequence of high cryptic simplicity, slippage mechanisms might be expected to scramble the intervening sequences, initially producing larger regions of homology surrounding small functionally important sequences such as binding sites, and gradually whittling these down to the smallest necessary size. This could result in a pattern whereby in comparisons between pairs of *Drosophila* species, these islands are sufficiently large to be recognised by dot plots, but in comparisons between *Drosophila* and *Musca* only the

actual binding sites remain - too small to be detectable by this method. Similarly, this could account for the pattern of detectable homology between the *Est6* promoters of closely related *melanogaster*-group *Drosophila* species, but absence of homology in the more divergent *D. pseudoobscura* (despite conserved *Est-6* expression patterns), described by Oakeshott *et al.* (1995).

Given the lack of broad regions of homology between the *D. melanogaster* and *M. domestica hb* upstream DNA sequences (and assuming that the sequence is contiguous), it was not possible to identify the bcd binding sites on the basis of sequence alignments. Therefore, the DNA sequence was searched for possible sites by examining it for matches to the *D. melanogaster* bcd consensus binding sequence - looking for a small sequence of *a priori* interest is statistically more meaningful than searching for random matches of similar size (Dickinson, 1991). No exact matches were found. However, when up to two mismatches were allowed, 20 sites were identified.

It is possible to derive an estimate of the expected frequency of TCTAATCCC sequences within a length of random DNA, assuming that the probability of finding each base at a given location is equal and independent of the type of the neighbouring bases. In which case, the probability of finding an exact match (in a single orientation) is $(1/4)^9 = 3.81 \times 10^{-6}$, or approximately 4 sites in every 1000 kb. If two mismatches are allowed, the probability increases to $(1/4)^7 = 6.10 \times 10^{-5}$, or roughly 6 times in every 100 kb. If both orientations are considered, then the expected frequency doubles to 12 per 100 kb - ie. 0.48 matches might be expected in the ~4 kb of upstream DNA examined. This is significantly fewer than the twenty observed. Taking into account biased base composition would result in a slightly increased expected frequency, but such considerations would be unable to account for more than a small portion of the 50-fold difference between the observed and expected number of two base pair mismatches to the *D. melanogaster* consensus bcd binding sequence. Therefore, it must be concluded that the distribution of these sites is highly unlikely to be merely random, and consequently that it is very likely that there is some cause for their significant frequency - one such cause could be selection for the presence of sequences that will bind bcd protein.

From the twenty sites showing up to 2 mismatches to the *D. melanogaster* consensus bcd-binding sequence, three candidate high-affinity bcd binding sites have been identified within the *M. domestica hb* gene, by taking into account the relative influences of particular nucleotides on the affinity of the bcd homeodomain -DNA interaction (Hanes and Brent, 1991) and assuming that the specificity of the bcd protein is unchanged in *M. domestica*. The three candidate sites define a consensus of TTTAATCC, as opposed to TCTAATCC for *Drosophila melanogaster*, interestingly with each of the three sites showing the same T to C change.

Homogenisation of variant bcd-binding sites

The three candidate *M. domestica* bcd binding sites define a consensus sequence of TTTAATCC, compared to TCTAATCC for the *D. melanogaster hb* promoter. That all three sites show the same C => T change suggests that this is unlikely to have occurred co-incidentally. Rather, an original site variant might have spread over time and replaced all three sites by slippage or other homogenising mechanisms, as discussed in Chapter 1.

Further evidence supporting the candidate sites

What evidence is there that these three candidate sites are authentic bcd binding sites involved in the regulation *M. domestica hb* expression? The binding sites were identified on the basis of their sequence, taking into account the known hierarchy of base preferences shown by *D. melanogaster* bcd. They would bind the *D. melanogaster* bcd protein - an artificial promoter consisting of multiple copies of a site corresponding to the *m. domestica* consensus binding sequence was shown to be activated by *D. melanogaster* bcd in yeast co-transfection assays (Hanes and Brent, 1991) albeit at reduced efficiency. Similarly, the immunoprecipitation experiments of Driever and Nüsslein-Volhard (1989) identified a bcd binding site with the sequence TTTAATCC.

However, this site was less well protected in footprinting assays with *D. melanogaster* bcd compared to the *D. melanogaster* consensus site.

As discussed in Chapter 2, the amino-acid sequence of the *M. domestica* bcd protein suggests that major differences in binding specificity compared to *D. melanogaster* are unlikely, although smaller, subtler changes cannot be ruled out. The studies of Hanes and Brent (1991) and Driever and Nüsslein-Volhard (1989) present some information on the relative affinities of various binding site sequences for the *D. melanogaster* bcd protein. These suggest that the presence of a TAAT core to the binding site is a prerequisite for high affinity binding, and that the identity of bases 3' to the TAAT core are more important in determining affinity than those 5' to the TAAT core. Oligonucleotide selection experiments performed using other Antp-class homeodomain proteins have similarly identified a critical role for the TAAT core, and the relative importance of the bases 3' to the TAAT core rather than those 5' to the core in determining the affinity of the binding site- homeodomain interaction (Ekker *et al.*, 1991,1992; Wilson *et al.*, 1993). This order of preference is also supported by the structural data which indicates that homeodomains make direct contacts with the major groove of the DNA binding site within the TAAT core, and to residues 3' to the core (Kissenger *et al.*, 1990; Billeter *et al.*, 1993; reviewed Gehring *et al.* 1994). Therefore, if the *M. domestica* bcd protein has an unchanged specificity, it will recognise sites containing a TAAT with higher affinity than those without. Furthermore, of those possessing a TAAT core, those with mismatches to the *D. melanogaster* consensus bcd binding site sequence located 5' to the core will be of higher affinity than those with mismatches 3' to the core. On this basis, the three TTTAATCC sites, of all the possible sites within the sequenced *M. domestica* hb upstream region, are likeliest to be the high affinity *M. domestica* bcd binding sites.

Independent evidence for the authenticity of the candidate *M. domestica* bcd binding sites comes from their roughly 100 bp spacing. A similar spacing is also found in *Drosophila melanogaster*, and is approximately conserved in *D. virilis*, despite the apparent occurrence of a number of insertion/deletion events (Driever and Nüsslein-Volhard, 1989; Treier *et al.*, 1989). It has been suggested that the spacing might be

important in allowing co-operative interactions (Driever and Nüsslein-Volhard, 1989) and that this is reflected in the conservation of the spacing (Treier *et al.*, 1989). Furthermore, in Chapter 5, experiments will be described that demonstrate that the *M. domestica* bcd homeodomain does bind to the *M. domestica* consensus bcd binding sequence as defined in this Chapter.

Chapter 5

BINDING ASSAYS

As discussed in Chapter 4, the putative bicoid binding sites identified in the sequenced fragment of the *M. domestica hunchback* gene define a consensus sequence of TTTAATCC, as opposed to TCTAATCC for *Drosophila*. Driever and Nüsslein-Volhard (1989) identified a site corresponding to the *M. domestica* consensus sequence, and found that this site would bind the *D. melanogaster* bcd protein, but with a reduced affinity when compared to that of the *D. melanogaster* consensus site. Similarly, Hanes and Brent (1991) found that *D. melanogaster* bcd would activate promoters consisting of multimers of a sequence corresponding to the *M. domestica* consensus, but less effectively than when the *D. melanogaster* consensus sequence was used. Therefore, if the *M. domestica* bcd protein is functionally identical to that of *D. melanogaster* bcd, then the affinity of the interaction between *M. domestica* bcd and the individual candidate sites within the sequenced portion of the *M. domestica* *hb* upstream region will be lower than that of *D. melanogaster* bcd for its *D. melanogaster* *hb* promoter sites.

As discussed in Chapter 1, the factors that determine the domain of early zygotic *hb* expression in *D. melanogaster* are the local concentration of bcd protein, the affinity of the interaction between bcd and its binding sites in the *hb* promoter, the number of these bcd binding sites. Also, a synergistic interaction between *hb* and bcd may be involved in regulation of early zygotic *hb* expression, in which case the number and affinity of *hb* binding sites within the *hb* promoter are also important in determining the extent of the *hb* expression domain. If no other factor in this regulatory interaction is altered between *M. domestica* and *D. melanogaster*, and the specificity of the bcd protein

is in fact conserved between the two species, then a reduced affinity of the *M. domestica* bcd binding sites relative to the *D. melanogaster* sites would be expected to lead to an anterior shift in the posterior boundary of *M. domestica* *hb* expression relative to that of *D. melanogaster*.

However, it is known that the early zygotic *hb* expression domain is conserved between *M. domestica* and *D. melanogaster*, extending over about 55% of egg length in both species (Sommer and Tautz, 1991). This suggests that there must have been a compensatory change in some component of the regulation of early zygotic *hb* expression. As bcd binds directly to these sites, one of the prime candidates for providing the compensatory change has to be the *M. domestica* bcd protein.

In Chapter 3, the comparison of the *bicoid* sequence between *M. domestica* and *D. melanogaster* suggested that it was possible that the specificity of the bicoid protein had changed subtly between the two species. A changed specificity of the *M. domestica* bcd homeodomain has also been indicated by transplantation and rescue experiments. Schröder and Sander (1993) used poleplasm from a number of insect species including *M. domestica* to rescue embryos from mutant *D. melanogaster* lacking a functional *bcd* gene product, and observed a qualitative failure of rescue. This is suggestive of an evolutionary divergence between *M. domestica* bcd and the regulatory regions of one or other of its' specific target genes, such that *M. domestica* bcd is unable to fully substitute for the *D. melanogaster* bcd function. This raises the interesting possibility that coordinated changes in the sequence of the three potential bcd binding sites in the putative *hb* promoter of *M. domestica* have elicited compensatory changes in the homeodomain of the *M. domestica* bcd protein in order to maintain the functional interaction.

The aim of the work described in this Chapter has been to express the *M. domestica* bcd homeodomain as a fusion protein, using this in gel retardation assays to determine its relative affinity to the two consensus bcd binding sequences, in a preliminary investigation of the conservation of the specificity of the bcd protein. Gel retardation assays are often used in investigations of DNA-protein binding interactions.

One particular virtue of these assays is that crude protein extract may be used to provide a rapid method of obtaining a comparison of the relative binding affinities of different proteins for a given sequence. Furthermore, this technique may be extended through the use of purified proteins to allow measurements of various biochemical constants, including the dissociation rate constant, thus enabling comparisons to be made of the affinity of different proteins for the same site. Therefore this technique provides a useful mechanism for quickly investigating the (qualitative) relative affinity of the *M. domestica* homeodomain for the *D. melanogaster* and candidate *M. domestica* bcd binding sites, whilst also facilitating a quantitative extension of the experiments in the future.

METHODS

Materials

BINDING SITES

Oligonucleotides were used to produce the probe binding sites for the gel retardation assays. The two strands of each probe possess 5'-AGCT-3' overhangs to permit end-labelling of the annealed oligonucleotides with DNA polymerase, cold dGTP, dTTP, and dATP and [α -³²P]-dCTP, as described in Chapter 2. The sequences of the probes are as follows:

<i>D. melanogaster</i> top strand	5' AGCTAGTCAT CTAATCC CTTTCTTGGTTC 3'
<i>D. melanogaster</i> bottom strand	5' ACGTGAACCAAGAAAG GGATTAG ATGACT 3'
<i>M. domestica</i> top strand	5' AGCTAGTCAT TTAATCC CTTTCTTGGTTC 3'
<i>M. domestica</i> bottom strand	5' ACGTGAACCAAGAAAG GGATTAAA TGACT 3'

BINDING BUFFER

1 x binding buffer is 20 mM Tris.Cl pH 7.5, 170 mM NaCl, 0.5 mM EDTA, 0.2 mM EGTA and 50% glycerol (Driever and Nüsslein-Volhard, 1989).

Methods

EXPRESSION OF *M. DOMESTICA* BCD HOMEODOMAIN FUSION PROTEIN

The *M. domestica bicoid* homeodomain was expressed as a fusion to maltose binding protein by ligation into the expression vector pMALc2 (New England Biolabs).

Expression vector construct

The primers BCD1 and BCD4 were used to PCR amplify (as described in Chapter 2) the *Musca domestica bcd* homeodomain from the plasmid bcdE11 (described in Chapter 4). The resulting fragment was digested with *Eco* RI and *Hind* III, and ligated into similarly cut pMALc2. The ligated DNA was transformed in the *E. coli* strain DH5 α F', and plated on Luria agar containing carbenicillin.

Replica plates were made from the resulting colonies. One replica was plated on Luria agar + carbenicillin containing 0.1 mM IPTG and 80 μ g/ml X-Gal, to enable blue/white colour selection. White colonies were identified, and the equivalent colonies from the plain Luria agar + carbenicillin replica plate were used to prepare plasmid DNA. Blue/white colour selection was not performed directly as the P_{tac} promoter is strong - transformants taken from a plate containing IPTG can contain mutant plasmids that have either lost part or all of the fusion gene, or no longer express it at high levels.

The plasmids were screened for the presence of the correct insert by preparing DNA and performing restriction digests with *Eco* RV, which cuts once within the vector and once within the insert to give two fragments, of 4.8 and 2 kb. The sequences of the inserts were confirmed by Taq cycle sequencing using the MAL primer, run on an

automated DNA sequencer as described in Chapter 2 (General methods). Plasmid pMALBCD is illustrated in Figure 5.1.

Production of crude extract and protein purification

Small scale pilot experiments were performed as described in Chapter 2. Subsequently, large scale preparations were made of crude protein extract from cultures of cells containing plasmid pMALBCD1, and also from cultures containing the pMAL plasmid. 10 mls of each crude extract was saved, and stored at - 20°C in aliquots. The remainders were run on amylose columns to produce the purified proteins (illustrated schematically in Figure 5.2). Large scale preparations and protein purification was performed as described in Chapter 2.

GEL SHIFT ASSAYS

Probe DNA

100 ng of probe (binding site) DNA was end-labelled with ^{32}P , as described in Chapter 2 (general methods). Unincorporated nucleotides were removed by ethanol precipitation, and the labelled DNA resuspended in 200 μl TE.

Binding reactions

The binding reactions proceed for 20 minutes at room temperature in 20 μl of 1 x binding buffer containing 1 mM DTT, 0.2 ng of labelled probe DNA and the specified quantities of crude protein extract. As a crude extract was used, the binding reactions also contained 1 $\mu\text{g}/\mu\text{l}$ of unlabelled competitor DNA (poly dI.dC, Sigma). The competitor DNA eliminates non-specific binding to the labelled probe DNA fragments by the different proteins contained in the crude extract.

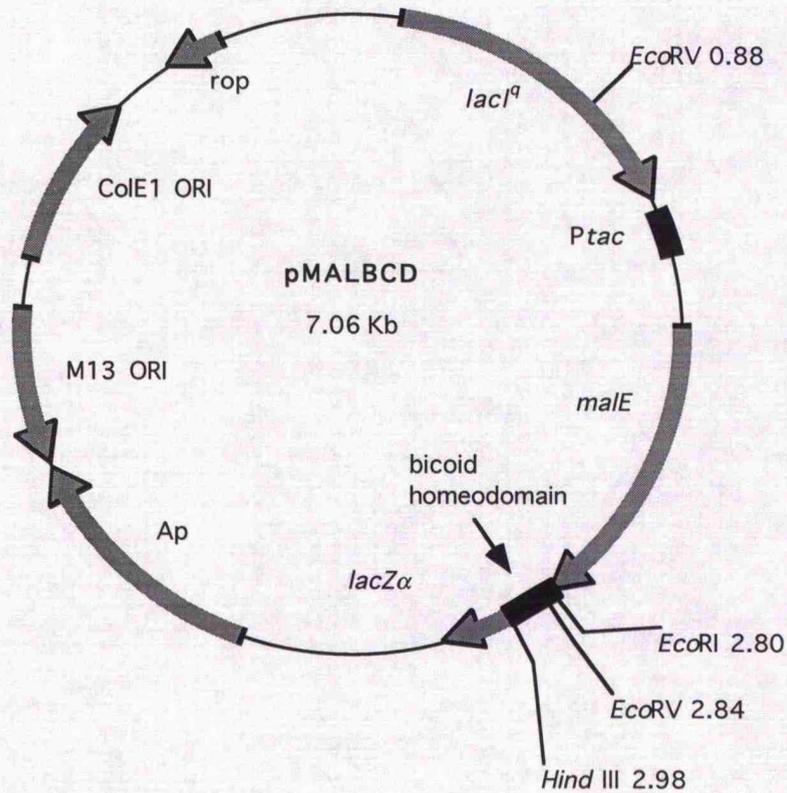
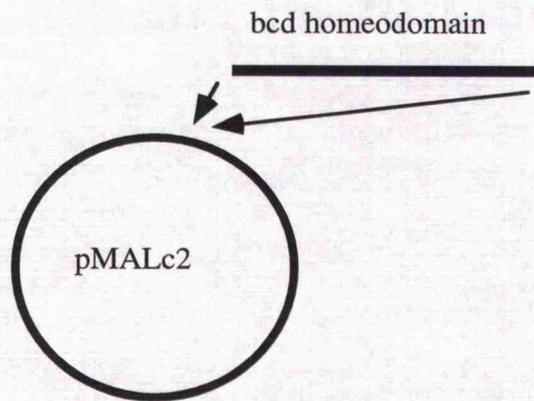


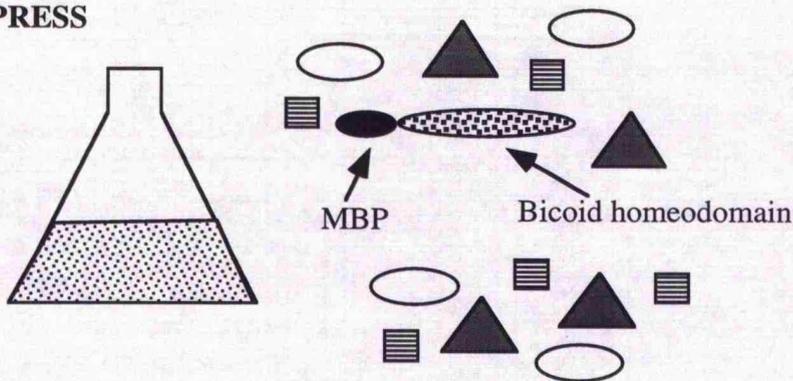
Figure 5.1. Restriction map of the plasmid pMALBCD. This contains the homeodomain of the *Musca domestica bicoid* gene ligated (in frame) into the polylinker of pMALc2 (New England Biolabs). pMALBCD expresses the bcd homeodomain as a fusion to the *E. coli malE* gene, which codes for maltose binding protein (MBP).

CLONE



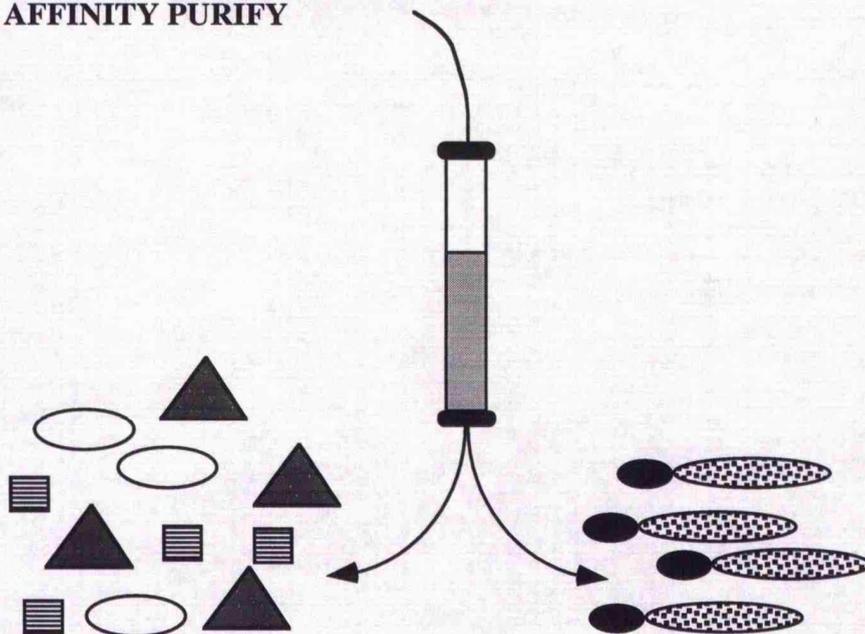
The bicoid homeodomain is cloned into the pMALc2 vector, creating a gene fusion with the maltose binding protein (MBP) - encoding *malE* gene

EXPRESS



Transformed *E. coli* is grown and the culture is induced to produce MBP fusion protein.

AFFINITY PURIFY



The crude cell extract is poured over the amylose column. The fusion protein is purified by binding to an amylose column, while all other proteins flow through. The fusion protein is then eluted in purified form with maltose.

Figure 5.2. Schematic diagram illustrating the expression and affinity purification of the *Musca domestica* bcd homeodomain - maltose binding protein fusion.

Electrophoresis

The samples from the binding reactions were electrophoresed for approximately 45 minutes at 100 V on small (8 cm x 7 cm) 4% polyacrylamide gels containing 1 x TBE and 2.5 % glycerol. Protein sample buffer was loaded in one of the outside lanes to enable the monitoring of the progress of the electrophoresis. After electrophoresis, the gel was placed on a sheet of Whatman 3MM paper, covered in Saran wrap and dried in a gel drier (BioRad).

Autoradiography and quantitation

Dried gels were subject to autoradiography at -80°C, with intensifying screens. Where required, bands corresponding to probe and free DNA were quantified by using a PhosphorImager and a storage phosphor screen (Molecular Dynamics) to acquire a digitised image of the gel to be analysed. Quantification then involved the definition of a window around each band, and the integration of the pixel values within these windows.

Relative binding affinity

The binding affinities of different DNA sites for a given protein were calculated from the quantity of bound DNA expressed as a percentage of the total DNA, for a given protein concentration, averaged over three independent replicate experiments.

RESULTS

To investigate the binding properties of the homeodomain of the *Musca domestica bicoid* gene *in vitro*, it was expressed in *Escherichia coli* as a fusion to the maltose binding protein (MBP), using the expression vector pMALc2 (New England Biolabs, Figure 5.1).

The pMALc2 vector uses the strong P_{tac} promoter and the translation initiation signals of MBP to express large amounts of the fusion protein. The maltose binding protein fusion facilitated subsequent purification by amylose-affinity chromatography. Figure 5.2 illustrates the expression and purification strategy.

PROTEIN EXPRESSION

A small scale pilot experiment was performed as described in the Methods section. This indicated that expression of the bcd homeodomain-maltose binding protein fusion is inducible, and that the fusion protein is soluble.

Large scale preparations were then made of both the bicoid fusion protein, and the maltose binding protein alone. Crude extract was obtained from both preparations (Figure 5.3a).

M. DOMESTICA BCD BINDS TO THE *M. DOMESTICA* SITE

Increasing concentrations of a crude extract of the bcd fusion protein were used in gel retardation assays against a constant concentration of the *Musca domestica* consensus bcd binding sequence (Figure 5.4a). These indicate that the *M. domestica* bcd homeodomain is capable of binding to the *M. domestica* sites.

A crude extract of cells expressing the maltose binding protein but no bcd fusion fails to shift the *M. domestica* site probe, indicating that the observed DNA binding is a property of the bcd homeodomain, and not due to the maltose binding protein part of the fusion or to the other proteins present within the crude extract (Figure 5.4b) - i.e. the observed band shift is specific to binding by the *M. domestica* bcd homeodomain.

RELATIVE AFFINITIES TO *M. DOMESTICA* AND *DROSOPHILA* SITES

The relative affinity of the *M. domestica* bcd homeodomain for the *M. domestica* and *D. melanogaster* sites was investigated by comparing the degree of binding to each

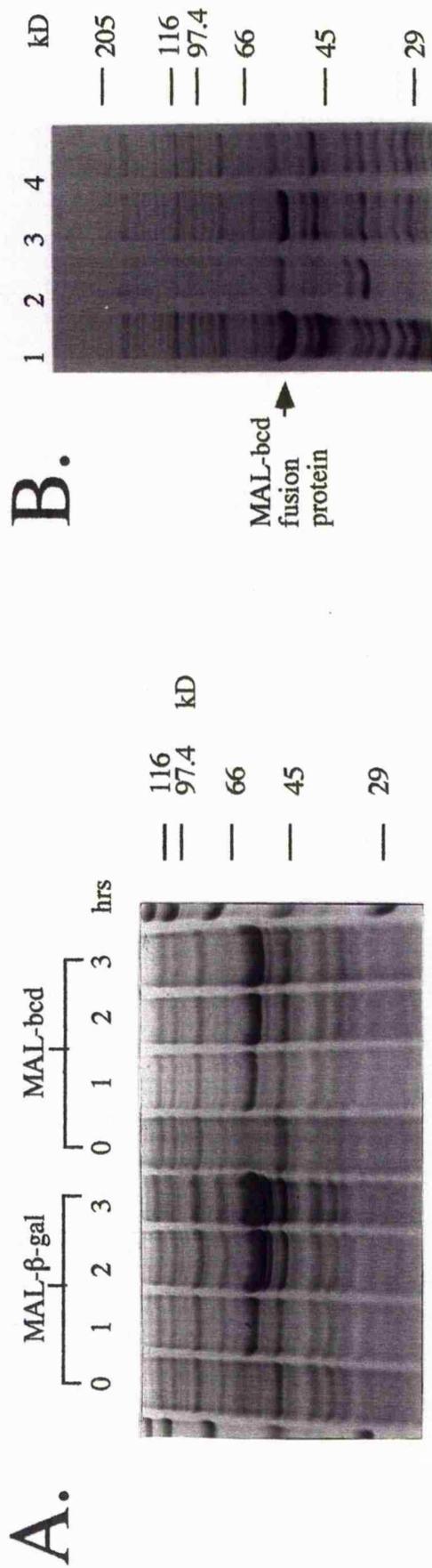


Figure 5.3. Expression and purification of the *M. domestica* bcd homeodomain fusion protein. The molecular weight of the maltose binding protein - β-gal fusion protein (MAL-β-gal) is 50 KDa. The expected molecular weight of the maltose binding protein - bcd homeodomain fusion protein (MAL-bcd) is 49 KDa (the MW of maltose binding protein alone is 42 KDa, and the estimated MW of the *M. domestica* homeodomain is 7 KDa).

A. Induction of fusion protein expression. Protein expression was induced in cells containing vector plasmid alone (MAL-β-gal) or pMALBCD (MAL-bcd) by the addition of IPTG, as described in chapter 2. The gel shows lysates of samples of cells taken 0, 1, 2 and 3 hours after induction.

B. Production of a crude extract of bcd fusion protein. Lane 1 - the crude extract of MAL-bcd fusion protein, 2 - insoluble material, 3 - induced cells, 4 - uninduced cells containing the expression plasmid pMALBCD.

C. Affinity purified proteins. Lane 1 - MAL-β-gal. Lane 2 - MAL-bcd fusion protein. The expression and purification strategy is outlined in Figure 5.2, and the methods are described in the text.



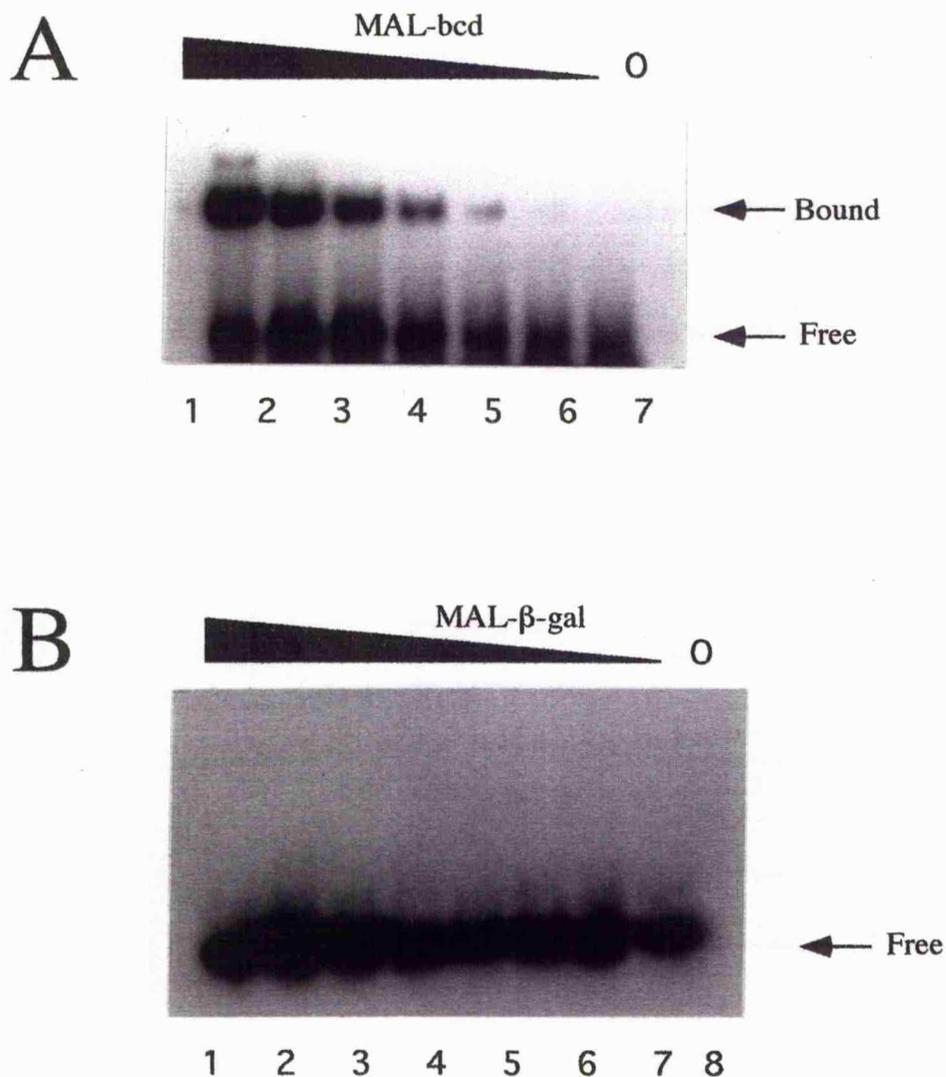


Figure 5.4.

A. Autoradiograph of a representative gel retardation assay indicating that the *M. domestica* bcd homeodomain fusion protein (MAL-bcd) does bind to the *M. domestica* consensus bcd binding sequence. Lane 1- 5 μ l of the crude extract of the MAL-bcd fusion protein , lane 2 - 1 μ l, 3 - 0.5 μ l , 4 - 0.1 μ l , 5 - 0.05 μ l , 6 - 0.01 μ l , 7 -no protein.

B. Autoradiograph of a gel retardation assay demonstrating that the maltose binding protein (MAL- β -gal) will not bind DNA. Lane 1- 5 μ l of the crude extract of the MAL- β -gal fusion protein , lane 2 - 1 μ l, 3 - 0.5 μ l , 4 - 0.1 μ l , 5 - 0.05 μ l , 6 - 0.01 μ l , 7 -no protein.

All assays were performed as described in the methods section of this chapter.

site in gel retardation assays. As it is possible to use crude protein extracts (thus avoiding the purification step), this is a rapid and easy method for making such preliminary comparisons.

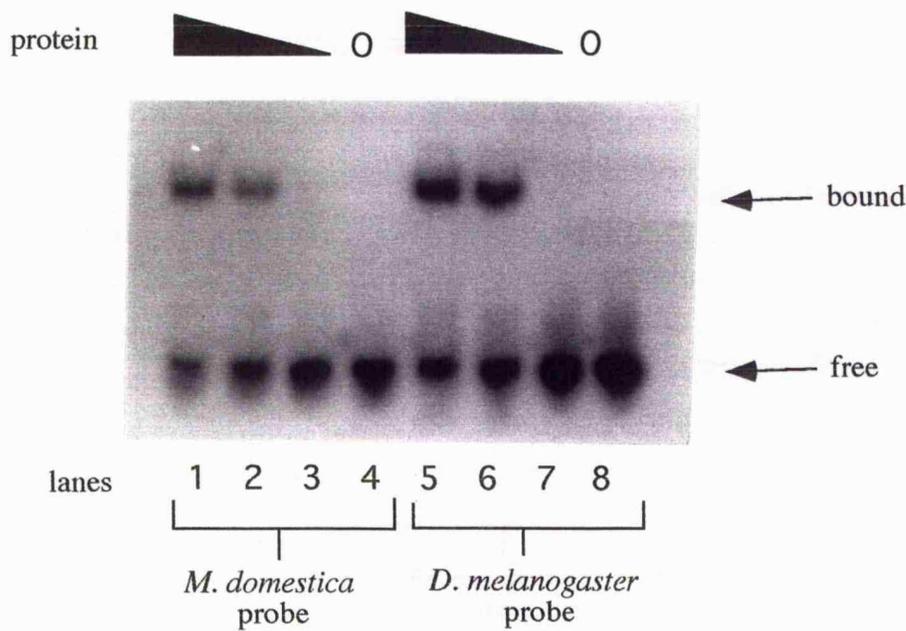
To facilitate direct comparisons between the two different sites, the sites were tested in parallel with equimolar concentrations labelled to approximately the same specific activity. Visual inspection of the autoradiographs suggested that the affinity of the *M. domestica* bcd protein is higher for the *D. melanogaster* sites compared to the *M. domestica* sites. This was confirmed by quantitation of the bands, using a Phosphor Imager, and calculation of the binding activity (% of bound DNA for a given protein concentration).

Figure 5.5 shows a typical gel, and gives the quantitated data from all three replicates. The quantitation confirmed the impression obtained from visual inspection of the autoradiographs - that the *M. domestica* bcd homeodomain has lower affinity for its own site. In fact, *M. domestica* bcd shows 60% binding activity to the *M. domestica* site relative to the *Drosophila* site, at low protein concentration. These experiments indicated that the *M. domestica* bcd homeodomain has a higher affinity for the *Drosophila* consensus site, with the differences in the degree of binding being most evident when binding assays were performed at lower concentrations (Figure 5.5).

PURIFICATION OF THE CRUDE EXTRACTS

The crude extract of the MALBCD fusion protein was purified by affinity chromatography on an amylose resin column (Figure 5.2). Figure 5.3c shows an SDS-PAGE gel of the purified homeodomain fusion protein. Should time have permitted, the purified protein would (as discussed below) have been used in further (quantitative) investigations of the DNA-binding interaction.

A



B

Probe site	Bound DNA (counts)	Free DNA (counts)	Total DNA (counts)	Binding affinity (%)	Average % binding affinity	Relative binding affinity
<i>M. domestica</i>	111084	781084	892168	12.451	14.991±1.768	60.3%
	111538	578022	689560	16.175		
	172750	883997	1056747	16.347		
<i>D. melanogaster</i>	213127	707003	920130	23.162	24.871±1.759	(100%)
	230316	613758	844074	27.286		
	229490	720233	949723	24.164		

Figure 5.5. Gel retardation assays indicate that the *M. domestica* bcd homeodomain has a higher affinity for the *D. melanogaster* consensus bcd binding sequence than for that of *M. domestica*.

A. Autoradiograph of a representative gel shift assay. Lanes 1-4, *M. domestica* site DNA; Lanes 5-8, *D. melanogaster* site DNA. Lanes 1 and 5, 0.5 μ l of the crude extract of the bcd homeodomain fusion protein; Lanes 2 and 6, 0.2 μ l; Lanes 3 and 7, 0.05 μ l. Lanes 4 and 8, no protein added.

B. Quantified results. Three independent repeats of the gel retardation assay were performed, and the intensity of the bands quantified using a PhosphorImager. The binding affinity of the *M. domestica* bcd homeodomain fusion protein is given by the quantity of DNA bound, expressed as a percentage of the total DNA (bound + free), for a given concentration of protein and DNA. The binding affinity was calculated for lanes 2 and 6 (0.2 μ l of protein). The relative binding affinity of the *M. domestica* bcd homeodomain for the *M. domestica* consensus bcd binding site is the binding affinity for this site expressed as a percentage of the binding affinity for the *D. melanogaster* site.

DISCUSSION

As discussed in the introduction to this Chapter, the sequence data from the *Musca domestica bicoid* and *hunchback* genes presented in Chapters 3 and 4 points to the interesting possibility that co-ordinated changes in the three potential bcd binding sites in the putative *hunchback* promoter might have elicited compensatory changes in the homeodomain of the *M. domestica* bcd protein in order to maintain this interactive function. The work described in this Chapter is intended as a preliminary investigation of the affinity of the *M. domestica* bcd homeodomain for the *D. melanogaster* and candidate *M. domestica* bcd binding sites.

M. DOMESTICA BCD BINDS TO THE CANDIDATE SITE SEQUENCE

The candidate bcd binding sites in the putative *M. domestica hb* upstream region were identified on the basis of the closeness of their match to the *D. melanogaster* consensus bcd binding site sequence. The results of the gel retardation assays presented in this chapter demonstrate that the *M. domestica* bcd homeodomain does bind to the candidate bcd binding site sequence, as defined in Chapter 4. This provides further evidence in favour of the candidate sites as the authentic *M. domestica* bcd binding sites. However, given that the *M. domestica* bcd homeodomain may have an altered DNA binding specificity (see Chapter 3), the existence of higher affinity sites of more divergent sequence also needs to be considered.

Mapping bcd binding sites

It would be interesting to map bcd binding sites within the *hb* regulatory region in a manner analogous to the immunoprecipitation experiments with which Driever and Nüsslein-Volhard (1989) originally identified the bcd sites within the *D. melanogaster* promoter. As the MAL half of the fusion binds to amylose resin in a very

similar buffer to that used in the binding assay, it may be possible to use amylose affinity chromatography to isolate bcd homeodomain-DNA complexes, therefore enabling the selection of those fragments of a restriction digest of the *hb* upstream DNA that bind to the fusion protein. Alternatively (or subsequently) footprinting experiments could be performed, or the approximate location of binding sites determined through the functional dissection of the upstream region using reporter gene systems.

RELATIVE AFFINITY TO THE TWO CONSENSUS SEQUENCES.

The expressed *M. domestica* bcd homeodomain fusion protein was found to have a lower affinity for the *M. domestica* consensus bicoid binding site in gel retardation assays than for the *D. melanogaster* consensus bicoid binding site. Therefore, both the *M. domestica* and the *D. melanogaster* bcd homeodomains show a higher affinity for the *D. melanogaster* consensus bcd binding site sequence.

This is an interesting result, given that the consensus bcd binding site sequence within the putative *M. domestica* *hb* promoter appears to have changed (Chapter 5), and yet the early zygotic expression domain of the *hb* gene is conserved between these species (Sommer and Tautz, 1991). If none of the changes observed in the *M. domestica* bcd homeodomain in Chapter 3 are functionally significant, then, like *D. melanogaster* bcd, it would be expected to have a lower affinity for the TTTAATCC sites found within the *M. domestica* upstream region than for the TCTAATCC sites in the *D. melanogaster* promoter. This would (if no other component of the bcd-*hb* interaction is changed) be expected to result in an anterior shift of the posterior border of *hb* expression in *M. domestica* relative to *D. melanogaster*.

So, does the conserved relative order of binding affinity of the bcd protein indicate an unchanged specificity in *M. domestica*?

A CHANGED SPECIFICITY OF BCD?

A reversed order of preference for the bcd binding sites in the gel retardation assay to that actually seen - that is, with *M. domestica* bcd showing a higher affinity for the *M. domestica* site - would have indicated a definite change in specificity from that of *D. melanogaster*. However, whilst both proteins show the same relative order of preference for the two sites, this does not necessarily indicate that the binding specificity of the bicoid protein is unchanged between the two species. One possibility is that regions of the bcd protein other than the homeodomain contribute to the specificity of the bcd protein, and that evolutionary changes in such regions have resulted in a changed specificity overall. A second possibility is that, whilst the optimal binding sequence (i.e. the sequence bound with highest affinity, in other words the specificity) has been conserved between the two bcd proteins, the differences between them result in *M. domestica* bcd having a higher absolute affinity for the *M. domestica* binding site than *D. melanogaster* bcd does for the same site. Furthermore, as discussed in the introduction to this chapter, the poleplasm transplantation and rescue experiments of Schröder and Sander (1993) are also suggestive of changed specificity of the bcd protein, as transplanted *M. domestica* poleplasm is capable of inducing partial, but not full, rescue of *bcd*⁻ *D. melanogaster* embryos.

Higher affinity of *M. domestica* bcd

The experiments described above indicate that the *M. domestica* bcd homeodomain has the same relative order of preference between the *M. domestica* and *D. melanogaster* sites as the *D. melanogaster* bcd protein. However, it is not possible to make direct comparisons of the affinities of the two different proteins for the same sites from these experiments. Therefore the possibility remains that whilst both *M. domestica* and *D. melanogaster* bcd show the same relative set of affinity for the two sites, the absolute binding affinities may be very different. The experiment described in this chapter does not rule out the possibility that changes have occurred such that the affinity of the *M. domestica* bcd protein for the *M. domestica* site is similar to that of the *D. melanogaster* protein for the *D. melanogaster* site - thus maintaining the same

functional interaction (expression of *hb* over 50% of egg length) between *bcd* and *hb* in both species despite the differences in binding site sequence, whilst incidentally raising the affinity of *M. domestica* *bcd* for the *D. melanogaster* site higher still. This possibility can be examined by comparing the dissociation rate constants (k_d) for the two pairs of interactions. The k_d can be measured through modified gel retardation assays (Stone *et al.*, 1991; Revzin, 1989).

Contribution of non-homeodomain regions of *bcd* protein

In this study, the *M. domestica* *bcd* homeodomain alone was expressed in *E. coli*, rather than the full length protein, for the practical reason that, as discussed in the Chapter 2, the first and fourth exons of the *M. domestica bcd* gene have yet to be identified. For *Ubx*, the homeodomain alone has been shown to have the same specificity as the full length protein *in vitro* (Ekker *et al.*, 1992), and comparative studies of binding affinities regularly use the homeodomain in isolation (for example, Ekker *et al.*, 1994; Smith and Johnston, 1994; Catron *et al.*, 1993). However, it is possible that regions of the *bcd* protein outside of the homeodomain could make a contribution to the specificity of DNA binding by the *bcd* protein. In which case, changes in these regions might result in a changed specificity of the *bcd* protein which would not be detected by gel retardation assays performed using only the homeodomain.

Both the POU and the 'paired' classes of homeodomain proteins contain additional non-homeodomain DNA binding motifs (Verrijzer *et al.*, 1992; Treisman *et al.*, 1991) which increase the DNA binding specificity of these proteins. However, as no additional DNA binding motifs have been identified within the *bcd* protein, a more reasonable possibility would be that interactions with other protein factors might affect the binding specificity of the *bcd* homeodomain. An example of the ability of other protein factors to modulate the binding specificity of a homeodomain protein is given by the yeast protein MAT α 2. The binding characteristics of this protein are dependent not only on its binding specificity, but also on its association with other transcription factors (reviewed Johnson, 1992). In combination with the MCM1 protein, MAT α 2

forms a heterotetramer (Keleher *et al.*, 1988) and recognises a set of operator sites associated with a-specific genes with improved specificity and affinity (Passmore *et al.*, 1989); whilst, in combination with the MATA1 protein, MAT α 2 forms a heterodimer which binds to a set of different operator sequences associated with haploid-specific genes (Dranginis, 1990; Goutte and Johnson, 1993).

Some evidence that other protein factors may be important in DNA binding by bcd comes from the work of Hanes *et al.*, (1994). Hanes *et al.* examined the expression of reporter genes with promoters made up of multimers of bcd binding sites in yeast and in *D. melanogaster*. They found that different site spacings were required for optimal expression levels in the two system, and suggest that this difference might be explained by an interaction with an ancillary protein(s), in a manner analogous to the yeast MAT α 2 protein. It is possible that, if bcd binding does involve ancillary protein factors, then evolutionary changes in this interaction between *M. domestica* and *D. melanogaster* might have resulted in a changed DNA-binding specificity of the bcd protein.

AN UNCHANGED SPECIFICITY?

In Chapter 3, the comparison of the amino acids sequences of the *M. domestica* and *D. melanogaster* bcd homeodomains suggested that it was possible that subtle changes in the specificity of the bcd homeodomain had occurred, but that major changes were unlikely as none of the amino acids at positions known to contact the DNA directly were different between the two species. Despite the possibilities discussed above, in the light of the observation in this chapter that both bcd homeodomains share the same order of preference for the two sites examined, it now seems even less likely that the specificity of the bcd homeodomain (and hence its affinity for the two sites) has changed, as measured by *in vitro* assays. As discussed previously, if no other component of the bcd-*hb* interaction has changed in compensation, an unchanged specificity of the bcd homeodomain when the binding sites in the putative *M. domestica* *hb* promoter are of lower affinity would result in an anterior shift of the posterior boundary of *hb* expression in *M. domestica* relative to *D.*

melanogaster - which, according to the work of Sommer and Tautz (1991), is not seen. Therefore, if the results of the comparative gel retardation assays are taken to indicate an unchanged specificity of the bcd protein between *D. melanogaster* and *M. domestica*, the conservation of the domain of *hb* expression in the light of the lower affinity bcd binding sites found within the *M. domestica hb* upstream region points to the possibility of a compensatory change in some other component of the bcd-*hb* interaction. Other components of the bcd-*hb* interaction that might have altered in compensation for the lower affinity of the sites found within the putative *M. domestica hb* promoter, in order to maintain the functional interaction between bcd and *hb*, include the number of the bcd binding sites within the *hb* promoter, the concentration of the bcd protein, and any contributions to the regulation of *hb* expression made by protein-protein interactions such as a synergistic interaction between the bcd and hb proteins. These possibilities are discussed below.

Number of sites

The domain of *hb* gene expression in *D. melanogaster* is determined by both the affinities of the bcd binding sites, and the number of these sites within the *hb* promoter (Struhl *et al.*, 1989; Driever and Nüsslein-Volhard, 1989; Driever *et al.*, 1989a). The number and quality of sites could be integrated by co-operative interactions. Beachy *et al.*, (1993), have proposed a model for the co-operative binding of regulatory proteins to multiple sites within regulatory elements in which co-operative binding integrates both the number and the affinity of multiple binding sites, allowing functionally equivalent regulatory elements to be produced from a few high affinity sites, many low affinity sites, or from sites of some indeterminate number and affinity. If bcd-binding does involve co-operative interactions, then in this view, the contribution of the low affinity "X" sites and higher affinity "B" sites of the *D. melanogaster hb* promoter (Figure 4.1) should also be considered. Therefore, the *M. domestica hb* promoter could contain a larger number of lower affinity sites in compensation for the lower affinity of the highest affinity sites. In this context it is important to recall that the X sites only show

5-6/9 bps match with the consensus sequence (Driever *et al.*, 1989a), so such sites would not be picked up in the search for matches. Possibly, larger number of lower affinity sites in conjunction with the three high affinity sites could form a regulatory element functionally equivalent to the *D. melanogaster* bcd binding sites.

Protein-protein interactions

The regulation of early zygotic *hb* expression may involve protein-protein interactions. As discussed in Chapter 1, there is some evidence that bcd binding may involve co-operative interactions (Driever and Nüsslein-Volhard, 1989; Hanes and Brent, 1989; Struhl *et al.*, 1989), possibly with auxiliary factors (Hanes *et al.*, 1994) including the hb protein itself (Simpson-Brose *et al.*, 1994).

If the synergism between bcd and hb (Simpson-Brose *et al.*, 1994; Small *et al.*, 1991) is important in regulating early zygotic *hb* expression (see Chapter 1) then the number and affinity of the hb binding sites within the *M. domestica* *hb* promoter would also contribute to determining the extent of the domain of *hb* expression. If this is the case, then an increase in the number or affinity of hb binding sites within the *M. domestica* *hb* promoter could compensate for a decreased affinity of the *M. domestica* bcd binding sites.

As discussed in Chapter 4, there are 23 sites, mostly falling within the AT-rich region flanking the candidate bcd binding sites, which show a single mis-match to the hb consensus binding sequence (variously defined as $C/G^A/CATAAAAA$ (Stanojevic *et al.*, 1989) and ACNCAAAAAANTA (Treisman and Desplan, 1989)). Given the disparate nature of these consensus sequences, the likelihood of identifying a site which could actually bind hb *in vivo* on the basis of matches to the consensus would be even less than for bcd.

Concentration of *bcd*

It is possible that, in *M. domestica*, the local concentration of *bcd* at 50% egg length (the posterior border of *hb* expression) is higher than at the 50% egg in *D. melanogaster*. A higher concentration of protein would compensate for the lower affinity of the *M. domestica* *bcd* homeodomain and *hb* promoter interaction. Such an increase in the concentration of *bcd* at 50% egg length could be achieved in one of two ways - by an increased concentration of *bcd* at the anterior pole with the same slope to the concentration gradient, or by a conserved concentration of *bcd* at the anterior pole with a shallower slope to the gradient (either due to faster diffusion or slower protein degradation). Both of these would result in higher concentrations of *bcd* elsewhere in the egg, and so would have knock on effects on the expression of other genes regulated by bicoid, and for that reason seems unlikely - unless these interactions too have coevolved.

The gel retardation assays described in this Chapter indicate a conserved order of preference of the *M. domestica* *bcd* protein for the two consensus sites. As described above, this does not rule out a changed specificity, but neither does it demonstrate a conserved specificity of the homeodomain. However, as the *hb* expression domain is conserved between *D. melanogaster* and *M. domestica* (Sommer and Tautz, 1991), even if the specificity of the *M. domestica* *bcd* homeodomain is unchanged, some alternative compensatory change seems likely - unless, that is, the changes seen in the binding site sequences are of no functional significance *in vivo*. Therefore, it is interesting to ask whether the *M. domestica* *hb* upstream sequence is functionally equivalent to the *bcd*-dependent P1 promoter of *D. melanogaster* - will it drive expression of the *M. domestica* *hb* gene in the authentic *D. melanogaster* expression pattern when transformed into *Drosophila*?

Chapter 6

TRANSFORMATIONS

As discussed in Chapter 1, the domain of *hb* expression is determined by the number and affinity of the *bcd* binding sites within its promoter, the number and affinity of the *hb* binding sites, the (putative) synergistic/co-operative interactions between the regulatory proteins and the local concentration of *bcd* protein.

THE RESULTS SO FAR

In Chapter 4, three candidate high affinity *bcd* binding sites were identified within the *M. domestica hunchback* upstream region. These sites have the consensus sequence TTTAATCC, as opposed to the *D. melanogaster* consensus *bcd* binding sequence of TCTAATCC (Driever and Nüsslein-Volhard, 1989). The *D. melanogaster* *bcd* protein has a lower affinity for sequences corresponding to the *M. domestica* consensus *bcd* binding site than for the *D. melanogaster* site (Driever and Nüsslein-Volhard, 1989; Hanes and Brent, 1991), and, in Chapter 5 it was demonstrated that the *M. domestica* *bcd* homeodomain possesses the same order of preference for the two sites. Whilst the sequence comparisons made in Chapter 3 indicated that the *M. domestica* *bcd* homeodomain could have a changed specificity, the conserved order of preference for the *bcd* binding sites, in conjunction with the lack of changes in amino acids within the homeodomain at positions known to contact the DNA directly, now suggests that other complicating factors might have a role to play.

If bcd has an unchanged specificity in *M. domestica*, but lower affinity sites within the putative *M. domestica hb* promoter, the expectation would be for a reduced domain of *hb* expression in *M. domestica* relative to *D. melanogaster*. However, *in situ* hybridisation experiments indicate that the extent of the primary (bcd dependent) domain of *hb* expression is conserved between the two species (Sommer and Tautz, 1991). This suggests that another component of the bcd and hb interaction must have changed in compensation. As discussed in Chapter 5, there are a number of possibilities - namely, an increased number of lower affinity bcd binding sites, such that the set of sites within the *M. domestica hb* promoter is functionally equivalent to the smaller set of higher affinity sites in the *D. melanogaster* promoter; an increase in the number of hb sites or in the affinity of hb for its sites within the *hb* promoter; an increase in the co-operativity between bcd and hb; or, an increase in the concentration of bcd at 50% egg length in *M. domestica* relative to *D. melanogaster*.

An alternative set of possibilities is that the observed changes are not functionally significant *in vivo* - that the redundancy within the interaction *in vivo* means that the observed changes in the binding site sequence make no difference to the outcome of the interaction. Is the putative *M. domestica hb* promoter capable of mediating expression over the authentic *hb* domain, *in vivo*, in conjunction with *D. melanogaster* proteins? And if it were not, would the differences in expression domain be sufficient to have an effect on the embryo, given the redundancy prevalent throughout the developmental program?

FUNCTIONAL ASSAYS

One way of testing whether homologous genes are functionally equivalent *in vivo* is through transformation into *D. melanogaster*. Examination of the domains of expression of the transformed and endogenous genes provides a simple assay of the equivalence of the promoters of the two genes, and (if the protein products are functionally interchangeable, or chimeric constructs are used) examination of the ability of the transformed gene to rescue mutant flies can provide a very sensitive assay for the

correct spatial, temporal and quantitative function of the promoter element. In this way, Lukowitz *et al.* (1994) have been able to demonstrate that the *D. melanogaster* and *D. virilis* P2 *hb* promoters are functionally equivalent. The approach of testing homologous genes from other species in the *D. melanogaster* background has also been used in a number of other cases (Mitsialis and Kafatos, 1985, Heberlein and Rubin, 1990; Maier *et al.*, 1990, Malicki *et al.*, 1992; Langeland and Carroll, 1993).

Initially, it is necessary to determine whether the *M. domestica hb* promoter is capable of driving expression over the authentic domain of *hb* expression when transformed into *D. melanogaster*. If so, then this would suggest that either the differences in the binding sites seen between the two *hb* promoters have no functional significance *in vivo*, or that the total number of bcd sites within the *M. domestica hb* promoter (or possibly, the number or affinity of the hb binding sites) compensates for the lower affinity of the bcd binding sites.

If the putative *M. domestica hb* promoter is not capable of driving expression over the full authentic *hb* expression domain in *D. melanogaster*, then this would suggest that the compensatory change has occurred in another component of the regulation of *hb*, such as the local concentration of bcd, or the postulated interaction between the bcd and hb proteins. However, it would still be important to see whether the changed expression pattern had any functional effect *in vivo*. It is necessary to examine the biological significance of any changes in the expression domain that are seen, as a certain degree of misexpression may be tolerated. For example, in *D. melanogaster* embryos containing either 1, 3 or 4 copies of the *bcd* gene, the expression domains of downstream genes are displaced, but the larvae hatch normally (Driever and Nüsslein-Volhard, 1988b), indicating that the changes in the expression patterns of downstream genes caused by altered bcd concentrations are not important biologically. Similarly, Schultz and Tautz (1994) created embryos with an altered shape and location of *hb* gradient, but found that the larvae still show normal abdominal segment pattern. This suggests that the correct relative order of the activation and repression of target genes may be more important than the expression of these genes at a defined location within the embryo (Schultz and Tautz, 1994). In the same manner, it is possible that a certain

degree of change in the *hb* expression domain might be tolerated. Therefore, if the transformed *M. domestica hb* gene fragment is not expressed in the full *D. melanogaster hb* expression domain, its ability to rescue the *Drosophila hb⁻* phenotype should be examined.

IN THIS CHAPTER

The work described in this Chapter is intended as a preliminary experiment to examine whether or not the *M. domestica hb* gene fragment (both putative promoter and inferred protein coding region) is functionally equivalent to the *D. melanogaster hb* gene. If functional equivalence could be demonstrated, then this would suggest that the transformed gene fragment contains the authentic *M. domestica hb* promoter, and that no molecular co-evolution had occurred. A lack of equivalence suggests that further investigations would be worthwhile.

The primary aim of the experiment described in this chapter is to compare the expression domains of the endogenous *hb* gene and the transformed gene fragment. A reporter gene construct was not used as the differences between these two expression domains are expected to be relatively small, and therefore it is necessary to compare like with like. The greater stability of the *Lac-Z* transcript usually used as the reporter gene might subtly alter the appearance of the extent of the expression domain and the level of expression, masking the effect of the changes in the *hb* binding sites.

A secondary consideration in the design of this experiment was to allow the determination of the *in vivo* significance of any altered expression domain which might be observed. Using the complete *M. domestica hb* gene allows this to be examined by crossing the *M. domestica hb* gene into a *hb⁻ D. melanogaster* background, although the interpretation of results from this would be complicated should evolutionary divergence of the *M. domestica hb* protein have occurred, affecting its function.

Therefore, in a preliminary test of whether, *in vivo*, the *M. domestica hb* gene is functionally equivalent to the *D. melanogaster hb* gene, a fragment of the *M. domestica*

gene has been transformed into *D. melanogaster* via P-element mediated transformation. In this chapter, the insertion of the *M. domestica hb* DNA into a P-element vector and its subsequent transformation into *D. melanogaster* is described. The expression of the transformed *M. domestica* gene is then compared with that of the endogenous *D. melanogaster hb* gene by *in situ* hybridisation.

METHODS

Materials

D. MELANOGASTER STRAINS

W; SbΔ2-3/TM6 is described in Robertson *et al.*, (1988).

All other strains used were obtained from A. Peixoto, and are described in Peixoto (1994).

Methods

CONSTRUCTION OF *M. DOMESTICA HB* TRANSFORMATION FRAGMENT

The *M. domestica hb* fragment used for germline transformation was cloned into the pW8 P-element transformation vector (Klemnez *et al.*, 1987), which utilises the *white* gene for the detection of transformants. The resulting construct is summarised in Figure 6.1.

The 11 kb *Sal* I - *Sma* I fragment of λhb3.13 (see Chapter 4) was ligated into pW8 cut with *Xho* I and *Hpa* I. The ligated plasmid was transformed into *E. coli* (DH5αF'). The resulting colonies were screened for plasmids containing the *M. domestica hb* insert by colony hybridisation using a probe to the first zinc finger domain (described below), and three positives were identified. Restriction digestion of the plasmid DNA indicated that two of these, pW8hb1 and pW8hb2, contained an intact 11 kb *hb* insert. Sequence

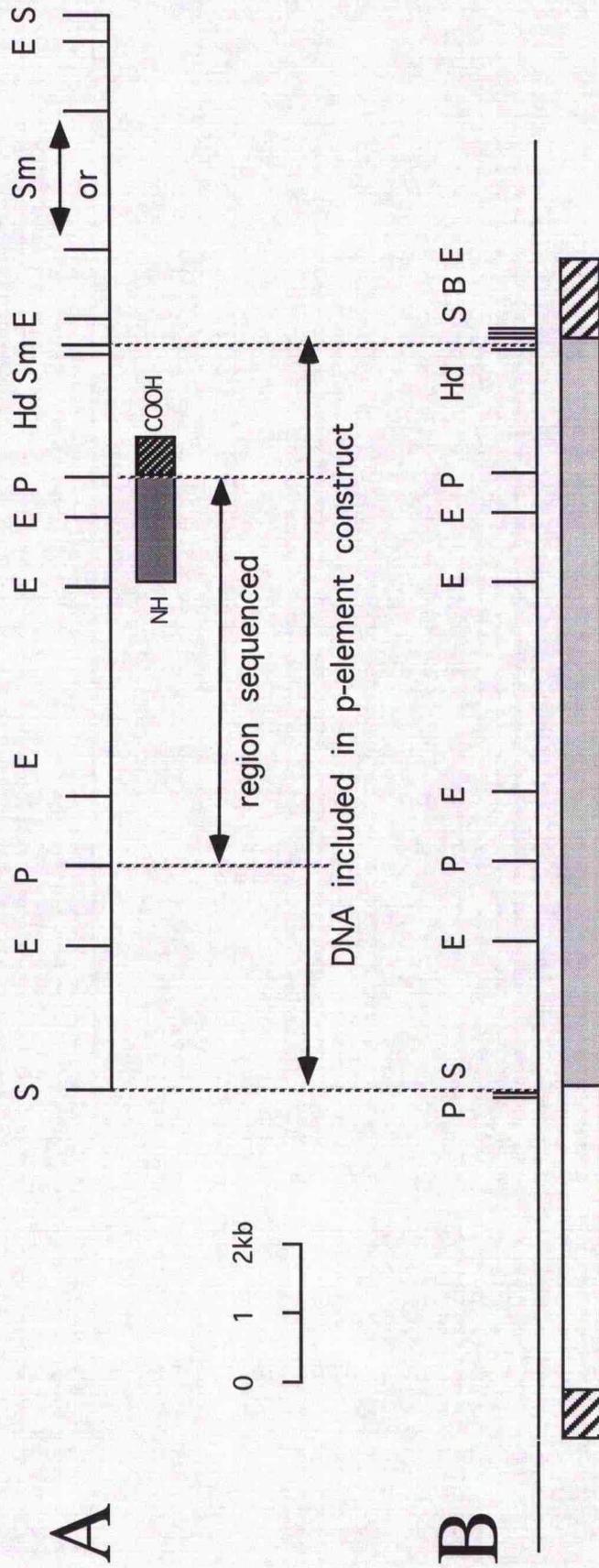


Figure 6.1. The plasmid p[Md**hb**], used to transform *Drosophila melanogaster* with the *Musca domestica hunchback* gene. **A.** a restriction map of the insert in λ **hb** 3.15 (Chapter 4), showing the region sequenced and the region included in the transformation construct. **B.** A restriction map of the construct itself, arbitrarily cut. Enzymes: S, *Sal* I; E, *Eco* RI; P, *Pst* I; Hd, *Hind* III; Sm, *Sma* I

P element
 M. domestica DNA
 sequenced coding region of *M. domestica hb* gene
 inferred location of C-terminal portion of *hb* coding sequence (see text)

was obtained from both plasmids using the MD2 primer to confirm the identity of the insert. Plasmid pW8hb1 was used for transformation.

P-ELEMENT MEDIATED TRANSFORMATION

This was performed as described by Rubin and Spradling (1982). The microinjection of *D. melanogaster* embryos is described in Chapter 2.

The *w; Δ2-3* strain of *D. melanogaster* was used as recipient for injections and as the source of P-element transposase (Robertson *et al.*, 1988).

Preparation of DNA for injection

A large scale preparation of the DNA to be injected was made using the Wizard Mega Prep plasmid purification system (Promega). The DNA was then ethanol precipitated and resuspended in 1 x injection buffer to a final concentration of 500 ng/μl.

The DNA was centrifuged briefly in a microcentrifuge before using, in order to remove any debris that might clog the needle, and back loaded into the needle by capillary action.

Crosses

As each adult fly derived from an injected embryo (the G₀ generation) ecloses, it was mated individually to several *w* strain flies. The parents were transferred to fresh vials every 4-5 days.

The G₁ offspring were examined for red eyed offspring. Any such offspring found were crossed individually with *w* flies, to establish a stock of the transformed line, and the pattern of segregation of the transformation marker observed. The injected embryos were *w; SbΔ2-3/ TM6*. Therefore both of the third chromosomes are

marked, and linkage of the red eye colour phenotype to either of these would indicate the presence of an insert on the respective third chromosomes. Insertions on the X chromosome were determined by crossing red eyed males to *w* females. If all red eyed offspring are female and all white eyed offspring male, this would indicate an insertion on the X chromosome. By elimination, any remaining insertions are likely to be on the second chromosome, and this was confirmed by crossing initially to *FM7; CyO/Sco* flies. In the next generation, *w*⁺ *CyO* flies are individually mated to *w* flies, and if all the *CyO* offspring have white eyes, this indicates an insertion on the second chromosome.

Transformed offspring of the G₀ injected flies fall into two classes distinguishable by their orange or red eye colour. Doubling of the gene dose in orange eyed fly stocks results in an easily visible darkening of eye colour, enabling the generation of homozygous transformants by simple inbreeding (Klemenez *et al.*, 1987). Transformed lines with red eyed heterozygote flies and containing an insert on the second chromosome were made homozygous by crossing to *FM7; CyO/Sco*, then crossing *w*⁺ *CyO* offspring together. Non-*CyO* flies were then used to establish the homozygous stock.

Southern analysis

Genomic DNA prepared from flies from the lines selected for *in situ* hybridisation analysis was used in Southern blots in order to confirm the integrity of the insert, and to check that each line contains a single insert (Figure 6.3). Southern blots of *Pst* I, *Bam* HI, *Eco* RI, and *Sal* I digests of genomic DNA were probed with the PCR amplified insert to plasmid hbE4.3Ss (Figure 6.3) cloned into similarly cut Bluescript II (Stratagene). The *Pst* I lane should contain a 5.9 kb fragment, and an additional band (greater than 3 kb) for each separate insert. The *Bam* HI lane should contain bands greater than 16.5 kb, with the number corresponding to the number of inserts. The *Eco* RI lane should contain two fragments, 1 kb and 2.7 kb in size, and the *Sal* I lane should give a single fragment of 11 kb.

IN SITU HYBRIDISATION

Whole mount *in situ* hybridisation of *D. melanogaster* embryos was essentially performed as described by Tautz and Pfeifle (1989). Details of the method are given in Chapter 2.

Probes

The *D. melanogaster* and the *M. domestica hb* probes consisted of the first zinc finger domains of the respective genes, amplified by PCR. The hybridisation conditions permit hybridisation only if the labelled fragment exceeds 94% complementarity with the mRNA expressed in the embryo (Tautz *et al.*, 1992). As the first zinc finger domain of *M. domestica* is only 78% identical to *D. melanogaster* at the nucleotide level, this allowed the first zinc finger region to be used as a probe. The *M. domestica hb* zinc finger probe was tested for cross hybridisation to the *D. melanogaster hb* gene by the *in situ* hybridisation of this probe to *D. melanogaster* embryos (Figure 6.3).

The *D. melanogaster hb* probe was amplified from the plasmid pBst-B-p2 (which contains the 3.35 kb *Eco* RI - *Xho* I fragment of the *D. melanogaster hb* gene; Tautz *et al.*, 1987), using the primers HB2 and HB3 (described in Chapter 2). The *M. domestica* probe was described in Chapter 4.

Both probes were labelled with digoxigenin (DIG labelling and detection kit, Boehringer Mannheim) by random priming according to the manufacturers instructions (described in Chapter 2), using the reagents supplied.

The labelling of the probes was tested by hybridisation to dot blots of serial dilutions of homologous plasmid DNA, as described in Chapter 2. Hybridisation was carried out with a probe concentration of 10 ng/ml, and colour development was allowed to proceed for 1 hr. After this time, 1 pg of homologous DNA could be detected. This is half the manufacturers claimed maximum specific activity, however,

probes with only a moderate specific activity (about 5 times less than the maximum) work well in *in situ* hybridisation experiments (Tautz and Pfeifle, 1989).

RESULTS

P-element mediated transformation (Rubin and Spradling, 1982) was used to transform *D. melanogaster* with the *M. domestica hunchback* gene. As discussed in the introduction to this Chapter, this potentially provides a sensitive *in vivo* assay for the ability of the *M. domestica hb* upstream regulatory region to provide the correct spatial, temporal and quantitative expression of the *hb* gene in conjunction with the *D. melanogaster* regulatory elements, and also an assay of the ability of the *M. domestica hb* protein to function in a *D. melanogaster* background.

Transformation with *M. domestica hunchback*

THE CONSTRUCT

D. melanogaster flies were transformed with a construct (pWhb1) containing an 11 kb *Sal* I - *Sma* I fragment from the insert of the library clone λ hb 3.15 (described in Chapter 4). This fragment has been ligated into the *Xho* I and *Hpa* I sites of the P-element vector pW8, which uses the *white (w)* gene for detection of transformants (Klemenez *et al.*, 1987).

As mentioned in Chapter 4, the integrity of the λ clone insert has yet to be confirmed using southern blots. If it is assumed that the insert contains a single continuous fragment of genomic DNA that has not undergone any rearrangement, then it can be calculated that the construct includes 7.5 kb of DNA upstream of the *hb* start site of translation. Sommer (1992) has amplified a fragment spanning both zinc finger domains of the *M. domestica hb* gene from genomic DNA by PCR. The sequence data presented in Chapter 4 (Figure 4.7) overlaps extensively with that of Sommer (1992),

and therefore it would appear that in *M. domestica* as in *D. melanogaster* the coding sequence is contained within a single exon and that the coding sequence is of conserved length. Therefore (assuming the integrity of the λ clone insert) the transformation construct may be inferred to contain approximately 1.5 kb of DNA downstream of the stop codon (end of zinc finger 2), as well as the protein coding region of the *M. domestica hb* gene (summarised in Figure 6.1).

INJECTION AND GENETIC ANALYSIS

The P-element construct containing the *M. domestica hb* gene, pWhb1, was introduced into *w; Sb Δ 2-3/TM6* flies by microinjection as described in the Methods section. The *Sb Δ 2-3* strain of flies contains the immobilised Δ 2-3 P-element integrated into the genome as a source of the transposase required for transposition of the construct element, rather than a helper plasmid (Robertson *et al.*, 1988). This should result in a higher rate of transposition as the transposase is constitutively expressed, rather than requiring transient expression of a helper plasmid before the construct plasmid can transpose.

Thirteen independent transformant lines were obtained from 127 fertile G₀ adults, a success rate of 10.2%. Each red G₁ offspring was individually crossed to *w* flies, and the pattern of inheritance of red eye colour observed. From this it was possible to determine, as described in the Methods section, the chromosome into which the *w* gene had inserted.

As many of the heterozygous transformed flies are orange rather than red-eyed, but darker eyed when homozygous, it was possible to obtain homozygous lines of such flies by simple inbreeding (Klemenez *et al.*, 1987). Homozygous stocks of lines 86MA and 62FA, which have red, rather than orange eyes, and insertions on the second chromosome, were established by crossing to *FM7; CyO/ScO* flies as described in the Methods section. Table 6.1 summarises the transformed lines obtained.

Table 6.1. A summary of the transformant lines obtained.

Line	heterozygous eye colour	Location
11FA*	orange	chromosome 2
15FA	red	Sb Δ 2-3 (chromosome 3)
37FA*	orange	chromosome 2
62FA*	red	chromosome 2
81FC	orange	X chromosome
86MA	red	chromosome 2
88FA	orange	TM6 (balancer chromosome 3)
99FB	red	Sb Δ 2-3
108FA	orange	X chromosome
110MA	red	multiple inserts, not characterised
122FC	both red	2 inserts, 1 on X chromosome, 1 on Sb Δ 2-3
130FA	orange	TM6
143MA	orange	TM6

An asterisk indicates a line used for *in situ* hybridisation analysis. The recipient strain for the injections was *w; SbD2-3/TM6*. The transformed descendants of the injected flies fall into two classes, orange and red eyed. As doubling the gene dose in orange eyed fly stocks results in an easily visible darkening of eye colour, such stocks were made homozygous (where possible) by inbreeding. Red eyed fly stocks with inserts on the second chromosome were made homozygous by crossing to *FM7; CyO/Sco* flies, as described in Methods. Lines with inserts on the X, *TM6* or *SbD2-3* chromosomes were not used for further analyses, for the reasons described in the Results.

SELECTION OF TRANSFORMED LINES FOR ANALYSIS

In order to make a comparison between the extent of the *D. melanogaster hb* expression domain and the expression pattern of the transformed *M. domestica hb* gene, it is necessary to compare equal gene doses. Therefore, lines containing inserts on the X chromosome, or on the *TM6* balancer chromosome were not used. As the *Sb Δ 2-3* chromosome contains the source of transposase, any P-elements within flies containing

this chromosome are potentially mobile. Whilst crossing over will remove the $\Delta 2-3$ gene, as a number of other alternative lines were available, there was no need to pursue this option. Therefore, three transformant lines, 11FA, 37FA and 62FA, each containing insertions on the second chromosome, were selected for further analysis.

Southern analysis of these lines confirmed that they contain intact single inserts (Figure 6.2).

in situ hybridisation

In situ hybridisation was carried out with digoxigenin (DIG) labelled probes according to the method of Tautz and Pfeifle (1989), as described in Chapter 2. All hybridisation and staining reactions were carried out in parallel, except that the colour development of embryos hybridised with the *D. melanogaster* probe were terminated after 20 minutes, whilst all other colour reactions were allowed to continue for 1 hour.

WITH THE *D. MELANOGASTER hb* PROBE

Figure 6.3 (central column) shows untransformed (*w* strain) *D. melanogaster* embryos hybridised with a probe to the *D. melanogaster hb* gene. Full signal development took 10-15 minutes. The expected pattern of *hb* mRNA expression (described in Chapter 4, see also Tautz and Pfeifle, 1989) is clearly seen.

WITH THE *M. DOMESTICA* PROBE

The results of the *in situ* hybridisation of the *M. domestica* probe to untransformed (*w* strain) *D. melanogaster* after 1 hours are given in Figure 6 (left hand column). This experiment acts as a control for cross hybridisation of the *M. domestica* probe to the *D. melanogaster hb* gene. From Figure 6.3 it can be seen that no cross hybridisation of the *M. domestica hb* probe occurs.

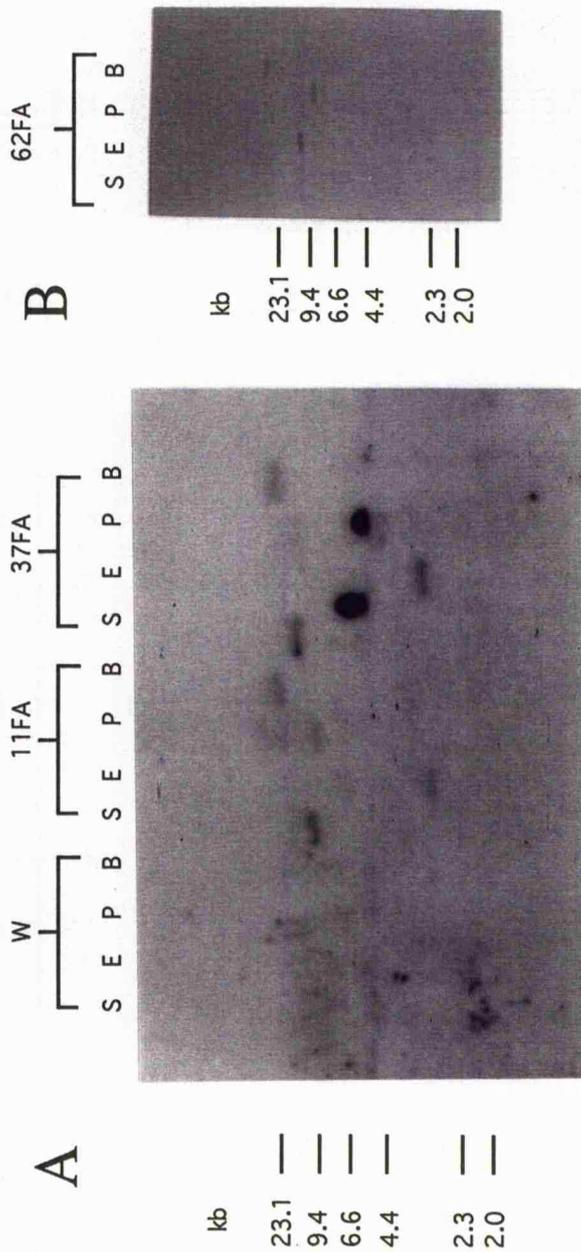


Figure 6.2. Southern blots of genomic DNA from three independent transformant lines of *Drosophila melanogaster*, indicating that each line contains a single, intact insert of the *Musca domestica hunchback* gene construct illustrated in Figure 6.1 w - negative control (not transformed); 11FA, 37FA & 62FA - transformed lines (see text for details). E - *Eco* RI digest, expected band size 2.7 kb; S - *Sal* I, expected fragment size 11.3 kb; P - *Pst* I, expected fragment size > 3.2 kb; B - *Bam* HI, expected fragment size > 16.5 kb. Both blots were probed with the 1.9 kb *Pst* I-*Sma* I fragment of the *M. domestica hb* gene (see Figure 6.1). For blot A the probe was labelled with 32 P, for blot B, the probe was labelled with DIG.

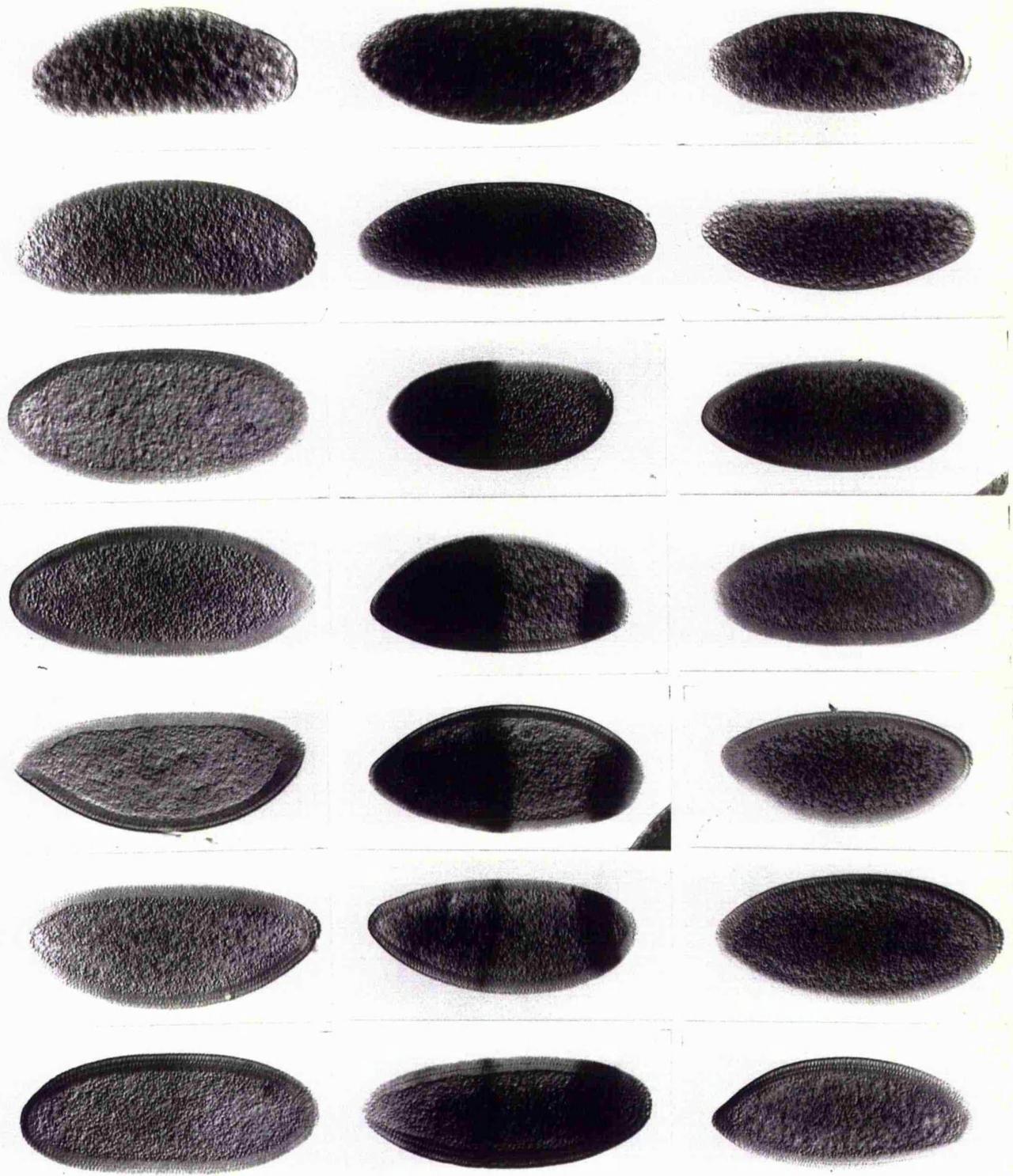


Figure 6.3. The expression pattern of the *Musca domestica hunchback* gene when transformed into *Drosophila melanogaster*. Left column - wildtype *D. melanogaster* embryos, hybridised with a probe specific to the *M. domestica hb* gene (negative control). Centre column - wildtype *D. melanogaster* embryos hybridised with a probe to the *D. melanogaster hb* gene. Right column - transformed *D. melanogaster* embryos, hybridised with a probe to the *M. domestica hb* gene. Equivalent stages are compared in each case. Embryos are oriented with the anterior to the left. See text for details.

Figure 6.3 (right hand column) shows the results of *in situ* hybridisation of the *M. domestica hb* probe to embryos from the transformant lines, allowing 1 hour colour development. Embryos from three homozygous transformant lines (11FA, 37FA, 62FA) were examined. No expression (maternal, early zygotic or later zygotic) was seen in any of the three transformant lines tested.

DISCUSSION

Transformation of *D. melanogaster*

Drosophila melanogaster were successfully transformed with a construct containing 11 kb of the *M. domestica hb* gene by P-element mediated transformation (Rubin and Spradling, 1982), and thirteen independent transformant lines obtained.

As mentioned in Chapter 4 and the results section of this chapter, the integrity of the original λ clone insert outside the sequenced section of the *hb* coding region is not known. It is therefore possible that it may contain multiple inserts or have undergone DNA rearrangements. However, if this is assumed not to be the case, then it may be inferred that the transformation construct contains the whole of the *M. domestica hb* coding region, and approximately 1.5 kb of DNA downstream from the stop codon, and 7 kb of DNA upstream of the start site of translation (see page 136). Sufficient upstream DNA was included to cover the *bcd* binding sites of the P2 promoter (as defined in Chapter 4), and possibly also any *M. domestica* equivalent of the P1 promoter (which drives later zygotic and maternal expression of *hb*, see Chapter 4).

The expression pattern of the transformed *M. domestica hb* construct was examined in three homozygous lines, 11FA, 37FA and 62FA, by *in situ* hybridisation of DIG-labelled probes to the *hb* mRNA (Tautz and Pfeifle, 1989). Early zygotic *bcd*-dependent expression over an anterior domain was expected, whilst any additional later zygotic or maternal expression would indicate the presence of elements related to the *D. melanogaster hb* P1 promoter within the construct.

Transformed *M. domestica hb* does not express

The expression pattern of the transformed gene fragment was analysed in three independent transformed lines (11FA, 37FA and 62FA) through the technique of *in situ* hybridisation, and compared to the expression pattern of the *D. melanogaster hb* gene.

The expected strong pattern of *hb* expression was observed when a probe to the *D. melanogaster* gene was used. Surprisingly, no expression of the *M. domestica hb* gene fragment was observed in any of the transformed lines examined (Figure 6.3). There are a number of possible explanations for the absence of *M. domestica hb* expression in the transformant lines, which are discussed below.

LOW CONCENTRATION OF TRANSCRIPTS?

It is possible that no expression of the *M. domestica hb* gene fragment was detected because the concentration of *hb* transcripts was too low to be detected. Such a low concentration of *M. domestica hb* transcripts could be caused by either a very low level of expression, or by high instability of the transcript. By raising the stringency of the hybridisation and washing procedures, it might be possible to reduce the level of the background sufficiently to allow a much longer staining period, and hence detection of very low levels of transcript. Possible causes of a low concentration of transcripts are outlined below.

Very low level expression

The level of expression of transformed genes varies according to the position of insert. Position effects could explain a low level of *hb* expression in one or two different lines, but seems unlikely to account for the lack of expression of the transformed gene in all three lines. The level of expression from artificial promoters containing *bcd* binding sites does vary according to the affinity of the sites (Driever *et al.*, 1989a),

however, it seems unlikely that this would be the explanation for the lack of the *M. domestica hb* expression as the bcd binding sites identified within the *M. domestica hb* upstream region in Chapter 4 are of higher affinity for *D. melanogaster bcd* than the "X" sites defined by Driever and Nüsslein-Volhard (1989, illustrated in Figure 4.1) which drive readily detectable levels of reporter gene expression when transformed into *D. melanogaster* (Driever *et al.* 1989a).

Another possibility is that the *M. domestica hb* protein is interfering with its own expression. If bcd and *hb* do interact synergistically with each other in directing *hb* expression, and this synergism is mediated by protein-protein interactions between bcd and *hb*, then co-evolution between the two proteins between *M. domestica* and *D. melanogaster* might result in the inability of the *M. domestica hb* protein to function in conjunction with *D. melanogaster bcd*. *M. domestica hb* expressed in the transgenic flies could interfere with *D. melanogaster bcd* in driving expression of the *M. domestica hb* gene in these flies. The *M. domestica hb* protein might be competing with the *D. melanogaster hb* protein for the occupancy of the *hb* binding sites in the *hb* promoters in transformed flies, and, if the *M. domestica hb* protein does not interact efficiently with the *D. melanogaster bcd* protein molecules, would disrupt the synergistic interaction, effectively repressing expression of the transformed gene. In this way a negative feedback loop would be formed, which could affect the level of expression of the transformed *M. domestica hb* gene, resulting in sufficiently low levels of *hb* expression as not to be detected in the *in situ* hybridisation experiments as performed in this chapter.

Not the authentic P2 promoter

A further possibility is that the bcd binding sites identified in Chapter 4 are not part of the authentic P2 promoter. As the integrity of the λ clone is not yet known, it is possible that the transformed "upstream regions" may not be related to the *M. domestica hb* gene. Consequently, the possibility that the construct does not contain the P2 promoter must be considered.

Similarly, even if the insert is integral, the genuine P2 promoter could be either composed of sites of very low affinity for the *D. melanogaster* bcd protein, or located so far upstream as not to be included within the construct, resulting in undetectable levels of *hb* expression (or no expression at all).

Transcript instability

The *D. melanogaster hb* mRNA transcript itself has a short half life, and a similar degree of instability might be expected for the *M. domestica hb* transcript. As the *D. melanogaster* transcript is easily detectable, a similar degree of instability of the *M. domestica* transcript on its own would not be expected to result in undetectable levels of *hb* expression. However, if the affinity of the bcd binding sites resulted in a lower level of expression anyway, an unstable transcript could reduce the concentration sufficiently that it falls to undetectable levels.

The construct contains 1.5 kb of DNA downstream from the stop codon. In *D. melanogaster*, there is a 3' UTR followed by a polyadenylation signal. If the 3' UTR is significantly longer in *M. domestica*, then it might not be fully included within the construct, which could result in a highly unstable transcript. As discussed in Chapter 4, Northern analysis of the P1 *hb* transcript in *M. domestica* indicates it to be 2 kb longer than expected.

Cis-acting factors

Possibly, because of the small amount of 3' sequence that has been included, *cis*-acting factors important in regulating *hb* expression have not been included in the construct, and this explains the low/non-existent expression of *hb*. However, reporter gene constructs containing the *D. melanogaster* P1 *hb* promoter are expressed correctly (Driever *et al.*, 1989a; Struhl *et al.*, 1989), suggesting that *cis*-acting factors are not required.

AFFINITY OF THE BINDING SITES

The *bcd* binding sites

As discussed in Chapter 1, the domain of *hb* expression in *D. melanogaster* is determined by the number and affinity of the *bcd* binding sites within its promoter. If the number remains constant, and the affinity is reduced, then the posterior border of the domain of *hb* expression is shifted towards the anterior. Therefore, one possible explanation for the lack of *hb* expression is that the affinity of *D. melanogaster* *bcd* for the sites within the *M. domestica* *hb* regulatory region is so low that there are no regions of sufficiently high *bcd* concentration in the *D. melanogaster* embryo to direct expression of the *M. domestica* gene - the domain of expression is so anteriorly restricted it is actually non-existent. This seems very unlikely.

In addition to the *bcd* binding sites identified by Driever and Nüsslein-Volhard (1989), Driever *et al.* identified three low affinity "X" sites within the *D. melanogaster* *hb* promoter. The *D. melanogaster* *bcd* protein is known to bind to the *M. domestica* consensus *bcd* binding sequence with an affinity higher than that for the "X" sites (Driever *et al.*, 1989a). As reporter gene constructs with promoters containing just the three "X" sites are expressed in *D. melanogaster* over a domain that is about half that of wildtype (100-79% egg length, compared to 100-50% egg length for wildtype, Driever *et al.*, 1989a), the expectation would be for the *M. domestica* *hb* gene to be expressed over a domain that extends to a position intermediate to the posterior boundaries of the X sites or wildtype promoter driven expression - to a point between 55 and 79% egg length - all other things being equal.

The *hb* binding sites

If, as suggested by Simpson-Brose *et al.* (1994), *hb* interacts synergistically with *bcd* in activating early zygotic *hb* gene expression, then the number and affinity of the *hb* binding sites may also play a role in determining the levels of *hb* expression. If the

affinity of *D. melanogaster hb* for the *M. domestica hb* promoter is sufficiently low, then this could explain the non-existent, or very low level, *hb* expression in the transformed lines. This would imply either that co-evolution had occurred between *hb* and its binding sites within the *hb* promoter between the two species, or that the regulation of early zygotic *hb* expression in *M. domestica* is different and does not require the *hb* protein.

Whilst the ideas discussed above may explain the lack of expression detected from the transformed fragment of the *M. domestica hb* gene, this apparent lack of expression raises two further issues. One is again the integrity of the λ clone insert. Before suggesting that the apparent absence of expression might be due to, for example, low levels of expression, unstable transcripts or even molecular coevolution, it is really necessary to rule out the possibility that the "upstream region" is composed of discontinuous fragments of DNA. The simplest explanation for the absence of expression would have to be the absence of a P2 promoter.

A second theory also suggests the absence of an *M. domestica* homologue of the *D. melanogaster hb* P2 promoter. It is possible that the interaction between *bcd* and *hb* is not conserved in *M. domestica*, and that the apparently conserved pattern of *hb* expression is produced using an alternate mechanism not able to function in the *D. melanogaster* embryo. Such an explanation would have been particularly attractive if maternal and secondary zygotic expression of the transformed construct had been seen in the absence of *bcd* dependent expression (though it could still have been explained - by for example, the rearrangement of a section of upstream DNA such that the P2 promoter was further upstream than the P1 and hence not included in the construct, or by differences in the structure of the *M. domestica hb* gene). The contention that there is

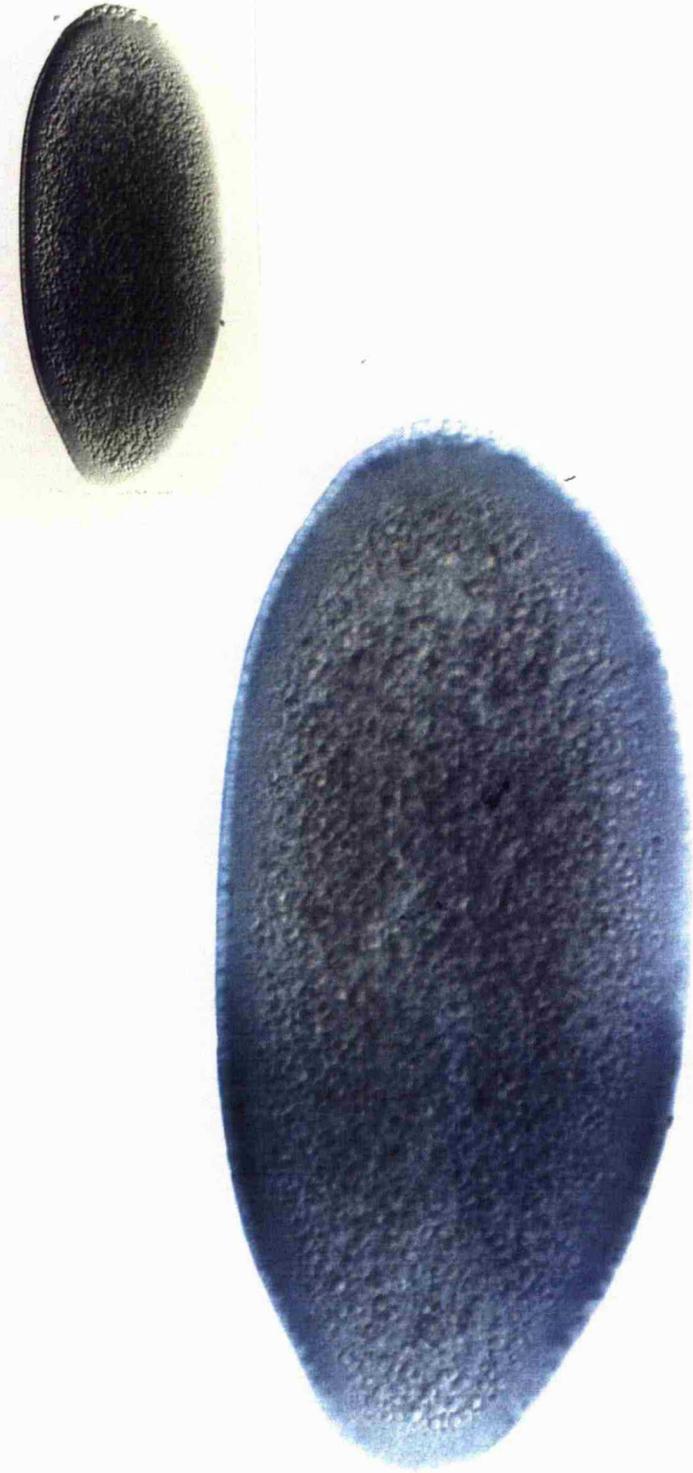


Figure 6.4. Expression of the *Musca domestica* hunchback gene fragment in a blastoderm stage *Drosophila melanogaster* embryo. The photographs show an embryo from the line 11FA, hybridised with a probe to the *M. domestica hb* gene, as described in the methods section.

Re-examination of the embryos produced in the *in situ* hybridisation experiments documented in Figure 6.3 revealed faint expression of the *M. domestica hb* gene in a few embryos derived from the transgenic lines. The embryo shown in the main colour photograph clearly demonstrates the observed expression pattern (F. Bonnerton, personal communication and photograph) which is less obvious in the original black and white photograph (inset).

no regulatory interaction between *bcd* and *hb* in *M. domestica* will be discussed further in Chapter 7.

As a postscript to the experiments described in this chapter, further *in situ* hybridisation experiments performed on the transgenic *Drosophila* lines described in this chapter have revealed expression of the transformed *M. domestica hb* gene fragment in an anterior domain, whilst no expression is detected in control embryos (F. Bonnerton, personal communication). Thus with the benefit of this information, it is in hindsight now possible to reinterpret the faint anterior shading in the embryo shown in the right hand column of row 3 of Figure 6.3 as expression of the transformed *M. domestica hb* gene fragment. Re-examination of the slides of embryos produced in the experiments described in this chapter has revealed one which shows this expression pattern particularly clearly (F. Bonnerton, personal communication; Figure 6.4). Whilst caveats regarding the integrity of the origin *M. domestica hb* λ clone must be taken into account, this result does suggest that a *M. domestica* homologue of the *D. melanogaster hb* P2 promoter is present within the transformed construct.

Chapter 7

DISCUSSION

Homologues of many developmental genes have been identified in a diverse range of species. From the comparisons of the conservation or divergence of individual gene sequences and expression patterns, inferences have been drawn as to the conservation or divergence of particular functional interactions.

The homeotic (*Hox*) genes, which specify segment identity (Figure 1.2), are widely conserved - homologues of many *Hox* genes have been identified in nematodes (reviewed Salser and Kenyon, 1994), leeches (reviewed Shankland, 1994), amphioxus (Holland *et al.*, 1992), and a variety of arthropods and vertebrates (reviewed Krumlauf, 1992, 1994). The *Hox* genes are clustered in the genome, and in *D. melanogaster*, and were found to be located in the same order within the cluster as their functional domains along the anterior-posterior axis of the embryo (Lewis, 1978). This expression order has been found to be conserved in other organisms and, where the genomic organisation of the *Hox* clusters are known - in nematodes, arthropods and vertebrates - the order of the genes has also been conserved (Kenyon, 1994). The spectacular conservation of the *Hox* gene clusters between nematodes, arthropods and vertebrates, including the spatial co-linearity of their expression with the order within the cluster has resulted in speculation that the homeotic complex represents a universal genetic mechanism responsible for the generation of segment identity in animals. This has led Slack *et al.*, (1993) to term this pattern of *Hox* gene expression the 'zootype' - the defining character of the kingdom Animalia; with the zootype being expressed at the 'phylotypic' stage - the point of greatest morphological similarity between the embryos of a phylum.

However, despite this widespread general conservation of the pattern of *Hox* gene expression, the precise manner in which it is achieved appears to vary. Early embryos are remarkable for their diversity. For example, *Drosophila* embryos undergo extensive patterning as a multinucleate syncytium; whereas in contrast, short germband insects such as the flour beetle *Tribolium* generate segments one by one from a posterior growth zone, and nematode embryos cleave asymmetrically to generate cells with different developmental potentials that undergo further diversification by cell-cell interactions. Given that early events in the establishment of *Hox* gene expression occur before cellularisation of the *Drosophila* embryo, whereas in vertebrates *Hox* gene expression is established in a cellular environment, it is often assumed that the upstream regulatory hierarchy of the *Drosophila* complex is not conserved in vertebrates. In *Drosophila*, maternal, gap and pair-rule genes are all involved in the establishment of *Hox* gene expression. So far, vertebrates do not appear to have homologues of many of these genes (Kenyon, 1994; Patel, 1994), suggesting that different strategies have evolved to set up the conserved pattern of *Hox* gene expression.

EARLY DEVELOPMENT IN INSECTS

Even within the insects, there appear to be differences in maternal, gap and pair-rule gene expression. Homologues of some of the early segmentation genes have been identified in both long and short germ band insects (reviewed Patel, 1994; Tautz and Sommer, 1995).

Homologues of the *Drosophila* pair-rule genes *eve*, *ftz* and *hairy* have been identified in the short germ band beetle *Tribolium* (Patel *et al.*, 1994; Brown *et al.*, 1994; Sommer and Tautz, 1993). These genes are basically expressed in the typical pair-rule pattern (see Figure 1.2), although appearing sequentially as the embryo elongates, rather than simultaneously as in long germ band insects such as *D. melanogaster*. This suggests that pair-rule genes still form part of the segmentation mechanism of the *Tribolium* embryo, although this occurs in a cellular, rather than syncytial,

environment. However, in the short germ band grasshopper, *Schistocerca gregaria*, *eve* and a potential homologue of *ftz* seem not be involved in the segmentation process - only later in development, during formation of the central nervous system, are their expression patterns homologous to those in *Drosophila* (Patel *et al.*, 1992; Dawes *et al.*, 1994).

Gap gene expression shows some similarities between *Tribolium* and *Drosophila*, with *Kr* being expressed in analogous domains in both species (Sommer and Tautz, 1993). In *Drosophila*, *Kr* is expressed in a central domain of the blastoderm. In *Tribolium*, *Kr* is expressed in a central domain only in early germ band embryos, whilst being expressed in a posterior cap at the blastoderm stage. This appears not to indicate a diverged expression pattern, but rather that the segments posterior to the *Kr* expression domain are not yet specified at the blastoderm stage - as indicated by comparisons of the relative expression patterns of the pair-rule gene *hairy* (Sommer and Tautz, 1993).

The original morphogen - *bcd* or *hb*?

In Chapter 1, it was stated that this analysis has been performed on the assumption that the fact of *bcd* dependent regulation of *hb* expression has been conserved in *M. domestica* (though its precise molecular basis may have changed). This assumption was justified by the conservation of the patterns of gene expression of *bcd*, *hb* and a number of other early developmental genes in *M. domestica*, coupled with the high degree of morphological similarities between *Musca* and *Drosophila*. However, the possibility that this interaction may not occur in *Musca* must also be considered.

In contrast to the apparently ancient origin of the homeotic gene clusters, a number of lines of evidence point towards the interaction between *bcd* and *hb* having arisen more recently. No gene homologous to *bcd* has been identified outside of the Diptera. Thus *bcd* is likely to be the result of a recent duplication within the *Antennapedia* complex of *Drosophila*, within which *bcd* is located. In contrast, the *hb* gene

is conserved, and may play an essential role in short germband insects and in more distantly related organisms (Sommer and Tautz, 1991; Sommer *et al.*, 1991).

The conservation of the maternal component of *hb* expression in *M. domestica* suggests that this expression may be evolutionarily older than the early zygotic *bcd* dependent expression (Sommer and Tautz, 1991). A gene sharing sequence homology to *nanos* (*nos*) is localised to the vegetal pole of the developing *Xenopus* oocyte (Mosquera *et al.*, 1993), and a sequence resembling the *nos* response element (NRE) of *hb* mRNA has been found to be essential for the localisation of the *glp-1* determinant of axis formation in the *Caenorhabditis elegans* embryo (Evans *et al.*, 1994). This suggests that the repression of expression at the posterior end may be the primordial morphogen (reviewed Kimble, 1994; Curtis, 1994). The presence of a functional NRE sequence in the *bcd* mRNA (Wharton and Struhl, 1991) suggests that at some point in evolution, *bcd* may have used the *nos* system to restrict the activity of its product to the anterior. It has therefore been suggested that in *Drosophila*, *bcd* is gradually assuming some of the morphogenetic functions once played by maternal *hb*, in particular by controlling early zygotic *hb* expression (Simpson-Brose *et al.*, 1994). Consequently, it could be suggested that control of the early zygotic *hb* expression pattern seen in *M. domestica* might not have been assumed by the *M. domestica bcd* gene.

If early zygotic *hb* expression is not under the control of *bcd* in *M. domestica*, the question is how is this apparently conserved expression pattern produced? The only simple explanation is that it could be generated through a positive feedback autoregulatory interaction in which protein translated from maternally transcribed *hb* mRNA (restricted to the anterior half of the embryo through the action of *nanos*) activates zygotic *hb* expression. Should this be the case, then (assuming no molecular coevolution of the *hb* protein and its own binding sites) the *M. domestica hb* promoter(s) would still be expected to function in *D. melanogaster* embryos - which do after all provide a good supply of *hb* proteins. (To distinguish between *hb*-dependent autoregulation and *bcd*-dependent activation of a transformed *M. domestica hb* gene, it would be necessary to examine the expression of constructs in both *bcd* and *hb* mutant backgrounds.) Therefore (notwithstanding caveats regarding the integrity of the λ

clone), divergent regulation of *hb* in *M. domestica* is not necessarily a satisfactory explanation for the absence of expression of the transformed gene fragment described in Chapter 6 - and nor is absence of detectable expression necessarily indicative of divergent regulation. In fact, the preliminary results of further *in situ* hybridisation¹ experiments performed on the transgenic *Drosophila* lines described in this thesis indicate that the transformed *M. domestica hb* gene fragment is expressed, at low levels, over an anterior domain in *D. melanogaster* embryos (F. Bonneton, personal communication), which is consistent with a *bcd*-dependent component of *hb* expression being maintained in *M. domestica*. Currently, the *M. domestica hb* gene fragment is being crossed into *bcd*- *Drosophila* (F. Bonneton, personal communication). Should expression be abolished in these lines, this would provide further evidence of the maintenance of a *bcd* dependent component to *M. domestica hb* expression.

Is there any evidence directly suggestive of divergent (i.e. non-*bcd* dependent) regulation of *hb* in *M. domestica*? It is known that *M. domestica* contains a *bcd* gene (Sommer and Tautz, 1991; Chapter 3), and that the *bcd* mRNA transcripts are anteriorly localised (Sommer and Tautz, 1991; Schröder and Sander, 1993). Therefore the existence of a *bcd* gradient in *M. domestica* can be inferred. Furthermore, anterior poleplasm from *M. domestica* embryos is capable of partial rescue of *bcd*- *Drosophila* embryos (Schröder and Sander, 1993) and in Chapter 4, the ability of the *Musca bcd* homeodomain to bind to the *Drosophila bcd* binding sequence was demonstrated. Therefore the presence of a *bcd* component within the regulatory interaction appears to have been conserved - any divergent regulation is not directly due to an absence of *bcd*-like activity in *Musca*. Therefore, if *hb* expression is not regulated by *bcd* in *Musca*, one would not expect to find sites capable of binding *bcd* within the *hb* promoter - as discussed in Chapter 4, sites with a 7/9 base-pairs match to the *bcd* consensus binding sequence would be expected to occur at a frequency of only 0.48 in 4 kb. In which case, (if the integrity of the *hb* λ clone insert is demonstrated, and the sequenced region can be demonstrated to contain regulatory elements) the presence of a significantly (50-fold) greater than expected number of close matches to the *Drosophila bcd*-binding site sequence upstream of the *M. domestica* coding region would require explanation. Whilst *bcd* may well have

come to usurp the function of maternally expressed *hb* as an anterior determinant in the lineage leading to *Drosophila*, there is no evidence so far to contradict the assumption (formed on the basis of conservation of morphology and patterns of gene expression) that the presence of a *bcd-hb* interaction - if not its precise molecular basis - has been conserved in *Musca*.

bcd and *hb* in *M. domestica* and *D. melanogaster*

As discussed in Chapter 1, examination of the fine grained differences that lead to this diversification of developmental programs requires the dissection and comparison of the interaction between two genes in closely related species. In this thesis, the *bcd* and *hb* genes of *M. domestica* have been partially sequenced (Chapters 3 and 4), and compared to those of *D. melanogaster*, in order to examine the possibility that this interaction may have been subject to molecular coevolution. Assays have been performed (Chapters 5 and 6) as a preliminary examination of the functional significance of the sequence changes observed.

The interaction between *bcd* and *hb* in *D. melanogaster* is determined by a number of factors, including the number and affinity of the *bcd* binding sites within the *hb* promoter, the concentration of *bcd* protein, and (possibly) a synergistic interaction between *bcd* and *hb* (see Chapter 1). Each of these factors is a potential substrate for evolutionary change. However, the outcome of the interaction between *bcd* and *hb* appears to be conserved in *M. domestica*, as both species express *hb* in an anterior domain extending from 100-50% of egg length (Sommer and Tautz, 1991). Therefore, changes in any one of the factors regulating *hb* expression in *M. domestica* must be compensated for by changes in others, in order that the same *hb* expression domain should result.

M. DOMESTICA HB

Three potential high affinity *bcd* binding sites have been identified within the region sequenced upstream of the *M. domestica hb* coding DNA. As discussed in

Chapter 4, these have a consensus sequence of TTTAATCC, as opposed to the TCTAATCC of *D. melanogaster*, with each of the three sites showing the same T to C change. If these do represent the homologues of the *D. melanogaster* bcd-binding sites, then rather than arising coincidentally, it would seem likely that this co-ordinated change would represent the spread of an original site variant - illustrating the ability of slippage and other homogenising mechanisms to propagate the accumulation of binding site variants within the regulatory regions of developmental genes.

D. melanogaster bcd has a lower affinity for a variant binding site sequence which corresponds to that of the *M. domestica* consensus candidate bcd binding site sequence than for the consensus *D. melanogaster* sequence *in vitro* (Driever *et al.*, 1989a; Hanes and Brent, 1991), suggesting the possibility that the co-ordinated changes in the binding site sequences might have elicited compensatory changes in the *M. domestica* bcd protein. It is possible that these differences in binding site affinity may be sufficiently small so as to cause negligible effects *in vivo*. Alternatively, if co-operative interactions are important in *hb* regulation, then the integration of other, lower affinity, sites within the *M. domestica* *hb* regulatory region with those of the three candidate high affinity sites by such interactions (Beachy *et al.*, 1994), could result in a *hb* promoter functionally equivalent to that of *D. melanogaster*. It will be important to ascertain the precise role of the postulated co-operative interactions between bcd (Driever and Nüsslein-Volhard, 1989, Struhl *et al.*, 1989) and synergistic interactions between bcd and hb (Small *et al.*, 1991; Simpson-Brose *et al.*, 1994) in the regulation of *D. melanogaster* *hb* expression in order to fully understand the changes that might have occurred in the bcd-*hb* interaction between *D. melanogaster* and *M. domestica*.

***M. DOMESTICA* BCD.**

The comparison of *bcd* in *M. domestica* and *D. melanogaster* revealed a number of interesting changes. Most notably, there were 5 changes (out of 60 amino acids) in the homeodomain sequence. This degree of divergence is striking, in the light of the higher degree of conservation typically seen in homeodomains such as those of the homeotic

genes over longer evolutionary periods - in several cases, the homeodomains of vertebrate *Antp*-homologous are virtually identical to that of *Antp* itself (59/60 amino acids; Akam, 1989). Whilst none of the amino acid changes are at positions implicated in determining sequence specificity of binding from the current structural and genetic data on homeodomain binding, and the *M. domestica* *bcd* protein shows the same relative order of preference for the two *bcd* binding sites, this, as discussed in Chapters 3 and 5, does not necessarily indicate a conserved specificity of the *bcd* protein. As the gel retardation assays described in Chapter 5 provide qualitative rather than quantitative measures of binding affinity, the possibility remains that despite maintaining the same order of preference for the two sites, *M. domestica* *bcd* binds with higher absolute affinity to the *M. domestica* site than does *D. melanogaster* *bcd*. Alternatively, it is possible that regions of the *bcd* protein outside of the homeodomain contribute to an altered binding specificity. Therefore the possibility that compensatory changes have accumulated in response to the spread of the variant binding sites cannot be ruled out. Further investigation of the specificity of the *M. domestica* *bcd* protein is required.

A DIVERGENT INTERACTION

The differences in binding site sequence in the putative *M. domestica* *hb* promoter, coupled with the conservation of the domain of *hb* expression (Sommer and Tautz, 1991), is suggestive of a coevolutionary change in this interaction. Furthermore, the failure of *M. domestica* *bcd* to fully rescue *bcd⁻* *D. melanogaster* embryos (Schröder and Sander, 1993) also suggests that the interaction between *bcd* and *hb* in both species is not completely functionally identical, as *M. domestica* *bcd* does not appear to be fully able to substitute for *D. melanogaster* *bcd*. This supports the contention that there has been some degree of change in at least some components of the *bcd-hb* interaction. However, at the very least (and regardless of the functional significance of these changes *in vivo*) provided that the sequenced DNA does contain the *M. domestica* homologue of the *D. melanogaster* P2 promoter, the differences in *bcd* binding site

sequence between the *M. domestica* and *D. melanogaster hb* promoters indicates that the precise molecular nature of the *bcd-hb* interaction, rather than being absolutely conserved, has in fact diverged between these two species.

The transformation experiments described in Chapter 6 aimed to provide an *in vivo* functional assay of the equivalence of the *M. domestica* upstream sequence with the *D. melanogaster hb* promoters. Until the experiment has been repeated (with higher stringency *in situ* hybridisations, or through using reporter gene constructs) and the integrity of the original λ clone insert demonstrated, the absence of detectable *M. domestica hb* expression in the transformed lines cannot be taken as demonstrating a non-equivalence of the *D. melanogaster* and *M. domestica hb* upstream sequences - and hence an incompatibility of *D. melanogaster bcd* with the *M. domestica hb* promoter. Nonetheless, if the sequenced 5.9 kb of *M. domestica* DNA can be shown to contain a *M. domestica* homologue of the *D. melanogaster hb* P2 promoter, then the divergent sequence of the three candidate high affinity *bcd* binding sites within the *hb* promoters would tend to indicate that there has been a divergence of the precise molecular basis of the interaction between *bcd* and *hb* during the time separating the two species.

In fact, further *in situ* hybridisation experiments performed on the transgenic lines described in Chapter 6 have revealed expression of the transformed *M. domestica hb* gene fragment in an anterior domain of blastoderm stage embryos (F. Bonneton, personal communication). This provides evidence that an *M. domestica* homologue of the *Drosophila* P2 promoter is in fact present within the transformed gene fragment. Interestingly, expression of the transgene appears to cover a very slightly more anteriorly restricted domain than the endogenous *hb* gene (compare the endogenous *hb* expression domain as seen in Figure 6.3, row 3 (centre) with the transgene expression domain as shown in Figure 6.3, row 3 (right) and in Figure 6.4. Should more rigorous analysis confirm this, it would be consistent with the expectations of molecular co-evolution. Assuming that the candidate *bcd*-binding sites described in Chapter 4 do mediate this aspect of *hb* expression, further evidence in favour of molecular co-evolution of the *bcd-hb* interaction would be provided if mutation of the *M. domestica*

candidate sites to match the *Drosophila* consensus sequence resulted in expression of the *M. domestica* transgene in the full endogenous *hb* anterior domain.

Flexibility and Constraints

It is commonly assumed that the requirement for a regulatory protein to interact with a number of genes places a constraint on the evolution of that interaction (for example, Scott, 1994). Once a relationship between a particular homeodomain protein and a specialised target gene has been established (termed a seminal regulatory interaction (SRI) by Scott, 1994), other useful genes might come under the influence of the regulator. Useful constellations of targets would be retained, along with neutral targets that might constitute working material for further evolution. Having multiple target genes under the control of one regulatory protein would then lock in the structure of both regulator and target, as neither could change without simultaneous compensatory changes in multiple other genes.

These constraints might be expected to stifle the molecular coevolution of interactions where regulatory proteins bind to multiple target genes. The *D. melanogaster* *bcd* protein is known to be directly involved in the regulation of a number of other genes, including *eve* (Small *et al.*, 1991; Small *et al.*, 1993), *Kr* (Hülkamp *et al.*, 1990; Hoch *et al.*, 1991,1992), *tailless* (Liaw and Lengyel, 1992) and *orthodenticle* (Finkelstein and Perrimon, 1990). However, as the analysis of the *M. domestica* *hb* upstream DNA shows, despite being one of many *bcd* target genes, the precise nature of the *bcd* binding sites within the putative *hb* promoter is different.

It is conceivable that the tolerance in the developmental program that has permitted the accumulation of these changes might also permit an altered specificity of the regulatory protein to drive the accumulation of compensatory coevolutionary changes in the regulatory regions of other target genes. Alternatively, where promoters are assembled from large numbers of low affinity binding sites for their regulatory

proteins, such target genes could remain unaffected by coevolutionary changes developed in interactions more stringently regulated by the same regulatory protein.

Subtle alterations in binding specificity may not necessarily affect the affinity of a regulatory protein for what are already low affinity binding sites. For example, the Ubx homeodomain binds optimally to the sequence T-T-A-A-T-G>T-G>A-C-C, whilst that of Dfd binds optimally to T/C-T-A-A-T-G>T-A>G-A-C. As the TAAT core plays the major role in determining the affinity of homeodomain-DNA binding (Ekker *et al.*, 1991; see Chapter 4), this represents a fairly subtle difference in binding specificity between the two proteins. This subtle difference results in appreciable differences in affinity for each others' optimal site. Whilst Ubx binds to its optimal site sequence of TTAATGGCC with a dissociation rate constant (k_D) of $-0.89 \times 10^{-2} \text{ min}^{-1}$, Dfd binds to the same site with a k_D of $-2.4 \times 10^{-2} \text{ min}^{-1}$ - which translates into protein-DNA complex half-lives of 78 and 29 minutes respectively. In contrast, both proteins bind to the low affinity site TTAATCGCT with similar affinity (k_D of $-5.4 \times 10^{-2} \text{ min}^{-1}$ for Ubx, and $-5.0 \times 10^{-2} \text{ min}^{-1}$ for Dfd, with half-lives of 13 and 14 minutes respectively; Ekker *et al.*, 1992). If many promoters contain larger numbers of lower affinity sites for their regulatory proteins, rather than a few very high affinity sites, this would suggest that the constraints due to interacting with multiple genes might be less imposing than previously suggested.

Therefore, it would be interesting to examine in the interactions between bcd, hb and some of their downstream target genes, to see if the changes seen in the putative bcd-binding sites upstream of the *M. domestica hb* coding sequence are mirrored in the regulatory regions of these others. A prime candidate for examination in this respect would be the gap gene *Krüppel (Kr)*.

THE BCD-HB-KR INTERACTION

Kr is expressed in a central domain of the early embryo (Knipple *et al.*, 1987; Gaul *et al.*, 1987). Both bcd and low concentrations of hb activate *Kr* expression by

binding to multiple sites within the *Kr* promoter, such that the posterior border of the *Kr* domain is initially set by limiting *bcd* and *hb* concentrations (Hülskamp *et al.*, 1990; Hoch *et al.*, 1991). The anterior border of the *Kr* domain is controlled by repression by several *bcd*-dependent gap proteins, including *knirps* (*kni*), *tailless* (*tll*), and high concentrations of *hb* (Gaul and Jäckle, 1987; Hoch *et al.*, 1992, Eldon and Pirrotta, 1991; Capovilla *et al.*, 1992; Kraut and Levine, 1991).

As described in Chapter 1, the activation of *Kr* expression is both functionally redundant (both *bcd* and *hb* are capable of driving *Kr* expression in a central domain) and genetically redundant (two separate elements are each capable of promoting the correct expression of *Kr*). Furthermore, each promoter element contains multiple regulatory protein binding sites - competition between activators and repressors within the *Kr* promoter represents a 'multi layered backup system' (Hoch *et al.*, 1992) that could provide a flexible and redundant system within which molecular coevolution could take hold.

Given the apparent fine-grained differences in the *bcd-hb* interaction between *D. melanogaster* and *M. domestica*, has the precise molecular basis of the interaction between the *Kr* promoter and *bcd* and *hb* also diverged? The *bcd* binding sites within the *Kr* promoter are of lower affinity than the *bcd* sites in the *D. melanogaster hb* promoter (Hoch *et al.*, 1991). It seems possible that this could confer a degree of tolerance for an evolving *bcd* protein, as subtle changes in the specificity of the *bcd* protein might not significantly alter its overall affinity for a set of these low affinity sites. A change in *bcd* specificity from TCTAATCC to TTTAATCC might result in a slightly increased affinity for some sites, and decreased binding to others, but would be unlikely to result in a changed affinity for a sequence such as the low affinity AATAATCC *bcd*-binding site (Hoch *et al.*, 1991) found within the *D. melanogaster Kr* promoter. This could allow coevolution to proceed between *bcd* and *hb*, without being constrained by the requirement for similar changes within the *bcd-Kr* interaction.

Molecular coevolution and the evolution of developmental genes.

Development is regulated through local, combinatorial interactions between independent genetic elements. These elements are often modular and redundant, as a consequence conferring properties of flexibility and tolerance on the developmental program. As a consequence of this tolerance the action of natural selection and genomic processes of mutation and turnover may drive the molecular coevolution of interacting molecules. Constraints imposed by the requirement of a regulatory protein to interact with multiple target genes may be reduced where functionally equivalent promoters are composed of many low affinity sites rather than small numbers of high affinity sites, allowing the coevolution of one interaction without necessarily forcing simultaneous changes in other target genes of the same regulatory protein.

Caveats regarding the assumption of an interaction between *bcd* and *hb* in *M. domestica* and the integrity of the *hb* λ clone insert notwithstanding, comparison of *bcd* and *hb* between *M. domestica* and *D. melanogaster* would suggest that, whilst the logical outcome of the interaction (the *hb* expression domain) is conserved in both species, the precise molecular basis by which this interaction is managed may have diverged, as the nature of the *bcd* binding sites appears to be different in the *hb* promoters of the two species. Whilst the functional significance of the observed changes in the *bcd* protein and the *bcd* binding site sequences in the *hb* gene has yet to be fully analysed, coevolutionary change in this interaction remains a possibility.

REFERENCES

- Akam, M.** (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**, 1-22.
- Akam, M.** (1989). *Hox* and *HOM*: homologous gene clusters in insects and vertebrates. *Cell* **57**, 347-349.
- Akam, M.** (1994). Is pairing the rule? *Nature* **367**, 410-411
- Akam, M., Averof, M., Castelli-Gair, J., Dawes, F., Falciani, F., and Ferrier, D.** (1994). The evolving role of Hox genes in arthropods. *Development (Supplement)* 209-215.
- Appel, B. and Sakonju, S.** (1993). Cell-type specific mechanisms of transcriptional repression by the homeotic gene products of Ubx and Abd-A in *Drosophila* embryos. *EMBO J.* **12**, 1099-1109.
- Arnheim, N.** (1983) Concerted evolution of multigene families. In *Evolution of Genes and Proteins* (Nei, M. and Koehn, R.K., eds.), pp. 38-61, Sinauer, Sunderland, U.S.A.
- Beachy, P.A., Varkey, J. Young, K.E., von Kessler, D., Sun, B.I. and Ekker, S.C.** (1993). Cooperative binding of an *Ultrabithorax* homeodomain protein to nearby and distant DNA sites. *Mol. Cell. Biol.* **13**, 6941-6956.
- Bell, S.P., Piknard, C.S., Reeder, R.H. and Tijan, R.** (1989). Molecular mechanisms governing species-specific transcription of rRNA. *Cell* **38**, 623-624.

- Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C.** (1988). The role of localisation of bicoid mRNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* **7**, 1749-1756.
- Bermingham, J.R. Jr, and Scott, M.P.** (1988). Developmentally regulated alternative splicing of transcripts from the *Drosophila* homeotic gene *Antennapedia* can produce four different proteins. *EMBO J.* **7**, 3211-3222
- Bender, M., Horikami, S., Cribbs, D., and Kaufman, T.C.** (1988). Identification and expression of the gap segmentation gene *hunchback* in *Drosophila melanogaster*. *Dev. Genet.* **9**, 715-732
- Beverley, S.M. and Wilson, A.C.** (1984). Molecular evolution in *Drosophila* and the higher Diptera II: A time scale for fly evolution. *J. Mol. Evol.* **21**, 1-13.
- Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K.** (1993). Determination of the NMR solution structure of an *Antennapedia* homeodomain-DNA complex. *J. Mol. Biol.* **234**, 1084-1093.
- Brent, S.J.** (1982). Regulation and autoregulation by the *lexA* protein. *Biochemie* **64**, 565-569.
- Brent, S.J. and Ptashne, M.** (1981). Mechanism of action of the *lexA* gene product. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4204-4208.
- Brönner, G. and Jäckle, H.** (1991). Control and function of terminal gap gene activity in the posterior pole region of the *Drosophila* embryo. *Mech. Dev.* **35**, 205-211.
- Brown, S.J., Hilgenfeld, R.B., and Denell, R.E.** (1995). The beetle *Tribolium castaneum* has a *fushi tarazu* homolog expressed in stripes during segmentation. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12922-12926.

Buluwela, L., Forster, A., Boehm, T. and Rabbitts, T.H. (1989). A rapid procedure for colony screening using nylon filters. *Nucleic Acids Res.* **17**, 452.

Burke, W.D. and Eickbush, T.H. (1986). The Silkmoth Late Chorion Locus.

I. Variation within two paired multigene families. *J. Mol. Biol.* **190**, 343-356.

Capovilla, M., Eldon, E. and Pirrotta, V. (1992). The *giant* gene of *Drosophila* encodes a b-ZIP DNA binding protein that regulates the expression of other segmentation gap genes. *Development* **114**, 99-112.

Carroll, S.B. (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479-485.

Casanova, J. (1990). Pattern formation under the control of the terminal system in the *Drosophila* embryo. *Development* **110**, 621-628.

Catron, M.K., Iler, N. and Abate, C. (1993). Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins. *Mol. Cell. Biol.* **13**, 2354-2365.

Chomczynski, P. and Sacchi, N. (1987). Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Chung, C.T. and Miller, R.H. (1988). A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Res.* **16**, 3580.

Cohen, S.M. and Jürgens, G. (1990). Mediation of *Drosophila* head development by gap-like segmentation genes. *Nature* **349**, 482-485.

- Curtis, D.** (1994). Translational repression as a conserved mechanism for the regulation of embryonic polarity. *BioEssays* **16**, 709-711.
- Dawes, R., Dawson, I., Falciani, F., Tear, G. and Akam, M.** (1994). *Dax*, a locust *hox* gene related to *fushi-tarazu* but showing no pair-rule expression. *Development* **120**, 1561-1572.
- Dickinson, W.J.** (1991). The evolution of regulatory genes and patterns in *Drosophila*. In *Evolutionary Biology (Vol 25)*. Hecht, M.K., Wallace, B. and MacIntyre, R.J. eds pp. 127-173. Plenum Press.
- Dover, G.A.** (1982). Molecular drive: a cohesive mode of species evolution. *Nature* **299**, 111-117.
- Dover, G.A.** (1989). Slips, strings and species. *Trends Genet.* **5**, 100-102
- Dover, G.A.** (1992). Observing development through evolutionary eyes: A practical approach. *BioEssays* **14**, 281-287
- Dover, G.A. and Flavell, R.B.** (1984). Molecular coevolution: DNA divergence and maintenance of function. *Cell* **38**, 622-623
- Dover, G.A. and Tautz, D.** (1986). Conservation and divergence in multigene families: alternative to selection and drift. *Phil. Trans. Roy. Soc. ser B*, **312**, 275-289.
- Dranginis, A.M.** (1990). The binding of yeast a1 and a2 as a heterodimer to the operator DNA of a haploid-specific gene. *Nature* **347**, 682-685.
- Dreiver, W., and Nüsslein-Volhard, C.** (1988a). A gradient of the *bicoid* protein in *Drosophila* embryos. *Cell* **54**, 83-93.

Driever, W. and Nüsslein-Volhard, C. (1988b). 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). bicoid protein, a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature* **337**, 138-143.

Driever, W., Thoma, G., and Nüsslein-Volhard, C. (1989a). 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

Driever, W., Ma, J., Nüsslein-Volhard, C., and Ptashne, M. (1989b). Rescue of *bicoid* mutant *Drosophila* embryos by Bicoid fusion proteins containing heterologous activating sequences. *Nature* **342**, 149-154.

Eggleston, P. (1992). Identification of the *abdominal-A* homologue from *Aedes aegypti* and structural comparisons among related genes. *Nucleic Acids Research* **20**, 4095-4096.

Eickbush, T.H. and Burke, W.D. (1986). The Silkworm Late Chorion Locus.

II. Gradients of gene conversion in two paired multigene families. *J. Mol. Evol.* **190**, 357-366.

Ekker, S.C., Young, K.E., von Kessler, D.P. and Beachy, P.A. (1991). Optimal DNA sequence recognition by the *Ultrabithorax* homeodomain of *Drosophila*. *EMBO J.* **10**, 1170-1186.

Ekker, S.C., von Kessler, D.P. and Beachy, P.A. (1992). Differential sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J.* **11**, 4059-4072.

- Ekker, S.C., Jackson, D.G., von Kessler, D.P., Sun, I.B., Young, K.E. and Beachy, P.A.** (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins, *EMBO J.* **13**, 3551-3560.
- Eldon, E.D. and Pirrotta, V.** (1991). Interactions of the *Drosophila* gap gene *giant* with maternal and zygotic pattern-forming genes. *Development* **111**, 367-378.
- Evans, T.C., Crittenden, S.L., Kodoyianni, V. and Kimble, J.** (1994). Translational control of maternal *glp-1* mRNA establishes an asymmetry in the *C. elegans* embryo. *Cell* **77**, 183-194.
- Falb, D. and Maniatis, T.** (1992). A conserved regulatory unit implicated in tissue-specific gene expression in *Drosophila* and man. *Genes Dev.* **6**, 454-465.
- Federoff, N.V.** (1979). On spacers. *Cell* **16**, 697-710
- Ferguson, E.L., and Horvitz, H.R.** (1989). The multivulva phenotype of certain *Caenorhabditis elegans* mutants results from defects in two functionally redundant pathways. *Genetics* **123**, 109-121.
- Ferrandon, D., Elphick, L., Nusslein-Volhard, C. and St Johnston, D.** (1994). Stauf protein associates with the 3' UTR of *bicoid* messenger RNA to form particles that move in a microtubule dependant manner. *Cell* **79**, 1221-1232.
- Finkelstein, R. and Perrimon, V.** (1991). Molecular genetics of head development in *Drosophila melanogaster*. *Development* **112**, 899-912.
- Frohman, M.A., Dush, M.K. and Martin, G.R.** (1988). Rapid production of full length cDNAs from rare transcripts: Amplification using a single gene-specific primer. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998-9002.

- Frohnhofer, 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.
- Frigerio, D., 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.
- Furukubo-Tokunaga, K., Flister, S. and Gehring, W.J.** (1993). Functional specificity of the *Antennapedia* homeodomain. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6360-6364.
- Gaul, U. and Jackle, H.** (1987). Pole region-dependent repression of the *Drosophila* gap gene *Krüppel* by maternal gene products. *Cell* **51**, 549-555.
- Gaul, U., Seifert, E., Schuh, K. and Jäckle, H.** (1987). Analysis of *Krüppel* protein distribution during early *Drosophila* development reveals post-transcriptional regulation. *Cell* **50**, 637-647.
- Gerber, H.-P., Seipel, K., Georgier, O., Höfferer, M., Hug, M., Rusconi, S. and Schaffner, W.** (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**, 808-811.
- Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G. and Wüthrich, K.** (1994). Homeodomain-DNA recognition. *Cell* **78**, 211-223.
- Genetics Computer Group** (1991). Program manual for the GCG package, Version 7. 575 Science Drive, Madison, Wisconsin. USA 53711
- Gottlieb, E.** (1992). The 3' untranslated region of localised maternal messages contains a conserved motif involved in mRNA localisation. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7164-7168.

- Gould, S.J.** (1977). *Ontogeny and Phylogeny*. Cambridge, M.A: Harvard University Press.
- Gould, S.J.** (1992). Ontogeny and phylogeny - revisited and reunited. *BioEssays* **14**, 275-279.
- Goutte, C. and Johnson, A.D.** (1993). yeast a1 and a2 homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein. *J. Mol. Biol.* **233**, 359-371.
- Grummt, I., Roth, E. and Paule, M.R.R.** (1982). Ribosomal RNA transcription *in vitro* is species-specific. *Nature* **296**, 173-176.
- 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.
- Hanes, S.D. and Brent, R.** (1989). DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell* **57**, 1275-1283.
- Hanes, S.D. and Brent, R.** (1991). A genetic model for interaction of the homeodomain recognition helix with DNA. *Science* **251**, 426-430.
- Hanes, S.D., Riddihough, G., Ish-Horowicz, D. and Brent, R.** (1994). Specific DNA recognition and intersite spacing are critical for action of the bcd morphogen. *Mol. Cell. Biol.* **14**, 3364-3375.
- Heberlein, U. and Rubin, G.M.** (1990). Structural and functional comparisons of the *Drosophila virilis* and *Drosophila melanogaster rough* genes. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5916-5920.

- Henikoff, S.** (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-359.
- Hoch, M., Schröder, C., Seifert, E. and Jäckle, H.** (1990). *Cis*-acting control elements for *Krüppel* expression in the *Drosophila* embryo. *EMBO J.* **9**, 2587-2595.
- Hoch, M., Seifert, E. and Jackle, H.** (1991). Gene expression mediated by *cis*-regulatory sequences of the *Krüppel* gene in response to the anterior morphogens *bicoid* and *hunchback*. *EMBO J.* **10**, 2267-2278.
- Hoch, M., Gerwin, N., Taubert, H. and Jackle, H.** (1992). Competition for overlapping sites in the regulatory region of the *Drosophila* gene *Krüppel*. *Science* **256**, 94-97.
- Holland, P.W.H., Holland, L.Z., Williams, N.A. and Holland, N.D.** (1992). Amphioxus homeobox genes - sequence conservation, spatial expression during development and insights into vertebrate evolution. *Development* **116**, 653-661.
- Hooper, J.E., Pérez-Alonso, M., Bermingham, J.R., Prout, M., Rocklein, B.A., Wagenbach, M., Edstrom, J.-E., de Frutos, R. and Scott, M.P.** (1992). Comparative studies of *Drosophila Antennapedia* genes. *Genetics* **132**, 453-469.
- Hülskamp, M., Schröder, C., Pfeifle, C., Jäckle, H., and Tautz, D.** (1989). Posterior segmentation of the *Drosophila* embryo in the absence of a maternal posterior organizer gene. *Nature* **338**, 629-632.
- Hülskamp, M., Pfeifle, C., and Tautz, D.** (1990). A morphogenetic gradient of *hunchback* protein organizes the expression of the gap genes *Krüppel* and *knirps* in the early *Drosophila* embryo. *Nature* **346**, 577-580.
- Ingham, P.** (1988). The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* **318**, 433-445.

- Irish, V., Lehmann, R., and Akam, M.** (1989). The *Drosophila* posterior-group gene *nanos* functions by repressing *hunchback* activity. *Nature*, **338**, 646-648.
- Irvine, K.D., Helfand, S.L. and Hogness, D.S.** (1991). The large upstream control region of the *Drosophila* homeotic gene *Ultrabithorax*. *Development* **111**, 407-424.
- Jantzen, H.M., Admon, A., Bell, S.P. and Tijan, R.** (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins.
- Jefferys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neill, D.L., and Armour, J.A.L.** (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics* **6**, 136-145.
- Johnson, A.D.** (1992). A combinatorial regulatory circuit in budding yeast. In: *Transcriptional regulation*, S.L. McKnight and K.R. Yamamoto, eds. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp. 975-1006.
- Jones, C.W. and Kafatos, F.C.** (1982). Accepted mutations in a gene family: Evolutionary diversification of duplicated DNA. *J. Mol. Evol.* **18**, 87-103.
- Joyner, A.L., Herrup, K., Auerbach, B.A., Davis, C.A. and Rossant, J.** (1991). Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the *En-2* homeobox. *Science* **251**, 1239-1243.
- Kassis, J.A.** (1990). Spatial and temporal control elements of the *Drosophila engrailed* gene. *Genes Dev.* **4**, 433-443.
- Kassis, J., Poole, S., Wright, D., and O'Farrell, P.** (1986). Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. *EMBO J.* **5**, 3583-3589.

- Kassis, J.A., Desplan, C., Wright, D.K. and O'Farrel, P.H.** (1989). Evolutionary conservation of homeo-domain binding sites and other sequences upstream and within the major transcription unit of the *Drosophila* segmentation gene *en*. *Mol. Cell. Biol.* **9**, 4304-4311.
- Keleher, C.A., Goutte, C. and Johnson, A.D.** (1988). The yeast cell-type-specific repressor $a2$ acts cooperatively with a non-cell-type-specific protein. *Cell* **53**, 927-935.
- Kenyon, C.** (1994). If birds can fly, why can't we? Homeotic genes and evolution. *Cell* **78**, 175-180.
- Kimble, J.** (1994). An ancient molecular mechanism for establishing embryonic polarity. *Science* **266**, 577-578.
- Kissinger, C.R., Liu, B., Martin-Blance, E., Kornberg, T.B. and Pabo, C.O.** (1990). Crystal structure of an *engrailed* homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-590.
- Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O.** (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21-32.
- Klemenez, 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**, 3947-3958.
- Knipple, D.C., Seifert, E., Rosenberg, U.B., Preiss, A., and Jäckle, H.** (1985). Spatial and temporal patterns of *Krüppel* gene expression in early *Drosophila* embryos. *Nature* **317**, 40-44.

- Kraut, R. and Levine, M.** (1991). Mutually repressive interactions between gap genes *giant* and *Krüppel* define middle body regions of the *Drosophila* embryo. *Development* **111**, 611-622.
- Krumlauf, R.** (1992). Evolution of the vertebrate Hox homeobox genes. *BioEssays* **14**, 245-252.
- Krumlauf, R.** (1994). Hox genes in vertebrate development. *Cell* **78**, 191-201.
- Laemmli, U.K.** (1975). Most commonly used discontinuous buffer system for SDS electrophoresis. *Nature* **277**, 680.
- Langeland, J.A. and Carroll, S.B.** (1993). Conservation of regulatory elements controlling pair-rule stripe formation. *Development* **117**, 585-596.
- Laurent, B.C. and Carlson, M.** (1992). Yeast SNF2/SWI2, SNF5 and SNF6 proteins function coordinately with the gene-specific transcriptional activators GAL4 and bicoid. *Genes Dev.* **6**, 1707-1715.
- Lawrence, P.** (1992). *The Making of a Fly*. Blackwell.
- Lehmann, R., and Nüsslein-Volhard, C.** (1987). *hunchback*, a gene required for segmentation of an anterior and posterior region of the *Drosophila* embryo. *Dev. Biol.* **119**, 402-417.
- Levinson, G. and Gutmann, G.A.** (1987). Slipped-strand mispairing- a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203-221.
- Lewis, E.** (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565-570.

- Li, X. and Noll, M.** (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature*, **367**, 83-87.
- Liaw G-J. and Lengyel, J.A.** (1992). Control of *tailless* expression by *bicoid*, *dorsal* and synergistically interacting terminal system regulatory elements. *Mech. Dev.* **40**, 47-61.
- Long, E.O. and Dawid, I.B.** (1980). Repeated genes in eukaryotes. *Annu. Rev. Biochem.* **49**, 727-764
- Luk, S.K.S., Kilpatrick, M., Kerr, K. and Macdonald, P.M.** (1994). Components acting in the localization of bicoid messenger-RNA are conserved among *Drosophila* species. *Genetics* **137**, 521-530.
- Lukowitz, W., Schröder, C., Glaser, G., Hülskamp, M., and Tautz, D.** (1994). Regulatory and coding regions of the segmentation gene *hunchback* are functionally conserved between *Drosophila virilis* and *Drosophila melanogaster*. *Mech. Dev.* **45**, 105-115.
- MacDonald, P.M.**, (1990). *bicoid* mRNA localization signal: phylogenetic conservation of function and RNA secondary structure. *Development* **110**, 161-171.
- MacDonald, P.M., and Struhl, G.**, (1988). Cis-acting sequences responsible for the anterior localization of *bicoid* mRNA in *Drosophila* embryos. *Nature* **336**, 595-598.
- MacDonald, P.M., Kerr, K., Smith, J.L., and Leask, A.**, (1993). RNA regulatory element BLE1 directs the early steps of *bicoid* mRNA localisation. *Development* **118**, 1233-1243.
- Maier, D., Sperlich, D and Powell, J.R.** (1993). Conservation and change of the developmentally crucial *fushi tarazu* gene in *Drosophila*. *J. Mol. Evol.* **36**, 315-326

- Maier, D., Preiss, A. and Powell, J.R.** (1990). Regulation of the segmentation gene *fushi tarazu* has been functionally conserved in *Drosophila*. *EMBO J.* **9**, 3957-3966.
- Malicki, J., Schughart, K. and McGinnis, W.** (1990). Mouse *Hox 2.2* specifies thoracic segmental identity in *Drosophila* embryos and larvae. *Cell* **63**, 961-967.
- Malicki, J., Cianetti, L.C., Peschle, C. and McGinnis, W.** (1992). A human *HOX4B* regulatory element provides head-specific expression in *Drosophila* embryos. *Nature* **358**, 345-347.
- Margolis, J.S., Borowsky, M.L., Wha Shim, C., and Posakony, J.W.,** (1994). A small region surrounding the distal promoter of the *hunchback* gene directs maternal expression. *Dev. Biol.* **163**, 381-388.
- Margolis, J.S., Borowsky, M.L., Steingrímsson, E., Wha Shim, C., Lengyel, J.A. and Posakony, J.W.** (1995). Posterior stripe expression of *hunchback* is driven from two promoters by a common enhancer element. *Development* **121**, 3067-3077.
- Mitchell, P.J. and Tijan, R.** (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371.
- Mitsialis, S.A. and Kafatos, F.C.** (1985). Regulatory elements controlling chorion gene expression are conserved between flies and moths. *Nature* **317**, 453-546.
- Mlodzik, M., Fjofe, A. and Gehring, W.** (1987). Molecular structure and spatial expression of a homeobox gene from the *labial* region of the Antennapedia-complex. *EMBO J.* **7**, 2569-2578.
- Moss, T., Mitchelson, K. and de Winter, R.** (1985). The promotion of ribosomal transcription in eukaryotes. In *Oxford Surveys of Eukaryotic Genes* (Mifflin, B., ed.) vol. 2, pp. 207-250, Oxford University Press, Oxford.

- Mosquera, L., Forristall, C., Zhou, Y. and King, M.L.** (1993). An mRNA localised to the vegetal cortex of *Xenopus* oocytes encodes a protein with a *nanos*-like zinc finger domain. *Development* **117**, 377-386.
- Müller, J. and Bienz, M.** (1991). Long range repression conferring boundaries of *Ultrabithorax* expression in the *Drosophila* embryo. *EMBO J.* **10**, 3147-3155.
- Nagy, L.M., Booker, R. and Riddiford, L.M.** (1991). Isolation and embryonic expression of an *abdominal-A*-like gene from the lepidopteran, *Manduca sexta*. *Development* **107**, 201-212.
- Needleman, S.B. and Wunsch, C.D.** (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443-453.
- Oakeshott, J.G., Boyce, T.M., Russell, R.J. and Healy, M.J.** (1995). Molecular insights into the evolution of an enzyme; esterase6 in *Drosophila*. *Trends Ecol. Evol.* **10**, 103-110.
- O'Connor M.B., Binari, R., Perkins, L.A. and Bender, W.** (1988). Alternative RNA products from the *Ultrabithorax* domain of the bithorax complex. *EMBO J.* **7**, 435-455.
- Ohno, S.** (1985). Dispensable genes. *Trends Genet.* **1**, 160-164.
- Ohta, T and Dover, G.A.** (1984). The cohesive population genetics of molecular drive. *Genetics* **108**, 501-521.
- Otting, G., Qian, Y.Q., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K.** (1988). Secondary structure determination for the Antennapedia homeodomain

by nuclear magnetic resonance and evidence for a helix-turn-helix motif. *EMBO J.* **7**, 4305-4309.

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085-3092.

Pabo, C. O. and Sauer, R.T. (1984). Protein-DNA recognition. *Annu. Rev. Biochem.* **53**, 293-321.

Patel, N.H. (1994). Developmental evolution: insights from studies of insect segmentation. *Science* **266**, 581-590.

Patel, N.H., Ball, E.E. and Goodman, C.S. (1992). Changing role of *even-skipped* during the evolution of insect pattern formation. *Nature* **357**, 339-342.

Patel, N.H., Condrón, B.G. and Zinn, K. (1994). Pair-rule expression patterns of *even-skipped* are found in both short and long germ beetles. *Nature* **367**, 429-434.

Passmore, S., Elble, R. and Tye, B.-R. (1989). A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes. *Genes Dev.* **3**, 921-935.

Peifer, M. and Weischaus, E. (1993). The product of the *Drosophila melanogaster* segment polarity gene *armadillo* is highly conserved in sequence and expression in the housefly *Musca domestica*. *J. Mol. Evol.* **36**, 224-233.

Percival-Smith, A., Müller, M., Affolter, M. and Gehring, W.J. (1990). The interaction with DNA of wildtype and mutant *fushi tarazu* homeodomains. *EMBO J.* **9**, 3967-3974. Corrigendum. *EMBO J.* **11**, 382.

- Peixoto, A., Campesan, S., Costa, R., Kyriacou, C.P.** (1993). Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. *Mol. Biol. Evol.* **10**, 127-139.
- Perler, F., Efstratiadis, A., Lomedica, P., Gilbert, W., Kolochner, R. and Dodgson, J.** (1990). The evolution of genes: the chicken preproinsulin gene. *Cell* **20**, 555-566.
- Puissant and Houdebine** (1990). *Biotechniques* **8**, 148-149.
- Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J. and Wüthrich, K.** (1990). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell* **59**, 573-580.
- Quiring, R., Walldorf, U., Kloter, U. and Gehring, W.J.** (1994). Homology of the *eyeless* gene of *Drosophila* to the *Small eye* gene in mice and *Aniridia* in humans. *Science* **265**, 785-789.
- Reeder, R.H.** (1984). Enhancers and ribosomal gene spacers. *Cell* **38**, 349-351.
- Revzin, A.** (1989). Gel electrophoresis assays for DNA-protein interactions. *BioTechniques* **7**, 346-355.
- Richsteiner, M.** (1987). Do *myc*, *fos* and E1A function as protein phosphatase inhibitors? *Biochem. Biophys. Res. Comm.* **143**, 194-198.
- Riddihough, G., and Ish-Horowicz D.** (1991). Individual stripe regulatory elements in the *Drosophila hairy* promoter respond to maternal, gap and pair-rule genes. *Genes Dev.* **5**, 840-854.

- 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**, 461-470.
- Rodgers, S., Wells, R., and Richsteiner M.**, (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**, 364-368.
- Ronchi, E., Treisman, J., Dostatni, N., Struhl, G. and Desplan, C.** (1993). Down-regulation of the *Drosophila* morphogen bicoid by the torso receptor-mediated signal transduction cascade. *Cell* **74**, 347-355.
- Rubin, G.M. and Spradling, A.C.** (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348-353.
- Salser, S.J. and Kenyon, C.** (1994). Patterning *C. elegans*: homeotic cluster genes, cell fates and cell migrations. *Trends. Genet.* **10**, 159-164.
- Sambrook, J., Fritsch, E.F. and Maniatis, T.** (1989). *Molecular cloning: a laboratory manual* (2nd edition). Cold Spring Harbor, New York: Cold Spring Harbor Press.
- Sanger, F., Nicklen, S. and Coulson, A.R.** (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.
- Schier, A.F. and Gehring, W.J.** (1992). Direct homeodomain-DNA interaction in the autoregulation of the *fushi-tarazu* gene. *Nature* **356**, 804-807.
- Schier, A.F. and Gehring, W.J.** (1993a). Functional specificity of the homeodomain protein *fushi tarazu*: the role of DNA binding specificity in vivo. *Proc. Natl. Acad. Sci. USA* **90**, 1450-1454.

- Schier, A.F. and Gehring, W.J.** (1993b). Analysis of the *fushi tarazu* autoregulatory element: multiple sequence elements contribute to enhancer activity. *EMBO J.* **12**, 1111-1119.
- Schröder, C., Tautz, D., Seifert, E., and Jäckle, H.** (1988). Differential regulation of the two transcripts from the *Drosophila* gap segmentation gene *hunchback*. *EMBO J.* **7**, 2882-2887.
- Schröder, R., and Sander, K.,** (1993). A comparison of transplantable *bicoid* activity and partial *bicoid* homeobox sequences in several *Drosophila* and blowfly species (*Calliphoridae*). *Roux's Arch. Dev. Biol.* (1993). **203**, 34-43.
- Schulz, C. and Tautz, D.** (1994). Autonomous concentration-dependent activation and repression of *Krüppel* by *hunchback* in the *Drosophila* embryo. *Development* **120**, 3043-3049.
- Scott, M.P.** (1994). Intimations of a creature. *Cell* **79**, 1121-1124.
- Seeger, M.A., and Kaufman, T.C.** (1990). Molecular analysis of the *bicoid* gene from *Drosophila pseudoobscura*: identification of conserved domains within coding and non-coding regions of the *bicoid* mRNA. *EMBO J.* **9**, 2977-2987.
- Shankland, M.** (1994). Leech segmentation: a molecular perspective. *BioEssays* **16**, 801-808.
- Simpson-Brose, M., Treisman, J. and Desplan, C.** (1994). Synergy between the *hunchback* and *bicoid* morphogens is required for anterior patterning in *Drosophila*. *Cell* **78**, 855-865.
- Slack, J.M.W., Holland, P.W.H., and Graham, C.F.** (1993). The zootype and the phylotypic stage. *Nature* **361**, 490-492

- Small, S., Kraut, R., Hoey, T., Warrior, R., and Levine, M.** (1991) Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* **5**, 827-839.
- Small, S., Blair, A. and Levine, M.** (1992). Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J.* **11**, 4047-4057.
- Smith, D.L. and Johnson, A.D.** (1994). Operator-constitutive mutations in a DNA sequence recognized by a yeast homeodomain. *EMBO J.* **13**, 2378-2387.
- Sommer, R.** (1992). Vergleichende Analyse von Segmentierungsgenen bei Insecten. Ph.D. thesis. Ludwig-Maximilians-Universität München.
- Sommer, R., and Tautz, D.** (1991). Segmentation gene expression in the housefly *Musca domestica*. *Development* **113**, 419-430.
- Sommer, R., and Tautz, D.** (1993) Involvement of an orthologue of the *Drosophila* pair-rule gene hairy in segment formation of the short germ-band embryo *Tribolium* (Coleoptera). *Nature* **361**, 448-450.
- Sommer, R., Retzlaff, M., Goerlich, K., Sander, K., and Tautz, D.** (1992). Evolutionary conservation pattern of zinc-finger domains of *Drosophila* segmentation genes. *Proc. Natl. Acad. Sci. USA* **89**, 10782-10786.
- Southern, E.M.** (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503.
- Stanojevic, D., Hoey, T. and Levine, M.** (1989). Sequence-specific DNA-binding activities of the gap proteins encoded by *hunchback* and *Krüppel* in *Drosophila*. *Nature* **341**, 331-335.

- Stevenson, E.C., Chao, Y., and Fackenthal, J.** (1988). Molecular analysis of the *swallow* gene of *Drosophila melanogaster*. *Genes. Dev.* **2**, 1655-1665.
- St. Johnston, D., Driever, W., Berleth, T., Richstein, S. and Nüsslein-Volhard, C.** (1989). Multiple steps in the localisation of *bicoid* mRNA to the anterior pole of the *Drosophila* oocyte. *Development* **107** (Suppl.), 13-19.
- St. Johnston, D. and Nusslein-Volhard, C.** (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201-219.
- Stone, R.S., Hughes, M.J. and Jost, J-P.** (1991). Studies of protein-DNA interactions by gel mobility-shift assay. *BioMethods*, Vol. 5, pp. 163-184.
- Stroeher, V.L., Gaiser, C. and Garber, R.L.** (1988). Alternative RNA splicing that is spatially regulate: generation of transcripts from the *Antennapedia* gene of *Drosophila melanogaster* with different protein coding regions. *Mol. Cell. Biol.* **8**, 4143-4154.
- Struhl, G.** (1989). Differing strategies for organizing anterior and posterior body pattern in *Drosophila* embryos. *Nature* **338**, 741-744.
- Struhl, G., Struhl, K., and Macdonald, P.M.** (1989). The gradient morphogen *bicoid* is a concentration-dependent transcriptional activator. *Cell* **57**, 1259-1273.
- Stuart, J.J., Brown, S.J., Beeman, R.W. and Denell, R.E.** (1993). The *Tribolium* homeotic gene *Abdominal* is homologous to *abdominal-A* of the *Drosophila* homeotic complex. *Development* **117**, 233-243.
- Tamkin, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C. and Kennison, J.A.** (1992). *brahma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**, 561-572.

- Tautz, D., and Pfeifle C.** (1989). A non-radioactive in situ hybridization method for the localization of specific RNA's in *Drosophila* reveals translational control of the segmentation gene *hunchback*. *Chromosoma* **98**, 81-85.
- Tautz, D.** (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres. *Nature* **332**, 281-284.
- Tautz, D., and Sommer, R.** (1995). Evolution of segmentation genes in insects. *Trends. Genet.* **11**, 23-27.
- Tautz, D., Trick, M., and Dover, G.A.** (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322**, 652-656.
- Tautz, D., Hulskamp, M., and Sommer, R.J.** (1992). Whole mount *in situ* hybridisation in *Drosophila*. In: Wilkinson, D.G. *In situ hybridization, a practical approach*. IRL press, Oxford, pp. 61-73.
- Tautz, D., Lehmann, R., Schuh, R., Schnürch, H., Seifert, E., Keinlin, A., Jones, K., and Jäckle, H.** (1987). Finger protein of novel structure encoded by *hunchback*, a second member of the gap gene class of *Drosophila* segmentation genes. *Nature* **327**, 383-389.
- Tear, G., Akam, M. and Martinez-Arias, A.** (1990). Isolation of an *Abdominal-A* gene from the locust *Schistocerca gregaria* and its expression during embryogenesis. *Development* **110**, 915-925.
- Treier, M., Pfeifle, C., and Tautz, D.** (1989). Comparison of the gap gene *hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. *EMBO J.* **8**, 1517-1525.

- Treisman, J. and Desplan, C.** (1989). The products of the *Drosophila* gap genes *hunchback* and *Krüppel* bind to the *hunchback* promoters. *Nature* **341**, 335-337.
- Treisman, J., Gönczy, P., Vashishtua, M., Harms, E. and Desplan, C.** (1989). a single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* **59**, 553-562.
- Turing, A.** (1952). The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. (Lond.) B* **237**, 37-72.
- Ueno, K., Hui, C.-C., Fukuta, M. and Suzuki, Y.** (1992). Molecular analysis of the deletion mutants in the E-homeotic complex of the silkworm *Bombyx mori*. *Development* **114**, 555-563.
- Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., van Leeuwen, H.C., Strating, M.J.J. and van der Vliet, P.C.** (1992). The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.* **11**, 4993-5003.
- Walldorf, U., Fleig, R. and Gehring, W.J.** (1989). Comparison of homeobox containing genes of the honeybee and *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9971-9975.
- Wang, C. and Lehmann, R.** (1991). *Nanos* is the localized posterior determinant in *Drosophila*. *Cell* **66**, 637-647.
- Warren, R.W., Nagy, L., Selegue, J., Gates, J. and Gehring, W.** (1994). Evolution of homeotic gene regulation and function in flies and butterflies. *Nature*, **372**, 458-61.
- Wharton, K.A., Yedvobnick, B., Finnerty, V. and Artavanis-Tsakonas, S.** (1985). *opa*: a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D. melanogaster*. *Cell* **40**, 55-62.

- Wharton, R.P. and Struhl, G.** (1991), RNA regulatory elements mediate control of the *Drosophila* body pattern by the posterior morphogen *nanos*. *Cell* **67**, 955-967.
- Wilde, D.C., and Akam, M.** (1987) Conserved sequence elements in the 5' region of the *Ultrabithorax* transcription unit. *EMBO J.* **6**, 1393-1401.
- Wilson, A.C., Maxson, L.R. and Sarich, V.M.** (1974). Two types of molecular evolution. Evidence from studeis of interspecific hybridisation. *Proc. Natl. Acad. Sci. USA* **71**, 2843-2847.
- Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. and Desplan, C.** (1993). Cooperative dimerization of Paired class homeodomains on DNA. *Genes Dev.* **7**, 2120-2134.
- Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O.** (1991). Crystal structure of a *MATa2* homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**, 517-528.
- Wolpert, L.** (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology* **25**, 1-47.
- Zuker, C.S.** (1994). On the evolution of eyes: would you like it simple or compound? *Science* **265**, 742-743.