THE ISOLATION AND CHARACTERIZATION

OF A

P. ANGULOSUS HOMEOBOX.

Peter Lance Pfeffer

Thesis submitted in fulfillment of the requirements for the degree of DOCTOR of PHILOSOPHY in the Department of Biochemistry, Faculty of Science,
University of Cape Town.

Cape Town, January 1991.

CERTIFICATE OF SUPERVISOR

In terms of paragraph eight of "General regulations for the degree of Ph.D.", I, as supervisor of the candidate P.L.Pfeffer, certify that I approve of the incorporation into this thesis of material that has already been published or submitted for publication.

Professor C. von Holt

Professor in the Department of Biochemistry and
Director of the UCT-FRD Research Center for Molecular Biology.

ACKNOWLEDGEMENTS

I wish to thank

Professor Claus von Holt, my supervisor, for his excellent guidance, support and encouragement during the course of this thesis.

Rob Warren for his helpful suggestions and many fruitful discussions.

My colleagues in lab 402 for their moral support and joviality, making the lab a second home. Special thanks to Sylva Schwager and Madhu Chauhan.

Dr. W.McGinnis for the kind donation of the *Antennapedia* homeobox-containing plasmid p903G.

The Foundation for Research Development and the University of Cape Town Research Committee for their generous financial support.

My wife Cynthia, my mother and my brother for their continual enthusiasm and interest in my work.

INDEX

THE ISOLATION AND CHARACTERIZATION OF A SEA URCHIN HOMEOBOX

CONTENTS PAGE

SUMMARY

THE ISOLATION AND CHARACTERIZATION OF A *P.ANGULOSUS* HOMEOBOX

The aim of this thesis was to isolate and characterize a homeobox-containing gene of the South African sea urchin *Parechinus angulosus*. This was achieved by constructing a genomic library of several individuals and screening this library using a probe containing the *Antennapedia* homeobox. Eight clones were isolated and shown to represent different alleles of the same gene. One clone was sequenced, revealing a homeobox which was termed *PaHbox1*. This homeobox was compared to published homeobox sequences and shown to be a member of the *Antp* (*Hox*1.1) subclass (table 1.1). A splice donor site was identified 23 bp upstream of the homeobox and the observation confirmed by RNAase mapping.

PaHbox1 is situated in a genomic area showing a significantly higher degree of restriction fragment polymorphism than expected. This was shown by a statistical analysis which should be of general value in the interpretation of such polymorphisms.

The expression of *PaHbox1* was examined by RNAase protection assays and Northern blotting. Two distinct phases of expression were observed - during embryogenesis *PaHbox1* is expressed transiently at low levels in 11,5 hr mesenchyme blastula stage embryos (44 ± 8 transcripts per embryo) with levels 3-5 fold lower 2,5 hr before and after this stage. Expression is observed again at up to 160 fold higher levels in the adult with maximal expression in testis (11 transcripts per 10 pg total RNA), and increasingly lower levels in intestines, ovary and Aristotle's lantern. Two transcripts of size 5,2 and 5,7 kbp were observed. Expression in Aristotle's lantern and embryonic stages could not be detected by Northern analysis.

ABBREVIATIONS

a.a.	amino acid(s)
ATP	adenosine 5'-triphosphate
bp	base pair(s)
BSA	bovine serum albumin
cpm	counts per minute
DMSO	dimethyl sulphoxide
dNTP	deoxyribonucleotide
DTT	dithiothreitol
kbp	kilo base pairs
LB	Luria-Bertrani (medium)
mRNA	messenger RNA
pfu	plaque forming units
rpm	revolutions per minute
TCA	trichloroacetic acid
TE	10 mM Tris.HCl; 0,1 mM EDTA

Genes

Abd-X	*Abdominal-X*
Antp	*Antennapedia*
bcd	*bicoid*
ddl	*distal-less*
Dfd	*Deformed*
en	*engrailed*
eve	*even-skipped*
ftz	*fushi tarazu*
gsb	*gooseberry*
lab	*labial*
Oct	*Octamer*
prd	*paired*
Scr	*Sex combs reduced*
Ubx	*Ultrabithorax*

One- and three-letter codes for amino acids

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic acid	P	Pro	Proline
E	Glu	Glutamic acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

CHAPTER 1 LITERATURE REVIEW

THE STRUCTURE AND FUNCTION OF HOMEOBOXES

1.0 Index

1.1 Historical perspective

Since the turn of the century, the fruit fly *Drosophila melanogaster* has been intensely studied by geneticists to yield a comprehensive genetic map. Certain mutations resulting in the replacement of one segment by another or a segmented structure by a homologous one (termed homeosis by Bateson, 1896) attracted attention with developmentalists. Most of these homeotic mutations were mapped to two adjacent clusters in the genome - the Antennapedia (ANT-C) and Bithorax (BX-C) complexes. With the advent of recombinant technology sections of both regions were cloned and sequenced and in 1984 a 180 bp (basepair) conserved DNA sequence was found in three genes of the complexes : *Antennapedia (Antp)* and *fushi tarazu (ftz)* of the ANT-C and *Ultrabithorax (Ubx)* of the BX-C complex (McGinnis 1984a, Scott 1984). Mutations in both *Antp* and *Ubx* cause homeotic transformations and hence this 180 bp region was termed the homeobox. Great excitement ensued when it was found that

this homeobox was ubiquitous not only in *Drosophila* developmental genes but also in vertebrate genes (Holland 1986).
Since then, progress has been rapid with multiple homeobox sequences now available. Homeoboxes have been detected in organisms ranging from yeast and nematode to insects, echinoderms and vertebrates with amino acid positional identities ranging from 20 % upwards.
In this review I shall refer to the nucleotide sequence when mentioning 'homeobox' and to the translated amino acid sequence as 'homeodomain'.

1.2 Classes of Homeoboxes

Well over a hundred homeobox sequences have been reported and by probing with homeobox sequences that are more and more divergent from the classical *Antp* homeobox, the rate of discovery is not declining. The sequences have been ordered into 10 distinct classes plus several unique examples in 1989 (Scott et al.) based solely on comparisons of the amino acid homology within the 60 amino acid (a.a.) homeodomain (Scott et al. also considered an additional N-terminal a.a. where available).
Acampora et al.(1989) have divided the 30 Antp-type human homeodomains together with 36 non-human homologous sequences into 13 (sub)classes taking into consideration the 6 a.a. C-terminal to the homeodomain.
Vertebrate genes with at least 55% homology to *Antp* were classified by pairwise sequence comparisons (Kappen 1989) and a phylogenetic tree derived by cladistic analyses of the subset of 21 mouse homeobox sequences (Schughart 1989).
Since these reports several additional classes of homeodomain sequences have been coined often with a *Drosophila* member giving the group its name.
I have compiled a table of homeobox classes together with the members ordered by species, using the above mentioned classifications and recent reports of new classes and sequences (see table 1.1). From this table it is clear that certain homeodomains show barely detectable similarity to *Antp*. Considering that early work (1984-1987) used mainly Antp-class homeoboxes to probe for new boxes and that 60 % nucleotide similarity represents the limit of detection by hybridization, it is probable that this table is distributionally slanted toward *Antp*-type homeobox classes.

Table 1.1 Classes of homeodomain proteins.

Class [a]	Subclass	organism [b]	Homeodomain [c]	%cf *Antp*[d]	Reference[e]
Antp	?	*Drosophila*	*Scr*	90,0	1.
	?		*Ubx*	90,0	1.
	?		*ftz*	81,7	1.
	(*Hox* 1.1)		*Antp*	100	1.
			abd-A	96,7	1.
		Mouse	*Hox*.1.1	98,3	1.
			Hox.2.3	96,7	1.
		Xenopus	*XlHbox*2 (MM3)	98,3	2.
			*XlHbox*3	98,3	2.
	(*Hox* 1.2)	Mouse	*Hox*.1.2	93,3	1.
			Hox.2.2	93,3	1.
			Hox.3.3 (6.1)	91,7	1.
		Xenopus	*XlHbox*1 (AC1)	91,7	1.
	?	Sea Urchin	*TgHbox1* (*hb*-1)	88,3	1.
	?		*hb3*	81,7	Dolecki 1988
Dfd	(*Hox* 1.3)	Mouse	*Hox*.1.3 (m2)	90,2	1.
			Hox.2.1	90,2	1.
		Human	*Hox*.3.D (6.2)	86,7	2.
		Xenopus	*XlHbox*4 (*XHox*1B)	88,3	2.
			*XlHbox*5	85,0	2.
	(*Hox* 1.4)	*Drosophila*	*Dfd*	81,7	1.
		Mouse	*Hox*.1.4	81,7	1.
			Hox.2.6	81,7	1.
			Hox.4.2 (5.1)	83,3	1.
		Human	*Hox*.3.E	83,3	2.
		Xenopus	*XHox* 1.A	78,4	1.
Hox.2.4		Mouse	*Hox*.2.4	85,0	1.
			Hox.3.1	83,3	1.
		Human	*Hox*.4.C	81,7	2.
Hox.1.5		Mouse	*Hox*.1.5	75,0	1.
			Hox.4.1	65,0	1.
		Human	*Hox*.2.G	70,0	2.
mab 5		Nematode	*mab*-5	73,4	Costa 1988
Abd-B	(*Hox*.1.7)	*Drosophila*	*Abd-B* (*iab*-7)	58,3	1.
		Mouse	*Hox*.1.7	68,3	1.
			Hox.3.2	70,0	1.
			Hox.4.4 (5.2)	68,3	Dollé 1989a
		Human	*Hox*.2.E	70,0	2.
		Xenopus	*XlHbox* 6	68,3	2.
		Sea Urchin	*hb4*	70,0	Dolecki 1988
	(*Hox*.1.H)	Mouse	*Hox*.4.5 (5.3)	63,4	Dollé 1989a
		Human	*Hox*.1.H	66,7	2.
labial		*Drosophila*	*labial*	66,7	1.
		Mouse	*Hox*.1.6	61,7	1.
			Hox.2.9	60,0	Wilkinson 89
		Nematode	*ceh* 11	56,6	Schaller 90
		Nematode	*ceh* 13	55,0	Schaller 90

continued

Table 1.1 Classes of homeodomain proteins *(continued, page 2)*

Class [a]	Subclass	organism [b]	Homeodomain [c]	%cf *Antp*[d]	Reference[e]
Hox.2.8		Mouse	*Hox*.2.8	65,0	Wilkinson 89
zen		*Drosophila*	*zen*	60,0	1.
			z2	61,7	1.
Hox.1.9		Human	*Hox*.1.I	53,3	2.
			Hox.4.F	55,0	2.
rough		*Drosophila*	*rough*	56,6	1.
caudal		*Drosophila*	*caudal*	53,4	1.
		Mouse	*Cdx*-1	60,0	Duprey 1988
		Nematode	*ceh*-3	53,8	Bürglin 1989
engrailed		*Drosophila*	*engrailed*	50,0	1.
			invected	48,4	1.
		Mouse	*En*-1	50,0	1.
			En-2	51,7	1.
		Sea Urchin	SU-*hb*-*en*	46,7	Dolecki 1988
even-skipped		*Drosophila*	*even-skipped*	50,0	1.
		Mouse	*Evx* 1	51,7	Bastian 1990
			Evx 2	51,7	Bastian 1990
		Xenopus	*XHox* 3	51,7	1.
msh		*Drosophila*	*msh*	55,0	Robert 1989
		Mouse	*Hox*.7.1	43,3	Hill 1989
Hox.3.F		Human	*Hox*.3.F	48,3	2.
		Mouse	*Hox*.4.7 (5.6)	?	Kessel 1990
Hox.3.G		Human	*Hox*.3.G	45,0	2.
			Hox.1.J	45,0	2.
NK 1		*Drosophila*	*NK* 1	38,6	Kim 1989
NK 2-4			*NK* 2	41,7	Kim 1989
			NK 3	48,3	Kim 1989
			NK 4	43,3	Kim 1989
bicoid		*Drosophila*	*bicoid*	41,7	1.
Dll		*Drosophila*	*Dll*	41,7	Cohen 1989
H2.0		*Drosophila*	H2.0	40,0	1.
		Nematode	*ceh* 5	40,0	Bürglin 1989
paired		*Drosophila*	*paired*	36,7	1.
			gsbBSH4	36,7	1.
			gsbBSH9	33,3	1.

continued

Table 1.1 Classes of homeodomain proteins *(continued, page 3)*

Class [a]	Subclass	organism [b]	Homeodomain [c]	%cf *Antp*[d]	Reference[e]
Mix 1		*Xenopus*	*Mix* 1	34,0	Rosa 1989
cut		*Drosophila*	*cut*	28,3	1.
POU		*Drosophila*	*Cf1a*	25,0	Johnson 1990
		Human	*Oct* 1	31,7	1.
			Oct 2	31,7	1.
		Mouse	*Oct* 3 (*Oct* 4)	25,0	Rosner 1990 Schöler 1990
		Human	*E2A*	25,8	Kamps 1990
			prl	25,8	Nourse 1990
		Bovine	*bGHF1*	28,3	Bodner 1989
		Rat	*LFB1*	20,0	Frain 1989
			ERA-1	61,7	LaRosa 1988
		Nematode	*unc*-86	30,0	Finney 1988
			ceh 6	25,0	Bürglin 1989
		Nematode	*mec*-3	25,0	Way 1988
			ceh 14	25,0	Bürglin 1989
		Nematode	*ceh* 7	25,0	Bürglin 1989

a. These classes are based on the listed homeodomain sequences in ref.1. and 2. The classes seperated by a dashed line could be grouped together as one greater *Antp* type class as they show greater than 75% a.a.homology. In the text, when referring to this greater class, I shall use the term '*Antp*-type' instead of '*Antp*-class'.
b. *Drosophila* homeodomains have been chosen as representative for the Phylum Arthropoda, and mouse and African toad homeodomains for Vertebrata. Homologous homeodomain sequences from other species within these Phyla (honeybee, human, cattle, sheep, pig, rat, zebrafish, trout, chicken) have not been listed.
c. Some still used alternative names are given.
d. These values refer to amino acid percentage positional identity to the sequence of the *Antp* homeodomain. Most values were determined using the GCG TFASTA (Devereux 1984) programme, the rest manually from the relevant papers.
e. The original references for these homeodomains are listed in Scott et al. 1989 (1.) and Acampora et *al.* (2.). The references for more recent publications can be found in the reference section.

Thus when drawing conclusions about common features of all homeodomains it is better to weigh classes and not individual sequences equally. Scott et al.(1989) determined the consensus sequences of the *Antp*, *Dfd*, *lab*, *Abd-B*, *en*, *eve*, *prd*, *Hox*1.5, *Hox*2.4 and POU classes as well as the unweighted consensus of all homeodomain sequences and the variability at each position. Four amino acids (a.a.) are invariant : **tryptophan 48, phenylalanine 49, asparagine 51** and **arginine 53**. Five further positions are practically invariant with only occasional replacement by a very similar a.a. (conservative a.a. shown in brackets) : **leucine 16 (valine), phenylalanine 20 (tyrosine), isoleucine 45 (val),** and **lysine at 55 and 57 (arginine)** (see fig.1.1). The numbering differs from Scott et al. by -1 to reflect a more general numbering system. Several other residues are highly conserved especially in the carboxy-terminal third of the protein. We shall return to these when discussing the structure of the homeodomain [1.3].

In general, the protein sequences outside different homeodomains are not conserved although certain motifs recur [1.3]. This observation together with the astounding conservation of the homeodomain in organisms that have diverged hundreds of millions of years ago, such as *Drosophila* and *Xenopus* which show 98% identity between the *Antp* and *XlHbox*2 homeodomains, indicates that this sequence element must play a functionally autonomous and crucial role. Furthermore if one compares the genomic complexity of a species with the number of *Antp*-type homeobox genes that it harbors, a distinct correlation is detected. Larger genome complexities tend to indicate a greater phylogenetic complexity and a larger number of structural genes, thus more intricate regulatory mechanism are required. As we shall see later, most genes containing homeoboxes are involved in development or cell differentiation in a regulatory manner and therefore the above mentioned correlation is not surprising.

The chromosomal organization of the *Antp*-like homeoboxes in mammals and insects is very suggestive.

In *Drosophila* many of the homeobox genes are arranged in two separate about 300 kbp large clusters called ANT-C and BX-C, whereas in the beetle *Trilobium* mutations analogous to those of both *Drosophila* complexes map to a single chromosomal locus, the HOM-C complex (Beeman 1987) indicating that the split of the ANT-C and BX-C complexes in *Drosophila* is a

relatively recent event in the evolutionary line leading to the fruit fly.

In the mouse homeoboxes are organized into 4 clusters each spanning more than 100 kbp : *Hox* 1, *Hox* 2, *Hox* 3 (previously *Hox* 3 and 6) and *Hox* 4 (previously *Hox* 4 and 5) (Kessel 1990). The members of each cluster are more similar to cognate (homologous) homeoboxes in corresponding position within other clusters than to each other (Kappen 1989) indicating that the 4 clusters arose by duplication of an entire primordial cluster. The human homeoboxes are arranged in a similar fashion (Acampora 1989).

The two *Drosophila* homeobox clusters (or HOM-C) can be aligned with the mammalian homeobox clusters (Duboule 1989, Graham 1989), the genes being physically linked in the same order. For example, *labial* and *Hox* 2.9, *deformed* and *Hox* 2.6, *Scr* and *Hox* 2.1, and *abdominal-B (Abd-B)* and *Hox* 2.5 share homeodomains of the same class, share some homology in regions outside the homeobox and are arranged in the same 5' to 3'order on the chromosome. Furthermore homeoboxes in the middle of the mouse clusters (*Hox* 2.2, 2.3, 2.4) are related to *Antp*, *Ubx*, *ftz* and *Abd-A* of HOM-C. The fruitfly and mammalian clusters differ in that all mammalian clustered homeoboxes are transcribed in the same direction and are intronless whereas this is not the case with the *Drosophila* genes (Kappen 1989). Apart from these differences which indicate a small degree of independent chromosomal rearrangements, it is clear that homeobox genes were already arranged in a cluster before the divergence of arthropods and chordates. Even more intriguing is the correlation between expression and chromosomal order which we shall address in section 1.7.

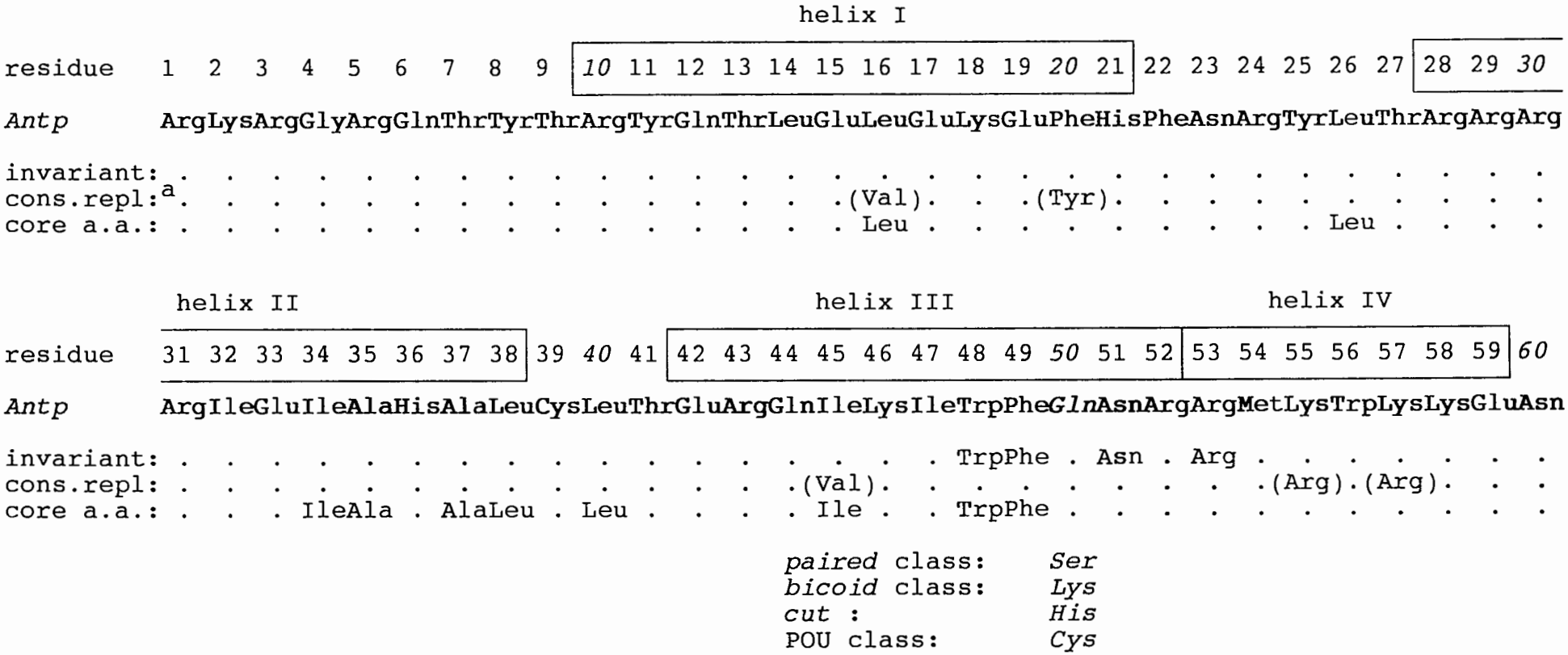


Figure 1.1 Antennapedia homeodomain sequence and salient features as discussed in sections 1.2, 1.3 and 1.4. The α-helices determined by Qian et al. (1989) are boxed. Helix III is termed the recognition helix. The sequence from helix II to III forms the helix-turn-helix motif.
a.) Conservative replacements refers to practically invariant positions that in certain cases are replaced by the similar amino acid shown in brackets.

1.3 Structure of the homeodomain

The first clues to the structure of the homeodomain came from protein databank comparisons which revealed sequence similarity between the C-terminal half of the homeodomain and the DNA binding domain of the yeast mating type proteins a1 and α2 (Laughon 1984, Shepherd 1984). It was proposed that the region from residues 32 to 51 formed a helix turn helix motif in analogy to that of bacterial DNA binding proteins with the second helix (43-51) corresponding to the 'recognition' helix of the bacterial proteins (Ohlendorf 1983).
The α-helix-turn-α-helix structure was found compatible with circular dichroism spectra of a chemically synthesized *Antp* homeodomain (Mihara 1988) and confirmed using nuclear magnetic resonance (NMR) of a 68 a.a. homeodomain produced by recombinant DNA technology (Otting 1988, Qian 1989, Affolter 1990).
The data from these studies (comparable to a 2,0 A resolution X-ray diffraction map) defined three distinct helices at residues 10-21 (helix I), 28-38 (II) and 42-52 (III) with a more flexible fourth helix at 53-59 (IV) (see fig.1.1). In a 3-dimensional view, helices I and II are aligned in an almost exactly antiparallel fashion with helices III and IV approximately perpendicular to them. The axis of helix IV forms an angle of about 30° with respect to helix III and it points away from the globular core of the protein. Helix II and III form the helix turn recognition helix motif respectively and a comparison of residues 30 to 50 with 21 a.a. of the bacterial DNA binding proteins lambda cro, phage 434 repressor and trp repressor reveals a backbone RMSD (pairwise root mean square deviation which is a measure of the differences in atomic positions of corresponding atoms in two structures after these structures have been superimposed to minimize the sum of the square of these differences) of between 0,8 and 1,0 A. For comparison, RMSD values for the helix turn helix motifs between lambda cro and several repressors lie in the range 0,4 to 0,9 A (Qian 1989).
The core of the homeodomain is formed by 10 hydrophobic residues (see fig.1.1): **Leu-16,26; Ile-34, Ala-35,37;** Leu-38,40; **Ile-45, Trp-48, Phe-49** with Thr-13 contributing a hydrogen bond to the indole ring of Trp-48. Leu-16, Ile-45, Trp-48 and Phe-49 are 4 of the 9 most highly conserved

residues amongst homeodomain [1.2]. Ala-35 and Leu-38,40 are also well conserved (Scott 1989) whereas position 26, 34 and 37 are quite variable. The residues at 26 and 34 are nonetheless mostly hydrophobic and the hydrophobic contribution of the Ala-37 residue is minimal (Qian 1989: figure 3), thus it is not surprising that this residue is not conserved.
Qian et al. point out that the homeodomain should have a global electric dipole moment because the negative charges are grouped on the helix I side whereas a total of 7 positive charges are distributed along helices III and IV which thus could electrostatically interact with the DNA backbone.
So far one interesting anomaly to the classical 60 a.a. homeodomain has been reported, namely the highly diverged homeodomain of mammalian liver specific transcription factor LFB1 (Frain 1989) which is 81 a.a. long. It has been demonstrated that 18 a.a. could be deleted without strongly affecting binding (Nicosia 1990) These 18 a.a. are postulated by alignment with other homeodomains to have been inserted between residues 33 and 34. Comparison with the 3-dimensional structure of *Antp* indicates that such an insertion would not directly interfere with the hydrophobic core (residue 34 contributes in the *Antp* homeodomain to the core) but possibly disrupt helix 2. The remaining three extra a.a. are postulated to be inserted between residues 38 and 39 which is in the turn of the helix-turn-helix motif. As LFB1 binds as a dimer (Frain 1989), the tertiary structure of its homeodomain may be quite different from that of *Antp*.

1.4 Binding studies

Apart from the vertebrate octamer-binding homeodomain proteins, most work in this field has been performed on *Drosophila* proteins. The cloned genes are expressed in bacteria and the homeodomain proteins, sometimes in the form of β-galactosidase (βgal) fusions (Desplan 1985,1988, Cho 1988, Laughon 1988) are fully purified (Beachy 1988, Cho 1988, Muller 1988, Biggin 1989a, Affolter 1990) or used as partially purified bacterial extracts (Desplan 1985,1988; Driever 1989a, Hoey 1988, Laughon 1988, Treisman 1989). Using increasing amounts of these proteins, and DNA sequences upstream of genes known by genetic analysis [1.7]

to be regulated by these proteins in footprinting studies, resulted in the definition of short (≈10 bp) consensus binding sequences (see above references). The results of these experiments are summarized in table 1.2 with sequences aligned in the 5′ to 3′ direction (sometimes using the antisense sequence) to reveal an ATTA/TAAT consensus that occurs in all *Antp*-class binding sequences. This tetramer sequence seems to be essential for binding because when the first adenosine is mutated to thymidine (Desplan 1988, 'LP'sequence see table 1.2), binding is greatly reduced. Similarly whereas $(TAA)_n$ is sufficient for *Ubx* and *Antp* binding, simple AT-rich regions are not (Beachy 1988).
The apparent dissociation constants lie around 10^{-8} M (table) but these values underestimate the real affinities because 1.the purification methods of these proteins from bacteria are harsh and the proteins may not be able to fully renature, 2.some proteins are used as fusions with β-galactosidase which may affect binding, 3.the studies with *Antp* use the 60 a.a. homeodomain in isolation and 4. the DNA sequences used may not contain the optimal binding sequence. Affolter et al. (1990) investigated the binding of the purified *Antp* homeodomain to an 18 mer oligonucleotide (BS2 see table 1.2) in detail and determined in saturation binding studies and competition experiments the equilibrium dissociation constant Kd to lie between $1,6x10^{-9}$ and $1,8x10^{-10}$ M with a half life of the DNA-Antp complex of 89 min. Non-specific DNA binding assayed using a random 26 or14 oligomer occurred with a Kd of 0,6 to $4x10^{-7}$ M. In comparison, prokaryotic helix-turn-helix DNA binding domains show a Kd of 10^{-5} to 10^{-6} M for their monomer (half-operator) binding sites with a half life of seconds or less (Affolter 1990). The high affinity of the *Antp* homeodomain for non-specific sites together with the low dissociation rate (long half-life) suggest that non-specific interactions are important for the high stability of the complex and that binding is quite distinct from the prokaryotic DNA binding proteins.
The difference may lie in *Antp* helix IV which has been postulated to follow the major groove of the DNA by projecting at 30° outward from the helix III axis (Qian 1989). This helix would thus extend the recognition helix by an additional 7 amino acids. This region is also highly conserved (Scott 1989, Garcia-Blanco 1989), contains several

Table 1.2 Homeodomain protein binding studies

Protein	Class	Descript[a]	Recognition sequence(s)[b]	Kd(M)	Reference
Antp	*Antp*	60 aa. synth.	(TAA)nTAAT AA 5′ to *Antp* [(TA)nTAAT (TA)x]3		Mihara 1988
		68 aa purified	CAGCG TAAT GTGTTTC BS1, 5′ to *en* CTC TAAT GGCTTTTTTC BS2, CTG TAAT TGCTC BSX, 5′ to *ftz*	2,4E-8 1,2E-8 <1,6E-9	Muller 1988 Affolter 1990
ftz	*Antp*	purified	BS1,2 see *Antp*		Muller 1988
		fusion	TCA ATTA AATGA **NP** k1-3 5′to *en* TCA ATTA ATTGA **LP** TCA TTTA AATGA **RP** not well bound		Desplan 1988
Ubx	*Antp*	purified	TT ATTAnTTA n≥2 5′ *AntpP1, Ubx*	≈1 E-8	Beachy 1988
		(1B)	(CG ATTAnCGATTA 5′ *AntpP1, Ubx*		Krasnow 1989
XlHbox1	*Antp*	fusion	GGCA ATTA AACT F1, 5′ to *XlHbox2* AACA ATTA CAAA F2, 5′ to *XlHbox2*		Cho 1988
Dfd	*Dfd*	fusion	5′to *Antp*P1,P2	3E-8	Laughon 1988
zen	*zen*	part.pur.	TCA ATTA AAT **NP**		Hoey 1988
en	*en*	fusion	TCA ATTA AAT NP, LP, lower affinity to RP	<2E-8	Desplan'85,'88
		part.pur.			Hoey 1988
			TATA AAT TATA boxes		Ohkuma 1990a
eve	*eve*	part.pur	1.TCA ATTA AAT **NP** k1-3 5′to *en*		Hoey 1988
			2.TCAGCACCG sites e4,5 5′ to *eve*		Hoey 1988
paired	*paired*	part.pur.	site 2. that *eve* binds		Hoey 1988
		purified[c]	1.TTTGACGT ϕ lambda,but not site 2.		Treisman 1989
			2. **NP6** with 20x lower affinity		
bicoid	*bicoid*	purified	GGG ATTA GA 5′ to *hunchback*		Driever 1989
caudal	*caudal*	purified	CATA AA 5′ to *ftz*		Dearolf 1989
Oct-1,2	POU	purified	ATGCA AAT octamer sequence		Ko 1988
			TCTC ATTA CC binds with 10x lower affinity to ATTA-sites		Garcia-Blanco 1989
GHF1/ *Pit*-1	POU	fusion & purified	TA AAT 5′ prolactin,GH ATG NATA(T/A)(T/A) 5′ to prolactin 10x lower aff.to octamer		Bodner 1988 Ingraham 1988
LFB-1	POU	purified	• palindromic TGGTTAATN ATTA ACAA		Frain 1989
Cf1a	POU	fusion	ATGGT CATA AATCAAATTGT 5′ to *Ddc*		Johnson 1990

a.) fusion = β-*gal*
b.) Sequences have been aligned (taking the appropiate DNA strand) to reveal the 5′-ATTA-3′ (or 5′-TAAT-3′) motif. For the POU and *caudal*-class homeodomains, the TAAAT motif has been highlighted. This motif appears frequently 3′ to the ATTA motif (examine table, **NP** sequence). •: dyad axis of symmetry.
c.) In this case, the carboxy terminus was removed, revealing a novel DNA-binding specificity. Site 2. was recognized independently of the *paired* homeodomain.

arginines and lysines and thus may be involved in non-specific electrostatic interactions with the DNA.
The question of how the homeodomain interacts with its recognition sequence has been investigated in two ways : by ethylation and methylation interference (Affolter 1990) and by mutating residues in the recognition helix of various homeodomains (Hanes 1989, Treisman 1989).
Ethylnitrosourea, an ethylating agent, was reacted with a DNA oligomer containing a BS2 binding site which was then incubated with the *Antp* homeodomain and the bound and unbound DNA separated and electrophoresed in parallel. A loss of a band in the bound lane indicated that *Antp* no longer bound due to either a steric hindrance or loss of a electrostatic interaction resulting from the cancelling of a negative charge by the ethyl addition to the phosphate backbone (Affolter 1990). Eleven positions inhibited binding (see table 1.2) and in 3 dimensions these define a patch across three sugar-phosphate backbones on one face of a DNA helix. For methylation interference experiments, the DNA was modified with dimethyl sulphate which methylates N-3 of adenosine and N-7 of guanosine only, so to increase the number of positions that could be modified, oligomers containing BS1 and BSX (table) were also used. The protected bases were found to fall within the patch defined by the ethylation experiments thereby confirming the importance of the TAAT sequence and flanking bases in the binding of the homeodomain.
The 3-dimensional structure derived by Qian et al. (1989, see [1.3]) indicated that residues **51 (asparagine)** and **53 (arginine)** which are two of the invariant residues, project outwards from helix III (the recognition helix) and are ideally situated for interacting with bases of the major groove. Arginine can interact most favorably and specifically with guanine, donating hydrogen bonds to both N-7 and O-6 of guanine (Harrison 1990).
Residue **50 (glutamine)** is also situated optimally for interaction with the DNA. This residue however, is not invariant : In the *paired,prd* class it is occupied by serine, the POU class features cysteine, *bicoid*, *bcd* lysine and *cut* histidine (fig.1.1). All other homeodomains share the *Antp* glutamine residue. An exciting finding by Hanes et al.(1989) and Treisman et al.(1989) was that residue 50 (position 9 in the recognition helix) can determine the DNA binding specificity of homeodomain proteins : changing only

the lysine at position 50 of *bicoid* to glutamine caused a switch from *bicoid* DNA recognition sites to *Antp* recognition sites whereas mutating this residue to alanine abolished DNA binding activity (Hanes 1989). In the same study it was found that position 42 and 46 could be mutated to alanine with little effect on binding. Using DNAase footprinting and mutations of the *paired* protein, Treisman et al. showed that *paired*'s distinct recognition specificity could be changed to that of the *Antp*-class *ftz* or that of *bicoid* by altering solely the serine at residue 50 to that of *ftz* (glutamine) or that of *bicoid* (lysine) respectively.
Glutamine, serine and lysine side chains can all accept/donate specific hydrogen bonds which may explain the change in DNA site recognition. Glutamine can specifically interact with an adenosine base, with the amino side chain donating a hydrogen bond to N-7 and the C=O accepting one from N-6 (Harrison 1990). The loss of a protein -nucleic acid hydrogen bond can cause a reduction in favorable binding free energy of up to 16,8 $kJmol^{-1}$ if there are no compensatory effects such as a rearrangement of the DNA-protein complex to allow compensatory interactions internally or with solvent (Berg 1988).

It is still curious that the change of a single amino acid that can interact with a maximum of 2 to 3 base pairs can cause such drastic changes in the recognition specificity : from **TCAATTAAAT** of *Antp* and *ftz* (Hoey 1988) to **TTTGACGT** of *paired* (Treisman 1989) and **GGGATTAGA** of *bicoid* (Driever 1989a). The *Oct-1* and *Oct-2* (POU class) proteins which have a cysteine at position 50 recognize the octamer sequence **ATGCAAAT** but also the *Antp* consensus and several other ATTA containing sites albeit with far lower affinity (Garcia-Blanco 1989). In spite of this wide specificity in binding, single point mutations within *Oct-1* binding sites can abolish DNA binding (Sturm 1988). The POU class probably interacts differently with the DNA as the POU-specific domain is needed in conjunction with the POU-homeodomain for binding [1.5] and replacement of the recognition helix of *paired* with that of *Oct*-2 did not allow recognition of the octamer sequence (unpublished observation from Treisman 1989).
In this context the double specificity of *even-skipped (eve)* should be pointed out. This protein can recognize the quite

different *Antp* binding site and an e4 (table 1.2) site, TCAGCACCG, with equal efficiency (Hoey 1988).
We can thus postulate that residue 50 at position 9 in the recognition helix (III) is crucial in determining the binding specificity of homeodomain proteins which implies that the recognition helix is situated in such a way that this residue can maximally interact with the DNA helix. It is probable that more than one mode of interaction with the DNA exists, thus allowing quite different sets of consensus sequences to be recognized. Thus mutations outside the recognition helix and even outside the homeodomain may subtly affect the preferred binding sequence (within or without a consensus family) by altering the relative orientation of the recognition helix.

During the purification of *Ubx* 1b protein (Beachy 1988) and the *Antp* homeodomain (Muller 1988), it was observed that these proteins existed as homodimers in solution. In the case of the *Antp* homeodomain, dimerization is mediated by disulphide bridges involving the cysteine residue at position 39 (Muller 1988). This position is fairly variable amongst homeodomains (Scott 1989) and mutation to serine does not affect binding to a single consensus site (Affolter 1990).
Dimerization may however play a role in binding because cooperativity has been seen in binding of homeodomain proteins to clustered recognition sites (Desplan 1988). These authors showed that all possible orientations of two NP (table 1.2) sequences were bound equally well by *engrailed* and *ftz* β-galactosidase fusion proteins and that as the number of NP sites increased, the concentration of fusion protein needed to produce protection from DNAase1 decreased. These results are ambiguous because β-galactosidase is normally a tetramer and thus lead to aggregates of fusion proteins which may then preferentially interact with multiple sites. A phase shift study of the same proteins purified without a foreign moiety has though also indicated cooperativity in binding (Ohkuma 1990a,b). The highly diverged homeodomain protein LFB1 was shown to bind as a dimer (Frain 1989) mediated by a short N-terminal myosin like dimerization helix (Nicosia 1990). However in this case the consensus sequence is palindromic containing two ATTA motifs with one motif insufficient for binding.

1.5 Structural features of Homeodomain-proteins

Having hitherto discussed the homeodomain on its own we shall now first examine other features of proteins containing homeodomains and then in section [1.6] the function of these proteins.
Initially homeoboxes were found in the exons closest to the 3' end of the gene (Laughon 1985) about 20 to 30 a.a. from the C-terminus of the protein and although this is still the case for most proteins with an *Antp*-type homeodomain, it no longer applies to more diverged homeodomain proteins (table 1.3).
Mavilio (1986) and Kessel et al.(1987) pointed out a conserved pentapeptide 5 to 15 a.a. N-terminal to the *Drosophila Antp, Dfd, cad, Xenopus XlHbox 4* and human *Hox* 2C, 3C AND 4B homeodomains. The sequence has since been detected in 15 human *Hox* genes (Acampora 1989), the corresponding clustered mouse *Hox* genes, the *Xenopus XlHbox* 1, 2, (4) and 5 genes (Fritz 1988), the sea urchin TgHbox *1* gene (Angerer 1989) and in all homeotic genes of the *Drosophila* ANT-C and BX-C (table 1.3). More diverged homeodomain proteins do not have this sequence, consisting of a hydrophobic residue (Val, Ile or Leu) and the well conserved **Tyr-Pro-Trp-Met** core followed in many cases by a basic residue (Arg or Lys). A more diverged version of this motif is found in haemoglobins and myoglobin (Ile-Tyr-Pro-Trp-Lys-Arg) where it acts as a bend separating two α-helical regions of the globin protein (Stryer 1981). It has been proposed (Fritz 1988) that this pentamer may fulfill a similar role separating the DNA binding function of the homeodomain from the rest of the protein. It is interesting that a correlation exists between homeodomains located near the carboxyl end of the protein and homeodomains preceded by the pentamer box (cf. table 1.3).
The *Drosophila paired* and *gsb* proteins (Bopp 1986) as well as the mouse Pax 3, 6 and 7 proteins (Kessel 1990) contain, in addition to a homeodomain a so-called paired domain of 129 amino acids. The function of this domain is not known but a.a. 80-105 exhibit a helix-turn-helix motif with an 8 a.a. turn and a highly amphipathic first helix, which led to the speculation that this domain may act as a DNA binding moiety (Burri 1989). It should be pointed out that Treisman et al.(1989) noted that the C-terminal part of *paired* confers a DNA binding specificity independent of a

functional homeodomain on this protein. The occurrence of two distinct DNA binding domains within a single protein would offer interesting regulatory possibilities.
The POU (coined from Pit1, Oct1,2 Unc-86 (Herr 1988)) homeodomain proteins contain a second conserved region upstream of the homeodomain, termed the POU-specific domain. The POU-specific domain can be divided into two regions both of which have been shown to be necessary for effective DNA binding *in vitro* (Stern 1989).
Oct-2 contains in addition, downstream of the POU domain (= POU-specific + POU-homeodomain) four leucines separated by exactly 7 residues (Clerc 1988). Such an arrangement has been found in the DNA binding proteins GCN4 (a yeast transcriptional activator), the *jun*, *fos* and *myc* oncoproteins and the C/EBP enhancer binding protein (Struhl 1989). Termed the leucine zipper, this motif is required for dimerization and DNA binding (C/EBP) or tetramerization (*myc*)(Mitchell 1989). Its function in *Oct*-2 is unknown especially since *Oct*-2 binds DNA as a monomer (Clerc 1988) but could possibly lie in the interaction of another leucine zipper protein.
The mammalian liver-specific transcription activator LFB1 (Frain 1989, Nicosia 1990) binds as a dimer [1.4] which is mediated by a myosin like N-terminal helix. Nicosia et al. propose a coiled-coiled interaction between two such α-helices as the mechanism for dimerization in a manner analogous to leucine zipper interactions.
A very recently discovered motif that occurs in conjunction with highly diverged homeodomains is the LIM motif, found in the nematode l*in-12*, m*ec-3* (Way 1988) and rat insulin 1 enhancer binding protein I*sl-1* (Karlsson 1990). This motif of consensus sequence C-X_2-C-X_{17-19}-H-X_2-(C-X_2)$_2$-C-X_{7-11}-(C)-X_8-C, C=cys, H=histidine, X=any residue; is found in two tandem copies N-terminal to all three homeodomains as well as in another nematode homeodomain protein, *ceh*-14 (Freyd 1990 unpublished). The motif bears similarity to metal binding regions of several proteins such as ferredoxin, steroid receptors and the zinc finger motif (Freyd 1990) and may thus function in metal binding.
A more general feature amongst homeodomain proteins is the occurrence of homoaminoacid stretches and regions rich in one or two amino acids (see table 1.3). The most common amino acids forming such regions are Ala, His, Glu/Asp, Gly, Gln and Ser/Thr. Ptashne (1988) pointed out that acidic

regions (Glu/Asp) or regions that could be phosphorylated (Ser/Thr) may be involved in transcriptional activation. In the transcriptional activator Sp1 the two most potent activating domains contain about 25% Gln. A Gln-rich stretch from *Antp* can partially substitute for the Sp1 activation domains (Mitchell 1989), thus the Gln rich areas may be involved in transcriptional activation. Proline rich areas are also able to activate transcription in mammalian cells (Mitchell 1989) whereas Ala-rich stretches have been shown to be involved in transcriptional repression (Licht 1990).

1.6 The function of homeodomain proteins

The nuclear location (Gehring 1987) and sequence specific DNA binding upstream of genes [1.4] together with the suggestive amino acid rich regions [1.5], strongly argue in favour of a gene regulatory role for homeodomain proteins. Since the discovery of the homeobox several experimental approaches have been used to ferret out the function of these proteins.

1.6.1 Experimental approaches

A first approach to elucidate the function of a cloned gene involves a comparison to DNA/protein data banks in an attempt to identify homology to elements with a known function. In the case of *Drosophila* extensive genetic information allowed the mapping of cloned homeobox genes to specific developmental mutants and thus the phenotypic effects of homozygous lethal to overexpressing mutations provided evidence for the function of these genes.
The next approach involved the determination of spatial and/or temporal expression patterns during embryogenesis and in the adult. This was done by 1. probing the RNA extracted from various tissues and stages, 2. *in situ* hybridization and 3. by immunocytochemical probing. As discussed in section 1.7, these studies revealed that all homeodomain proteins with an homeodomain related to that of *Antp* by at least 50% were expressed during embryogenesis in a spatially and temporally controlled fashion while some proteins with more divergent homeodomains are instead expressed in a tissue or cell specific fashion (eg. *Oct*-2, LFB1, several nematode homeodomain proteins, see [1.7]).

Table 1.3 *continued*

Protein	Mol. Wgt.	Rel.homeodomain position	Other motifs	Reference
Vertebrates				
c-13	30 K	carboxy	5-mer	Simeone 1988
Hox 1.1	26 K	amino	5-mer, Pro-, Gln-rich	Kessel 1987
Hox 3.1	28 K	amino	5-mer, Gln-rich	Breier 1988
Hox 7.1	35 K	middle	divergent 5-mer, Cys-rich	Hill '89, Robert '89
isl-1			LIM-domain, Gln-rich	Karlsson 1990
LFB1	69 K	middle	POU-domain, Myosin-like dimeris., Ser-, Pro-rich	Nicosia 1990
Mix 1	42 K	carboxy	Pro-rich, acidic region	Rosa 1989
Oct-1	90 K	middle	POU-domain, 2x Gln-, Ser/Thr-rich	Stern '89, Sturm '88
Oct-2	60 K	middle	POU-domain, leucine-zipper, 3 x Gln-, Asp/Glu-rich (acidic)	Tanaka 1990
Oct-3(4)	42 K	middle	POU-domain, 2 x Pro/Gly-, Ser/ Thr-rich	Posner 1990
Pit1/GHF1	33 K	carboxy		Ingraham 1988 Bodner 1988
XlHox2	25 K		5-mer, Gln-rich	Wright 1987

Table 1.3 Characteristics of homeodomain-containing proteins

Protein	Mol. Wgt.	Rel.homeodomain position	Other motifs	Reference
Drosophila				
Antp	43 K	carboxy	5-mer, 2 x Gln-rich	Laughon 1986
Bcd	54 K	amino	no 5-mer, $(His-Pro)_n$, 10% Pro phosphorylated	Berleth 1988, Driever 1988a, Frigerio 1986
caudal			5-mer, poly-Arg, $(GCX)_n = Ala_n$	Mlodzik 1988
cut	240 K		3 cut repeats, 4 acidic regions Gln/His-, Pro/Ala-, Ala-, Asn-rich	Blochlinger 1988
engrailed	60 K	carboxy	poly-Gln, 2 x poly-Ala, Glu/Asp-(acidic), Ser-rich ,forms 10000 K complex	Laughon 1986
eve	40 K	amino	$(GCX)_n$ = poly-Ala	Macdonald 1986
ftz	48 K	carboxy	diverged 5-mer, Gln-rich phosphorylated	Laughon 1986 Krause 1989
H2.0	41 K		OPA = $(CAX)_{25}$: His, Gln-rich	
Dfd	64 K	carboxy	5-mer, Gly-, His-, Gly-rich poly-Gln, -Asn, acidic region (Glu, Asp)	Regulski 1987
labial	(3 kbp)	carboxyl	5-mer, intron at residue 44 of homeobox, 4 x $(CAX)_n$ (see Bcd) Ser/Thr-, Pro/Gln-, His/Pro-rich.	Mlodzik 1988
paired	67 K	middle	$(His-Pro)_n$	Friferio 1986
Scr	45 K	carboxy	5-mer, poly-Gly, Gln-, Ser-, Asn-rich	LeMotte 1989
Ubx	40 K	carboxy	5-mer, poly-Gly	Beachy 1990

continued

In *Drosophila* these results were extended using various mutants thereby establishing a hierarchy of homeobox gene product interactions as well as cross-regulatory and auto-regulatory properties [1.7]. However, these experiments do not reveal whether homeodomain proteins exert their effect directly or via intermediate proteins/factors. To circumvent this problem, 'producer-responder' constructs were made and introduced into *Drosophila* tissue culture cells or embryos or into yeast cells. These constructs entail a 'producer' plasmid consisting of a homeobox gene (modified or not) fused to a constitutively expressed (or inducible) promoter (usually the actin 5C promoter) and a 'responder' plasmid containing various arrangements of *in vitro* determined homeodomain binding sites fused to various promoters driving an easily assayable reporter gene such as β-galactosidase or chloramphenicol acetyltransferase. Such experiments were first reported for *ftz* and *engrailed* (Jaynes 1988) and *bicoid* (Driever 1989a) using cultured Schneider cells, and extended to embryos by P-element mediated transformation of the *Drosophila* germ line (Driever 1989b). Results from these experiments clearly established *Drosophila* homeodomain proteins as transcriptional regulators (see below) and it was shown that this effect is independent of *other Drosophila*-specific proteins by using producer-responder constructs in yeast (Samson 1989, Hanes 1989) and by purifying homeodomain proteins (*en*, *ftz*) and assaying their function in an *in vitro* transcription system from human cells (Ohkuma 1990a,b).
Vertebrate and echinoderm studies have been hampered by the lack of suitable genetically described mutations, and specifically in the case of vertebrates, the great genomic and developmental complexity. Apart from the POU family of transcription factors and LFB1 for which the protein was characterized before the cloning of the respective genes, there is no direct evidence for homeodomain proteins as transcriptional regulators in echinoderms and vertebrates

1.6.2 Transcription factors

Using the above mentioned producer-responder arrangement in *Drosophila* tissue culture cells, Jaynes and O'Farrell (1988) demonstrated that a responder consisting of six NP (see table 1.2 for sequence) homeodomain binding sites upstream of several different promoters, showed a 30 to 300 fold

increase in activity when cotransfected with an *ftz* producer. These binding sites were orientated in either direction and 33 to 386 bp upstream from the promoter indicating their status as enhancer elements. When the number of NP sites was reduced from 6 to 4 to 2, a large decrease in *ftz*-dependent activation from 60 to 10 to 5 fold over basal level was observed. There is evidence for co-operative binding of *ftz* [1.4] but it is not yet clear whether the large increase in activation associated with 6 NP copies relative to 4 represents increased site occupancy (due to co-operative binding) or a synergistic action on activation as more sites are occupied. Using the mutated NP sequences LP and RP (table 1.2) these authors showed that binding site alterations have parallel effects on *in vitro* binding affinity and transcriptional activation.
Fitzpatrick and Ingles (1989) fused the N-terminal 147 a.a. comprising the DNA binding domain of the yeast transcription factor GAL4 to *ftz* and found that this construct was almost as efficient as GAL4 itself in activating transcription in yeast cells of a responder plasmid containing GAL4 binding elements. All deletions of parts of the *ftz* protein reduced transcriptional activation : deletion of the N-terminal 170 a.a. or C-terminal 100 a.a. reduced activation 4 to 5 fold whereas more extensive deletions of the 413 a.a. *ftz* protein increasingly abolished activation. A puzzling observation is that a single point mutation in the homeodomain (Ala to Val291) which results in a temperature sensitive *ftz* phenotype, drastically reduces activation in yeast (70 fold). All these fusion proteins were shown to be equally stably expressed and thus one can conclude that the activation region of *ftz* is spread over the protein. *Ftz* is known to contain a Gln-rich stretch at its carboxyl end (Laughon 1985) and Krause et al.(1989) reported stage specific phosphorylation at 19 of the possible 41 Ser and 41 Thr residues during *Drosophila* development. Both negatively charged and Gln-rich regions have been implemented in transcriptional activation [1.5].
Winslow et al.(1989) reported a 20 fold induction by *ftz* using a AT-rich 75-mer binding site derived from upstream of the *Antp*-P1 promoter fused to a truncated *hsp70* promoter. This 75-mer contained three ATTATTA elements.
The *ftz* protein was recently purified and used in an *in vitro* transcription system from human cells (Ohkuma 1990b). *Ftz* activated transcription with an optimum stimulation at 1

mM Mg^{2+} whereas at 8 mM Mg^{2+}, which is normally used to achieve optimum TFIID (human general transcription factor IID) binding activity in cellfree systems, *ftz* did not activate. This *ftz* activation was dependent on NP binding sites upstream of the *hsp70* promoter.

Having established *ftz* as a transcriptional activator, how is this protein affected by other homeodomain proteins that recognize the same binding sequence, for example *en*, *eve* and *zen* (table 1.2)?

Using an *engrailed (en)* producer plasmid cotransfected with the responders used for *ftz*, no activation occurred (Jaynes 1988). However, when both *ftz* and *en* producers were cotransfected these authors observed a 30 to 100 fold reduction in the *ftz* mediated activation. This reduction was also shown to correlate with the amount of *en* producer relative to *ftz* producer used for transfection leading the authors to propose a simple competition model, whereby *en* represses *ftz* activation by competing for the same binding sites. This model is supported by the results of Ohkuma et al.(1990b) who demonstrated that when equimolar amounts of *en* and *ftz* were added in an *in vitro* transcription system, the promoter was expressed at its basal level. A promoter not containing the 6 NP sites was unaffected by these amounts of *en* (a 1:3 molar ratio of *en* to binding sites). *En* was shown to be able to compete with *ftz* prebound to the NP sites indicating that *ftz* does not form a very stable protein-DNA complex in this system. The authors claim that *en* binds 4 to 5 times more strongly to the NP sites than does *ftz* and interpret bandshift assays as indicative of co-operativity in the binding of two but not three *ftz* or *en* molecules. They observe no mixed complexes where *ftz* and *en* have bound adjacent sites.

The repression by *en* can also be mediated by competition with the TATA-box binding transcription factor TFIID albeit only at 5 fold higher amounts of *en* (but 30 fold higher *en* : binding site ratios, namely 10 : 1)(Ohkuma 1990a). This repression can be halved by preincubation of the promoter sequences with TFIID and the authors showed that this correlated with a reduced ability to compete with TFIID for the promoter proximal TATA binding region. The *ftz* promoter was repressed far less than the *hsp70* and adenovirus 2ML promoters. This correlated to the binding affinity and the authors note that the sequence context of the TATA boxes differ, being TATATA in the *ftz* promoter and TATAAA in the

other promoters. A comparison with the binding sequences listed in table 1.2 indicates that the latter conforms more closely to homeodomain binding consensus sequences : compare the *cad*, *cf1a* **CATAAA** as well as the 3' end of the NP sequence : TCAA**TTAAA**T.
Synergistic interactions between the homeodomain proteins *ftz*, *zen*, *z2*, *prd*, *eve* and *en* in various actin 5C driven producers in cotransfections into *Drosophila* tissue culture cells have been reported (Han 1989) but some of these interactions are difficult to interpret because complex binding regions were used and the relative binding affinity of the homeodomain proteins to any given site were not examined. The most informative results came from responder plasmids containing 1 to 5 k'copies inserted upstream of the metallothionein promoter driving the chloramphenicol acetyltransferase gene. The k' site consisted of the k1 and k2 sites of the *en* upstream region containing 3 and 2 copies respectively of a NP-like sequence (Hoey 1988). *Ftz* activated transcription of such a responder with 5 k' regions only 15 fold, whereas a responder with one k' sequence was not activated at all. Cotransfection with either *zen*, *paired* or both resulted in synergistic activation of transcription in all combinations. *Zen* and *ftz* bind to NP sequences (Hoey 1988, Desplan 1988) whereas *paired* does not (Treisman 1989) so the mechanism of this synergism may differ for these proteins.
Han et al (1989) observed the greatest transcriptional activation of their 5 k' responder plasmid with *z2* (related to *zen*) and *zen* (Rushlow 1987), whereas *eve* and *en* had no effect on basal transcription levels, although being able to bind the NP site (Hoey 1988, Desplan 1988, table 1.2). Removal of the C-terminal acidic domain of *z2* destroyed activation (Han 1989). Cotransfection of *z2* and *eve* or *en* resulted respectively in a 4 and 7 fold reduction in *z2* mediated activation. *Eve* own was also shown to reduce *ftz* and *paired* mediated activation 6 fold when cotransfected and to reduce *ftz*/*paired* synergistic activation 10 fold. The transcriptional repression mechanism of *en* has already been discussed; that of *eve* has been investigated in a *Drosophila* cell free transcriptional extract using purified *eve* protein as well as in an *in vivo* producer responder assay (Biggin 1989). This study indicated that *eve* reduces the basal transcriptional level of a promoter about 5 fold in a binding site dependent manner in both assays. This binding

region contained both type 1 and 2 *eve* consensus binding sites (table 1.2). In the Han (1989) experiments no effect on the basal transcriptional level was observed. It is interesting to note that neither purified *zen* nor transfected *zen* have any effect on transcription (Biggin 1989) again in marked contrast to the Han experiments. Possibly the different responder constructs are responsible for such differences especially since the binding regions are complex and differ as do the promoters used.
The Biggin and Tjian *in vitro* transcription results were independent of the orientation and location of the DNA binding sites, and without these sites no repression occurred shown for both *Ubx* and dADH promoters. It thus seems that *eve* directly interacts with some component of the transcriptional machinery or interferes with a transcriptional activator protein bound at another site. A second mode of repression involving competition with other homeodomain proteins cannot be ruled out (Han 1989). Results from genetic studies have indicated both positive and negative regulatory roles for *eve* and *en* [1.7].
Ubx and *Abd-A*, both of which have *Antp*-class homeodomains (the *Abd-A* domain differs only at a.a.11 where Phe replaces Tyr), were expressed as fusion proteins with the DNA binding domain of the bacterial transcriptional repressor Lex A in yeast (Samson 1989). Both proteins stimulated expression of a reporter gene only if a lex A or Ubx (table 1.2) binding site was present but this site could be at a variable distance up to 350 bp from the promoter (dADH and *Ubx* promoters) and in either orientation. Neither homeodomain nor the C-terminal most 130 a.a. of *Ubx* or 55 a.a. of *Abd-A* were required for the activation of lex A targets. Deletion of a further C-terminal 100 a.a. of *Abd-A* abolished all activation. This region contains a Gln-rich stretch which may be responsible for the activation. One should interpret these mutations with caution as gene activation by lex-A fusion proteins is a complex process involving entering into the nucleus and dimer formation and these processes may have been affected by mutations. It is intriguing that *in vivo* both proteins have been shown to repress transcription though *Ubx* could also activate its own gene (Bienz 1988). Repression may thus require interactions with *Drosophila* proteins not present in yeast.
In cotransfection assays of *Drosophila* tissue culture cells, *Ubx* was shown to both activate and repress transcription

(Krasnow 1989). A DNA region upstream of the *Ubx* promoter containing two clusters of *Ubx* binding sites with the $(TAA)_x$ motif was necessary for the activation of a *Ubx*-promoter driven responder plasmid. Insertion of these clusters upstream of an ADH promoter resulted in a 100 fold transcriptional activation. In contrast a 10 to 40 fold reduction in transcriptional activity was observed when cotransfecting a responder construct containing a 6 kbp *Ant* P1 promoter region. This region has also been shown to contain *Ubx* binding sites (Beachy 1988). Disruption of the *Ubx* homeodomain abolished repression (Krasnow 1989). A N-terminal deletion (a.a.36 to 226) abolished *Ubx* activation from the *Ubx* promoter construct but left the repression of the *Antp*P1 construct intact (Krasnow 1989) thereby implicating the N-terminal end of *Ubx* inactivation in accordance of the report by Samson et al. Repression is thus mediated by the C-terminal region (aa 226 -389) which includes the homeodomain (aa 294-354) and /or the first 35 a.a. which were still present in the deleted protein. Fusing the *Ubx* homeodomain and the residues C-terminal to it to the *Antp* N-terminal 295 amino acids (does not include the *Antp* homeodomain) gave a hybrid protein that functioned in a similar way to full-length *Antp* and *ftz* in cotransfection assays using *Antp* P1 and *Ubx* promoter regions : that is, it activated transcription in both cases (Winslow 1989). This implies that the *Ubx* mediated repression of the *Antp* P1 promoter must reside either in the N-terminal most 35 a.a. or between a.a.226-294. The most likely mechanism of repression by *Ubx* seems to be an interaction with a second *Drosophila* protein which 1. binds to a site in the 6 kbp *Antp* P1 promoter region that is absent from the *Ubx* promoter region and 2. which specifically interacts with a region of *Ubx* which *Antp* and *ftz* do not share.

Studies of *bicoid* (divergent class of homeodomains) indicate that it is a DNA binding dependent transcriptional activator (Driever 1989a,b; Hanes 1989; Struhl 1989a). The transcriptional activation domain was shown to lie in the N-terminal half of the protein using lexA-*bicoid* fusion producers and *bicoid* binding site containing responders in yeast (Hanes 1989, Struhl 1989a). A deletion leaving residues 3-251 but removing an acidic and a Gln-rich region left activation intact. Deletion of a further 100 C-terminal a.a. removed the activation function which could be restored by a fusion adding the GCN4 activation domain (Struhl

1989a). This *bicoid* region (a.a. 158-250) has an unusually high Ser and Thr content and it has been shown that *bicoid* is heavily phosphorylated *in vivo* in *Drosophila* and in yeast (Driever 1989a), which results in a high negative charge. Thus this area may function similar to the acidic GCN4 activating region. The proline rich region at the N-terminal side is another candidate for mediating activation [1.4] but was shown to be dispensable (Struhl 1989a).
The activation mechanism of the mammalian POU-domain proteins *Oct*-1 and 2 and LFB1 has been analyzed (Nicosia 1990, Muller-Immerglück 1990, Tanaka 1990). *Oct*-1 and *Oct*-2 are transcriptional activators but unlike *Oct*-2, *Oct*-1 does not activate transcription of most promoters that contain a TATA box. This was shown in HeLa cells which do express *Oct*-1 (Schaffner 1989).
Oct-2 activation seems to be dependent on 1.a Gln rich domain near the N-terminus which is also present in *Oct*-1, and 2. the C-terminal patches of multiple Pro and Ser/Thr, while 3. the C-terminal leucine zipper is dispensable (Tanaka 1990). Muller-Immergluck et al.(1990) find that either region is sufficient on its own. The N-terminal third of *Oct*-1 and 2 is interchangeable but interchanging the C-terminal regions results in a reversal of the activation ability (Tanaka 1990). The centrally located POU-domain was found to be solely responsible for DNA binding but not transcriptional activation though Muller-I.et al. report some activation by the POU-domain itself when the octamer sequence is located very close to the TATA box.
An interesting observation is that only active fusion proteins show a shift in electrophoretic mobility when phosphorylated (Tanaka 1990) suggesting that a phosphorylation dependent change in conformation is taking place before transcriptional activation occurs. Such an electrophoretic shift has also been observed for *ftz* (Krause 1989) and *bicoid* (Driever 1989a).
The distance of the *Oct*-2 binding site (the octamer) from the promoter severely affects the activation potential of *Oct*-2 in HeLa cells (Muller 1990, Tanaka 1990) but not in B-lymphoid cells (Muller 1990). As the identical responder construct was used this implies that *Oct*-2 can interact with cell-specific proteins which may be absent in HeLa cells. *Oct*-1 was shown to be able to interact with the herpes simplex virus transactivator VP16 via amino acid residues 32, 33 and 36 of helix II of the *Oct*-1 homeodomain, while

Oct-2 which differs at these (and 4 other) positions of the homeodomain can not interact with VP16 (Stern 1989). This interaction confers on *Oct*-1 the ability to activate the same octamer-motif containing β-globin promoters that are activated by *Oct*-2.
Nicosia et al.(1990) have used deletion mutants to pinpoint the activation domains of the LFB1 (POU-class homeodomain) transcription factor to a serine rich and a proline rich sequence. The proline rich area shows significant sequence similarity to the activation domain of CTF1.

1.6.3 Conclusions

We can summarize and expand on these results to come to several conclusions:
1. Homeodomain proteins can function as transcriptional activators (*ftz*, *Antp*, *bicoid*, *z2*, *zen*, *paired*, LFB1, *Oct*-1 and 2), repressors (*eve*, *en*) or both (*Ubx*, *Abd-A*).
2. Activation is dependent on binding sites which act as enhancer elements, that is, they can be at variable distances from the TATA box containing promoter and in either orientation as shown for *Antp*, *Abd-A*, *Ubx*, *ftz*, *zen* and *bicoid*. Whereas binding is mediated by the homeodomain, the activation function resides in separate domains characterized by glutamine (*Oct*-1,2; *Antp*, *Abd-A*, *ftz*), proline (*Oct*-2, LFB1), acidic regions (*z2*) or/and serine and threonine rich areas (*bicoid*, *ftz*, *Ubx*, *Oct*-2, LFB1) that have been shown to be phosphorylated (*bicoid*, *ftz*, *Ubx*, *Oct*-2). If phosphorylation is necessary for activation, then this represents an interesting mechanism for regulating the effect of transcriptional regulators. In one case activation was shown to be dependent on binding to a transactivator (*Oct*-1 : VP16).
Very little is known about the mechanism of activation. Interactions of transcription factors with the TFIID TATA-box binding protein or the C-terminal domain of the largest RNA polymerase subunit which contains multiple repeats of (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) which are phosphorylated at transcriptional initiation, have been proposed to play a part (Sawadago 1990, Lillie 1989).
3. Transcriptional activation / repression is promoter independent. The following TATA-box containing promoters have been interchangeably used : *hsp70*, *Antp*P1, *ftz*, *ubx*, *hunchback*, distalADH, proximalADH. Most studies used the

hsp70 (heat shock protein 70) promoter amongst others, and tissue culture cells (*Drosophila* Schneider 2, HeLa, yeast). Recent reports suggest that this promoter independence may not be mirrored in the *in vivo* situation. Driever et al (1989b) showed that the replacement of the *hunchback* promoter (-50 to +90) with the *hsp70* promoter (-50 to +90) in a responder construct in *Drosophila* embryos resulted in a lower level of responder expression. Vincent et al (1990) inserted 4 copies of the NP-homeodomain binding sequence upstream of the *hsp70* and *ftz* (-39 to +73) promoters driving a *lacZ* gene and transformed *Drosophila* embryos. They found glial-cell specific activation in the *hsp70* construct but not in the *ftz* promoter construct. That this activation is mediated by a homeodomain containing factor is supported by the observation that mutating the NP sites to give RP sites which greatly depresses binding (Desplan 1988) and activation in tissue culture cells (Jaynes 1988), abolishes glial-specific expression in the embryo. In an addendum Vincent et al. report that the insertion of 6 NP copies does result in *ftz* promoter specific expression in a different set of cells. No such promoter specificity is observed in tissue culture cells (Jaynes 1988).

It should be recalled that at high enough concentrations, *en* competes with TFIID for binding to the TATA box (Ohkkuma 1990a) This binding was promoter specific, depending on the sequence context of the TATA box.

The most plausible interpretation of these findings is that other cell or embryonic stage specific factors which can bind to sequences associated with specific promoters must be present or absent for homeodomain proteins to exert their *in vivo* function.

4. Co-operativity

Whether or not homeodomain proteins bind DNA as dimers or co-operatively [1.4], transcriptional activation can occur co-operatively (*ftz*, *bicoid*), with an increase in the number of recognition sequences leading to a disproportional increase in the extent of activation. In addition, *paired*, *zen* and *ftz* appear to function in a synergistic manner (Han 1989).

Homeodomain binding sites tend to be clustered with more than one cluster upstream of target genes (eg. the k1 to k5 sites upstream of *en*, A1-3 and B1,2 sites of *hunchback*, e4,5 of *eve* and the clustered sites upstream of the *Antp* P1 and *Ubx* promoters; see also [1.7]). This clustering together

with cooperativity can set various threshold levels of homeodomain protein concentrations below which no activation response is mediated. Such a mechanism of setting threshold levels is used in genes interpreting the *bicoid* gradient in early *Drosophila* embryos (discussed in the next section).

It is important to realize that a single binding site of 6 to 10 bp is not sufficient to assure occupancy by a regulatory protein especially when this protein can also recognize sites that deviate from the 'consensus' binding site. This is because in a typical genome of 10^8 - 10^9 base pairs the number of chance occurrences of sites that resemble biologically important recognition sites is very large (eg. an 8 bp sequence will occur about 1 : 4^8 times or 15000 times per 10^9 random bp and as more mismatches are allowed this number will increase rapidly).Thus to achieve functional specificity, either a large amount of protein has to be wasted by non-productive binding at sites occurring by a chance, or functionally important sites have to be clustered ensuring occupancy of at least one site. If bound proteins can interact to stabilize the protein-DNA complex, specificity is increased dramatically. For completeness it should be mentioned that making regulatory regions more accessible (eg. by a specific chromatin structure) would also aid functionally significant binding.

5. Repression

Transcriptional repression by homeodomain proteins may occur by several mechanisms though all seem to rely on DNA binding. There are three categories of DNA binding-dependent transcriptional repression :1.competition, 2.interaction with transcriptional activators, and 3. interaction with the basal transcription complex (see Levine 1989, Renkawitz 1990 for reviews).

En was shown to fit into category 1., competitively competing with homeodomain containing activators (*ftz*) for homeodomain binding sites, or , at high concentrations, with TFIID for the TATA box. When using overlapping octamer and spH motifs at a distance from the promoter, *Oct*-2 could reduce expression by competitively binding to the octamer site to the exclusion of the activating spH-binding factor. *Oct*-2 is normally an activator but cannot enhance transcription from a distant site in non-lymphoid cells (Tanaka 1990). Whether such repression is significant *in vivo* remains to be seen.

How *eve* represses is not known but the mechanism may fall into category 3, because *eve* can reduce the basal level of transcription. This interaction may involve the alanine rich region as deletion of this region impairs the ability of *eve* function as a transcriptional repressor (unpublished, mentioned in Licht 1990). A zinc finger protein, *Kruppel*, has been shown to repress via an Ala-rich region that can be aligned with *eve* (Licht 1990).
The category of *Ubx* repression is unclear but depends on certain sequence elements present in the *Antp* P1 but not *Ubx* promoter (Krasnow 1989).
6. RNA binding ?
Rebagliati (1989) discovered an RNA binding motif (RNA-CS type) in the *bicoid* protein but whether this is involved in RNA binding remains to be investigated.

1.7 Homeobox genes and Development

The discovery of the first homeoboxes in genes involved in development was not co-incidental. All of the genes containing an *Antp* type homeobox and for which a phenotype is known, either map to genetically determined developmental mutants or/and show temporally and spatially restricted patterns of expression during embryogenesis. More diverged homeoboxes are often expressed in a tissue specific manner and may be required for differentiation which is an integral process of development. *Oct*-1, ubiquitously expressed in adult mammalian tissues, seems to form an exception, acting simply as a promoter specific transcription factor.
Having described so far only the biochemical aspects of homeobox genes and their products, we shall now turn to their *in vivo* roles, taking examples from several phyla.

1.7.1 *Drosophila*

1.7.1.1 Early embryogenesis

The nucleus of the *Drosophila* egg undergoes 13 syncitial cleavage divisions during the last three of which the nuclei migrate to the periphery. At the 14 th division, membranes grow down from the cortex, enveloping each nucleus, so forming the cellular blastoderm. This process takes about 2,5 hr with 5000 cells being formed. By the completion of this event, the segmented prepattern of the fruitfly is already established. After this, gastrulation occurs, and the future abdomen folds over forming a double layered metameric structure with the posterior most segments lying adjacent to the head region .
Systematic genetic analysis revealed numerous mutations that affected this sequence of events (Akam 1987, Ingham 1988b), and these were classified into four main classes :
1. Coordinate mutants (usually maternally expressed) showing large alterations of body structure (eg. *bicoid*, *cad*).
2. Gap gene mutants where groups of segments are missing or altered (eg *hunchback*, *Kruppel*, *empty spiracles*).
3. Segmentation genes divided into a. Pair-rule genes with mutants showing changes in the periodic segmentation pattern (eg. *even-skipped*, *ftz*, *paired*) and b. segment polarity genes : mutants show changes in each segment (*engrailed*, *gooseberry*).
4. Homeotic gene mutations result in the change of identity of a single type of segment (most genes of the ANT-C and BX-C).
These genes act in a distinct temporal order corresponding to the order listed above during early embryogenesis as inferred from their expression patterns. Apart from the gap genes *hunchback* and *Krüppel*, all the listed examples are homeobox containing genes.

1.7.1.2 The *bicoid* protein gradient

The coordinate gene *bicoid* is required for the entire anterior half of the embryo and homozygous lethal mutants miss all head and thoracic structures (Frohnhofer 1986). *Bicoid* is transcribed during oogenesis and transported into the anterior egg pole where it is trapped (Berleth 1988, Driever 1988a). On egg deposition (independent of

fertilisation), *bicoid* RNA is translated and forms a stable concentration gradient with a maximum at the anterior tip followed by a two to three order of magnitude exponential decrease over slightly more than half the egg length with detection limits at 70 % of the egg length (Driever 1988a). The *bicoid* protein was found to be localized in the nuclei. By manipulating the density and distribution of *bicoid* mRNA genetically, using *Drosophila* females with one to four copies of the *bicoid* gene, and double mutants, it was shown that the height and shape of the *bicoid* protein gradient could be altered. Increases or decreases in the *bicoid* protein concentration in a given region resulted in a anterior or posterior shift respectively of anterior characteristics (Driever 1988b).The implication was that *bicoid* protein is a morphogen (a substance directing shape or morphology) providing positional information (Wolpert 1969) via an anterior-posterior concentration gradient, established by diffusion and degradation. The syncitial nature of the early *Drosophila* embryo makes the diffusion of such a large molecule (55 K) possible. The positional information laid down by a gradient has to be interpreted by cells. One such mechanism for the differential response of cells - or nuclei - would involve threshold levels (Wolpert 1969) : below a morphogen concentration x_i, gene i no longer is activated or repressed. *Bicoid* has been shown to contain a homeobox and to activate transcription in a binding site dependent manner [1.6]. That threshold levels for *bicoid* activation exist, was shown convincingly by Struhl et al.(1989a) and Driever et al.(1989b). These authors showed that the expression of the gap gene *hunchback* which is known to be positively regulated by *bicoid* (Tautz 1987), depends on both the *bicoid* gradient profile and the number and quality of the *bicoid* binding sites. Low affinity *bicoid* sites (deviating by 4 bases from the consensus sequence; see table 1.2) resulted in reporter gene expression only down to 33 % of the egg length whether there were 3, 6 or 9 copies while consensus binding sites resulted in gene expression in 35 % up to 49% down the egg for 1 to 6 consensus sites respectively (Driever 1989b, Struhl 1989a). Increasing the *bicoid* gene dosage in the embryo shifted the expression domain in a posterior direction (Struhl 1989a). These experiments demonstrate the potential of a single homeodomain protein gradient to selectively switch on multiple genes, depending simply on the number and nature of

binding sites in the upstream promoter region of the target gene [1.6].
Genes fulfilling the role of target genes at high *bicoid* concentration are the homeobox genes *empty spiracles* and *orthodenticle* which have expression patterns dependent on *bicoid* copy number (Dalton 1989, Finkelstein 1990, Cohen 1990). These genes, as well as *buttonhead*, are expressed in head segments in overlapping domains out of phase by one segment.
Apart from the gap genes involved in head segmentation, no other gap genes have been found to contain homeoboxes. However the three main gap genes *hunchback*, *Kruppel* and *knirps* contain a zinc finger motif which has been proposed to serve in DNA binding (Tautz 1987, Rosenberg 1986, Ingham 1988b) and these genes have been shown by genetic analysis to regulate each other (Jackle 1986), the primary pair-rule genes *even-skipped* (*eve*), *hairy* and *runt* (Gergen 1986, Macdonald 1986, Ingham 1988b, Goto 1989) as well as homeotic genes (Harding 1988).

1.7.1.3 The segmented subdivision of the embryo

The primary pair-rule genes are called primary because they are unaffected by mutations in *ftz* and other secondary pair-rule genes but mutations in them disrupt the 2° pair-rule expression patterns. The expression pattern of all pair-rule genes consists of a periodic arrangement of 7 transverse stripes with a repeat unit of 2 segmented units.
Both *eve* and *hairy* have complex upstream regulatory regions and expression in individual stripes has been correlated to different upstream regulatory elements, each of which responds to nonperiodic cues provided at least in part by the gap genes (Ingham 1988b, Goto 1989). At the 14 th cleavage division which corresponds to the cellular blastoderm stage, *eve* expression develops as 7 stripes for both mRNA and protein. These stripes which are out of phase with *hairy* and *runt* are individually regulated while an additional 7 'late' *eve* stripes, which appear between the initial stripes during gastrulation, have been shown to be under the control of an 800 bp region about 5 kbp upstream of the *eve* start site. This region responds to the periodic cues from all three 1° pair-rule genes : *hairy* induces, *eve* enhances and *runt* represses transcription (Goto 1989). *Eve*'s second phase of expression thus resembles that of secondary

pair-rule genes (see below). It is interesting that the autocatalytic activation function of *eve* has not been observed in transfection or *in vitro* experiments [1.6] and one should bear in mind that in the above mentioned genetic experiments one cannot unequivocally deduce that observed effects are direct. Unidentified intermediate proteins may mediate the observed effects. We shall later discuss further roles of *eve* in neuronal differentiation and in the development of the visceral mesoderm.
The main function of the 1° pair-rule genes is to interpret the broadly dispersed overlapping domains of the gap genes (by combinatorial mechanisms ?) and divide the embryo into double segmented units. The 2° pair-rule genes interpret these out of phase periodic patterns to establish the 13 parasegmental borders. A parasegment consists of the posterior compartment (at blastoderm stage 1-cell wide) of one future segment and the anterior compartment (3 cell wide) of the following segment. Parasegments are thus out of phase with the final segmented subdivision of the embryo.
It is interesting that three of the 2° pair-rule genes ('late' *eve*, *ftz* and *paired*) contain homeoboxes all of which recognize different consensus sequences (table 1.2) though *eve* can also bind to *ftz* sites. These three proteins can act in a combinatorial fashion (Han 1989, [1.6]) giving a large spectrum of regulatory possibilities.
Ftz mRNA is first seen as a continuous band that resolves into 7 stripes at two segment intervals by the 13th nuclear division and *ftz* protein is detected between cellular blastoderm formation and gastrulation, and persists throughout gastrulation (Hafen 1984a, Carroll 1985). A second phase of expression is seen during the development of the central nervous system (CNS). Three controlling elements in the 5' flanking region of *ftz* have been identified : the zebra element which is necessary for all 7 stripes and which is controlled by *hairy* and *runt*, an upstream enhancer element which requires the *ftz* protein itself for its effect and a neurogenic element which is involved in the second phase of *ftz* expression (Hiromi 1987). The upstream enhancer element was shown to contain NP-type homeodomain binding sequences and to be bound by numerous proteins, protecting 50 % of the 2,6 kbp region investigated (Harrison 1988). The zebra element is also a mosaic of regulatory elements containing a region that contains several activator elements which allow *ftz-lacZ* constructs to be expressed in

transformed embryos in a continuous band. It also contains pair-rule repressor elements which are necessary for the inhibition of *ftz* expression in the interstripe regions (Dearolf 1990). *Caudal*, a maternal effect coordinate gene expressed in a shallow posterior high- anterior low gradient in the early embryo, has been shown to be one of the activators of ftz expression (Dearolf 1989). The short half life of ftz mRNA of 6-8 min (Edgar 1986) suggests that the ftz striped pattern is produced by activation of ftz throughout most of the embryo followed by repression of transcription in the interstripe regions (possibly by the out of phase 1° pair-rule gene hairy). Rapid RNA degradation sharpens the ftz mRNA stripes which are then maintained throughout gastrulation by the autocatalytic activation function of the ftz protein (Edgar 1986, Dearolf 1990). The 2° pair-rule genes function to establish the expression domains of the segment polarity genes (engrailed [en], wingless and gooseberry) and of certain homeotic genes. At the end of cellularisation and the onset of gastrulation en and wingless mRNA accumulates in 14 narrow stripes (DiNardo 1985, Ingham 1988a). En protein appears shortly thereafter in bands only a single cell wide, representing the anterior limit of a parasegment (or posterior compartment of a future segment, Ingham 1988b). Genetic evidence indicates that both eve and paired are required for the odd-numbered stripes while ftz and odd-paired or odd-shaped are both required for en expression in even-numbered stripes (Ingham 1988a, DiNardo 1988). The posterior parasegment border is characterized by wingless expression and seems to be regulated in a negative fashion by ftz and eve (Ingham 1988b). It is thus clear that a combinatorial action of the pair-rule gene products is required to establish the limits of each parasegment by the end of the blastoderm stage. Three of the mentioned pair-rule genes (ftz, eve, paired) as well as two of the segment polarity genes (en and gooseberry) contain homeoboxes and the upstream regulatory region of en contains multiple ftz and eve binding regions (table 1.2) but again the in vivo postulated transcriptional activation role of eve does not agree with in vitro results.

1.7.1.4 The homeotic genes

Simultaneously with the establishment of parasegments, the regional identity of individual parasegments is determined by the homeotic genes mainly of the ANT-C and BX-C, most of which contain an *Antp* type homeobox. Genetic analysis of single and double mutants implicated mainly gap genes and 2° pair-rule genes in the control of homeotic genes, though the homeotic gene *deformed* has been reported to be activated by the coordinate gene *bicoid* (Jack 1990). Homeotic genes also show extensive crossregulation amongst themselves.
Deformed is negatively regulated by *ftz* and positively by the combination of *bicoid*, the gap gene *hunchback* and *eve*, though *eve* may be replaced by a combination of other pair-rule genes (Jack 1990).
Antennapedia (*Antp*) has two independently controlled promoters, P1 and P2. P1 is positively regulated by the gap gene *Kruppel*, P2 by *hunchback* and *ftz* (Irish 1989, Harding 1988). *Antp*, *Ubx* and *abdominal*-B (*Abd*-B) have massive regulatory regions. The regulatory regions of Ubx entail a 25 kbp region encompassing the *abx* and *bx* mutations which affect expression in parasegment 5 (PS 5) while the 40 kbp *bxd/pbx* region is responsible for expression in PS 6 (Beachy 1990). The *bxd* promoter is repressed by *hunchback* while the more proximal *ubx* promoter (*abx/bx*) is in addition positively regulated by *Kruppel* and *ftz* (Irish 1989). *Eve* and *en* may also positively regulate *Ubx* transcription (Biggin 1988a,b, Beachy 1990) and purified *zeste* and a GAGA-binding protein have been shown to bind to and activate the *Ubx* promoter (Biggin 1988a,b). *Abd*-B is negatively regulated by the gap genes *knirps* and possibly *Kruppel* (Harding 1988). These genes thus establish the initial pattern of homeotic gene expression but during later stages (germ band extension), coordinate, gap and many pair-rule gene products are no longer detectable. During this phase interactions between homeotic genes (Lewis 1978, Struhl 1982, Akam 1987,1988), segment polarity genes (Akam 1987) as well as general repressors such as *Polycomb* (Lewis 1978) refine and maintain the expression domains.
Lewis (1978) uncovered a remarkable phenomenon amongst the genes of the BX-C complex that has since been found to be applicable to all the ANT-C and BX-C homeotic genes : the expression domain of these genes correlates with the order in which they are arranged on the chromosome. Thus *labial*

the 3' most gene is expressed anterior of *dfd* in the head (Mlodzik 1988). *Dfd* is responsible for the mandibular and maxillary segments (PS 0 and 1, Regulski 1987), *scr* the next gene in line is expressed in labial and 1st thoracic (T1) segments (PS 2 and 3; LeMotte 1989), *Antp* specifies T2 and T3 (or PS 4 and 5; Schneuwly 1987), *Ubx* the next homeotic gene in sequence though in a cluster that is a third of a chromosome away specifies T3 and the first abdominal segment A1 (PS 5 and 6; Beachy 1990), *abd*-A is expressed in A2 to A4 (PS 7-9) and *Abd*-B in A5 to A9 (or PS 10-14)(Casanova 1986, Gonzales-Reyes 1990). Furthermore, each one of the more posteriorly expressed genes represses the expression of anterior genes initially active within their domains (Lewis 1978, Hafen 1984b, Gonzales-Reyes 1990).

Thus while the anterior margin of expression is established by the gap and pair-rule genes, the posterior margin is set through the repression by more posterior homeotic genes. In the case of *dfd* (Kuziora 1988) and *Ubx* (only in the visceral mesoderm) autoregulation has been shown to maintain expression.

An alternate model for the establishment of the anterior margin of expression has been proposed (Peifer 1987, Akam 1988, Gaunt 1990). This model proposes two states for a gene :'open for business' being a state in which the gene can be activated whilst 'closed' refers to a state insensitive to activation. Furthermore as one progresses to more anterior regions of the organism, more of the linearly arranged homeotic genes are 'closed'. Hence *Ubx* for example cannot be expressed in T2 (or PS 4) because its regulatory areas are 'closed' in this anterior domain. *Antp* lying more 3' on the chromosome is 'open'in this segment (and up to T1) and thus can be expressed; and so on. This model would explain the correlation between chromosome location and expression domain along the anterior-posterior axis. Gaunt et al. (1990) propose that the 'open and closed for business' states refer to active (expressible) and inactive chromatin (Reeves 1984) and that the *Polycomb* product, the absence of which causes anterior segments to be transformed into more posterior segments (Lewis 1978), mediates the heritable spreading of a heterochromatin structure along the homeogene clusters. It should be noted that the *ftz* gene required for *Scr* and *Antp* expression, is located between these genes on the chromosome and possibly would be shut down prematurely in this interpretation. Stubblefield (1986) implicated cell

division as the mechanism that passes gene control sequentially from one control unit to the next.
At gastrulation the identity of segments is established by the specific expression of homeotic homeobox genes while cell identities within each segment are further defined by the expression and interaction of the segment polarity genes (Akam 1988, Ingham 1988a).

1.7.1.5 Interpretations

It is clear from the above descriptions that the biochemical results do not always correlate with the postulated *in vivo* roles of homeodomain proteins. This is not surprising in view of the following considerations :
1. Genetic studies seldom unambiguously show whether an observed regulation is direct or mediated by other products.
2. The principle of 'double insurance' which states that any important event in development is regulated independently by several separate processes (see Gurdon 1988) may seriously affect the interpretations of experiments since changing one factor may have little effect on a process simply because it is not the sole factor.
3. The biochemical binding and activation studies themselves give different results depending on the size of the upstream regulatory region used and the context of the reaction.
4. The whole picture is complicated by the fact that many homeodomain proteins are able to bind the same sequence such that *in vitro* binding studies usually cannot predict *in vivo* significance.
The most clear correlation between biochemical and genetic data undoubtedly is that of *bicoid* activation of the *hunchback* gene. At later stages usually more than one homeobox gene is expressed simultaneously in one cell and the effect of these homeodomains on different regulatory regions may differ. For example, *Ubx* is activated by *ftz* and repressed by *Abd*-B while *ftz* is activated by its own product but *Abd*-B shows no effect (Biggin 1989b).
These effects may be explained by, firstly, the large regulatory areas of these genes which contain multiple clustered binding sites thereby allowing combinatorial interactions of homeodomain proteins.
Secondly, the affinity of homeodomain proteins for sites diverging from the optimal sequence is different [1.6] such that one homeodomain may outcompete another.

Thirdly, there are numerous other factors which also bind in the regulatory area and interact with specific homeodomains to change or enhance their regulatory effects. The factors binding to the *Ubx* and *ftz* regulatory area have already been mentioned and *en* has been shown to interact with at least 12 distinct proteins (Gay 1988).
A fourth possibility of attaining differential regulatory effects is by posttranslational modification of which phosphorylation is the most widely observed [1.6].

1.7.1.6 Later embryogenesis

Many homeobox genes are deployed in a second or even third phase during development, often under the control of distinct regulatory regions : the 'neurogenic' element of *ftz* has already been mentioned. This may reflect a principle of the process of evolution - it is easier to generate new controlling regions (note that typical consensus sequences are only around 10 bp long) than new DNA binding regulatory proteins which anyhow would need their own regulatory regions. We shall briefly look at the deployment of homeobox genes in the visceral mesoderm, the central nervous system (CNS) and in the specification of the dorsal-ventral embryonic axis.

After establishing segment identity shortly after the cellular blastoderm stage, the homeotic genes are expressed in somewhat different domains in the visceral mesoderm (Tremml 1989a,b), neural ectoderm (CNS) and ectoderm (Levine 1985, LeMotte 1989) but the order of expression along the anterior-posterior axis remains the same. In the visceral mesoderm homeotic gene expression is independent of gap genes (Tremml 1989b) and the anterior limit of each homeodomain is shifted posteriorly by one segment compared to the ectoderm (Tremml 1989a) possibly through interaction with one of the genes establishing the embryonic dorsal-ventral axis. *Zerknullt* (*zen*) and *z2* are so far the only homeobox genes known to affect this axis (Doyle 1986, Rushlow 1987). The three constrictions that form in the visceral mesoderm-adjoining midgut coincide with *Antp*, *Ubx/Abd*-A boundary and Abd-A expression and mutations of the respective gene(s) results in the failure of these constrictions to form (Tremml 1989a).

Ftz and *eve* are expressed in every segment in the CNS. *Ftz* is transiently expressed in a specific subset of segmentally repeated neuronal precursor cells, neurons and glia cells (Doe 1988a). Some of these neurons also express *eve*, *en* and, in T2 to A6, also *Ubx*. Doe et al. (1988a) showed that *ftz* is required for the normal expression of *eve* and *Ubx* but not *en*. This control differs from that in the blastoderm where *ftz* regulates *en* and *Ubx* but not *eve*. Mutations in either *eve* or *ftz* causes the transformation of a certain type of neuron (RP2) in which both genes are normally expressed, to that of neuron RP1 (Doe 1988b).
Other genes with highly diverged homeoboxes also control the fate of specific neurons. *Cut*, coding for a massive protein of 2175 a.a. whose homeodomain contains a histidine at position 50 in the recognition helix [1.4], has been shown to be involved in the development of es sensory organs (Blochlinger 1988). *Rough*, which has a homeodomain showing 57 % homology to *Antp* is required for the proper differentiation of photoreceptor cells that comprise the omnatidia of the compound eye (Saint 1988). *Rough* is also expressed in the larval brain but not during early embryogenesis. A highly conserved POU-domain is found in the *Cf1a* protein (Johnson 1990). This protein was shown to bind to a DNA element that is necessary for the expression of the dopa decarboxylase gene in specific dopaminergic neurons, and thus may be a neuron specific transcription factor (Johnson 1990).

1.7.2 Vertebrate

Although many vertebrate homeobox genes have been isolated none have so far been mapped to a known developmental mutant. Therefore the powerful genetic techniques which have clarified the role and interactions of *Drosophila* homeobox genes, cannot be applied to vertebrates. Hence functions had to be inferred from the temporal and spatial expression patterns. In *Xenopus*, injection studies with homeobox mRNA and antibodies against homeodomain proteins have been performed. Transgenic mice overexpressing homeobox genes have yielded further insights while the most advanced technique of *in vivo* mutagenesis (Zimmer 1989, Joyner 1989) by which genes can be specifically mutated, represents a powerful tool for dissecting developmental pathways.

The first homeodomain protein to be detected during vertebrate development is *Oct*-3, a POU-domain mouse transcription factor which is already expressed in maturing oocytes (Rosner 1990). Zygotic *Oct*-3 expression is activated between morula 12,5 days *post coitum* (d.p.c.) and blastocyst (3,5 d.p.c.) formation in several germ layers but is down-regulated at 7 d.p.c. except in the primordial germ cells. This expression pattern differs considerably from all other vertebrate homeobox genes.
In the frog *Xenopus laevis* the homeobox gene *Xhox*3 was shown to be involved in pattern formation along the anterior posterior axis (Ruiz i Altaba 1989a,b). *Xhox*3 is one of the first genes to be transcribed in the early embryo and is expressed in a graded fashion (anterior low) along its anterior-posterior axis in the axial mesoderm (Ruiz i Altaba 1989a). Injection of *Xhox*3 mRNA into fertilized eggs results in a uniform high level of mRNA up to neurula stage and such embryos fail to develop correctly in anterior regions where the gene is normally expressed at a low level. *Xhox*3 thus seems to be involved in interpreting positional information of the anterior-posterior axis although it is not sufficient on its own to convert anterior structures into posterior ones. Ruiz i Altaba and Melton (1989b) showed that *Xhox*3 expression is induced by maternally encoded peptide growth factors. Both XTC-MIF, a mesoderm inducing factor of the transforming growth factor-β family and bFGF, a basic fibroblast growth factor can influence, in a concentration dependent manner both *Xhox*3 mRNA levels and the anterior-posterior character of the mesoderm. Retinoic acid or platelet growth factor do not produce this effect.
Another homeobox gene, *mix1*, which is expressed even earlier is also inducible by XTC-MIF or FGF (Rosa 1989). This supports the view that these growth factors represent morphogens and mediate their effect, at least partly, via homeobox genes. The specification of positional information in the amphibian mesoderm is also crucial to that of the ectoderm which responds to inducing signals from the mesoderm (Spemann, see in DeRobertis 1989).
Other *Xenopus* homeobox genes involved in anterior-posterior development are *Xl*Hbox1 and *Xhox*1A. Injection of *Xhox*1A (Harvey 1988) and the longer species of *Xl*Hbox1 mRNA (Wright 1989) disrupt the segmentation pattern of the somitic mesoderm while the short species of *Xl*Hbox1 mRNA and antibodies against the long *Xl*Hbox1 protein cause spinal

chord malformation in the cervical region where *XlHbox*1 is normally expressed (Cho 1988, Wright 1989).
As already mentioned [1.2], practically all of the mouse and human *Antp*-type homeoboxes are arranged in 4 clusters, termed *Hox*, on separate chromosomes. These clusters have arisen by duplication of an ancestral cluster which may also have given rise to the insect HOM cluster (Akam 1989b, Schughard 1989, Kappen 1989). On the basis of sequence homology individual members of different clusters can be grouped into subfamilies though each subfamily may not be represented in every cluster (Gaunt 1988, Acampora 1989). So far each *Hox* member has been shown to be expressed in the central nervous system, and the linear order of the genes along the chromosome correlates with the spatial order of their anterior borders of expression (Gaunt 1988, Acampora 1989, Duboule 1989, Graham 1989). This is very reminiscent of the *Drosophila* situation and may reflect an ancient 'cassette' of homeobox genes that specified anterior to posterior domains (see also section 1.7.4 on nematodes). The fact that vertebrates contain 4 such cassettes or clusters may be a consequence of the greater complexity of these organisms.
Analysis of the expression of homeobox genes in the mouse CNS indicated a segmented distribution of these transcripts in the vertebrate hindbrain (Murphy 1989, Wilkinson 1989b). The hindbrain shows bulges early in development which have been termed rhombomeres. These are believed to be segmented units based on their pattern of nerve formation in the developing chick (Lumsden 1988). Wilkinson et al.(1989b) showed that the anterior boundaries of expression of the *Hox* 2.6, 2.7 and 2.8 genes coincide with the anterior boundaries of rhombomeres r7, r5 and r3 respectively in the 9,5 day mouse embryo, whereas *Hox* 2.1 was expressed up to the hindbrain-spinal chord boundary which lies at the posterior boundary of r8. These four genes are thus expressed at two segment intervals with the order of anterior most expression correlating with the order of the gene position on the chromosome. The *Hox* 2 order is : 5'-*Hox* 2.5, 2.4, 2.3, 2.2, 2.1, 2.6, 2.7, 2.8, 2.9 -3'. The 3'-most gene, *Hox* 2.9, unlike the other *Hox* genes which are expressed down to the posterior end of the neural tube, is expressed precisely and specifically in rhombomere r4 (Murphy 1989). Another regulatory gene, *Krox20*, that has been shown to bind *in vitro* to upstream regions of the *Hox* 1.4 gene (Chavrier

1990), is expressed specifically in rhombomeres r3 and r5 framing the domain of *Hox* 2.9 expression (Wilkinson 1989a). It will be interesting to see whether these genes interact. The anterior limit of *Hox* 1.5 expression, like *Hox* 2.7, coincides with that of rhombomere r5 (Gaunt 1986, Murphy 1989). Both genes belong to the same homeobox subfamily. Such correlation amongst genes of different clusters but the same subfamily is not always observed (Duboule 1989). Another point of interest is that genes expressed more anteriorly (and located more 3') are also expressed earlier during development (Duboule 1989).

The functions of these genes is still unclear. Overexpression of *Hox* 1.1 (same subfamily as *Hox* 2.3) in transgenic mice resulted in craniofacial abnormalities (Balling 1989). Expression normally occurs posterior to the level of the fourth ganglion in the CNS but in the transgenic mice the gene was expressed ubiquitously, that is, also ectopically in the anterior CNS. The authors point out that retinoic acid administration during pregnancy produces similar effects and it has been shown that retinoic acid induces the expression of many homeobox genes, including *Hox* 1.1, in teratocarcinoma cells (Colberg-Poley 1985, Deschamps 1987). They suggest that overexpression of *Hox* 1.1 (by retinoic acid administration or by directly increasing *Hox* 1.1 expression levels) in anterior regions leads to the observed abnormalities.

The *Hox* cluster genes are also expressed outside the CNS in an equivalent anterior-posterior manner. In the mesodermally derived prevertebral column the anterior expression boundaries are more posterior than in the CNS (Gaunt 1988). Expression in the mesoderm of organs correlates to the presumed position along the anterior-posterior axis of their founder cells in spite of these structures not showing any segmentation (Gaunt 1988, Duboule 1989, Graham 1989). The observation that homeobox gene expression domains in the mesoderm are posteriorly displaced with respect to those of the CNS is again reminiscent of *Drosophila* where expression of homeotic homeobox genes in the visceral mesoderm is shifted posteriorly [1.7.1].

It is intriguing that a correlation between expression domains and location of a *Hox* gene in the cluster also exists in the developing limb. Transplantation and grafting studies (see Lewis 1989a for review) have indicated two mechanisms defining the proximo-distal (base to tip) and

anterior-posterior (thumb to little finger) axes of development. The proximo-distal axis depends on the distally located apical ectodermal ridge whereas the anterior-posterior axis is specified by the zone of polarizing activity (ZPA) found at the posterior margin of the limb bud and moving distally as the limb grows. It is believed that the ZPA emits a morphogen, thought to be retinoic acid, which specifies anterior-posterior identity in a concentration dependent manner (see Eichele 1989 for review). Dolle et al (1989a,b) showed that the *Hox* 5 (since renamed *Hox* 4) homeobox genes show expression domains that are restricted along the proximo-distal axis of the developing limb according to their position in the complex (5'-*Hox* 5.6, 5.5, 5.3, 5.2 -/25 kbp containing *Hox* 4.3/- *Hox* 5.1 -3'). The more 3' a gene, the more distally restricted is its expression domain in the posterior area of the limb and the later it is expressed. *Hox* 5.1 is an exception in that it is expressed in the whole limb and flanking mesenchyme. Whether there is a causal relationship between retinoic acid emitted by the ZPA and homeobox gene expression remains to be investigated.

Another homeobox gene involved in limb development is *Hox* 7.1 (*msh* class homeodomain), which is initially expressed throughout the limb bud but becomes restricted to the most distal portion of the ectoderm and underlying mesenchyme. This is the region where the apical ectodermal ridge is located (Hill 1989, Robert 1989).

In contrast, the newt homologue of *XlHbox*1 and *Hox* 6.1 (or *Hox* 3.3), called NvHbox1 which is expressed during newt limb regeneration, is expressed at a higher level in proximal regions and unaffected by retinoic acid (Savard 1988, Tabin 1988). Oliver et al., using antibodies against the *Xenopus* *XlHbox*1 protein, demonstrated a gradient of *XlHbox*1 protein with the maximum in the proximal anterior region of the forelimb bud. Hindlimb bud mesoderm was devoid of *XlHbox*1 indicating an early molecular difference between arm and leg.

Homeobox gene products show a lineage restricted expression in mammalian haematopoietic cell lines. Four human *Hox* 2 genes (*Hox* 2.1, 2.2, 2.3, 2.6) and a clone described as PL1 were used to screen 18 human leukaemic cell lines representing erythroid, myeloid, and T- and B-cell lineages (Shen 1989). Each cell type exhibited expression of at least one homeobox gene and when differentiation was induced in

two of the cell lines, the levels of homeobox mRNA changed. *Hox* 2.2 was observed to be expressed also in normal bone marrow cells. Similar results were obtained by Kongsuwan et al (1988) who isolated homeobox cDNA clones from bone marrow and spleen of adult mice and a human leukaemic cell line. These authors and Blatt et al. (1988) observed a viral insertion 5' to *Hox* 2.4 resulting in constitutive expression of the gene in myeloid leukaemic cells. Expression of *Hox* 2.4 could not be detected in 7 other myeloid leukaemia cell lines (Blatt 1988) nor 28 other haematopoetic cell lines (Kongsuwan 1988, 1989), raising the possibility of a connection between the viral insertion and the production of this leukaemic cell line. Whether a causal relationship in addition to the correlative one exists between cellular differentiation during haematopoiesis and homeobox gene expression, remains to be established.

1.7.3 Echinoderms

The embryological stages of the sea urchin are readily accessible in large quantities making this organism ideally suited for developmental studies. However, as in the case of *Xenopus*, genetically very little is known, making the assignment of phenotype to specific genes very difficult. As the echinoderms represent an intermediate branch point on the evolutionary line leading to vertebrates and have been well researched by classical embryology (Giudice 1973, Hörstadius 1939), it is of value to examine their relationship to homeobox genes.
To date four homeobox genes of the sea urchin *Tripneustes gratilla* and a homologue of TgHbox1 in *Strongylocentrotus purpuratus*, SpHbox1, have been described (Dolecki 1986, 1988a,b, Angerer 1989) and it is known that *T.gratilla* contains at least another homeobox (Dolecki 1986). TgHbox1 and hb3 are of the *Antp*-class (Dolecki 1986, 1988a), hb4 the *Abd*-B class (Dolecki 1988a) and SU-hb-en of the *engrailed* class (Dolecki 1988b). All four genes show stage and/or tissue specific expression in sea urchin embryos and adults. Hb3 is expressed at low levels in gastrula and adult small intestine and at high levels in adult gonads (Dolecki 1988a). Hb4 shows two transcripts that are developmentally regulated in different ways. A 3,7 kb transcript increases in abundance from blastula to pluteus while a 4,4 kb transcript reaches a maximum at gastrula and is hardly

detectable by pluteus (Dolecki 1988a). Most adult tissues show expression. SU-hb-en could not be detected in embryogenesis and was found mainly expressed in Aristotle's lantern (mouth apparatus) of adults, with weak expression in gonads (Dolecki 1988b).
Most work has been done on TgHbox1 which is expressed from blastula to pluteus with a maximum at gastrula. A second larger sized transcript is only present at gastrula stage (Dolecki 1986). Expression is found in adult tissues at 20 to 80 fold lower levels than during gastrula. Using in situ hybridization, Angerer et al (1989) showed that Hbox1 in *T.gratilla* and *S.purpuratus* embryos is initially expressed in cells of the aboral ectoderm but the domain of expression becomes progressively more restricted toward the vertex as the embryo develops from gastrula to pluteus. Comparison with the temporal expression pattern of aboral-specific genes (*Spec1* and *actin* CYIIIa) revealed that Hbox1 was expressed later which implies that it is not involved in the initial determination or differentiation of aboral ectoderm cells.

1.7.4 Nematodes

Nematode homeodomains were first isolated in 1988 (Way 1988, Finney 1988, Costa 1988) but recently another 11 sequences of *Caenorhabditis elegans* have been reported (Bürglin 1989, Kamb 1989, Schaller 1990, Freyd 1990) and it is believed that *C.elegans* contains about 60 different homeobox genes representing 1% of the genome (Bürglin 1989). Several of these genes are involved in the differentiation of specific cell types : *lin*-11 is required for the asymmetric division of the 2° vulval blast cells to produce daughter cells of different fate (Freyd 1990), *mec*-3 specifies the differentiation of the touch receptor neurons (Way 1988) and *unc*-86 (Finney 1988). It is exciting that both *labial* and *Antp*-class homeodomains have been found : *ceh*-13 and *mab*-5 respectively (see table 1.1, Schaller 1990, Costa 1988). Furthermore, *ceh*-11 shows 57 % identity (Schaller 1990) and *ceh* 15 shows 71 % identity (only 31 a.a. sequenced) to *Antp* (Kamb 1989). In addition, these 4 genes are in close proximity forming a type of cluster (Schaller 1990). These findings suggest that the common ancestor of nematodes, arthropods and vertebrates may already have contained a homeobox cassette. Other homeodomain classes are also

represented in the nematode: *caudal* (*ceh*-3) and POU (*unc*-86 and *ceh*-6)(see table 1.1).
Considering 1. the number of *C.elegant* mutants available, 2. the fact that a library of ordered cosmid clones representing the entire genome of only 8×10^7 bp is near completion and, 3. that a complete cell-by-cell description of the development of *C.elegans* has been published (see Emmons 1987), one can expect a rapid elucidation of the role of homeodomain proteins in nematodes.

CHAPTER 2

THE ISOLATION OF A SEA URCHIN HOMEOBOX

This section describes the construction of a sea urchin genomic library and the screening thereof resulting in the isolation of a homeobox containing gene fragment.

2.0 Index

2.1 Library construction

Dependant on the starting material, one can choose between cDNA and genomic DNA when constructing a recombinant DNA library. Factors influencing the choice include spatial and or temporal information on the expression of the gene to be isolated and the purpose of the isolation.
Preliminary experiments indicated the presence of genomic fragments hybridizing to an Antennapedia (Antp) homeobox probe at low stringency whereas no expression during embryogenesis could be determined with this probe. Hence the course to follow lay in the construction of a genomic library.
The size of a library of completely random fragments of genomic DNA that is necessary to ensure representation of a particular sequence is determined by the size of the genome and the average size of the cloned fragments. Clarke and Carbon (1976) have derived the following relationship :
$N = \ln(1-P) / \ln(1-[F/G])$
where N gives the number of independent clones that must be screened to isolate a particular single copy fragment with

probability, P; F is the size (in bp) of the average cloned fragment and G is the size of the genome (in bp). Most species of sea urchin contain ≈ 1 pg DNA per haploid cell (Fasman 1976) which is equivalent to 9 x 10^8 bp. Modern vectors such as lambda EMBL3 can take inserts of the size 9 - 23 kbp (average 16000 bp). Thus, for a 99% probability of isolating a single copy gene of interest, N = ln(1-0,99) / ln(1-[16000/9x10^8]) = 260000.
EMBL3 DNA digested to completion with BamH1 and EcoR1, followed by treatment with alkaline phosphatase, was purchased from Promega.
Parechinus angulosus sea urchin genomic DNA was prepared from 14 hr embryo nuclei of an unknown number of individuals [7.2.1]. Such a library derived from multiple individuals allows the investigation of gene variants at the level of the DNA sequence. To avoid any bias in the library it is essential to ensure that cleavage of the genomic DNA occurs as randomly as possible. This was achieved by partially digesting the DNA with a restriction enzyme cutting frequently. NdeII was chosen because 1. it has a 4 bp recognition sequence (occurring 1 : 4^4 bp) with no reported bias in selecting one site over another and 2. the ends generated are compatible with the BamH1 generated ends of the EMBL3 vector used.

NdeII partially restricted DNA was electrophoresed on a 0,5% agarose gel (fig.2.1) and scanned (fig.2.2). Randomness is maximized by partial digestion to an extent where the size of the numerically most abundant class of partial products equals the vector capacity (fig.2 in Seed 1982). The relationship between the length corresponding to the maximum fluorescence intensity and the number average length of the DNA fragment distribution is given by $x_m = 2/L$ where x_m is fragment size at maximum fluorescence and L is the frequency of cleavage (which is the inverse of the number average length) (Seed 1982, Pfeffer 1985). From fig 2.2 digestion conditions were chosen where $x_m \approx 30000$ bp giving a number average length (1/L) of 15000 bp which falls into the middle

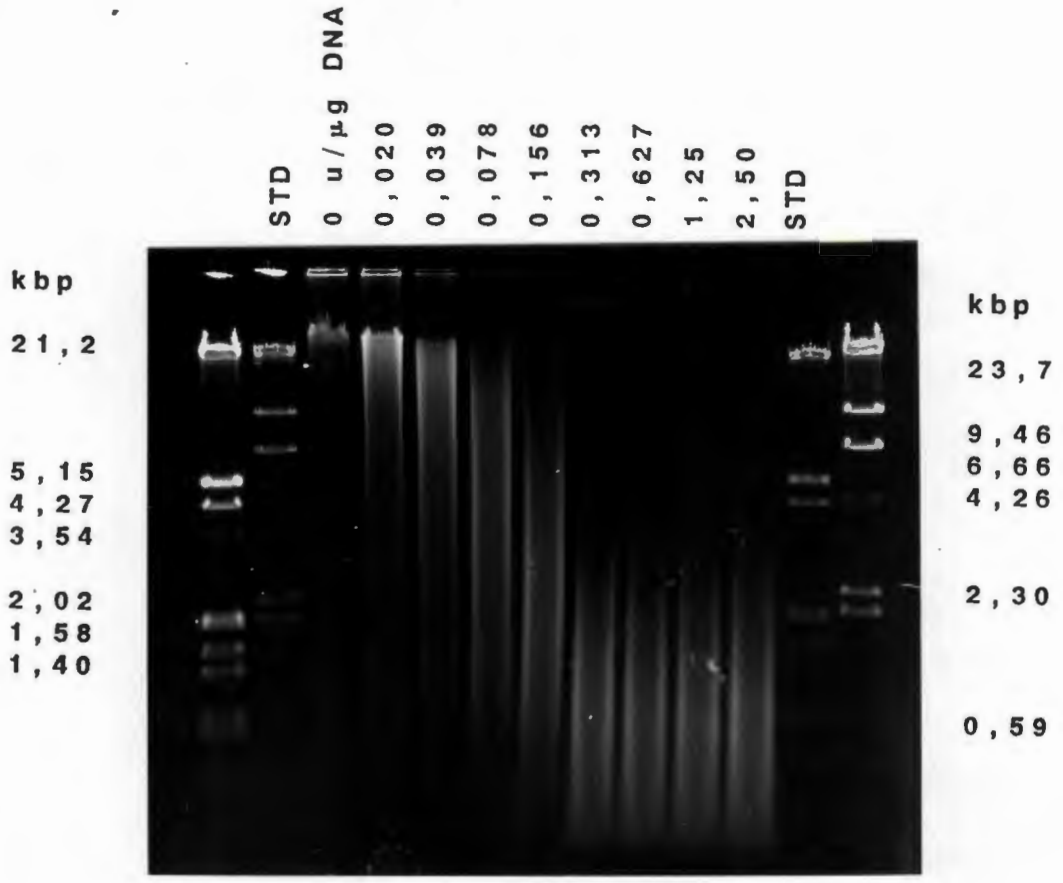


Figure 2.1. Nde II restriction kinetics. Genomic DNA was digested with 0 to 2,5 units Nde II (BRL) per µg DNA for 2 hr in 100 mM Tris.HCl pH 7,60, 10 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT at 37°C. The reaction was stopped with 25 mM EDTA (final concentration), and the samples electrophoresed at 1 Volt/cm on a 0,5% agarose TBE gel for 18 hr. The size standards (STD) are from left to right: lambda DNA digested with HindIII and EcoRI and lamda digested with HindIII only. Scans of the first five lanes are shown in fig.2.2.

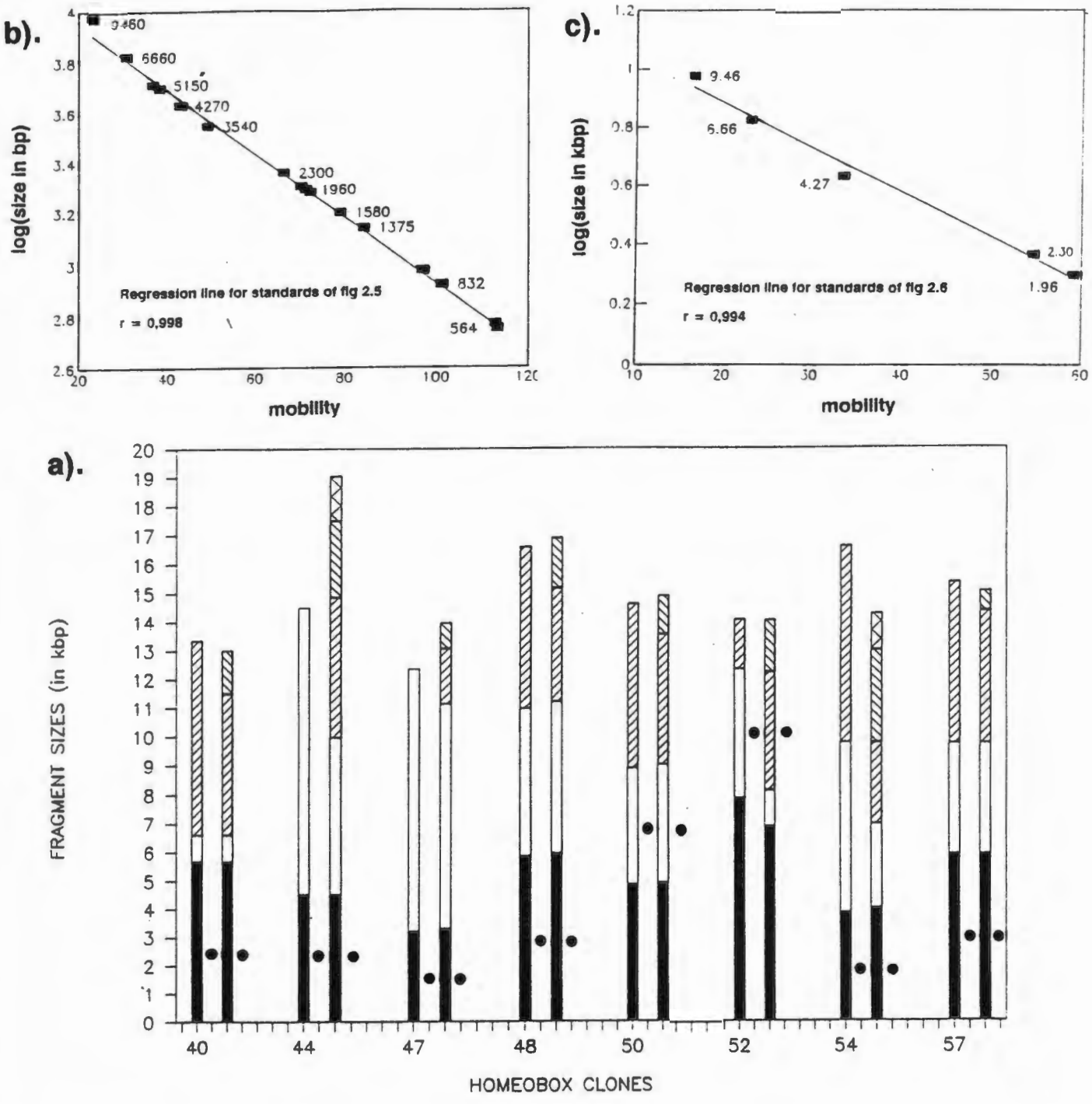


Figure 2.7. Histogram of restriction fragment sizes. EMBL3- clones 40 to 57 were digested with SalI (left) and SalI + EcoRI (right) and electrophoretically separated as shown in fig. 2.5 and 2.6. The sizes of fragments was determined using the regression lines depicted in *b* (for clones 40, 47, 48, 50, 52 and 54; see fig. 2.5) and *c* (for clones 44 and 57; see fig. 2.6). The EMBL3 arms are not indicated. The size of fragments greater than 8 kbp is significantly underestimated because the relationship of mobility α log(fragment size) no longer holds for large fragments (see correlation line and Seed, 1982). This explains discrepancies in the difference in total stack height for SalI and SalI + EcoRI for some of the clones. Individual restriction fragments are indicated by different types of hatching and the homeobox-containing fragments are marked with a filled circle. r = regression coefficient, RSD = residual standard deviation of log(size).

of the vector capacity size range. Using these conditions, 300 μg of genomic DNA from multiple individuals was digested, analyzed on a 0,4% agarose gel and fractionated by equilibrium centrifugation on a 10% to 40% sucrose gradient [7.5.3.6]. Fractions were electrophoresed on a 0,5% agarose-TBE gel (fig.2.3) and aliquots 28 to 34 which were in the appropriate size range (10-20 kbp) were pooled.

2.2 Ligation

To ensure efficient packaging, the bacteriophage lambda arms should be ligated to inserts in the form of concatemers $(\text{left arm-insert-right arm})_n$. Both the concentration of each of the DNA species and the ration of insert to lambda arms have to be in the correct range. Although the theoretical optimal molar ratio of arms to inserts is 2:1 (Maniatis 1982), some of the DNA arms may lack cohesive ends thereby changing the effective concentrations. Hence a series of test ligations was set up with molar ratios (arm:insert) of 1:1, 2:1, 4:1 and 8:1. The concentration of lambda arms was kept constant at 40 ng/μl giving an i to j ratio of ≈3:1 where j describes the effective concentration of the two ends of the same DNA molecule and i describes the concentration of all complementary ends in the solution (Dugaiczyk 1975, Maniatis 1982). When i>j, concatemers are favoured.

Table 2.1 Test ligation results

Library	1.1	1.2	1.3	1.4	2
ratio[a.]	1,1:1	2,2:1	3,3:1	6,6:1	2:1
[EMBL](ng/μL)	40	40	40	40	80
[insert](ng/μl)	72	36	24	12	77
pfu/μg arms[b.]	$1,1\times10^6$	$1,0\times10^6$	$0,55\times10^6$	$0,42\times10^6$	$1,9\times10^6$
total pfu	230000	200000	110000	83000	$1,5\times10^6$

a.) molar ratio of insert to EMBL3 arms DNA
b.) pfu = plaque forming units

After overnight ligation at 12°C, the mixtures (library 1.1 to 1.4) were packaged using a commercial packaging system (Promega;'Packagene') and titrated the same day (see table 2.1). Altogether 620 000 recombinants were obtained (non-recombinant background = 320 pfu/μg). As the titre and thus the complexity of the freshly packaged library drops several fold within the first few days, libraries 1.1 to 1.4 were amplified [7.3.5] as soon as the titration results were available. The titre of these amplified libraries lay between 2 and 4 x 10^9 pfu/ml.

From table 2.1 it was concluded that an insert : arms ratio of 1:1 to 2:1 was optimal. A large scale library (library 2) was thus prepared using an insert : arm molar ratio of 2:1 and increasing the DNA concentration to 80 ng/μl giving an i:j ratio of 6:1 which strongly favours concatenation. This resulted in a complexity of 1,5 x 10^6 pfu and a packaging efficiency of 1,9 x 10^6 pfu/μg of EMBL3 arms. 100 000 pfu were screened immediately, the rest amplified.

2.3 Library evaluation

Library 2 was tested for intactness by screening 10000 pfu from the unamplified library with a *P.angulosus* probe covering the ubiquitous histone quintet (probe h22, isolated by J.Rees, unpublished). Sea urchins contain several hundred copies of this quintet and although the exact number is not known for *P.angulosus*, one can estimate the percentage of the genome that can hybridize to the histone probe by assuming that it is similar to *P.miliaris* which would thus give a value of 600 x 6000 bp / 9 x 10^8bp = 0,40%.
67 pfu out of 10000 hybridized to the h22 probe as detected on duplicate filters (fig 2.4) which gives a prevalence value of 0,67% of the genome. This is very close to the estimated 0,40% and one can thus conclude that the frequency of genomic DNA fragments in the library closely parallels that in the genome.

Figure 2.4. About 10000 pfu replica plated onto a nylon membrane and hybridized at low stringency to the histone quintet-containing h22 probe.

2.4 Screening

There are both advantages and disadvantages to the amplification of a library. The number of clones that need to be screened to achieve a given probability of finding a clone is increased by an indeterminate amount because the inserts will lead to differential phage replication rates and thus certain sequences may be over- or under-represented. A general rule is to screen 30 to 40% more clones (Ausubel 1987) than calculated in section 2.1. The advantages are the stability and reusability of the amplified library - the same library can be rescreened many times with different probes. As increased screening is more economical than library preparation it was decided to amplify libraries.

For a 99% probability of finding a single copy homeobox gene in *P.angulosus*, 260000 pfu or 360000 pfu from an amplified library have to be screened.

I used large (150 mm diameter) Petri dishes and plated phage at a density of 10000 to 20000 pfu per dish [7.3.6]. *In situ* plaque hybridization was performed using nylon filters in duplicate and the orientation recorded by stabbing a needle dipped in indelible ink through the filters into the agarose. Denaturation of the transferred DNA (unpackaged and packaged into capsids), UV-crosslinking and low stringency hybridization with an Antp homeobox are described in section 7. The number of pfu screened and positives found is given in table 2.2.

Table 2.2 Screening results

Library	complexity	pfu screened	positive pfu
1.1a	230000	200000	0
1.2a	200000	150000	0
1.3a	110000	150000	0
1.4a	83000	100000	0
2	1500000	50000	0
		10000	67(h22 probe)
2a	630000	600000	8(Antp probe)

'a' refers to an amplified library

It is odd that no pfu hybridizing to Antp were found in library 1 (test ligations) which may be explained by a bias in the original unamplified library or a bias during amplification. For this reason the large scale second library was evaluated using a high copy number sequence (see above) and a fraction of the library screened before amplification.
The frequency of positive clones in amplified library 2 is 8/600000 or one clone per 1,1 x 10^9 bp of cloned DNA. The genome size is around 0,8 x 10^9 bp. We thus find clones hybridizing to the Antennapedia homeobox at the frequency of 1 per haploid genome equivalent of DNA.
Clones hybridizing to the Antp homeobox probe were picked using a Pasteur pipette, spread on secondary plates at low density (20 to 100 pfu/82 mm diameter plate) and rescreened. This procedure was repeated at least twice more to reduce the possibility of contamination by a second phage. Once pure, a high titre stock was prepared.

2.5 Characterization

The DNA of eight uncontaminated clones (lambda 40, 44, 47, 48, 50, 52, 54, 57) was purified [7.2.3] and digested with all combinations of the restriction enzymes Sal1 and EcoR1

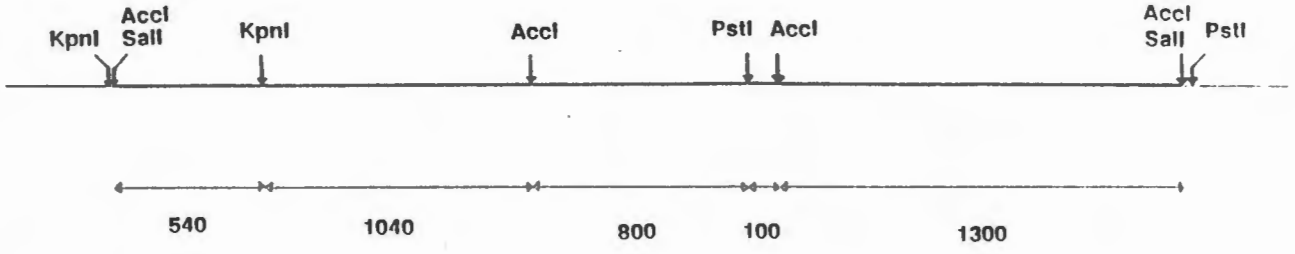


Figure 2.11. Restriction map of p54. The map is drawn to scale. The dark solid line represents the insert, the lighter line the Bluescript-SK vector. Below the map distances between restriction sites are given in base pairs. The 800 bp AccI-PstI fragment contains the homeobox and was subcloned [3.1].

[7.5.1.1]. SalI cuts at either end of the inserted fragment but not elsewhere in EMBL3 whereas EcoRI does not cut EMBL3 at all. It is thus possible to obtain information about the size of the insert (by using SalI alone) and of the relative position of a sea urchin homeobox-containing fragment within the insert.
The cut DNA was electrophoresed on a 0,7% agarose gel, stained with ethidium bromide (fig 2.5a, 2.6a) then Southern blotted onto a nylon filter and hybridized at low stringency to the Antp probe (fig 2.5b, 2.6b). The fragment sizes were determined using the lambda size standards and are shown in histogram format in fig 2.7. The size of these clones varies from 13 to 17 kbp (table 2.3) whereas the size of the homeobox containing fragment varies between 3,2 and 5,9 kbp.

Table 2.3 Fragment sizes of homeobox-containing lambda EMBL3-clones.

clone	size (kbp)	size of homeobox fragment (kbp)
40	13	5,7 SalI
44	14	4,5 SalI
47	15	3,2 SalI
48	17	5,9 SalI
50	15	4,1 SalI
52	14	4,5 SalI; 4,1 SalI/EcoRI
54	17	3,8 SalI
57	15	5,9 SalI

The question of whether any of these 8 clones represent the same gene is not easily answered because
1. the DNA is derived from a partial digestion and thus clones overlap and
2. multiple individuals were used in the construction of the embryonal genomic library (sperm of one individual was used to fertilize eggs derived from numerous females).

The second argument becomes important if the restriction fragment polymorphism is large in the vicinity of the homeobox(es).

The question was answered once the homeobox of one of these isolates was purified. This box was hybridized at high stringency to all eight isolates as well as a non-homeobox containing clone, EMBL-55 (fig.2.8). All 8 isolates hybridized under the high stringency conditions [7.7.4] indicating a high degree of homology. These isolates thus represent the same gene situated in a region exhibiting a high degree of restriction fragment polymorphism. Because half the embryonal genomic DNA is derived from one individual (sperm: 2 haploid genome equivalents), this polymorphism cannot be quantitated. In chapter 4, I have addressed the question of polymorphism directly.

2.6 Subcloning

One clone, lambda 54, was chosen at random for further investigation. The 3,8 kbp homeobox-containing SalI fragment of lambda 54 was cut out of a 0,8% low melting agarose gel and purified using a Quiagen column [7.5.3.4].
Bluescript SK was chosen as a vector because of 1) its large polylinker with 21 unique restriction enzyme sites, 2) the possibility of making sense and antisense transcripts using T_3 and T_7 RNA polymerase, 3) the ability to perform unidirectional deletions with Exonuclease III (ExoIII) for DNA sequencing and 4) the ease of detection of recombinants (white colour selection). Bluescript was digested to completion with SalI and treated with calf intestinal phosphatase to prevent self-ligation.
Ligation was done at a vector to insert ratio of 2,5 to 1.
The recommended host for Bluescript is Xl1-Blue (*recA*$^-$).
Three procedures to obtain competent cells were tested and the Chung-Miller (1988) method found to be optimal for the transformation of Xl1-Blue cells [7.4.1.2].

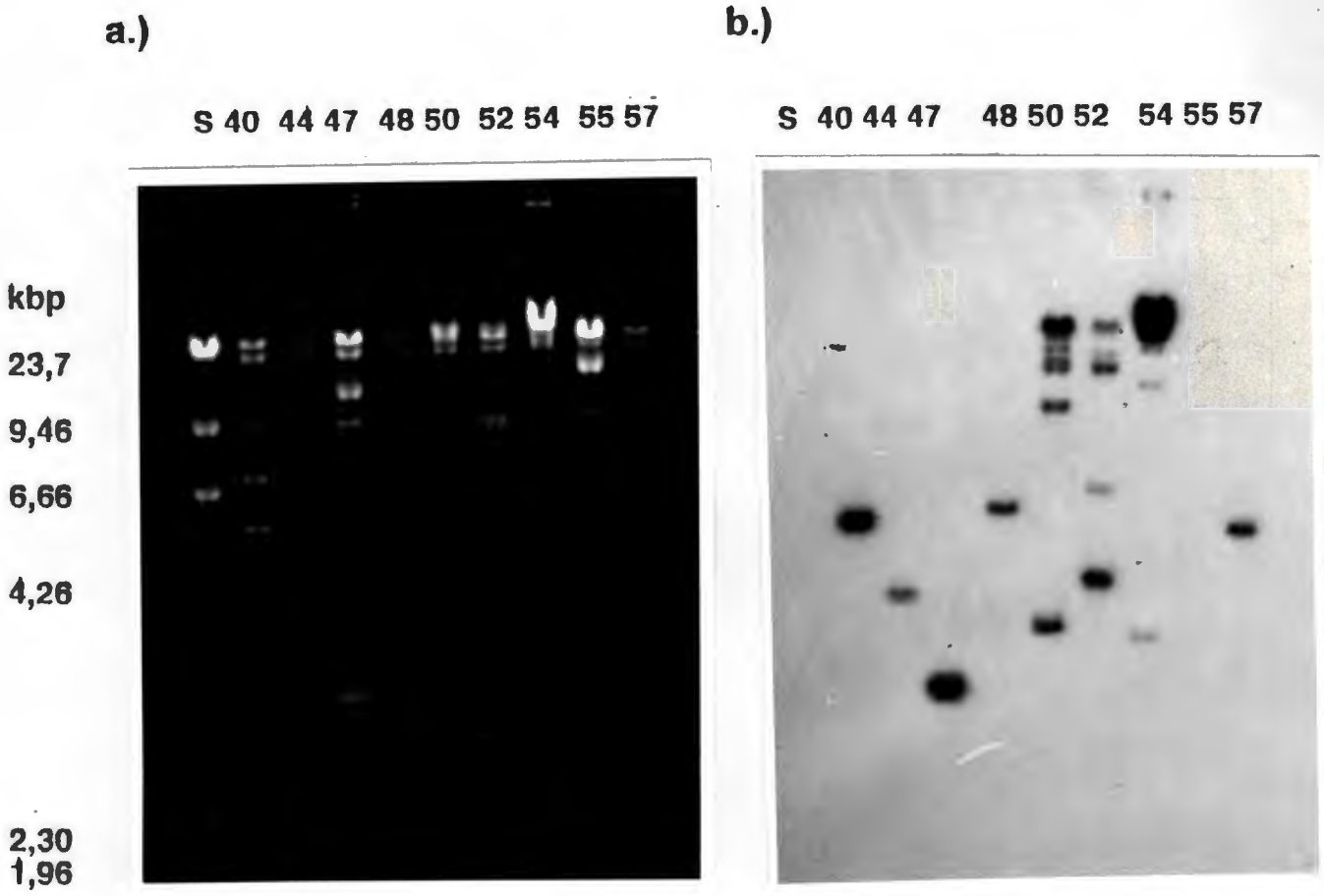


Figure 2.8. Characterization of clones by high stringency hybridization to PaHbox1 (derived from EMBL3 clone 54). a.) DNA from EMBL3 clones 40, 44, 47, 48, 50, 54, 57 as well as a negative control (clone 55), was digested with SalI and an aliquot electrophoresed on a 0,7% agarose-TBE gel for 16 hr at 1 Volt/cm. S = lambda DNA, HindIII digested. b.) Hybridization at high stringency to PaHbox1. The multiple bands for clones 50, 52 and 54 arise due to incomplete SalI digestion and do not affect the qualitative result of the experiment. The results can be compared to those of fig.2.5 and 2.6.

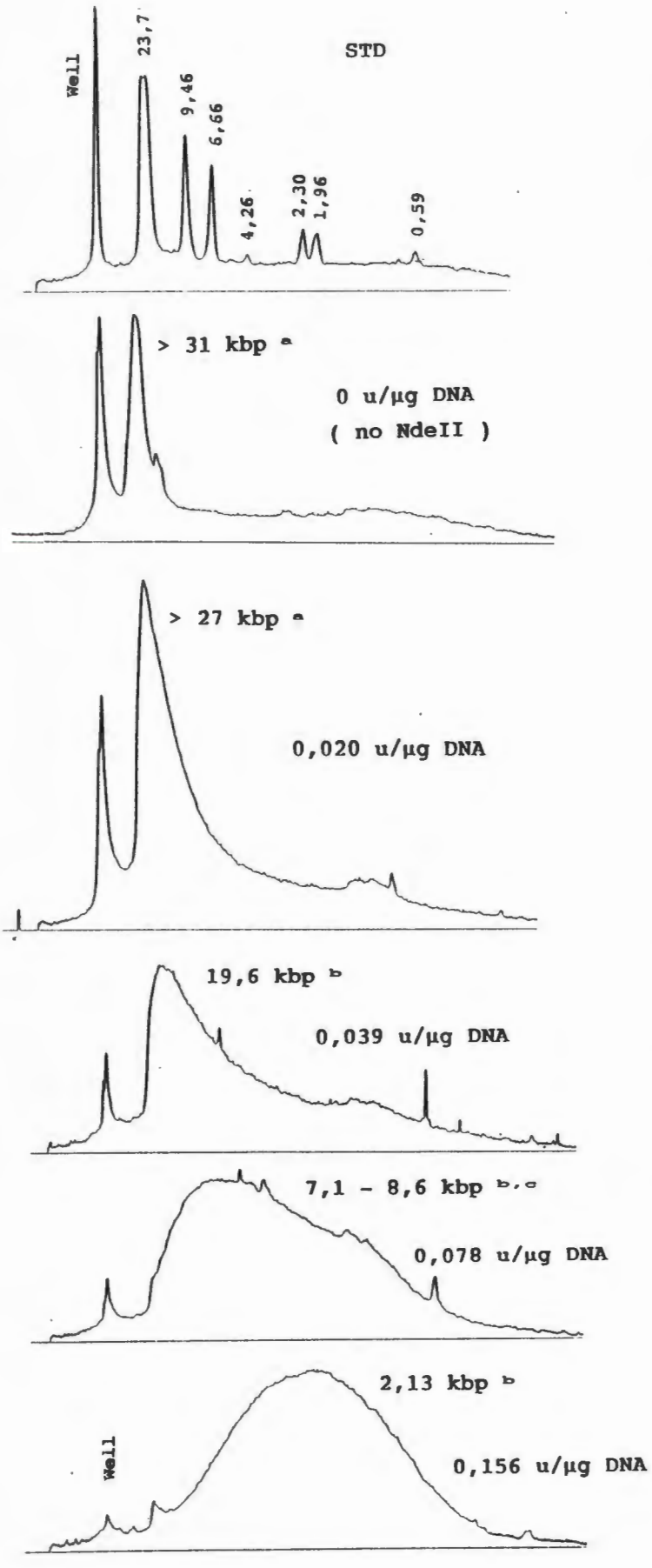


Figure 2.2. Densitometric scans of lanes 2 to 7 of the gel depicted in fig. 2.1. Scans of lambda DNA, HindIII digested (STD) and of sea urchin genomic DNA digested with 0 to 0,156 units NdeII per µg of DNA; the size of the DNA with the maximum fluorescence is given in kbp. Digestion conditions of 0,020 units NdeII / µg DNA were chosen (2 hr at 37°C).

a. Size is calculated from the 23,7 and 9,46 kbp points of the standard curve by extrapolation:

$$\frac{\log(\text{size unknown})-\log(23,7)}{d_{unknown} - d_{23,7}} = \frac{\log(23,7)-\log(9,46)}{d_{23,7} - d_{9,46}}$$

where d represents the distance migrated. The size thus derived is an underestimation because the relationship d α log(size) no longer holds for very long DNA molecules separated on dilute agarose gels (Seed 1982).

b. These values are calculated using a correlation equation derived from all standard points (correlation coefficient = 0,988).

c. As the peak is very broad, a range of sizes is given.

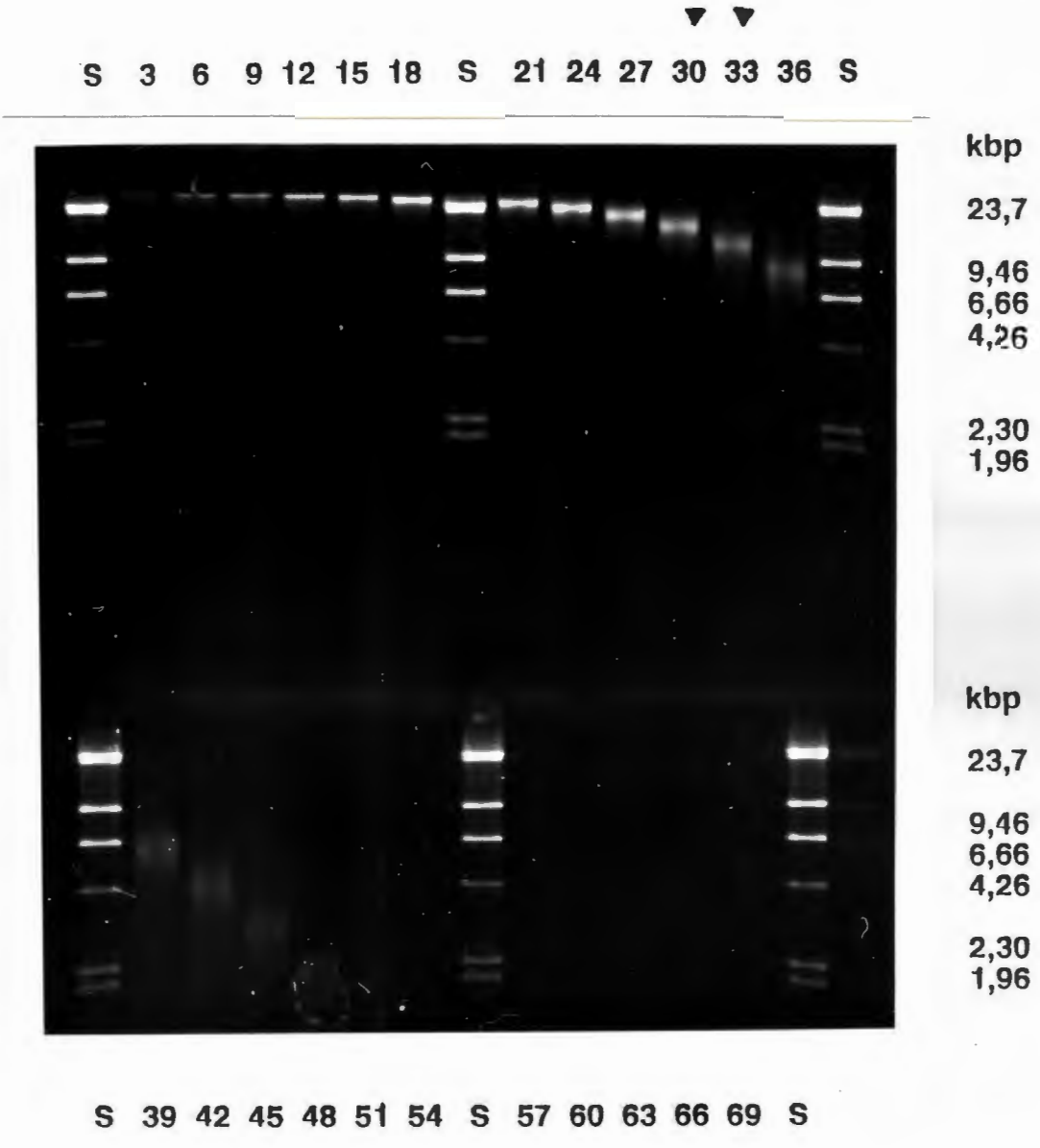


Figure 2.3. Sucrose gradient fractions. A 4% (v/v) aliquot of every third fraction of NdeII partially digested genomic DNA was separated on a 10 to 40% sucrose gradient and electrophoresed on a 0,5% agarose TBE gel at 1 Volt/cm for 18 hr. Fractions pooled are shown by arrowheads.

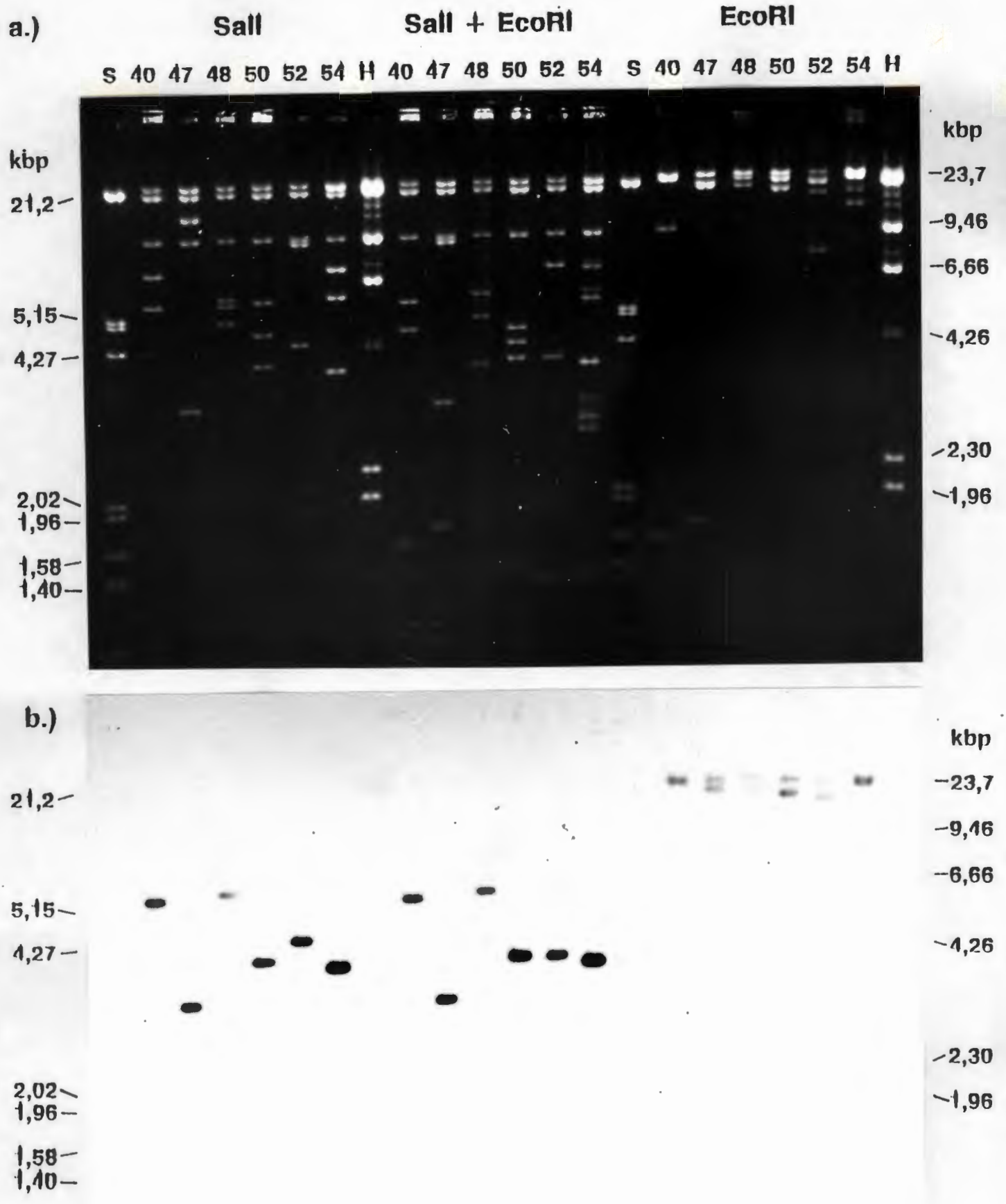


Figure 2.5. Restriction mapping of lambda EMBL3 clones 40, 47, 48, 50, 52 and 54. a.) 300 ng of DNA samples were restricted overnight with 1 unit of SalI, EcoRI or both and electrophoresed on a 0,7% agarose-TBE gel for 20 hr at 1 Volt/cm. S = lambda restricted with HindIII and EcoRI, H = lambda restricted with HindIII. b.) Autoradiograph of gel 2.5a. The DNA was transferred to a nylon membrane and hybridized at low stringency to the *Antp* homeobox fragment as described in Methods.

Transformed cells were plated in selective media, a recombinant (p54) picked and the DNA isolated and characterized by digestion with restriction enzymes: p54 was digested with AccI, BamHI, EcoRI, HindIII, KpnI, PstI, SalI and XhoI and fragments separated by electrophoresis on a 1% agarose gel. All these enzymes cut once in Bluescript; BamHI, EcoRI, HindIII and XhoI did not cut in the insert, SalI cut out the insert (partially); KpnI and PstI cut once in the insert and AccI cut out the insert and cut twice within it. The enzymes AccI, KpnI and PstI were used in different combination to map the restriction sites. Figure 2.9 shows the fragments that hybridized to the Antp probe at low stringency. The fragment sizes are shown in fig 2.10 and the restriction map in fig 2.11. The homeobox resides in a 800 bp AccI/PstI fragment.

For sequencing purposes the 800 bp AccI/PstI fragment was subcloned into the EcoRV/PstI sites of Bluescript SK. It was necessary to use one blunt end (EcoRV) because AccI has a degenerate recognition specificity of 5'GAT(A/C)(G/T)AC and therefore an unpredictable overhang. The vector was prepared by sequential digestion with EcoRV, PstI and EcoRI. EcoRI cuts between the EcoRV and PstI sites which are separated by 12 bp, thereby halving the size of the cut-out piece. The DNA was purified by phenol and chloroform extractions and passed over a G-50 spin column to remove the cut-out oligonucleotide [7.5.3.2].

The insert was derived by digesting p54 with AccI to completion; the DNA was purified and the AccI overhangs filled in using the Klenow fragment. This was followed by digestion with PstI. The fragments were separated on a 3,5% nondenaturing acrylamide gel and the appropriate band cut out and eluted [7.5.3.4].

For ligations a 3:1 molar ratio of insert to vector was used. Theoretically, no self-ligation could occur as the two ends of each DNA species differed. Transformation was done as explained previously, using X11-Blue cells as host.

The resulting clone was termed pHB1 (plasmid HomeoBox 1) and used for further experiments.

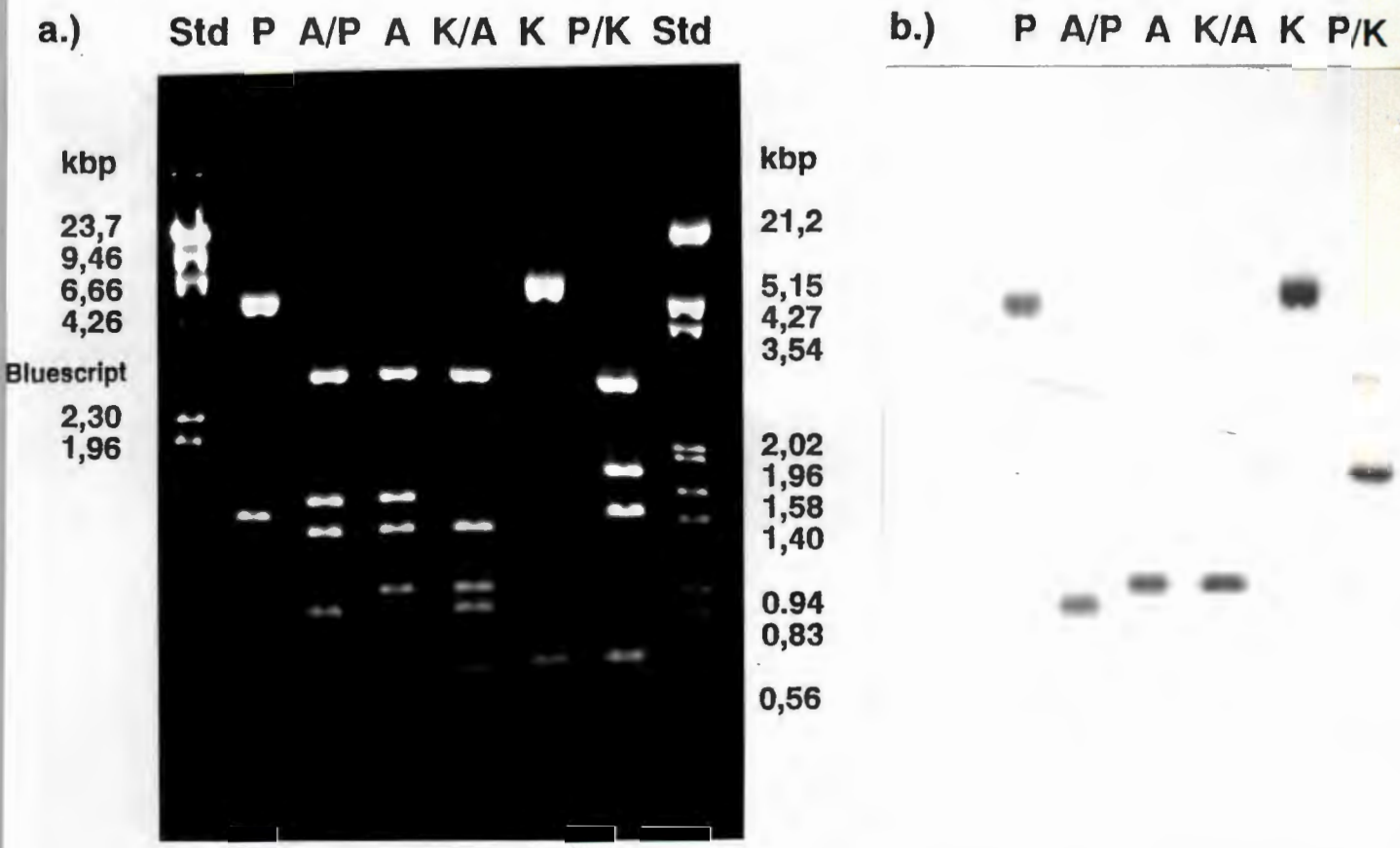


Figure 2.9. Restriction mapping of clone p54. a.) Bluescript-SK+ containing the 3,8 kbp SalI fragment from clone EMBL3-54 was fully digested with various combinations of AccI (A), KpnI (K) and PstI (P) and electrophoresed on a 1% agarose-TBE minigel at 0,5 Volt/cm for 14 hr. b.) The transferred DNA was hybridized at low stringency to the *Antp* homeobox.

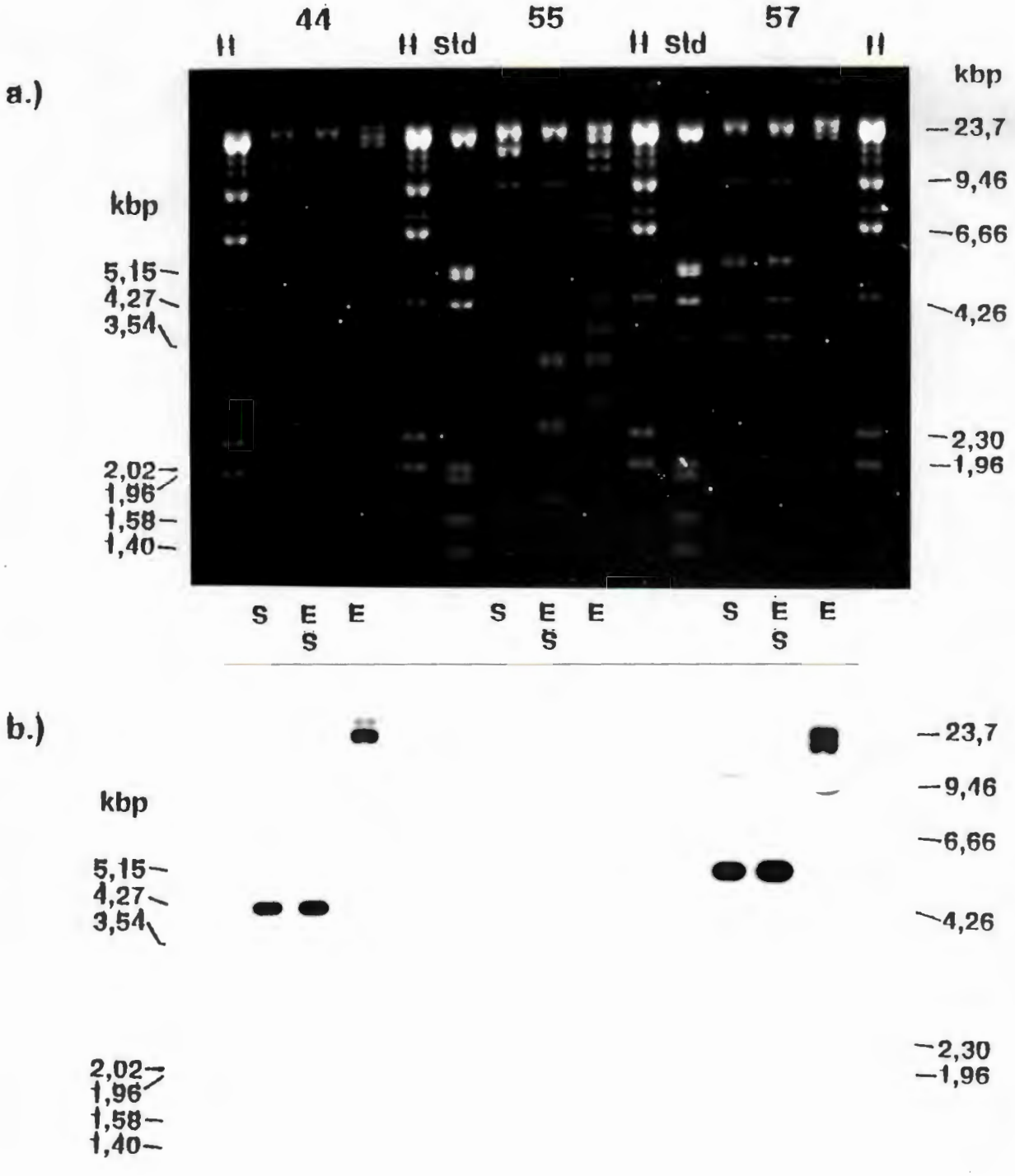


Figure 2.6. Restriction mapping of lambda EMBL3 clones 44, 55 and 57. a.) 300 ng of DNA samples were restricted overnight with 1 unit of SalI, EcoRI or both and electrophoresed on a 0,7% agarose-TBE gel for 20 hr at 1 Volt/cm. Std = lambda restricted with HindIII and EcoRI, H = lambda restricted with HindIII, S = SalI, E = EcoRI. b.) Autoradiograph of gel 2.5a. The DNA was transferred to a nylon membrane and hybridized at low stringency to the *Antp* homeobox fragment as described in Methods. Clone EMBL3-55 is an isolate of a non-hybridizing plaque and represents a negative control.

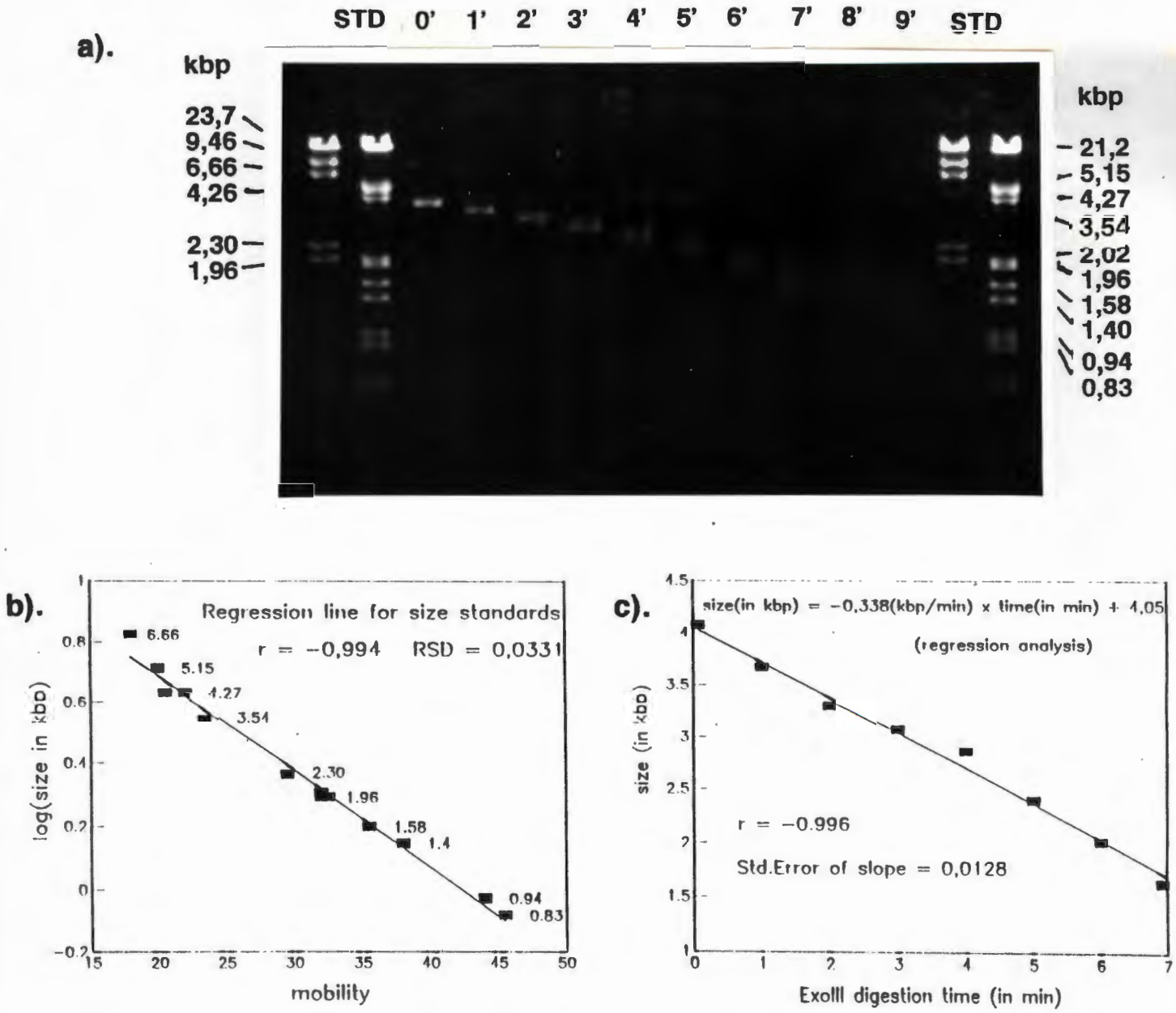


Figure 3.2. Exonuclease III kinetics. pHB1 was digested with KpnI and HindIII, then in 30 sec intervals for 0 to 10 min with Exo III [3.1]. a.) The DNA was separated on a 1% agarose-TBE gel at 2 Volt/cm for 15 hr. STD = lambda DNA digested with HindIII (left) and HindIII + EcoRI (right). b.) Regression line for molecular size standards of the gel, used to determine the size of the ExoIII digestion products. c.) Determination of the rate of ExoIII digestion: as can be seen from the figure, this rate was 340 ± 13 bp/min.

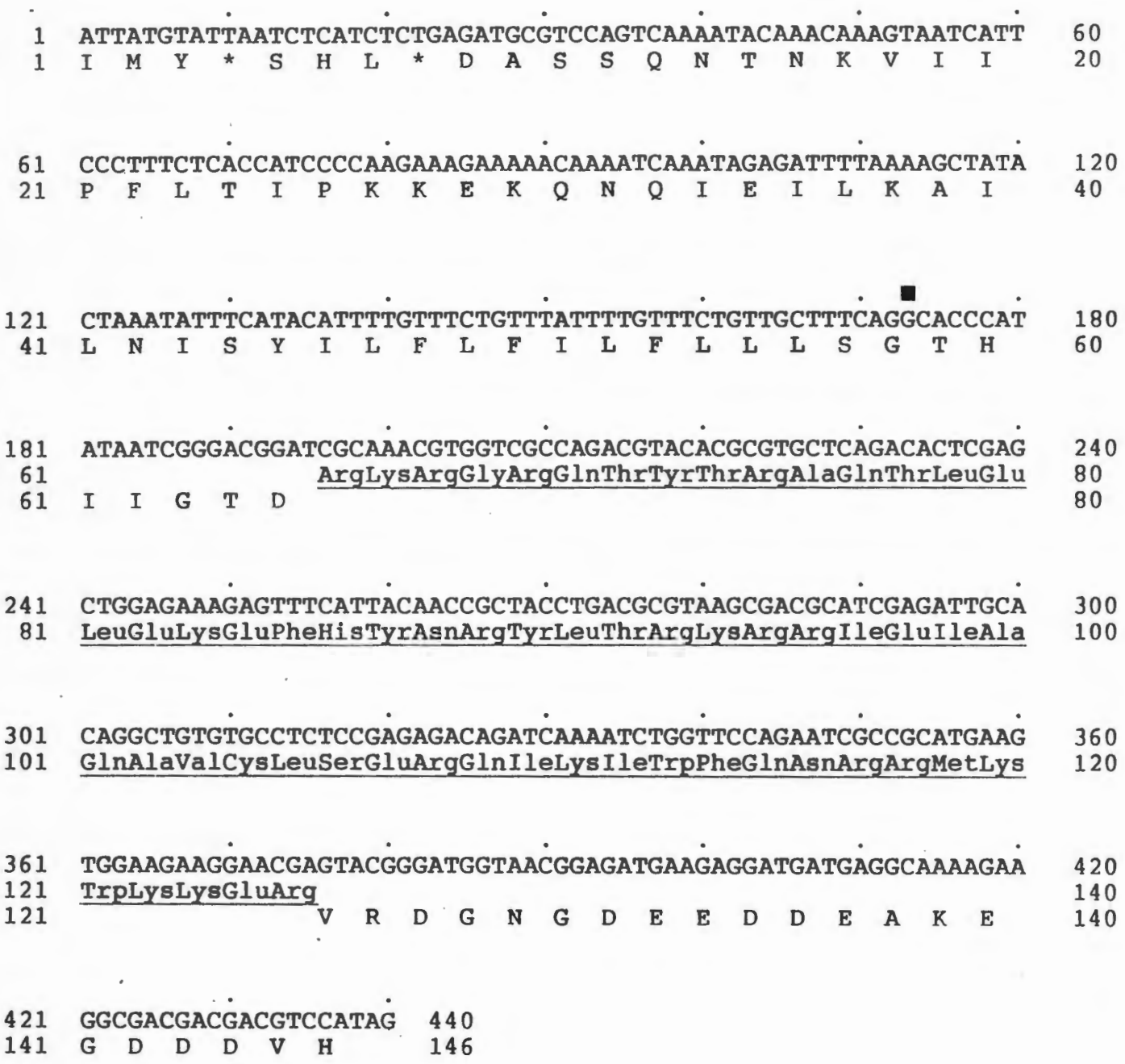


Figure 3.3. Sequence of PaHbox1 and flanking regions. 440 bp and their conceptual translation is shown. The homeobox is underlined. A possible splice site (see text) is indicated by a '■'.

sequence subsequently termed *PaHbox1* and the 9 bp immediately preceding it. Comparisons of the conceptual translation of *PaHbox1* again showed amino acid (a.a.) homology only within and immediately preceding (3 a.a) the 60 residue homeodomain. Figure 3.4 displays the homology between the conceptual *PaHbox1* homeodomain, the most closely related homeodomains, the sea urchin homeodomains and representatives from other homeodomain-classes. It is evident that *PaHbox1* is a member of the *Antp*-class. The most closely related homeobox is that of the human *Hox* 2.C (formerly hu-c1; Boncinelli,1985) gene which shows 80% nucleotide and 90% a.a. positional identity. In addition, the amino acid homology extends to the 3 a.a. preceding the homeobox. The other vertebrate genes homologous to *Hox* 2.C, namely the mouse *Hox* 2.3 and *Hox* 1.1 and African toad *XlHbox* 2 and 3, have very similar positional identity figures. The insect Antennapedia homeobox shows 78% nucleotide and 88% a.a. positional identity to *PaHbox1* but in contrast to the vertebrate homeoboxes the homology does not extend to the 3 a.a. preceding the homeobox.

The *Tripneustes gratilla* sea urchin homeoboxes hb 1 and hb 3 are also of the *Antp* class whereas hb 4 belongs to the AbdB class (Dolecki 1988a). It is interesting that hb 1 shows 78% and hb-3 80% a.a. positional identity but only 67% (hb 1) and 71% (hb 3) nucleotide positional identity. The a.a. immediately preceding hb 3 is also identical to *PaHbox1*. From these comparisons (see also fig 3.5) it is clear that *PaHbox1* is more related to vertebrate and protostomate homeobox(es) than it is to any published sea urchin homeoboxes. As will be discussed later, hb 3 and *PaHbox1* show intriguing similarities in terms of their expression [section 5] and DNA context [section 4] and may indeed be related.

When comparing a.a. and nucleotide identities it is apparent that the sea urchin homeoboxes hb 1 and 3 are 11% and 9% respectively more similar to *PaHbox1* on the peptide than the

	1 10 20 30 40 50 60	identity a.a.(%)	class
PaHbox1	THIIGTD RKRGRQTYTRAQTLELEKEFHYNRYLTRKRRIEIAQAVCLSERQIKIWFQNRRMKWKKER VRD		
Hox 2.C	MRSS... .....Y.....R.....H.L..T.....N KTA	90	*Antp*(*Hox*1.1)
Hox 2.3	MRSS.P. .....Y.....R.....HTL..T.....N KTS	88	*Antp*(*Hox*1.1)
Hox 1.1	MRSS.P. .....Y.....F.....R.....H.L..T.....H KDE	88	*Antp*(*Hox*1.1)
Xlhbox2	MRSA.S. .....Y.....F.....R.....HTL..T.....N KAS	87	*Antp*(*Hox*1.1)
Xlhbox3	LXRK.P. .....Y.....F.....R.....H.L..T.....H KEE	88	*Antp*(*Hox*1.1)
Antp	QFGKCQE .....Y.....F.....R.....H.L..T.....N KTK	88	*Antp*(*Hox*1.1)
hb-1	NVGLEVG ...C.....Y.....F.....R....LSHLLG.T.....Y...S KNK	78	*Antp*
hb-3	HVSYLA. G.....Q.....FS..V..R..F....SLG.....R.H GSN	80	*Antp*
Dfd	SYQP.M. P..Q.TA...H.I.....R.....HTLV.....DN KLP	78	*Dfd* (*Hox*1.4)
Hox 2.4	PPFPAAG .R.....S.Y.....LF.P.....VSH.LG.T...V.....N NK.	77	*Hox* 2.4
Hox 1.5	GPPGQAS S....TA...P.LV.....F....M.P..V.M.NLLN.....Y..DQ KGK	70	*Hox* 1.5
hb-4	NWLSA.S GRKK.CP..KF.....LF.M....D..L...RLLS.T...V.....M..QN RAQ	63	*Abd-B*(*Hox*1.7)
Abd-B	EWTGQVS VRKK.KP.SKF.....LF.A.VSKQK.W.L.RNLQ.T...V.....N..NS Q.Q	53	*Abd-B*(*Hox*1.7)
Labial	CSLSSNT NNS..TNF.NK.LT.....F.....A.....NTLQ.N.T.V.....Q..RV KEG	63	*labial*
caudal	QPGKTRT KDKY.VV..DF.R.....YCTS..I.IR.KS.L..TLS.....V.....A.ERTSN KKG	52	*caudal*
Eve	IPADDPS VR.Y.TAF..D.LGR.....YKEN.VS.P..C.L.AQLN.P.ST..V.....D.RQ. IAV	50	*even-skipped*
Suhb-*En*	RREKKA. E..P.TAFSAS.LQR.KQ..QQSN...EQ..RSL.KELT...S.....K.A.I..AS GLK	47	*engrailed*
Bcd	DSLVMRR PR.T.T.F.SS.IA...QH.LQG....AP.LADLSAKLA.GTA.V....K...RRH.IQS XXX	40	*bicoid*
H2-0	GSNGKRK .SWS.AVFSNL.RKG..IQ.QQQK.I.KPD.RKL.ARLN.TDA.V.V.....RHT. ENL	38	H2.0
Paired	GIALKRK QR.C.T.FSAS.LD...RA.ERTQ.PDIYT.E.L.QRTN.T.AR.QV..S...ARLR.QH TSV	35	*paired*

Figure 3.4 Comparison of the conceptual *PaHbox1* homeodomain sequence with the most similar homeodomains as well as with members of other homeodomain classes. The amino acid positional identities percentages were calculated over the 60 a.a. homeodomain only. References for these sequences can be found in table 1.1.

A. Nucleotide positional identity

	PaHbox1	*hb* 1	*hb* 3	*hb* 4	*Antp*	*Hox* 2.C
PaHbox1	100					
hb 1	67	100				
hb 3	71	70	100			
hb 4	63	63	65	100		
Antp	78	71	73	66	100	
Hox 2.C	80	74	77	67	84	100

B. Amino acid positional identity

	PaHbox1	*hb* 1	*hb* 3	*hb* 4	*Antp*	*Hox* 2.C
PaHbox1	100					
hb 1	78	100				
hb 3	80	77	100			
hb 4	63	67	63	100		
Antp	88	88	82	70	100	
Hox 2.C	90	87	80	68	98	100

Figure 3.5. Nucleotide and amino acid matrices of percentage positional identities among six 180 bp homeoboxes and corresponding 60 amino acid homeodomains.

nucleotide level, indicative of a large fraction of silent mutations. Comparing the percentage mutations at the different codon positions (table 3.1) indicates an overwhelming fraction of mutations at the third "wobble" (Crick 1976) position. This position shows as expected the greatest degree of degeneracy.

Table 3.1 Number of positional differences from *PaHbox1* at all 3 codon positions and the percentage of the total number of differences

Codon position:	1	2	3
hb 3	13 (24%)	8 (15%)	32 (60%)
hb 1	17 (29%)	6 (10%)	35 (60%)
Antp	9 (22%)	4 (10%)	27 (68%)
Hox 2.C	5 (14%)	3 (8%)	29 (78%)

The conceptual translation of *PaHbox1* shows several differences to the other *Antp*-class homeodomains: The alanine residue at position 11 is uncommon and found mainly in the *Hox* 1.5 - class of homeodomains (see fig.3.4). However, this position is generally very variable amongst homeodomains. The lysine at residue 29 is a conservative change from the usual arginine at this position, as are the valine at position 38 (usually leucine) and serine at 41 (threonine). The glutamine at residue 36, also found in the paired and *engrailed* homeobox classes, is at a not well conserved position. As all the residues that are most conserved among homeodomains (see [1.2]) are also conserved in *PaHbox1*, and as *PaHbox1* shows such high sequence homology to *Antp* with mainly conservative a.a. changes, one can assume that the discussion of the *Antp* homeodomain structure [1.3] will be equally applicable to the *PaHbox1* homeodomain.

Examination of the DNA sequence for potential splice acceptor/donor sites and polyadenylation signals (Trifonov,1986) indicates none downstream of *PaHbox1* but one splice donor site at position 170 to 173 that conforms to the consensus requirements of such a site (consensus of(C/T)AGG with a pyrimidine rich stretch and no (AG)-dinucleotides upstream)(Mount,1982). RNAase mapping [7.8] indicates a protected fragment of 273 bp (see fig 5.4). This would make the potential splice site at position 173 the most likely candidate for the anterior boundary of the exon. This could mean that only 7 a.a. preceding the *PaHbox1* homeodomain form part of the homeobox containing exon. This sequence bears no relationship to the pentapeptide motif (or a truncated pentapeptide motif) found 5 to 20 a.a. upstream of several homeodomains [1.5]. Such pentapeptide motifs generally have been found in the same 3' exon as the homeodomain ([1.5], Acampora 1989). It is also clear that the *PaHbox1* homeobox is not interrupted by an intron.

CHAPTER 4

RESTRICTION FRAGMENT LENGTH POLYMORPHISM

4.0. Index

4.1. Restriction fragments

The DNA from the sperm cells of eleven *P.angulosus* individuals was digested to completion with EcoR1 and the Southern transfer probed at high stringency with *PaHbox1* and its flanking 600 bp of sequence. As seen in fig.4.1, a high degree of fragment length polymorphism is evident with ten different fragments amongst the 22 haploid genomes. These fragments are between 7,5 and 20 kbp in size and are listed according to size and frequency of occurrence in table 4.1. At this point, two questions arise :

1. Are we looking at a single copy or multiple copy gene ?
2. Is this restriction fragment length polymorphism unusually high ?

Table 4.1. Restriction fragments of fig.4.1 arranged according to size

size (in kbp)	Number of fragments (of 22)
21	2
19	3
17	2
15	2
14	1
13	4
12	2
10	3
9,1	1
7,3	2

4.2. Copy number

The answer to the first question is fairly staightforward. Sea urchin cells are diploid hence every individual will have two copies of all somatic genes. These alleles of a gene need not be identical, so we expect one or two restriction fragments per individual to hybridize to a probe for a single copy gene. This is what is observed (fig 4.1). More fragments may arise if one or more restriction sites occur within the nucleotide stretch hybridizing to the probe. In this study a short (800 bp) probe containing no EcoR1 restriction site was used and it is therefore unlikely to see many instances where more than two restriction fragments per individual are observed.
Likewise, it is unlikely that there are two (or more) copies of this gene per haploid genome because then we would expect between 1 and 4 fragments (or more, if internal restriction sites). One fragment would be seen only if the two copies and their alleles were identical. This required homozygosity does not correspond to the observed overall level of heterozygosity of fragment sizes in the population as a whole. We can therefore conclude that *PaHbox1* is a single copy gene.

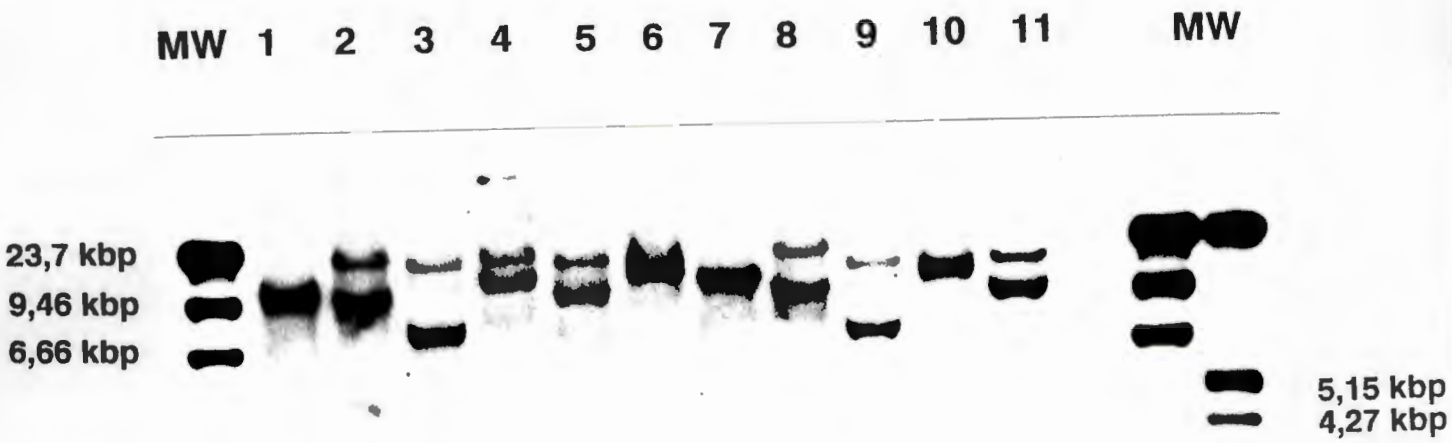
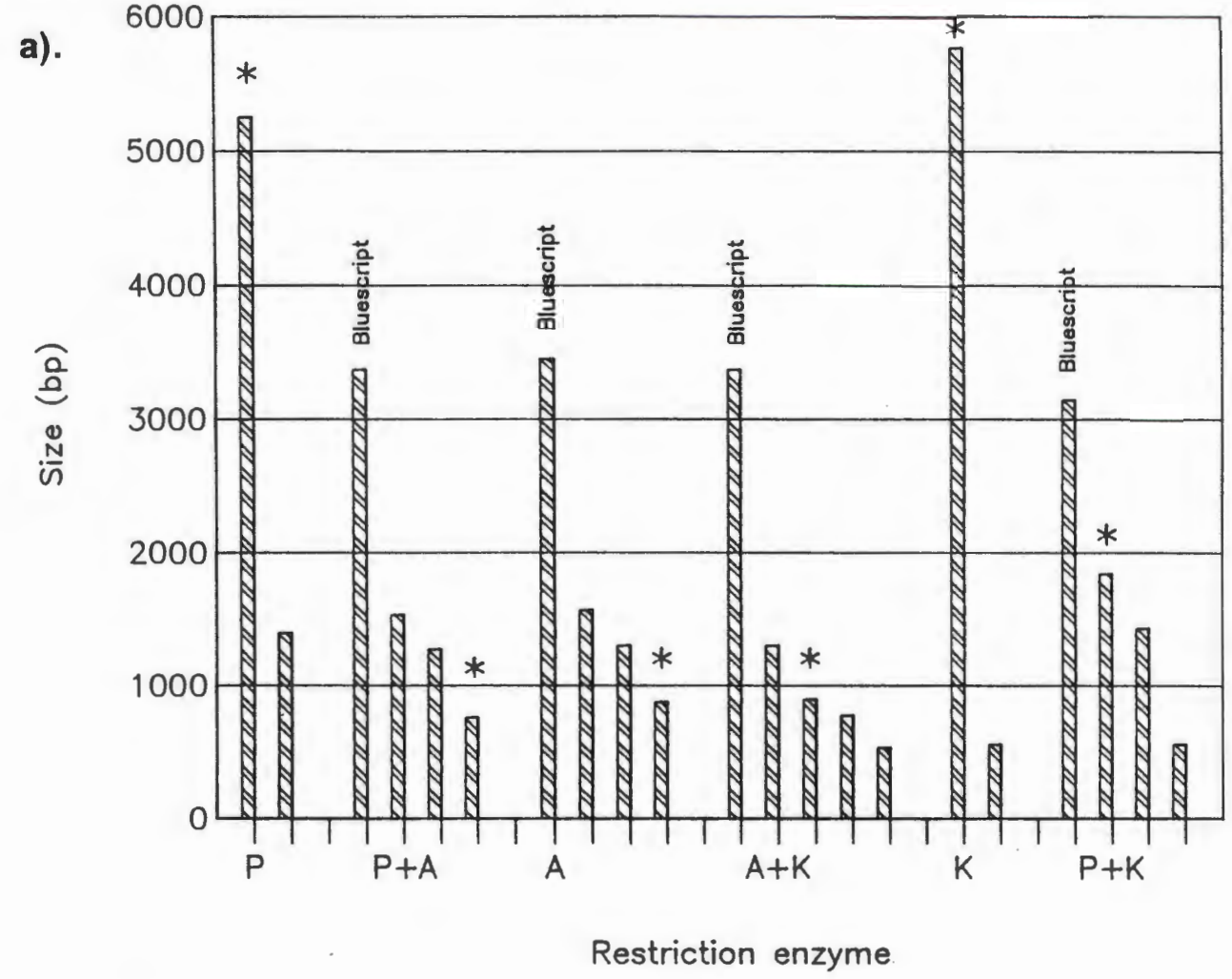


Figure 4.1. Hybridization of PaHbox1 to genomic DNA of 11 *P. angulosus* individuals. 10 µg DNA was digested to completion with EcoRI, electrophoresed on 0,6% agarose gels and Southern transferred onto nylon membranes. The probe and high stringency hybridization condition are described in section 7. MW : a lambda probe was hybridized separately to show up the molecular weight markers: lambda DNA cut with HinDIII or HinDIII and EcoRI (far right).



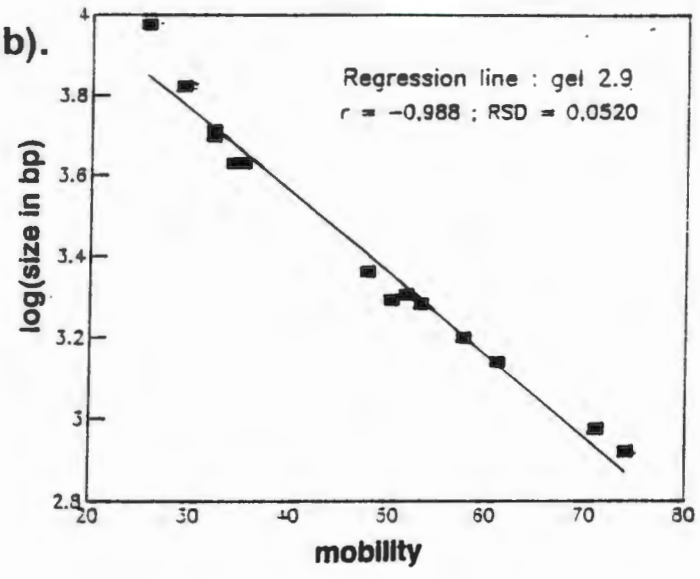


Figure 2.10. *a*.) Histogram of restriction fragments of p54 digested with AccI (A), KpnI (K), PstI (P) or combinations of two of these enzymes. The sizes were determined from gel 2.9 using the regression line depicted in *b*.). Homeobox-containing fragments are depicted with a star and the cut-out vector (Bluescript-SK) is labled. r = regression coefficient, RSD = residual standard deviation.

CHAPTER 3

SEQUENCE CHARACTERIZATION

3.0 Index

3.1 Method

pHB 1 was sequenced using the Sanger dideoxy method (Sanger 1977) but using [α-35S]thio-dATP, "Sequenase" instead of the Klenow fragment (Tabor 1987) and double stranded templates.

2 µg of plasmid DNA (0,376 pmol/µg), purified by CsCl gradient centrifugation [7.2.2] was added to 4 pmol of an 17 base primer complementary to the Bluescript vector sequence abutting the insert, in a total volume of 18 µl. 2 µl of 0,2 N NaOH were added and the mixture heated at 85°C for 5 min to denature the plasmid DNA. The mixture was placed on ice and 10 µl 5 M ammonium acetate (pH 7,50), 100 µl absolute ethanol added. The ethanol precipitate was recovered by centrifugation, washed in 75% ethanol and redissolved in 1 x "sequenase" buffer (40 mM Tris.HCl pH 7,5; 20 mM $MgCl_2$; 50 mM NaCl). Annealing was performed by heating to 40°C and allowing to cool to room temperature over a period of at least 15 min.

Labeling and termination reactions were performed as described in the "Sequenase" manual (USBC,1988) except that all solutions were made 10% (v/v) in DMSO.

Before loading, samples were heated to 80°C for 2 min and 2,5 µl loaded on a 6% polyacrylamide 7,7 M urea sequencing

gel [7.6.4) and electrophoresed for 2 hr at 55^{0}C and 1500 Volts. After this time another 2,5 μl was loaded in adjacent lanes and electrophoresis continued for another 2-3 hr. Gels were soaked in 10% acetic acid, 12% methanol for 5 min, then in water for another 5 min before drying at 80^{0}C on a modified hottray linked to a vacuum pump. Gels were exposed in direct contact with Cronex-4 X-ray film for 1 to 7 days.
Sequences were read manually three times and analyzed using the Genetics Computer Group's Sequence Analysis Software Package (abbreviated GCG; Devereux,1984)

3.1.1. Sequencing strategy

The sequencing strategy is shown in fig 3.1. Deletion mutants were constructed as follows.
1.pHB 1 was digested to completion with HinDIII and KpnI. KpnI cleaves to produce a 3' overhang which is refractory to exonuclease III digestion. Both restriction enzymes have unique recognition sites on the T_7 polymerase side of the insert, with the KpnI site closer to the T_7 primer sequence. This means that exonuclease III (Exo III) will only digest in the direction of the insert.
2. 5 μg of linearized pHB1 DNA (2 pmol of exonuclease III accessible ends) was prewarmed to 37^{0}C in 50 μl of 66 mM Tris.HCl pH 8,00; 6,6 mM $MgCl_2$ for 5 min. 60 units of ExoIII (Amersham) were added at time = 0 and 2,5 μl aliquots removed at 30 second intervals, diluted with 5 μl H_2O and inactivated by incubation at 70^{0}C for 10 min.
3. To each sample 12,5 μl of 2 x S1-Buffer (2 mM sodium acetate, pH 4,60; 500 mM NaCl; 2 mM $ZnSO_4$, 10% [v/v] glycerol) and 5 μl (20 units) of S1-Nuclease was added and the mixture incubated at room temperature for 20 min. The reaction was stopped by the addition of 5 μl 0,8M Tris.HCl pH8,00; 20 mM EDTA, 80 mM $MgCl_2$.

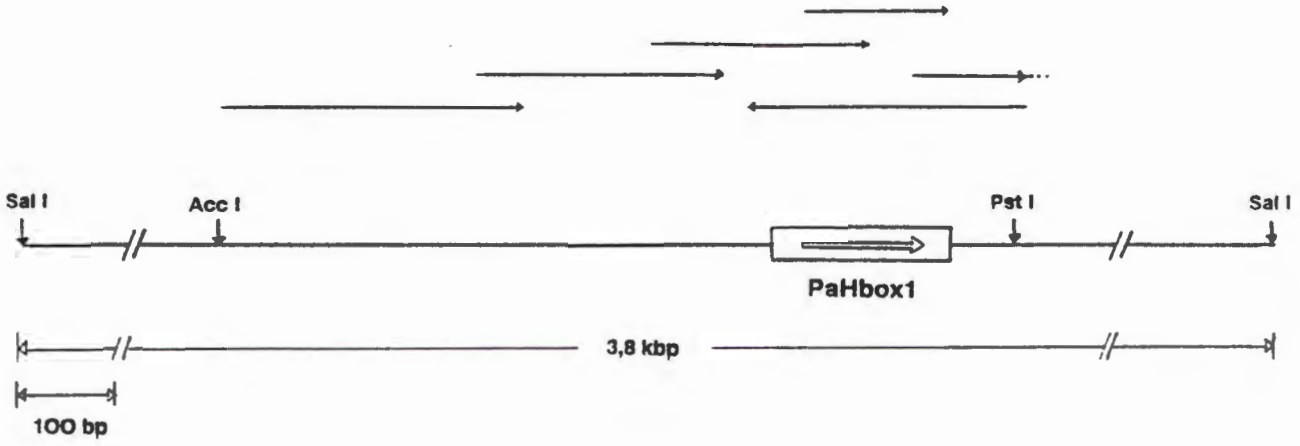


Figure 3.1. Restriction map of pHB1 and sequencing strategy. The homeobox is shown as an open box with the direction of transcription indicated by an open arrow. Arrows above represent the sequencing strategy (see text for construction of truncated clones).

4. 7 µl of every second sample was electrophoresed on a 1% agarose gel (fig 3.2a) and the rate of ExoIII digestion determined (fig 3.2c). The rate was 340 ± 13 bp/min and therefore the aliquots from time 30, 60 and 90 sec were used for further experimentation.
5. Any remaining 3' overhangs (KpnI generated) were removed using the 3' to 5' exonuclease activity of Klenow enzyme: 2 min incubation at 37^0C with 5 units Klenow. 5' overhangs were filled by adding free nucleotides (2 µl of 0,125 mM dNTP mixture) and incubating for a further 10 min at 37^0C.
6. Samples were phenol:chloroform (1:1) extracted and ethanol precipitated. The DNA pellets were resuspended in 0,1 x TE pH 7,50 and ligated overnight [7.5.1.4]. Xl-1 Blue cells were transformed [7.4.2], colonies picked and the size of the DNA (extracted using a 'miniprep' protocol [7.2.2]) determined by gel electrophoresis. Suitable deletion mutants (deletions of 250, 440, 570 and 690 bp) were cultured and the DNA isolated [7.2.2] and sequenced as described above.

3.2 Results and Discussion

Employing the sequencing strategy described above, 803 bp of pHB 1 were sequenced, the most 3' stretch of 440 bp being verified by sequencing of at least three independent clones. This 440 bp sequence contains a clearly identifiable *Antennapedia*-class homeobox with 195 bp of upstream, 65 bp of downstream sequence. The sequence has been submitted to the EMBL Nucleotide Sequence Data Bank and has been issued with the accession number X54494.
The sequence together with its conceptual translation is shown in fig 3.3. I have, in agreement with current terminology, termed the homeobox *PaHboxl* (*Parechinus angulosus* Homeobox, 1st homeobox found).
Database (GenBank, EMBL, NBRF-Nucleic) searches using FASTA and WordSearch (Devereux,1984) revealed sequence homology with other homeobox containing genes only within the 180 bp

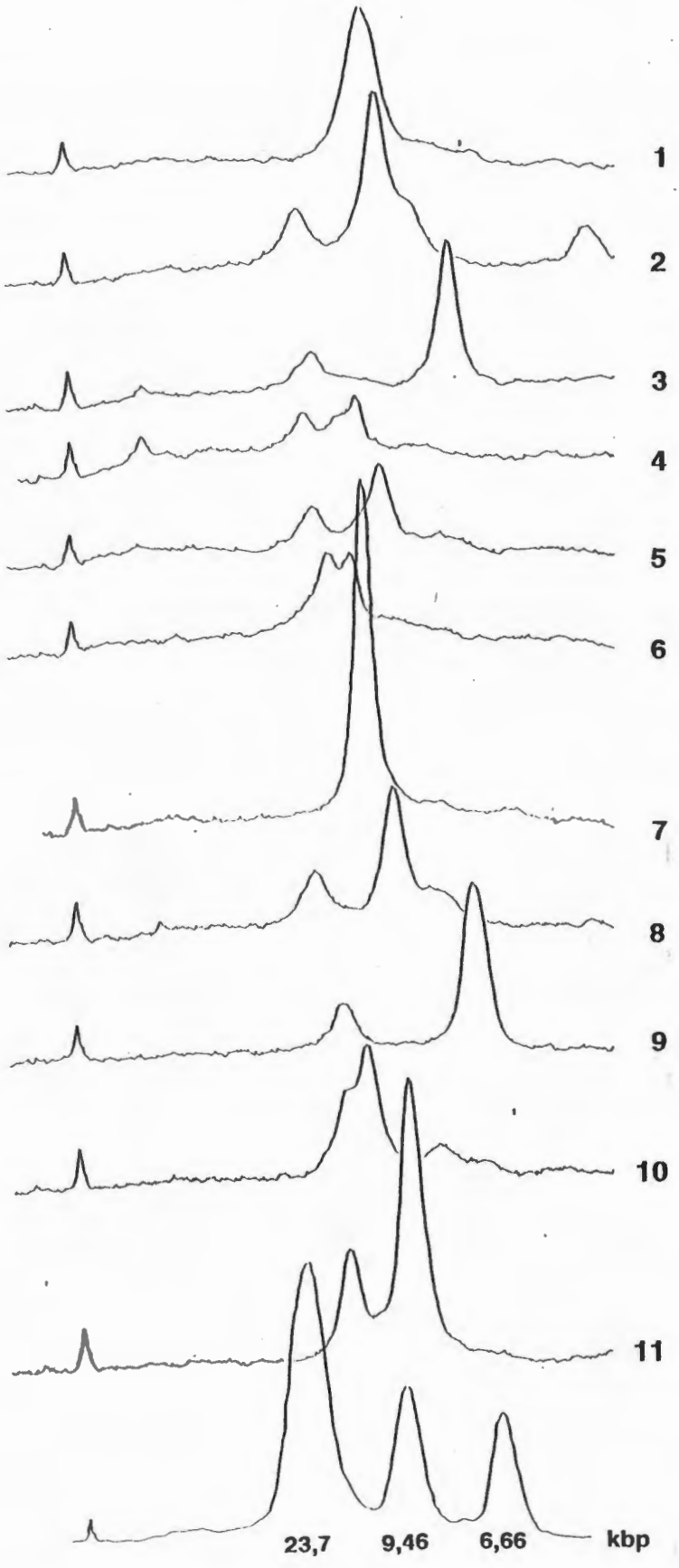


Figure 4.1.1. Microdensitometer scan of the gel depicted above. The distance migrated was accurately measured and the molecular sizes assigned using a linear regression (ln[size] Á migration) with correlation coefficient of 0,97.

4.3. Degree of polymorphism

The second question cannot be answered by mere inspection. Restriction fragment polymorphism is rare amongst the sea urchin homeobox containing genes but for hb-3 (Dolecki 1986, 1988a, 1988b). For other sea urchin genes though, polymorphism is quite common (Posakony 1983; Fregien 1983). This may be explained by the high sequence divergence. Britten et al. (1978) determined the single copy sequence difference between *S.purpuratus* individuals to be 4% by using thermal stability measurements.
To achieve equal restriction fragments from different haploid genomes two conditions have to be met. Firstly the restriction sites defining the fragment ends have to be conserved and secondly, no new sites may arise between these end sites. The second condition is fragment length dependent. Given a sequence independent rate of divergence, the larger the restriction fragment the greater the possibility of a new site arising and thus the greater the polymorphism observed.
Reexamining polymorphism patterns one notices that the polymorphism of hb-3 (Dolecki 1986), the HindIII digest of Su-en (Dolecki 1988b) and *PaHbox1* occurs with large (> 8 kbp) fragments whereas less polymorphism is observed when the fragments are small (Dolecki 1986,1988a,b).
A second factor that favours little polymorphism in short fragments concerns the nature of the area probed. In most cases the probe hybridizes to a coding sequence. Such a sequence is under evolutionary constraint and hence cannot diverge as much and rapidly as non-essential flanking regions. There is on average a 5 fold difference between the average substitution rates, s, in various gene regions: coding non-degenerate sites versus fully degenerate sites (mainly third codon positions),flanking regions, introns and pseudogenes (Li 1985, see also table 3.1). These data were extracted from mammalian sequences but should be generally applicable. One thus expects short restriction fragments covering mainly coding areas to show about 2 fold lower

levels of divergence (average of 2 nondegenerate and one degenerate sites per codon) than the genomic average which is mainly that of non-coding areas (C-value paradox). Thus one can reformulate the second question as "Is the observed polymorphism significantly higher than expected ?" To answer this, we have to derive a statistical method for examining and predicting the restriction fragment pattern.

4.3.1. Statistical method for predicting the frequency of restriction fragments.

We need to determine the probability, P, that an ancestral restriction fragment having undergone a constant rate of nucleotide substitution, s, is still found at time, t. As already mentioned, we have to meet two independent conditions

$$P = (A)^2 * B \qquad (1)$$

where
A is the probability of no change occurring in one of the restriction sites; A^2 : both sites
B is the probability of no restriction site appearing between the end sites.

For this discussion we shall neglect the rather unpredictable and in most cases selectively disadvantageous possibilities of genomic rearrangements (insertions, deletions, inversions) and assume that most changes occur by nucleotide substitutions. Small insertions, deletions may be regarded as substitutions whereas large (kbp) changes would not have been detected in Britten et al's divergence measurements as these authors used DNA fragmented to a length of 800 bases.

$$A = (1-d)^r \qquad \text{(Upholt 1977)} \qquad (2)$$

where d, the degree of sequence divergence = st;
r = the number of base pairs of the restriction enzyme recognition sequence

B is derived by considering a site of r bp. The probability of a new restriction site being formed at this site is the probability of one or more bp changing, times the probability, z, of this new random sequence being a restriction site.
The probability of one or more bp changing is simply one minus the probability of no change occurring:

$$1 - A \qquad (3)$$

whereas the chance of a new site arising (a), is

$$z = (1/4)^r \qquad \text{(Bothan 1973)} \qquad (4)$$

(This formula assumes a GC content of 50%).
Thus the probability of a new restriction site being formed can be expressed as

$$(z) \times (1 - A) \qquad (5)$$

and the probability of NO new restriction site arising at a site of r bp as

$$1 - \{z(1 - A)\} \qquad (6)$$

Because there are L - r + 1 such sites in a fragment of length L (in bp), the probability of no restriction site appearing between the end sites, is given by

$$B = [\ 1 - z(1 - A)\ ]^{L-r+1} \qquad \text{(Upholt 1977; Nei 1979)} \qquad (7)$$

It should be mentioned that the last equation is not strictly correct because the L-r+1 sites of size r overlap each other and thus cannot be considered to be fully independent (McNeill, personal communication).

Thus, from equations (1), (2) and (6) :

$$P = ([1-d]^6)^2\ \{1-(1/4)^6(1-[1-d]^6)\}^{L-5} \qquad (8)$$

(see table 4.2)

This gives the probability of finding a restriction fragment after time t {d=st}. We can safely assume that all the sea urchins collected from the same site share a common ancestor so $t_i = t_j$ where i,j represent two haploid genomes. As both

i and j have diverged for time t from their common ancestor, their relative divergence to each other is $\lesssim$ 2st. The $\lesssim$ function is used to indicate the possibility of parallel and back mutations which become significant with increasing divergence, d.
The divergence measured by Britten et al (1978) is that between individual haploid genomes, hence the value of 4% $\lesssim$ 2st and so d = st $\gtrsim$ 2%.

4.3.2. Comparing predicted and observed polymorphism

We now have to apply the above considerations to the observed data. There are only two possibilities for any genome: to have the ancestral fragment (probability P, eqn.8) or not to have it (1-P). Hence we can use the binomial distribution to derive the expected number of ancestral fragments

$$f(x) = \binom{n}{x} (P)^x (1-P)^{n-x} \qquad (9)$$

where $f(x)$ = probability that a given fragment occurs x times in a population of n individuals.
The properties of the binomial equation (9) allow us to determine the expected value of x, (X), as well as the variance of this value:

$$X, \text{ the expected number of fragments } = \sum_{x=0}^{n} x\, f(x) \qquad (10)$$

$$v, \text{ the variance, } = \left(\Sigma\, x^2\, f(x) \right) - X^2 \qquad (11)$$

Equations 10 and 11 have been calculated for various values of d and L (see table 4.3.1 and 4.3.2) and graphically displayed in fig.4.2.
Examining table 4.1., it is by no means clear which fragment, if any, represents the hypothetical ancestral one. However, the most likely candidate is the one occurring at the highest frequency.

Table 4.2 P(equal) for values of d and L (in bp)

d (ratio)	P(equal) for L = 5000	10000	13000	16000	19000
0.005	0.9082	0.8759	0.8571	0.8387	0.8207
0.010	0.8253	0.7684	0.7362	0.7053	0.6757
0.015	0.7504	0.6751	0.6336	0.5946	0.5580
0.020	0.6827	0.5939	0.5463	0.5025	0.4622
0.030	0.5660	0.4616	0.4084	0.3614	0.3198
0.040	0.4701	0.3606	0.3076	0.2623	0.2237
0.060	0.3260	0.2233	0.1779	0.1418	0.1130
0.080	0.2275	0.1407	0.1054	0.0790	0.0592
0.100	0.1595	0.0900	0.0639	0.0453	0.0321

Table 4.3.1 The expected (mean) number, X, of fragments of size L (in bp) in a population of size n = 22, for various values of divergence, d.

d (ratio)	X for L = 5000	10000	13000	16000	19000
0.005	19.981	19.271	18.857	18.452	18.056
0.010	18.157	16.905	16.196	15.517	14.866
0.015	16.510	14.852	13.938	13.081	12.276
0.020	15.020	13.066	12.018	11.054	10.168
0.030	12.451	10.155	8.985	7.951	7.035
0.040	10.342	7.933	6.766	5.771	4.922
0.060	7.173	4.912	3.914	3.119	2.485
0.080	5.005	3.095	2.320	1.739	1.303
0.100	3.509	1.980	1.405	0.997	0.707

Table 4.3.2 The standard deviation of X (see table 4.3.1) for different values of d and L.

d (ratio)	Std. dev. of X for L = 5000	10000	13000	16000	19000
0.005	1.354	1.546	1.641	1.725	1.799
0.010	1.781	1.979	2.067	2.138	2.196
0.015	2.030	2.197	2.260	2.303	2.329
0.020	2.183	2.303	2.335	2.345	2.338
0.030	2.325	2.338	2.306	2.253	2.188
0.040	2.341	2.252	2.165	2.063	1.955
0.080	1.966	1.631	1.441	1.265	1.107
0.100	1.717	1.342	1.147	0.976	0.827

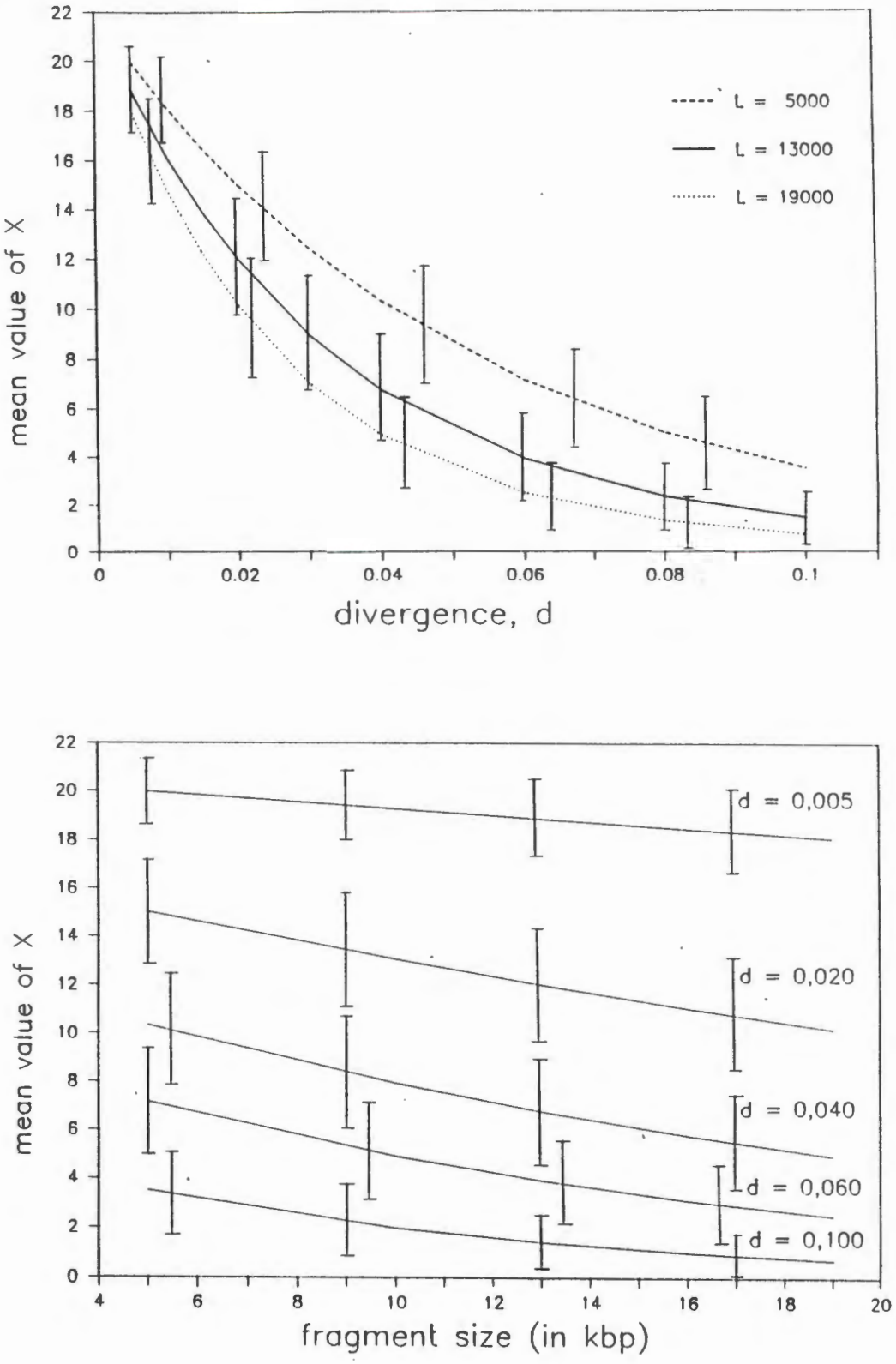


Figure 4.2. Expected frequency of occurrence (mean value of X) of fragments of size L (in kbp) at a sequence divergence, d, for a population sample of size n = 22 (see text and tables 4.3.1 and 4.3.2).

If we take the fragment occurring with the highest frequency as the ancestral one, then L = 13000 (bp) and the observed frequency = 4. The 95% confidence interval
$= 4 \pm 1{,}96\sqrt{4(1-4/22)}$
$= 4 \pm 3{,}5.$
This large confidence interval implies that the sample size is too small to determine the true population frequency of this fragment or to unambiguously pinpoint an ancestral fragment. However, we can answer the question we set out to solve: Is the observed polymorphism significantly higher than expected? The answer is yes :
The observed number of fragments of size L = 13000 is $4 \pm 3{,}5$ (95 % confidence). However, the expected number of fragments of size L = 13000 given d = 0,02 is $12{,}0 \pm 2{,}3$. This comparison can be repeated taking different fragments sizes as "ancestral"; in each case the observed and expected ranges do not overlap.
We can thus conclude that *PaHbox1* is situated in a genomic area displaying higher than average (the average for single copy DNA) evolutionary change. Taking L = 13000 and the observed frequency as 4, the predicted divergence from an ancestral fragment would be 6% (cf. 2%) and between haploid genomes $\leq$ 12%.

CHAPTER 5

EXPRESSION

5.0 Index

5.1 RNA dot blots

Most homeoboxes are expressed at some stage during embryogenesis and therefore RNA was isolated from several embryonic stages (egg, 4 hr = 128-cell morula, 9 hr hatching blastula, 14 hr mid-gastrula, 21 hr prism and 30 hr pluteus stage embryos) and probed with *PaHbox1* at high stringency [7.7.3]. RNA:DNA hybrids are more stable in 50% formamide than DNA:DNA hybrids (Casey 1977), hence the high stringency conditions used for DNA to DNA hybridizations would not result in any signal being missed. Nonetheless no hybridization signal was obtained. This meant that either *PaHbox1* was not expressed during embryogenesis or it was expressed transiently during a time intermediate to those sampled. For this reason 2, 6, 11.5, 17 and 25 hr embryonal RNA was isolated. In addition adult tissues were examined and autoradiographic exposures increased to 14 days using two intensifying screens.

Figure 5.1 shows a RNA dot blot with all the above mentioned RNA samples. 100 μg of total RNA was loaded as well as 100 to 0,032 pg of pHB1 (*PaHbox1* in Bluescript) as a hybridization control. Total RNA was used in preference to poly-A^+ mRNA because the isolation of the latter may introduce a bias in the quantity of *PaHbox1* messenger RNA. It can be seen that *PaHbox1* is expressed profusely in adult

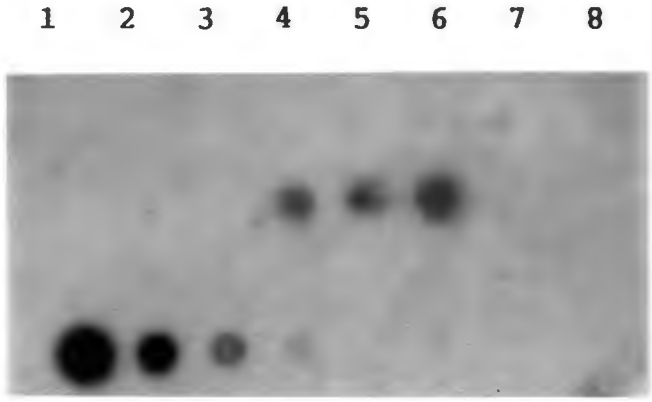
tissues (ovary, testis, intestine) but at hardly detectable levels in 11,5 and 14 hr embryos.

5.2 RNAase titrations

The most accurate means of quantitating the level of a particular RNA species is that of RNAase protection titrations. This method is superior to methods based on the transfer of RNA to a solid support, followed by hybridization to a probe and subsequent autoradiography, as these are limited in their quantitative aspects by the inability to control the efficiency of transfer and binding to the support as well as the uncertainties involved in the hybridization kinetics of filterbound nucleic acids (Beltz 1983). It is also useful to know that RNA samples degraded to an average length of 500-1000 nucleotides are still useful substrates for this assay (Lee 1987).

The RNAase protection assay is described in section 7.8. Basically, a series of hybridization reactions is performed under conditions of equal amounts of probe of at least ten fold excess and increasing amounts of total cellular RNA. The hybridization reaction is thus pseudo first order and is carried out to completion (>95%) by maximizing the probe concentration. Samples are RNAase digested and the resistant fragments TCA precipitated and scintillation counted. The observed ribonuclease background is generally ≤ 0,5% of the input probe (Lee 1987) and thus the lower sensitivity limit of the method is determined by the lowest probe concentration that can be used for the reaction to go to completion in a reasonable time (≤ 2 days). Longer reaction times may lead to unacceptable levels of radiochemical damage of the RNA (Sambrook 1989).

The amount of probe needed to achieve a ten fold excess was estimated by comparing the signal obtained with 100 μg of adult tissue RNA to the pHB1 series in the RNA dot blot (fig 5.1). This estimate was 10 pg per 100 μg. A trial RNAase



Key:

1	2	3	4	5	6	7	8	
egg	2hr	4hr	6hr	9hr	11,5 hr	14 hr	17 hr	A
21hr	25hr	30hr	tes.	ovary	int.			B
PaHbox1:						Antp:		C
100pg	20pg	4 pg	0,8pg	0,16 pg	0,032 pg	100pg	20pg	D

Figure 5.1. RNA dot blots of total RNA samples of embryonic atages and adult tissues. The Northern blot was prepared as described in section 7.7.3. High stringency hybridization conditions and a 800 bp *PaHbox1*-containing probe were used [7.7.4]. On quadrant 1A to 6B, 100 µg of total RNA was loaded (Tes = testes, Ov = ovary, Int = intestines). Row D represent hybridization controls: quadrants 1D to 6D represent amounts of the *PaHbox1* homeobox-containing pHB1 DNA, 7D and 8D amounts of the *Antp* homeobox-containing p903G DNA [7.5.1.3]. The stringency conditions used were able to differentiate between the 78% identical homeoboxes.

titration experiment with ovary RNA gave a value of 5 pg. For the following experiment a minimal amount of probe was used (50 pg/ 30 μl reaction) to allow the detection of embryonic levels above background, and the maximum amounts of testes and intestine RNA reduced to 50 μg. The rate constant, k = (5400/L) x 170 M^{-1}/sec under standard conditions (Galau 1976). In my experiments, L, the probe length, is 200 nucleotides, the monovalent cation concentration is 0,4 M which gives an acceleration factor of 3,4 fold (Britten 1974) over standard conditions whereas the effect of the 80% formamide used, is to reduce the rate by about 80% (Young 1985). The rate constant k is thus 3,1 x 10^3 liter.mol^{-1}.sec^{-1}.
For a pseudo first order reaction, $S = e^{-kCot}$
where S is the fraction of complementary RNA that remains unpaired, C_0 the probe concentration in moles of nucleotide per liter and t the time in seconds (Lee 1987). The reaction is thus 95% complete after t = ln(.05) ÷ (-3,1 x 10^3 liter.mol^{-1}.sec^{-1}) ÷ (1,7 x 10^{-3} μg/ml) x (83 hr.μg/sec.mol) = 47 hours (90% complete after 36 hrs).

The data obtained from these experiments (μg RNA versus counts precipitated) was used in a least squares analysis (table 5.1). The regression line was used to determine the Y-axis intercept which represents the background level of counts. This value was subtracted from all the readings and the data replotted as shown in figure 5.2.

The limit of the sensitivity of this method was approached by the readings obtained with the RNA samples from Aristotle's lantern. This is reflected in the slightly lower correlation coefficient (r = 0,971). No protection could be detected using embryonal RNA.

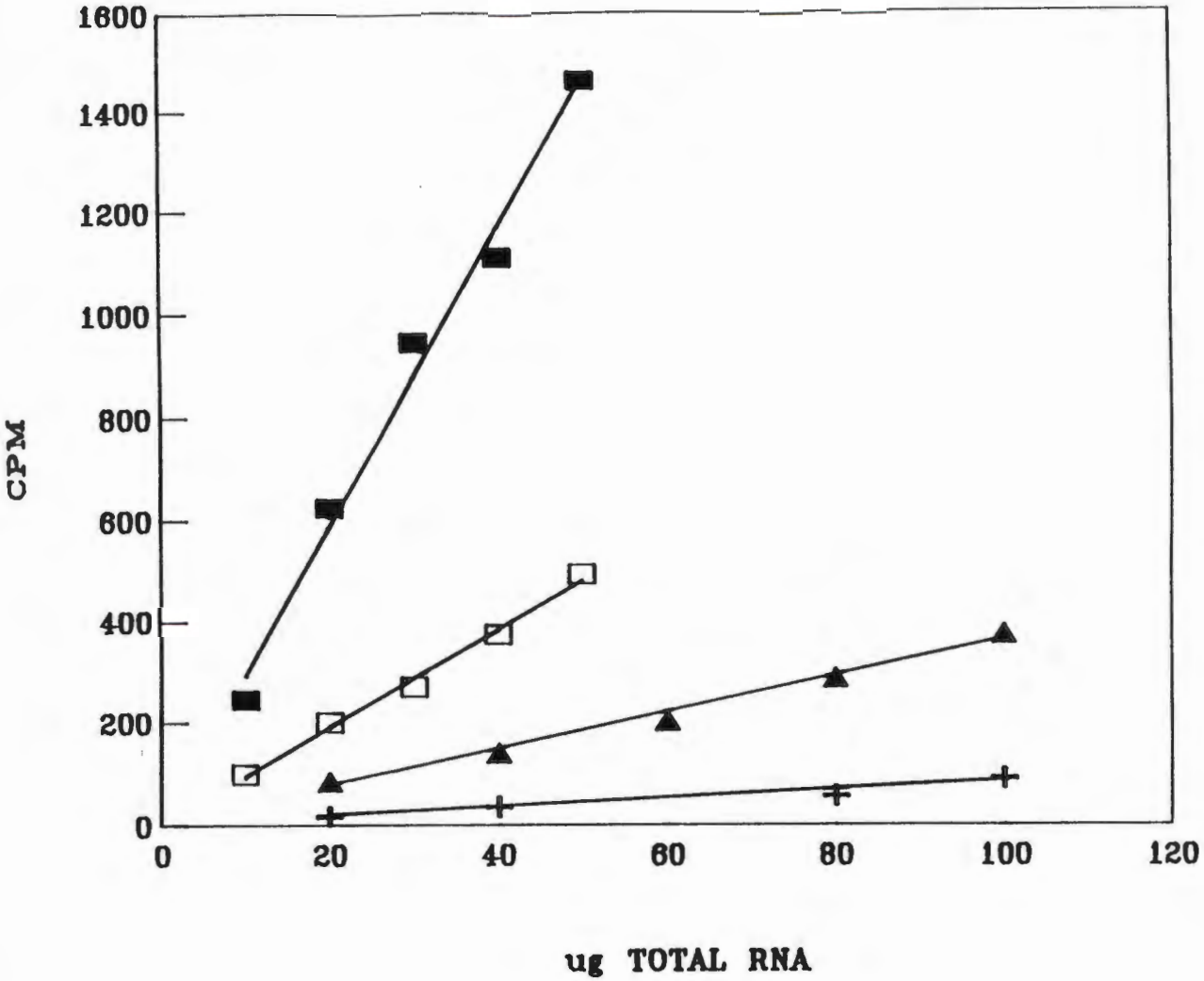


Figure 5.2. RNAase titration of PaHbox1 RNA levels in P.angulosus. Raw data was obtained as described in section 7.8 and plotted to determine the background counts from the y-axis intercept of the regression line. This value was subtracted from each reading and replotted, shown above. 11,5 hr embryo RNA and control yeast tRNA gave no increase above background and are not shown. The number of transcripts, n, was calculated using the relationship
n = (mass of RNA) * (slope of titration curve) * (Avogadro's Constant) / [(probe specific activity) * (α)]
where α = fraction of probe that is can hybridize
= 240 bp / 800 bp
Resulting values are shown in fig.5.5 and table 5.1. Symbols and correlation coefficients (r): closed rectangle = testis, r = 0,993; open rectangle = intestines, r = 0,996; triangle = ovaries, r = 0,995; and "+" = Aristotle's lantern, r = 0,971.

One can now estimate the number of transcripts, n, for a given mass of RNA:

$$n = m \cdot s \cdot N_0 \div (S_a \cdot \alpha)$$

where

m = mass of RNA (in μg)

s = slope of the regression line (cpm/μg)

N_0= Avogadro's constant

S_a= specific activity of probe (cpm/mol)

α = fraction of probe (counts) that can hybridize
= 270 bases / 800 bases (see [5.3] for the protected size)

Table 5.1 Analysis of RNAase protection titration assay

tissue	testes	ovary	intestine	Aristotle's lantern
slope of regression line (cpm/μg)	29,21	3,59	9,56	0,83
Std.err.of slope	1,98	0,21	0,48	0,14
corr.coeff.(r)	0,993	0,995	0,996	0,971
$n/10^6$ per 100 μg^a	110	13	42	3,7

a. See text for calculation

To determine the number of transcripts per cell or embryo, the amount of RNA per cell or embryo has to be known. This information is not available for the South African sea urchin. We can estimate the amount of RNA per cell as being 10 pg (Maniatis 1984, Lee 1987). The calculated number of transcripts per 10 pg of total RNA is given in table 5.3.

5.3 RNAase mapping

The titration assay was not sensitive enough to allow the detection and quantitation of embryonal *PaHbox1* mRNA levels.

To this end, as well as to obtain information as to the exon borders, RNAase mapping was done.

The RNAase mapping methodology is described in section 7.8. It is identical to the RNAase titration assay until after the RNAase digestion step whereafter the samples were purified and electrophoresed on a sequencing gel. The dried gel was autoradiographed for up to 14 days and the size of the protected bands determined by comparison to molecular size standards and quantitated by microphotometric scanning [7.9.3].

The autoradiograph film was preexposed to an absorbance increase of 0,15 OD_{540nm} units and exposed at -70°C. This extends the range over which the signal is proportional to the amount of radiation to 0,1 - 1,5 OD_{540nm} units (Laskey 1980).

Figure 5.3 shows an autoradiograph of adult and embryonic RNA samples exposed for 14 days. The lanes within the linear response range of the film were scanned and the peaks quantitated by cutting out and weighing. These weights were converted to absolute values using a conversion factor derived from the readings for Aristotle's lantern. Representative scans are shown in fig.5.4 and the peak values and calculations given in table 5.2.

I have determined the total amount of RNA per 11,5 hr embryo to be 6,4 ± 0,8 ng as outlined in detail in section 7.2.6. For comparison, the total RNA content of *S.purpuratus* eggs and embryos is 2,8 - 3,3 ng/embryo (Goustin 1981). It is thus possible to calculate the number of *PaHbox1* transcripts per embryo. The results for adult tissue and embryonal *PaHbox1* mRNA levels are summarized in table 5.3 and fig.5.5.

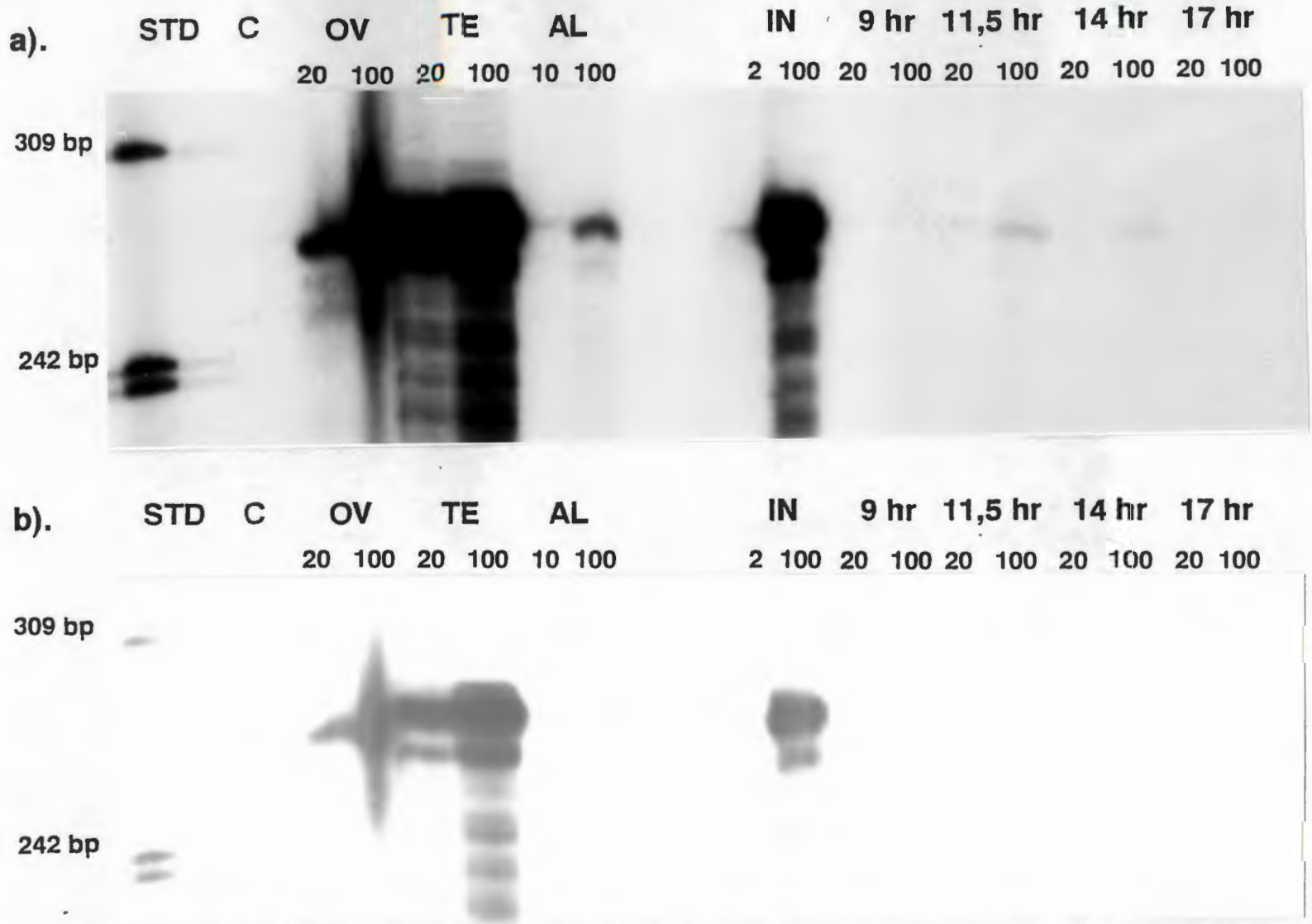


Figure 5.3. Expression of PaHbox1 in adult tissues and during embryogenesis. Total RNA samples were hybridized to PaHbox1 for 36 hours, digested with RNAase A and T1 and electrophoresed on a 6% sequencing gel as described in sections 7.6.4 and 7.8. To show up the faint bands of the embryonic stages, the gel was exposed for 14 days with an intensifying screen. The Aristotle's lantern signal fell within the linear range of the film's response curve and hence the scan could be directly compared with scans of the embryonic stages (see fig.5.4). a.), b.) are different exposures of the same gel. STD = HpaII-cut pBR322 labeled with ^{32}P-dCTP using the Klenow fragment, C = control (100 µg yeast tRNA), OV = ovaries, TE = testes, AL = Aristotle's lanterns, IN = intestines, x hr refers to embryos of age 'x'. Numbers above the lanes indicate the amount (in µg) of total RNA used. The RNA samples for fig.5.2 and 5.3 are from the same isolation.

a).

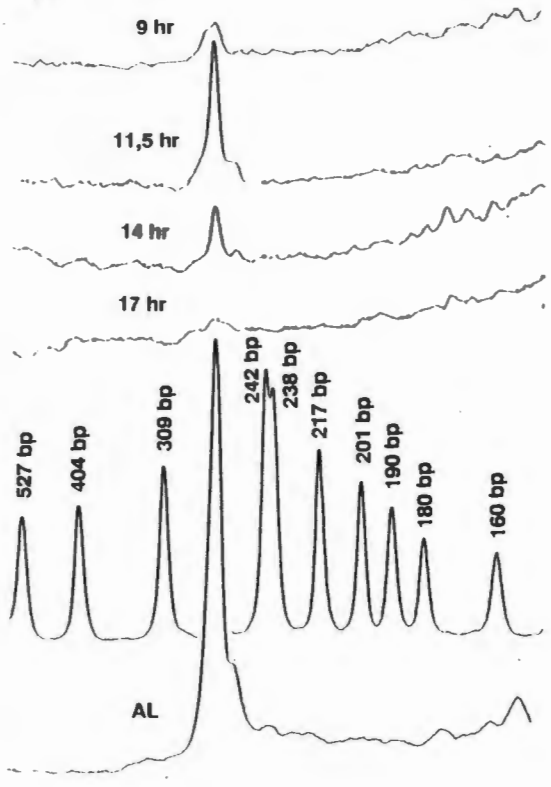


Figure 5.4. *a.)* Densitometric scans of significant lanes of the gel depicted in fig.5.4. See table 5.2 for the quantitation of peak areas. *b.)* Determination of the size of the protected band. The regression equation is log(size in bp) = 2,797 - 0,00188 x (migration). The measured migration was 192 ± 0.816 units giving a size of 272,7 ± 1,0 bp.

b)

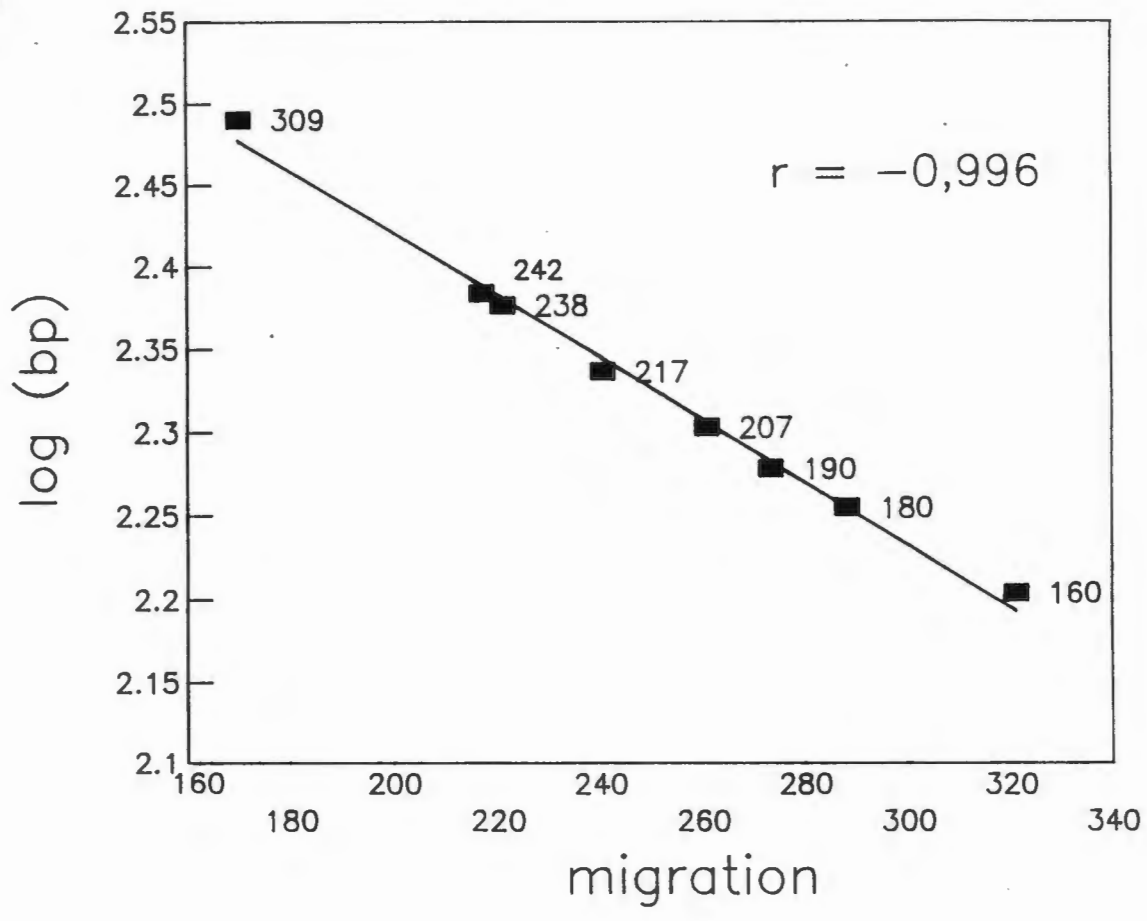


Table 5.2 Analysis of Scanning Data

sample	Aristotle's lantern	9 hr	11,5 hr	14 hr	17 hr
Weight / 100 µg (in mg x scale)	14,78	5,14	27,6	8,71	2,38
Std.error[a]	6,2%	10%	5,1%	6,1%	13%
n per 100 µg (in thousands)[b]	3700	130	690	220	60
Std.error of n[c]	17%	21%	19%	19%	22%

a. Sample sizes between n = 4 for A.L. and 3.

b. Number of transcripts calculated as follows:
$n = (\text{weight of i})/(\text{weight of AL}) \times n_{AL}$
where i is any sample, AL iis Aristotle's lantern,
n_{AL} = number of AL transcripts per 100 µg total RNA
$= 3,7 \times 10^6$ (from table 5.1).

c. Combined error calculated as square root of the sums of the squares of the errors listed in *a* for A.L. and the respective sample and the error in the transcript number estimation of A.L.(see table 5.1)

It is of interest to point out that the sensitivity of this method is very high. In the case of 17 hr embryonal RNA, as little as 60000 transcripts per 100 µg RNA could be detected. This represents $1,0 \times 10^{-19}$mol or 9,5 fg of RNA which converts to about 1,6 cpm given $S_a = 5 \times 10^8$ cpm/µg and α = 270bases/800bases. Detectable images (OD_{250nm} = 0,02) can be achieved with 0,5 dpm/mm²/day using Kodak XAR-2 film (Laskey 1980). The area of the above band is about 5 mm² giving an exposure of 4,5 dpm/mm²/day (Cronex-4 film).

The size of the protected fragments was determined to be 273 nucleotides for all RNA samples tested (fig.5.3 and 5.4b). This defines the 5' exon border as discussed in section 3.2.

Table 5.3. Quantitation of PaHbox1 mRNA levels in *P.angulosus*

Sample[a]	Embryonic stage	transcripts /10 pg RNA[b]	levels rel. to 11,5 hr	transcripts per embryo [c]
ovary	adult	1,3 ± 0,08	19	
testes	adult	11 ± 0,7	160	
intestine	adult	4,2 ± 0,2	61	
Aristotle's lantern	adult	0,37 ± 0,06	5,4	
9 hr embryo	hatched blastula	0,013 ± 0,003	0,2	8,3 ± 1,7
11,5hr embryo	mesenchyme blastula	0,069 ± 0,013	1	44 ± 8,4
14 hr embryo	mid - gastrula	0,022 ± 0,004	0,3	14 ± 2,7
17 hr embryo	late gastrula	0,006 ± 0,001	0,1	3,4 ± 0,7

a). Expression was not detectable in egg, 2, 4, 6, 21, 25 and 30 hr embryos.
b). Adult tissues - values from table 5.1.and fig 5.2: RNAase titrations.
Embryonic levels - see table 5.2.: RNAase mapping.
10 pg represents a rough approximation to the amount of total RNA present in an adult cell (Maniatis 1982).
c). The amount of total RNA per *P.angulosus* 11,5 hr embryo is 6,4 ± 0,8 ng (see section 7.2.6). For the calculations of transcript per embryo 6,4 ng of total RNA / embryo was taken for all embryological stages.

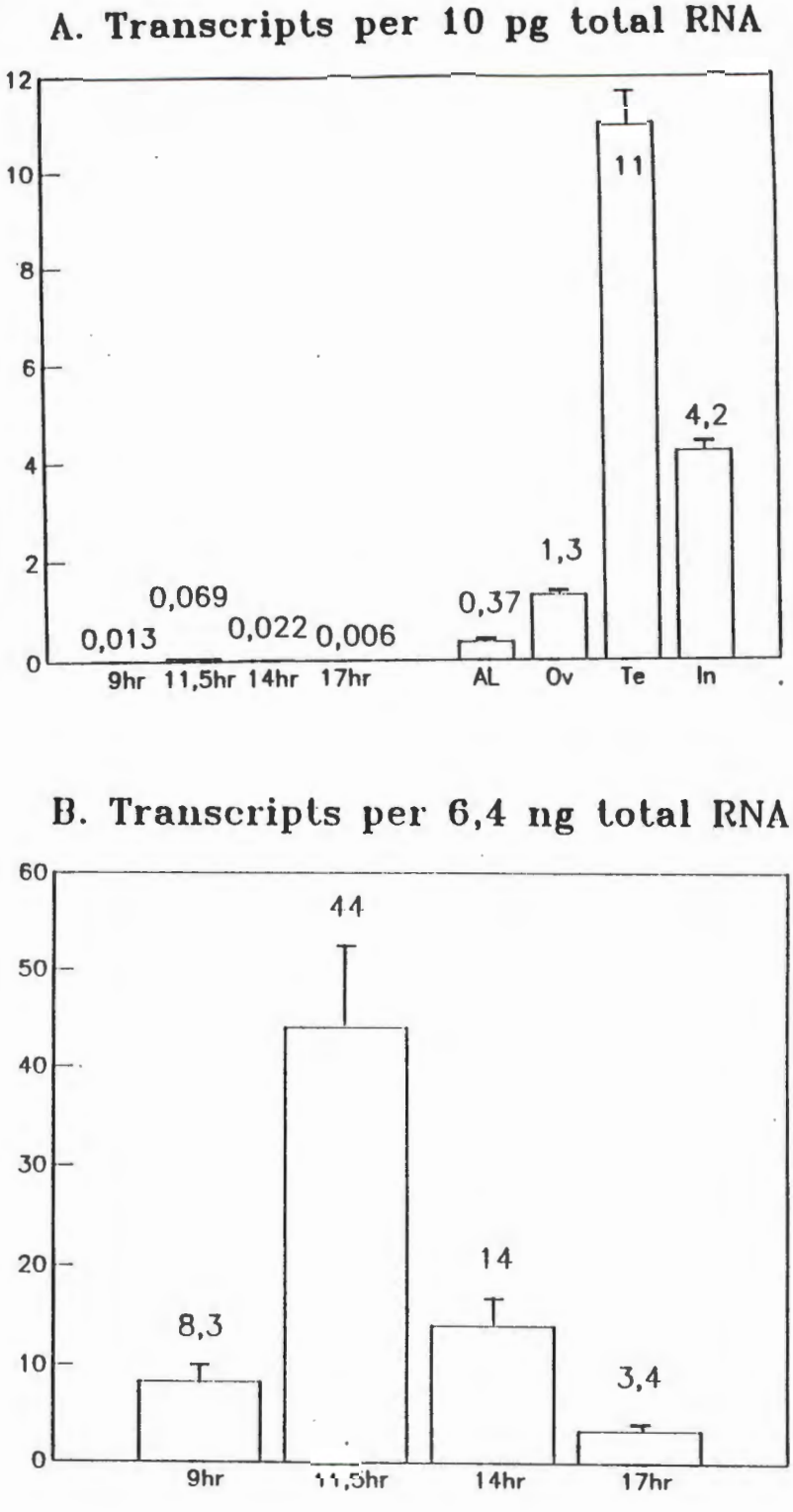


Figure 5.5. Histograms of PaHbox1 transcript levels. (A) Transcripts per cell assuming a constant total RNA content of 10 pg per cell in *P.angulosus* embryos and adult tissues. (B) Transcripts per *P.angulosus* embryos. The error bars represent half standard errors.

The additional minor bands seen mainly in the overexposed lanes of RNA from adult tissue in fig.5.3, are caused by a small percentage of less than full length probe transcripts and can be disregarded.

5.4 Northern blots

Poly-A^+ messenger RNA was isolated by standard procedures [7.2.5] and 2 to 10 μg was glyoxal denatured [7.5.2.3] and electrophoresed on 1 to 1,1% agarose gels [7.6.1]. Removal of the glyoxal adduct, Northern transfer onto nylon membranes and hybridization conditions under high stringency are discussed in detail in section 7.7.2. Lambda HindIII/EcoR1 cut size standards were co-electrophoresed and visualized by a separate hybridization to nicktranslated lambda DNA.

An example of such a gel is shown in fig 5.6. Intestine RNA was consistently partly degraded in spite of freezing the tissue immediately after dissection and thereafter exposing it to a 4 M solution of guanidinium isothiocyanate. I attribute this to a high concentration of RNAases in the intestinal tract with degradation occurring in the time taken to dissect the animal.

As seen in fig 5.6 *PaHbox1* transcripts could only be detected in adult testes, ovary and intestine. In all three tissues a transcript of 5,7 kbases is detected but in testes an additional band at 5,2 kbases appears to be present. It should be pointed out that the purpose of this experiment was qualitative and not quantitative and the levels seen in this Northern blot may not be strictly equal to those determined accurately by the RNAase protection assays because a bias may have been introduced by different purification yields during the preparation of the poly-A^+ mRNA from total RNA. During embryogenesis and in different adult tissues the levels of household genes such as actin also varies as shown in fig.5.7. Visual examination of

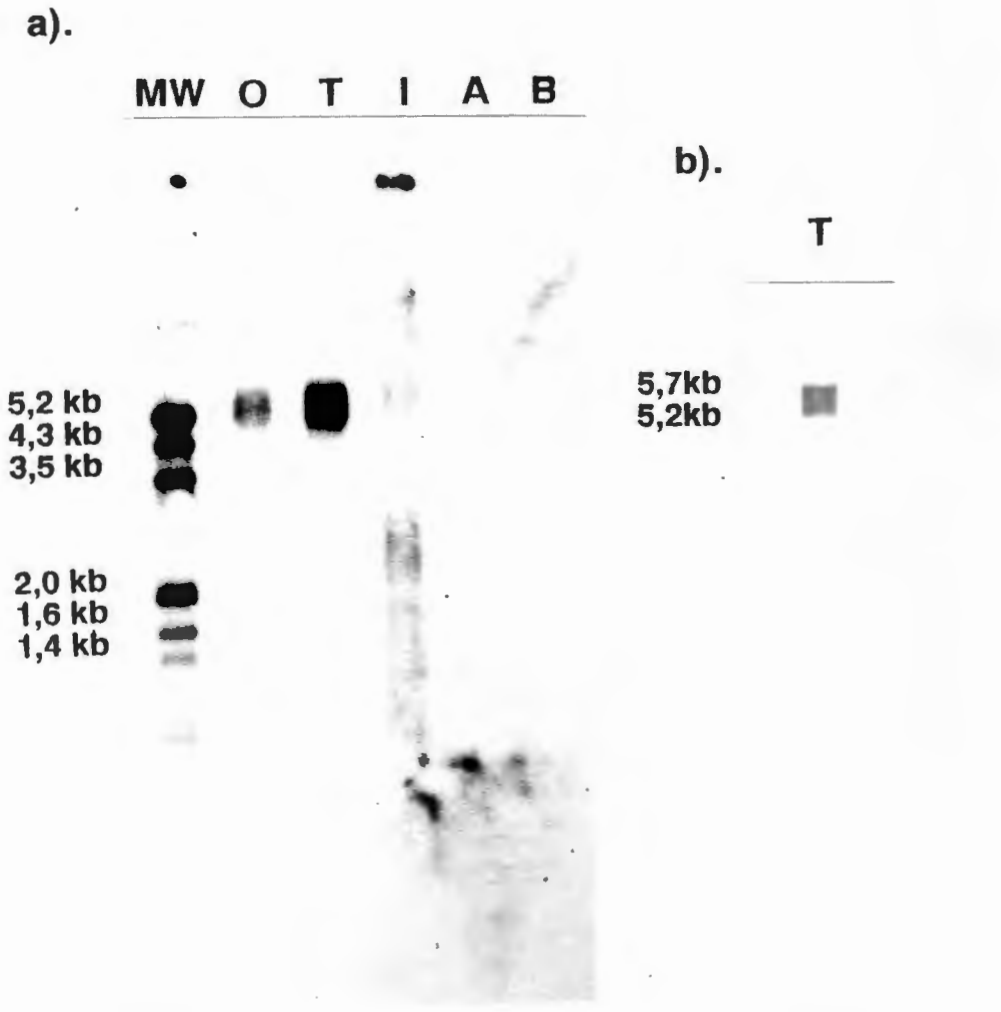


Figure 5.6. Northern transfer of poly(A)$^+$RNA from sea urchin embryos and adult tissues hybridized to *PaHbox1*. a.) 10 µg of poly(A)$^+$RNA from the indicated samples was glyoxal denatured and run on a 1,1% agarose gel, transferred to nylon membrane and hybridized to PaHbox1 DNA probe under high stringency conditions. The filter was exposed at -70°C with an intensifying screen for 10 days. The HinDIII/EcoRI-cut lambda standards (MW) were shown up by a seperate hybridization to a lambda probe. Intestine RNA is somewhat degraded but still shows the characteristic band(s) at 5,2 / 5,7 kilobases. b.) Shorter photographic exposure of the testis lane to show the two separate bands of size 5,2 and 5,7 kilobases. O = ovary, T = testis, I = intestine, A = Aristotle's lantern, B = 11,5 hr mesenchyme blastula embryo.

acridine orange-stained RNA samples indicates similar levels of poly-A^+ RNA and, except for intestinal RNA, intactness as judged by the sharpness of the poly-A^- RNA bands (fig.5.7).

5.5 Discussion

The expression data is summarized in table 5.3 and fig.5.5. It is evident that the *PaHbox1* gene is expressed in two distinct temporal phases - transiently during embryogenesis and at higher levels in adults. Using a value of 6,4 ng total RNA per embryo, the abundance of *PaHbox1* is only 43 transcripts per embryo at mesenchyme blastula stage during which expression is maximal. At this stage the embryo has between 500 and 1000 cells and it is very likely that expression is restricted to a subpopulation of cells as otherwise these expression levels are too low to bear any biological function (see column 3 of table 5.3). *PaHbox1* levels fall off rapidly (3 to 5 fold) over only 2,5 hr of development to either side of 11,5 hr blastula stage, indicating a tight temporal control of expression.

During the second phase of expression, in adults, levels are higher than during embryogenesis in all four tissues tested. The highest levels are found in testes (11 transcripts per 10 pg of total RNA) which is 160 times as high as in mesenchyme blastula. There are two transcripts of size 5,2 and 5,7 kbases (fig.5.6). Ovary, intestines and Aristotle's lantern show expression at several fold lower levels (table 5.3).

The second phase of *PaHbox*1 expression is typical for homeobox-containing genes of echinoderms, vertebrates and *Drosophila* as discussed in section 1 and may reflect a principle of evolution - it is easier to generate new controlling regions (typical DNA recognition consensus sequences are only around 10 bp long) than new DNA binding regulatory proteins which would anyhow need their own target recognition regions.

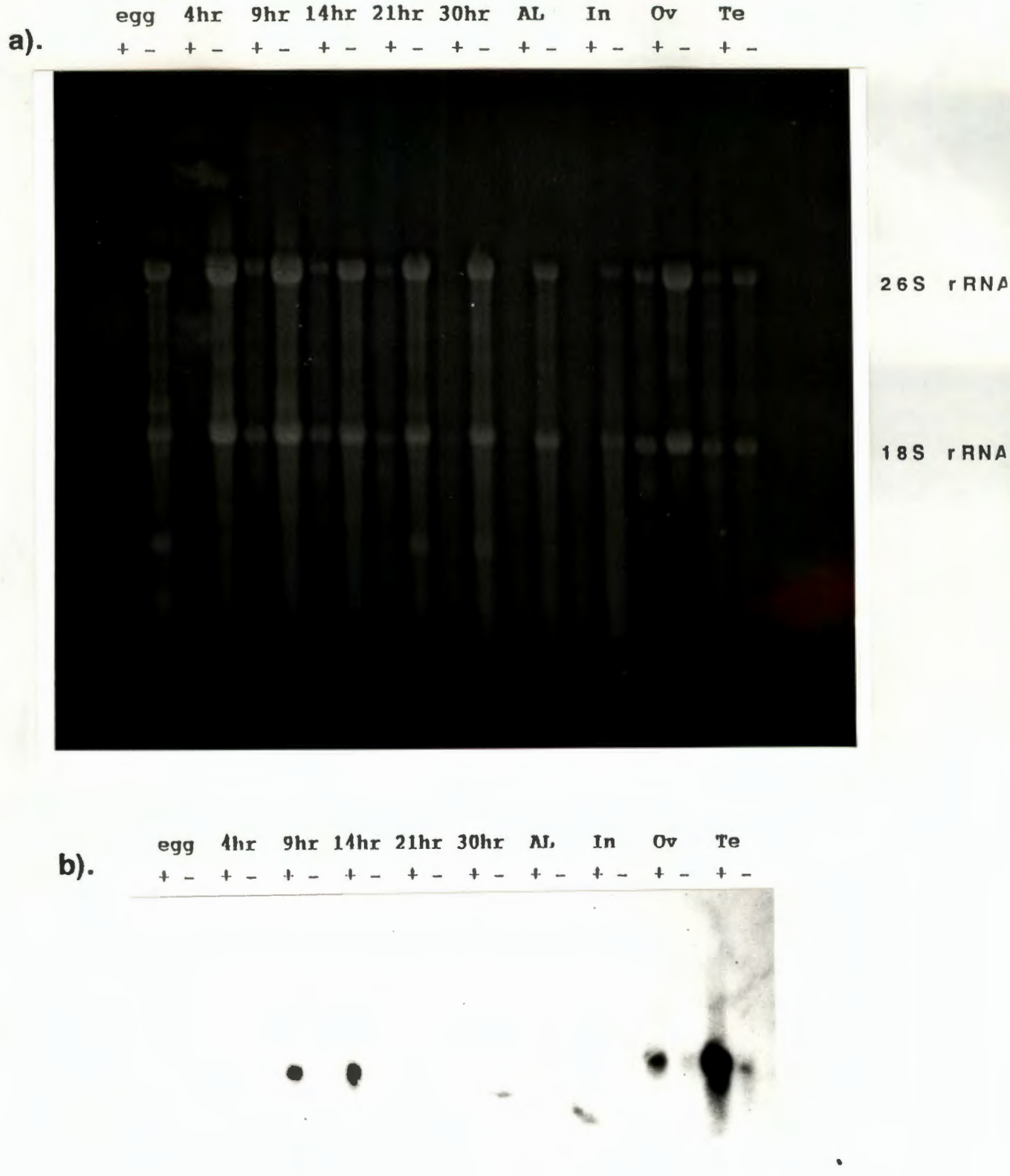


Figure 5.7. *a*.) Acridine orange-stained gel of poly-A$^-$ and poly-A$^+$ RNA samples. 2 µg poly-A$^+$ and 10 µg poly-A$^-$ RNA was loaded except for the following samples: egg, 4 hr = 0,8 µg poly-A$^+$; Aristotle's lantern = 0,6 µg poly-A$^+$. Intestine RNA is slightly degraded as can be seen by the increase in smearing relative to the rRNA bands in the poly-A$^-$ sample. A.L. = Aristotle's lantern, In = intestines, Ov = ovary, Te = testes. *b*.) The gel depicted in *a*.) was Northern transferred [7.7.2] and hybridized to a rat β-actin probe at low stringency [7.7.4]. Levels during embryogenesis and in different adult tissues vary greatly but where present, the single band seen indicates intactness of the poly-A$^+$ RNA. Neither Aristotle's lantern nor intestine tissues show actin expression.

CHAPTER 6

CONCLUSION

PaHbox1 represents a novel sea urchin homeobox of the *Antennapedia-Hox1.1* subclass. Of the four published sea urchin homeoboxes of *T.gratilla* [1.7.3], *hb*-3 is the most similar to *PaHbox1* in terms of sequence, expression pattern and genomic context. However, *hb*-3 and *PaHbox1* do not appear to be homologous genes although these two homeoboxes share 80% amino acid positional identity (71% nucleotide identity), *PaHbox1* shares 90% a.a. identity to the vertebrate *Hox* 2.C (2.3) and protostomate *Antp* homeoboxes with nucleotide positional identities of around 80%. It is unlikely that there should be a greater sequence difference between two sea urchin species than between sea urchins and species belonging to different phyla.

Nonetheless, *hb*-3 and *PaHbox1* both show high levels of restriction fragment polymorphism. The number of observations for *hb*-3 (Dolecki 1988a) are too few to allow any statistical analysis but the observed polymorphism stands in marked contrast to that of the other three sea urchin homeobox genes [1.7.3].

Secondly, *hb*-3 and *PaHbox1* are both expressed at low levels during embryogenesis (*PaHbox1* at mesenchyme blastula, *hb*-3 at gastrula) and at higher levels in adult gonads. *Hb*-3 shows low expression in adult intestines whereas *PaHbox1* is expressed at high levels in this tissue and is also detectable in Aristotle's lanterns.

It thus seems that although these two genes are not homologous, they are closely related and may represent an evolutionary relatively recent duplication event of an ancestral gene and its tissue and stage specific regulatory regions.

The early sea urchin embryo can be divided into five embryonic territories based on specific patterns of gene expression and the fate of the constituent cells. One of these territories is the vegetal plate. Cells of the vegetal plate remain as an indifferent single-cell-thick layer until the end of blastula stage, but then give rise to many different late embryonic cell types and internal larval structures which include mesenchymal derivatives, the gut and coeloms (Davidson 1989). One of the coeloms, the left posterior somatocoel, gives rise to 1) the genitial chord from which the gonads develop and 2) the 5 dental sacs from which part of the Aristotle's lantern arises (Hyman 1955).

Thus we can conclude that the vegetal plate gives rise to the adult tissues in which *PaHbox1* expression is observed. Furthermore, as Aristotle's lantern consists of several different tissues, many of which are derived from the oral ectoderm (Davidson 1989), the low levels of *PaHbox1* expression in this structure may be a reflection of localized tissue-specific expression.

The second point of interest concerns the advent of vegetal plate differentiation, namely the end of blastula stage. This coincides with the peak of *PaHbox1* expression: expression of *PaHbox1* is first detectable at hatching blastula stage (9 hr) and peaks at mesenchyme blastula stage (11,5 hr). At this stage the embryo consists of between 500 and 1000 cells and it is very likely that expression of *PaHbox1* is restricted to a subpopulation of cells as otherwise the observed expression levels (40 transcripts per embryo) are too low to bear any biological function.

It is tempting to speculate that *PaHbox1* is involved in the determination of the vegetal plate cells of which there are about 60 at mesenchyme blastula stage (Cameron 1987) or a subset of these cells. This would have to be verified by *in situ* hybridisation and will be an area for future research.

CHAPTER 7

METHODS AND MATERIALS

7.0 Index

7.1 Sea urchins

7.1.1. Embryos

P.angulosus were collected along the coast of the Cape Peninsula and induced to spawn by injection with 2 ml 0,55 M KCl. Gametes were collected by inverting males over plastic trays and females over trays fitted with a metal grid and partially filled with sea water. Sperm were stored undiluted on ice until use whereas eggs were filtered through gauze and washed by settling with 4 changes of 3MM paper (Whatman) filtered sea water. The suspension was made 4% with respect to packed sea urchin embryos and Pennicillin-G and Streptomycin sulfate (Sigma) were added to a final concentration of 100 and 50 mg/l respectively. 100 ml of the egg suspension was fertilized with 0,1 ml of 1:100 diluted sperm and only batches showing greater than 95% fertilization were used for further experiments. Embryos were grown at 20°C in conical flasks on an orbital shaker at 180 rpm.
Embryos were collected by centrifugation at 4°C (1000 g, 5 min) washed 4 times with 0,55 M KCl and if not used immediately, frozen rapidly in microcentrifuge tubes in liquid nitrogen and stored at -70°C.

7.1.2 Adult tissues

Adults were dissected by cutting the test along the greatest circumference line. Organs were washed in 0,5 M NaCl; 2 mM $NaHCO_3$; 10 mM KCL; 0,01% EDTA pH 8,2 and immediately immersed and then stored in liquid nitrogen.

7.2 Nucleic acid isolation

7.2.1 Genomic DNA

Sea urchin genomic DNA was isolated from *P.angulosus* sperm as described by Kedes (1975) or from sea urchin embryo nuclei by overnight proteinase K digestion, repeated phenol and chloroform:isoamylalcohol extractions followed by extensive dialysis against 10 mM Tris; 1 mM EDTA; pH 8,00 (TE).

7.2.2 Plasmid DNA

Plasmid DNA was isolated using the Triton X100 method followed by CsCl equilibrium centrifugation (Maniatis 1982). Rapid small scale isolations were done using the alkali lysis method (Ausubel 1987).

7.2.3 Lambda DNA

Lambda DNA was prepared from plate lysates and purified using DEAE-cellulose and CTAB as described by Manfioletti (1988).

7.2.4 Total RNA

1,5 ml to 2 ml of 0,55 M KCl-washed packed embryos were added to 10 ml of guanidinium isothiocyanate solution (4 M GITC, 5 mM Sodium Citrate, 0,5% sarcosyl (w/v), 100 mM β-mercaptoethanol, pH 7,00), homogenized with a Dounce and centrifuged at 15000 g for 10 minutes. CsCl was added to the supernatant (1 g per 2,5 ml) and the mixture applied to a CsCl cushion (5,7 M CsCl; 0,1 M EDTA pH 7,50) and centrifuged as described by Glisin (1974).
1,5 to 2 g adult testes, ovary or intestine was homogenized in 10 ml guanidinium isothiocyanate solution and treated as above. Aristotle's lantern from 3 to 4 individuals were chopped up in a coffee grinder under liquid nitrogen before being treated as above.
After centrifugation the RNA pellets were resuspended in 0,1% SDS; TE, pH 8,00 and twice ethanol precipitated. The RNA was checked for intactness and DNA contamination by treating an aliquot of each sample with RNAase A (1μg RNAase A / 1μg total RNA, 37°C for 15 min) and electrophoresing it with an untreated aliquot on an acrylamide gel [ ]. Staining

with acridine orange (30 µg/ml) was used to reveal the presence of any contaminating DNA. Samples that showed contamination or degradation (untreated lanes) were discarded. Intestinal RNA showing no degradation could not be obtained, in this case the least degraded sample was used.

7.2.5 Poly A^+ RNA

Poly A+ RNA was purified by one to two rounds of oligo dT-cellulose chromatography as described by Maniatis (1982), except that samples were passed three times over each column. Both poly A^+ and poly A^- RNA was ethanol precipitated twice. For long term storage (> 1 month) the RNA was stored as an ethanol precipitate, for short term it was kept frozen in RNAase free water.

7.2.6 RNA content per embryo

The total RNA content per 11,5 hr *P.angulosus* embryo was determined as follows: A 25 ml 4% embryo culture was grown [7.1.1] for 10,5 hr, when 25 - 50 µCi of [5-^{3}H]-uridine (Amersham 26 Ci/mmol) was added. At 11 hr this was chased with 100 nmol of non-radioactive uridine.
At 11,5 hr three 1 ml aliquots were removed, the embryos pelleted and resuspended in 700 µl of TE ph 7,50; 0,5% SDS and the nucleic acids and proteins pelleted by the addition of an equal amount of 10% ice-cold TCA. The precipitate was pelleted and resuspended in 5% TCA, then 100% ethanol. The pellet was subsequently heated to 95°C for 15 min in 0,5 ml 10% TCA which quantitatively resolubilizes nucleic acids (Schneider 1957). 5 ml of scintillation mixture was added and the tubes scintillation counted giving the amount of radiation incorporated into the RNA.
At 11,5 hr the remaining 22 ml of labled embryos were pelleted and resuspended in 3ml of guanidinium isothiocyanate solution (see section 7.2.4). RNA was isolated as described above [7.2.4] and 1.) scintillation counted in triplicate; 2.) quantitated by spectroscopy using a conversion factor of 37 µg RNA per 1 absorbance unit at 260 nm. In 3 separate RNA isolations the 260 nm /280 nm absorbance ratio lay consistently at 1,9. DNA contamination lay below the sensitivity (<5%) of a Hoechst H33258 agarose plate assay (Rieber 1990).

From the percentage recovery of RNA and spectroscopic quantitation the amount of RNA per ml of 4% culture was determined to be 197 ± 14 µg (3 separate RNA isolations of the parallel cultures, individual measurements were 196, 184 and 212 µg).
The number of embryos per ml of the above embryo cultures was determined by counting 20 µl aliquots on a gridded slide. This number was determined to be 30 800 ± 3060 embryos/ml (individual measurements per 20 µl: 598, 707, 614, 536, 620).
The amount of RNA per 11,5 hr embryo was thus calculated to be 6,4 ± 0,8 ng.

7.3 Sea urchin genomic library

7.3.1 Vector and Host

The vector used was lambda EMBL3 which was bought from Promega and had been digested with BamHI and EcoRI, followed by treatment with alkaline phosphatase. The background level (arms ligated and packaged without insert) using the 'Packagene' packaging system and the bacterial host *E.coli* LE392 [F-, *hsd*R514 (rk-, mk+), *sup*E44, *sup*F58, *lac*Y1, *gal*K2, *gal*T22, *met*B1, *trp*R55, lambda-] is 320 pfu/µg.

7.3.2 Ligation

The ratio of vector to insert used is discussed in detail in part 2 (see also table 2.1). Reaction volumes were dictated by the DNA concentrations needed to ensure optimal concatemerisation ligations (table 2.1).
The ligation buffer consisted of 40 mM Tris-HCl, pH7,50; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP; 50 µg/ml BSA. 1,25 units of T_4 DNA ligase (Amersham) per 5 µl reaction was used.
The reaction was performed at 12°C for 14 to 16 hr.

7.3.3 Packaging

The ligated products were packaged without prior ethanol precipitation because this would cause shearing of the long DNA molecules. A commercial packaging system, 'Packagene' (Promega) derived from a *E.coli* C extract described by Rosenberg (1985) was used as described by the manufacturers. After 2 hr incubation at 20 to 22°C, 500 µl of ice-cold SM buffer (100 mM NaCl, 8 mM $MgCl_2$, 50 mM Tris-HCl pH7,50,

0,01% gelatin) and 20 μl of chloroform was added. The packaged mixtures were stored at 4°C for a maximum of 18 hr before being amplified or screened.

7.3.4 Titration

To avoid titre drops of the freshly packaged lambda recombinants, titrations were performed immediately after packaging. The previous day a freshly plated LE392 colony was used to inoculate 10 ml of LB medium (Luria-Bertrani : 10 g Bacto-tryptone, 5 g Bacto yeast extract, 5 g NaCl) and grown overnight at 37°C. 1 ml of this culture was used to inoculate 50 ml of LB medium supplemented with 0,5 ml of 20% maltose and 1 M $MgSO_4$ and shaken at 37°C until the $O.D._{260nm}$ had reached 0,5 to 0,6. Cells were cooled at 4°C, centrifuged at 3000 g for 10 min at 4°C and gently resuspended in 1/3 volume of ice-cold 10 mM $MgSO_4$. 100 μl were mixed with 10 to 100 μl of appropriate dilutions of the packaged DNA and incubated for 30 min at 37°C for the phage to absorb to the bacteria. 2 to 3 ml of molten (45 - 46°C) NZY agarose (NZY consists of 5 g NaCl, 2 g $MgSO_4$, 5 g yeast extract, 10 g casein hydrolysate per litre; NZY-agarose is NZY to which 0.7% agarose has been added) were added and the mixture poured onto a NZY agar plate (NZY to which 15% Difco agar has been added). After the top agarose had dried, the plates were transferred to 37°C and plaques grown for 6 to 8 hr and counted.

7.3.5 Amplification

The packaged phage was plated at a concentration of approximately 50000 plaques per freshly poured 137 mm diameter NZY agar plate. 300 μl of host LE392 cells prepared as described above were used and the amount of top NZY agarose increased to 6 - 8 ml. Plates were incubated at 37°C until plaques were visible and touching but not confluent (≈ 5 hr). 10 ml of cold SM buffer was added and the plates shaken gently (100 rpm) at 4°C overnight. The bacteriophage suspension was pooled, chloroform added to 5% and after a 15 min incubation at room temperature the cell debris was removed by centrifugation at 4000 g for 5 min. The supernatant was stored in glass and microfuge tubes at 4°C in the presence of 0,3% chloroform. A 1 ml aliquot of each amplified library was stored at -70°C in the presence of 7%

(v/v) DMSO (no chloroform added). Each library was titrated as described above.

7.3.6 Screening

Bacteriophage were plated at a density of 10000 to 20000 pfu/ 137mm plate as described above except that 2 to 3 day old NZY plates were used. Phage were grown until plaques were easily visible but still well separated (≈6 hours). The plates were transferred for 30 min to 4°C to prevent the agarose from lifting during subsequent procedures. Plaques were transferred to labeled nylon filters for 2 min and to duplicate filters for 4 min. The filters were pricked at three or more asymmetric locations with a narrow gauge needle that had been dipped in waterproof filter sterilized ink for orientation. Filters were dried for at least 10 min. One stack (3 layers) of absorbent paper was soaked in denaturing (0,5 M NaOH, 1,5 M NaCl), two in neutralizing (0,5 M Tris-HCl pH7,5; 1,5 M NaCl) and one in rinsing solution (2 x SSC [7.7.4] , 0,2 M Tris-HCl pH7,5). The nylon filters were placed for 4 min on each of the stacks in the above order, then dried. The denatured DNA was crosslinked to the filters by UV irradiation (5 min). Filters were either directly hybridized or stored dessicated at 4°C.

7.4 Microbial techniques

7.4.1 Competent cells

After obtaining insufficiently many transformants using the calcium chloride procedure described by Maniatis et al (1982), the method proposed by Hanahan (1983) was tried. This did not give the expected transformation efficiency (colonies formed per μg of DNA) using XL-1-Blue cells which was determined to be caused by too low a cell viability. This most likely was caused by impurities in one or more of the numerous chemicals that are used in this elaborate protocol. In lieu of determining the factor responsible for the low transformation efficiencies in the Hanahan assay, a then novel simple method for the preparation of competent bacterial cells was tested (Chung 1988):

XL-1-Blue recA$^-$ (*rec*A1, *lac*$^-$, *end*A$^-$, *gyr*A96, *thi*, *hsd*R17, *sup*E44, *rel*A1, {F' *pro*AB, *lac*I^q, *lac*ZdelM15, Tn10}) (Stratagene) cells were streaked onto Luria Bertrani [LB,

see 7.3.1] plates containing 10 μg/ml tetracycline to select for the F' episome. A single colony was grown overnight in LB medium and 1 ml of the overnight culture used to inoculate 50 ml of LB medium. The bacteria were harvested by 10 min centrifugation at 1000 g, 4°C after reaching an $O.D._{600nm}$ of 0,35 to 0,45. The cells were resuspended in 4,2 ml of cold (4°C) transformation buffer (TSB): LB medium (pH 6,1) containing 10% polyethylene glycol 3350 (Sigma) is autoclaved, then distilled DMSO is added to a concentration of 5% (v/v) and filter sterilized 1 M stock solutions of $MgCl_2$ and $MgSO_4$ are added to a final concentration of 10 mM each. Cells were incubated on ice for 10 to 20 min and used for transformations. Competent cells were always freshly prepared.

7.4.2 Transformation

For transformation, 100 μl of competent cells were pipetted into cold microfuge tubes and 2 -4 μl of the ligation mixture (100 pg DNA) was added and the cells kept on ice for 40 min. To allow expression of the antibiotic resistance genes (ampicillin, $tetracyclin^r$), 900 μl of TSB containing 20 mM glucose was added and the cells incubated at 37°C, 225 rpm for 1 hr. 200 μl was plated onto AXI plates: LB plates containing 100 μg/ml ampicillin, 80 μg/ml X-gal (5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside, Sigma) and 20 mM IPTG (isopropyl β-D-thiogalactopyranoside, Sigma). This method consistently gave 1 x 10^6 colonies per μg input DNA.

7.5 Nucleic acid manipulation

7.5.1 DNA : Enzymatic manipulation

7.5.1.1 Restriction enzymes

Enzymes were obtained from Anglian, Boehringer Mannheim (NdeII) and Amersham (Kpn1). Digestions of long DNA (genomic, lambda) were done overnight at final DNA concentrations of between 0,1 and 0,5 μg/μl in the presence of 0,1 μg/μl BSA (nuclease-free, Boehringer Mannheim) at an enzyme concentration of 1 unit/μg DNA at the recommended temperature. Except for NdeII which was supplied with a buffer, the Boehringer Mannheim set of restriction enzyme buffers were employed as recommended. For plasmid DNA digestions, no BSA was added, digestion times were reduced

to two hours and DNA concentrations increased to 1 to 2 μg/μl.
For double digestions either a buffer compatible with both enzymes was chosen or the sample was digested with the enzyme requiring the lower salt concentration and then the salt concentration adjusted to the requirements of the second enzyme.

7.5.1.2 Klenow fragment

Klenow fragment (Anglian) was used to fill in 5′ overhangs generated by restriction enzymes to produce blunt ends required for certain ligations or to radioactively label the fragments.
A typical reaction contains 1 μg DNA in 20 μl restriction buffer containing 0,5 pmol of each deoxyribonucleotide. 1 μl Klenow (2,5 units, Amersham) is added and the reaction allowed to proceed for 15 min at 30°C. The reaction is stopped by the addition of 1 μl of 0,5 M EDTA.
When radioactively labeling restriction fragments, 1 μl of ^{32}P-α-dCTP (3000 Ci/mmol) was added instead of cold dCTP.
After fill-in reactions, the DNA was purified by passage over a spin column [7.5.3.2] and ethanol precipitation.

7.5.1.3 Nick translations

DNA probes: Antp homeobox probe - 600 bp BamHI-PvuII fragment from p903G (McGinnis 1984).The Pahbox1 homeobox-containing - 800 bp PstI-HindIII fragment from pHB1.
Fragments were purified from 3,5% nondenaturing polyacrylamide gels [7.6.2] and nick-translated as described below.

For nick translations the following mixture was prepared :

- 5 μl 10 x buffer (0,5 M Tris.HCl, pH8.00
  0,1 M $MgCl_2$
  10 mM DTT
  0,5 mg/ml BSA)
- water to a final volume of 50 μl
- 1 μl of each of 1 mM dATP, dGTP, dTTP
- 1 μl of DNA probe (50 - 100 ng)
- 5 μl of 10 μCi/μl ^{32}P-α-dCTP (3000 Ci/mmol; Amersham)
- 1 μl of 1:1000 dilution of 0,4 units/μl DNAase1 (Promega)
- 1 μl of 5 units/μl *E.coli* DNA polymerase 1 (Amersham).

The reaction was incubated at 15-16°C for 60 min and stopped by the addition of 2 µl of 0,5 M EDTA. Unincorporated nucleotides were removed using a spin column [7.5.3.2] and 1 µl of the flowthrough scintillation counted to determine the specific activity of the probe. Specific activities of 5 x 10^8cpm/µg were routinely achieved.

7.5.1.4 Ligations

The following mixture was prepared :

1 µl 10x buffer (200 mM Tris.HCl pH 7.60
50 mM $MgCl_2$
50 mM DTT)
1 µl 5 mM ATP
1 µl 0,5 mg/ml BSA
100 pg of vector (phosphatase treated)
at least equimolar amount of insert
water to final volume of 10 µl (including enzyme)
1 µl T_4 DNA ligase (2,5 Weiss units/µl; Amersham).

The mixture was incubated at 12°C overnight and used directly for transformations [7.4.2].
The exact amounts of vector and insert DNA used are described in the text at the relevant positions.

7.5.1.5 Dephosphorylations

Calf intestinal phosphatase (Boehringer Mannheim) was used as outlined by Ausubel et al 1987.

7.5.2 RNA Manipulations

7.5.2.1 RNAases

When working with RNA precautions were taken to avoid contamination of solutions and glassware with RNAases because these enzymes, once present are extremely difficult to destroy. All glassware was baked (4 hr 240°C), tips and microfuge tubes (freshly opened bags) packed with flamed tweezers and autoclaved for 30 min at 120°C. Large plastic containers (electrophoresis chambers etc.) were treated for 10 min with 10% peroxide.Water was collected directly from the distiller and autoclaved in baked bottles. Such 'RNAase-free' water was used to make up solutions. Nuclease-free chemicals or chemicals reserved for RNA work were used and the solutions autoclaved or, if heat labile, filtered (0,22

μm). For certain reactions human placental ribonuclease inhibitor (HPRI, Amersham) was used as an additional precaution.

7.5.2.2 Transcriptions

Bluescript contains T_3 and T_7 RNA polymerase promotors flanking the insertion site enabling the production of both both sense and antisense transcripts. The plasmid template, pHB1, was restricted downstream of the insert (far side of RNA polymerase promotor used) in the multiple cloning site. Proteinase K (Boehringer Mannheim) to a final concentration of 50 μg/ml was added and digestion allowed for 30 min at 37°C. The DNA was then twice phenol:chloroform (1:1) extracted, ethanol precipitated and redissolved at 1 μg/μl in 10 mM Tris.HCl pH7,4; 0,1 mM EDTA.

RNA probes were prepared as follows:

5 μl 5x buffer (200 mM Tris.HCl pH 8,00
250 mM NaCl
40 mM $MgCl_2$
10 mM spermidine)
1 μl each of 10 mM ATP, CTP, GTP
1 μl of 0,1 mM UTP
7 μl H_2O
1 μl 0,75 M DTT
1 μl 2,5 units/μl HPRI
1 μl DNA template (1 μg)
5 μl 10 μCi/μl ^{32}P-α-UTP (3000 Ci/mmol, Amersham)
1 μl (10 units) T_3or T_7 RNA polymerase (Promega)

This mixture was incubated at 37°C for 45 min after which 2 units of RNAase-free DNAasel (Promega) was added and incubation continued for another 15 min.
23 μl TE pH 7,50 and 2 μl 10 μg/μl tRNA
were added and the unincorporated nucleotides removed by passage over a spin column [7.5.3.2]. The eluant was phenol:chloroform (1:1) extracted and the RNA ethanol precipitated and redissolved in water.
For an 800 nucleotide long transcript (PaHbox1 and flanking regions ex. pHB1) the specific activity of the RNA was $2,2x10^8$ dpm/pmol or $7,8x10^8$ dpm/μg and the yield about 80 ng.

7.5.2.3 Glyoxal denaturation

For RNA dot blots and Northern blots, RNA samples were glyoxal denatured (McMaster 1977). Glyoxal (7 M) was deionized by repeated passage through a mixed bed resin (Unilab Amberlite MB-1) until the pH reached 6 and stored at -70°C in small aliquots.
Two glyoxal mixes were prepared: with or without DMSO. The mixture without DMSO was used when it was essential to keep the volume to a minimum. DMSO is not essential for denaturation (Thomas 1983).

DMSO glyoxal mix: 15 µl 7 M glyoxal
50 µl DMSO
2 µl 0,5 M phosphate pH 7,00
Two volumes of this mix were added to one volume of RNA solution (≤ 3,5 µg/µl).

Glyoxal mix : 15 µl 7 M glyoxal
9 µl H_2O
1 µl 0,5 M phosphate pH 7,00
One volume of this mix was added to three volumes of RNA solution (≤ 3,5 µg/µl).
The solutions were incubated for 1 hr at 50°C and kept on ice until used.

7.5.3 General nucleic acid manipulations

7.5.3.1 Ethanol precipitation

Nucleic acids were ethanol precipitated by adjusting the sodium acetate (pH 5,2) concentration to 0,3 M and adding 2,5 volumes (DNA) or 3 volumes (RNA) of absolute ethanol kept at -20°C. After 30 min at -20°C, samples were centrifuged at 16000 g for 15 min and the precipitates washed with 75% ethanol and recentrifuged. Ethanol was carefully aspired and the nucleic acids allowed to airdry at room temperature.
Dilute solutions of short fragments (≤200 bases) were incubated for 3 hr at -20°C and centrifuged at 22000 g for 30 min.

7.5.3.2 Spin columns

Spin columns were prepared and run as described in Maniatis (1982).

7.5.3.3 Concentration determinations

The concentration and purity of nucleic acid solutions was determined spectroscopically using the following conversions:
DNA : 50 μg has absorbance at 260 nm of 1 O.D.
RNA : 37 μg has absorbance at 260 nm of 1 O.D.
For DNA the absorbance ratio at 260 nm : 280 nm should be greater than 1,75, for RNA $\geq$ 2,0.
The concentration of dilute DNA solutions was determined using a fluorescence assay as described in Ausubel (1987): 3 μl aliquots of standard DNA solutions (0 - 20 μg/ml) and various dilutions of the unknown were mixed with 3 μl of a 1 μg/ml solution of ethidium bromide and spotted onto a household packaging polyethylene foil ('gladwrap') covered UV-transilluminator. The spots were photographed using a videocamera and the unknown concentration estimated by visual comparison.

7.5.3.4 Gel electrophoretic purification

Small DNA fragments ($\leq$ 1000 bp) were separated using nondenaturing acrylamide gel electrophoresis [7.6.2] and recovered by overnight elution in 0,5 M ammonium acetate, 1 mM EDTA, pH 8,00 as described by Ausubel et al 1987. Larger DNA fragments were separated on low melting agarose gels [,(Sigma Type XI)], the bands cut out and the DNA recovered using heat and urea to dissolve the agarose followed by passage over an ion-exchange matrix as described in the Quiagen manual (Diagen).

7.5.3.5 TCA precipitation

For rapid estimation of ^{32}P labeled TCA precipitable nucleic acids, an aliquot was spotted onto glassfiber circles (Whatman) and the Cerenkov-radiation counted. The glassfiber circles were then washed on a filter funnel with 2 ml 10% TCA and 10 ml 5% TCA and recounted.
For tritium labeled nucleic acids and more accurate quantitation, an aliquot was spotted onto glassfiber circles

and washed 10 min in 10 ml of icecold 10% TCA (in scintillation vials on an orbital shaker at 200 rpm) then twice 2 min each in 5% icecold 5% TCA, then absolute ethanol. The filters were dried and counted in 10 ml of scintillation fluid.

7.5.3.6 Sucrose gradient size fractionation

A 10 to 40% sucrose gradient was poured in a Beckman SW-28 tube, 300 µg of partially restricted genomic DNA layered on top and the tube centrifuged in a SW-28 rotor at 25000 rpm for 24,0 hr (20°C). The gradient was prepared using a gradient former and filter-sterilized stock solutions of 10 and 45% sucrose in 1,0 M NaCl, 20 mM Tris.HCl pH 8,00, 5 mM EDTA. The tube was filled to the 38 ml mark to give a 10 to 40% gradient.
After centrifugation the gradient was fractionated into seventy 540 µl fractions and every third fraction examined electrophoretically.

7.6 Electrophoresis

7.6.1 Agarose

Agarose gels were prepared as described in Maniatis, 1982. DNA samples were electrophoresed at 1 Volt/cm for 16 hr in 1x TBE buffer (0,089 M Tris-borate
0,089 M boric acid
2 mM EDTA pH 8,00)
with 0,5 µg/ml ethidium bromide.

RNA fragments larger than 1000 nucleotides were electrophoresed on 1 - 1,1% agarose gels in 10 mM phosphate buffer pH7,00 at 3 to 5 volts/cm (≤44 mAmp) for 4 to 7 hr. During this time the buffer at each chamber end was continually stirred (magnetic stirrers) and the buffer recirculated. Gels were 12,5 cm long and 3-4 mm thick. RNA samples were denatured with glyoxal [7.5.2.3] and loaded with bromophenol blue as marker. Gel equipment for RNA electrophoresis was treated 10 min with 10% peroxide to destroy RNAases and used solely for RNA work.

7.6.2 Non-denaturing acrylamide

Non denaturing 3,5% polyacrylamide (29:1 of acryl:bisacrylamide) gels were poured and run in 1x TBE (see above) as described by Ausubel et al 1987. Gels were 200 mm long, 1,5 mm thick.

7.6.3 Denaturing acrylamide RNA gels

Small RNA (<1000 nucleotides) were separated on 10% polyacrylamide, 7 M urea gels. Gels were prepared according to standard techniques using the following solution:

30 g acrylamide
1 g bisacrylamide
42 g urea
10 ml 10x TBE buffer [7.6.1]
water to final volume of 100 ml.

After degassing, 30 µl TEMED and 700 µl freshly prepared 10% ammonium persulphate was added and the gel poured.
The RNA sample was mixed with two volumes of sample buffer: 10 M urea, 0,1x TBE, 0,001% of bromophenol blue and xylene cyanol, heated to 90°C for 5 min and loaded onto the preelectrophoresed (hand hot) gel. Electrophoresis was done at 50 mAmp (600 Volt) at a temperature of 50 - 60°C.

7.6.4 Sequencing gels

Gel solutions consisted of 6% acrylamide (20:1 acryl:bisacrylamide), 8 M urea and 1x TBE. This solution was deionized with a mixed bed resin (5 g/100 ml of Amberlite MB-1, Unilab) for 2 hr before the addition of 0,5 ml 10% ammonium persulphate and 25 µl TEMED and subsequent pouring. Gels were 450 mm long and 0,4 mm thick.
The apparatus was 'home-built' with a waterbased cooling system. Samples were electrophoresed at 1000 to 1500 Volts, 55°C, for 3 or 6 hr. Gels were soaked in 10% ethanoic acid, 12% methanol for 5 min, then deionized water for a further 5 min before being dried at 80°C under vacuum.

7.7 Nucleic acid transfer and hybridization

7.7.1 Southern blots

After completion of gel electrophoresis, the ethidium bromide-stained gel was photographed on a short wavelength UV-transilluminator and transferred to a dish containing at least 4 gel volumes of denaturing solution (1,5 M NaCl, 0,5 M NaOH). It was gently shaken on an orbital shaker at 70 rpm for 1 hr, rinsed with distilled water and soaked twice for 30 min in neutralizing solution (1,5 M NaCl, 0,5 M Tris.HCl, pH 7,5).
The gel was subsequently transferred overnight onto a nylon membrane (Hybond-N: Amersham) using 10 x SSC as described in Maniatis, 1982.
The following morning, the well positions were marked on the membrane, the membrane air-dried and the DNA crosslinked by UV-irradiation (5 min under lamp used; determined empirically). Membranes if not used immediately, were stored dessicated at 4°C.

7.7.2 Northern blots

RNA samples were glyoxal denatured [7.5.2.3] before electrophoresis. After electrophoresis [7.6.1], the RNA was transferred onto nylon membranes (Hybond-N: Amersham) without prior treatment (Thomas 1983). The method was identical to the one used for DNA samples [7.7.1] except for the following points: 1. 20 x SSC buffer was used for transferral. 2. All equipment was made as RNAase free as possible. 3. The RNA was UV-crosslinked for only 60 sec (empirically determined; the time is shorter than for DNA possibly owing to the presence of glyoxal). After UV-crosslinking the glyoxal adduct was removed by soaking the membrane in 20 mM Tris.HCl, pH 8,0 (pH at room temperature) at 100°C for 10 min.
The intactness of RNA was determined by electrophoresis on separate gels or on duplicate lanes followed by acridine orange staining and photography [7.9.1]. This strategy avoided RNA degradation during acridine orange staining and destaining.

7.7.3 RNA Dot blots

RNA was glyoxal denatured without DMSO [7.5.2.3]. Samples were spotted onto nylon membranes pretreated with 10 x SSC, using a home built vacuum-driven slot- and dot-blotting apparatus similar to commercial units. After the samples had been sucked through (1-5 sec/μl) the membranes were dried, UV-irradiated and boiled as described above [7.7.2].

7.7.4 Hybridization

Membranes were placed into polyethylene plastic bags with 80 μl of prewarmed prehybridization fluid per cm^2 of membrane area. Air bubbles were removed and the bags heat-sealed. After 2 to 3 hr incubation with gentle shaking in a waterbath at the required temperature, the bag was cut open, the contents squeezed out and replaced with 50 $\mu l/cm^2$ of prewarmed hybridization fluid containing the freshly denatured radioactive probe. The bag was heat sealed, air bubbles driven to the top and sealed off, and the membrane incubated with gentle shaking for 24 to 48 hr at the required temperature.
Hybridization conditions were based on those reported by McGinnis (1984).

Hybridization solution:	Low stringency	High stringency
Deionized formamide	43 % v/v	50 % v/v
SSC *a*	5 x	5 x
Phosphate buffer. pH 6,80	50 mM	50 mM
Denhardt's solution [b]	5 x	5 x
Herring testes DNA [c]	250 μg/ml	250 μg/ml
SDS	0,1 %	0,1 %

a. 20 x SSC solution consists of 3,0 M NaCl, 0,3 M sodium citrate, pH 7,00.
b 100 x Denhardt's solution consists of 2,0 % w/v of each of Ficoll 400 (Sigma), polyvinylpyrrolidone-40 (Sigma) and bovine albumin fraction V (Seravac).
c. DNA (Sigma, type XIV) was sonicated to an average length of 400 bp and boiled for 10 min followed by rapid cooling, then stored at 4°C.

Low stringency hybridizations were performed at 37°C, high at 42°C. Prehybridization and hybridization solutions were identical except for the addition of probe to the latter.

Radioactively labeled probes were added at a final concentration of 2-10 ng/ml, 10^6 cpm/ml. They were denatured by heating to 100°C for 5 min followed by rapid cooling on ice.

After hybridization, membranes were rinsed twice for 5 min in 2 x SSC, 0,1 % SDS at room temperature, twice for 15 min in the same solution at 45-50°C for low stringency conditions and at 60-65 °C followed by an additional 20 min in 0,2 x SSC, 0,1 % SDS at 60-65°C for high stringency conditions. Membranes were rinsed in 2 x SSC, blotted dry (avoiding drying-out) and covered in 'gladwrap'. Autoradiography was done as described in [7.9.2].

7.8 RNAase titration and mapping

The methods outlined in Sambrook (1989) and Lee (1987) were used as guidelines.
A ^{32}P labeled 800 nucleotide transcript was made as described above to a specific activity of 7 x 10^5 dpm/ng. The amount of RNA was kept constant at 100 μg by the addition of yeast tRNA. To aid redissolving in the hybridization buffer, RNA samples were rehydrated in 30 μl of distilled water and evaporated to near dryness in a vacuum centrifuge. 50 fg of transcribed probe in 30 μl hybridization buffer (80% v/v formamide; 40 mM PIPES pH 6,40; 0,4 M NaCl; 1 mM EDTA pH 7,50) was added to the 100 μg of ethanol precipitated and rehydrated RNA.
Samples were heated 10 min to 85°C and incubated for 36 hours at 45-46°C. 300 μl of RNAase digestion mix was added (30 mM NaCl; 10 mM Tris pH 7,40; 5 mM EDTA pH 7,50; 210 units RNAase T1 and 12 μg RNAase A from Boehringer Mannheim) and the mixture digested for 60 min at 30°C.
For titrations 330 μl 10% ice-cold TCA was added and the precipitate collected on glass fiber filters, washed first with 5% TCA followed by ethanol, dried and counted [7.5.3.5].
For mapping, samples were digested with proteinase K (Boehringer Mannheim), phenol-chloroform (1:1 v/v) extracted and ethanol precipitated. Electrophoresis was done on 6% polyacrylamide, 7 M urea sequencing gels at 50^0C [7.6.4]. Gels were dried and autoradiographed. Bands were quantitated by densitometric scanning [7.9.3] of autoradiographs (preflashed as described above) exposed for varying lengths

of time so as to remain in the linear response range of the film.

7.9 Photographic techniques

7.9.1 Photography

Ethidium bromide and acridine orange stained gels were photographed on a short wavelength UV-transilluminator using Ilford Pan-F film and a red filter. Autoradiographs were captured on Kodak Technical Pan film using a light box and a combination of a UV-haze and a 80A (blue) filter.
Film was developed using Rodinal (Agfa) according to the manufacturers instruction.

7.9.2 Autoradiography

Membranes were exposed to Cronex-4 film (preflashed to an absorbance increase of 0,15 OD units) with intensifying screens (Dupont Plus) at -70^{o}C for up to 14 days. After reaching room temperature, the film was developed With Agfa G127c developer for 5 min (1:6 dilution), rinsed in 2,5 % ethanoic acid, fixed in a 1:5 dilution of Agfa graphic fixer G333c and washed for 10 min in running tap water.

7.9.3 Microdensitometer scanning

Negatives and autoradiographs were scanned using a home-built (K.Achleitner, H.Alk, J.Duncan; departmental workshop) microdensitometer and analyzed and plotted with software written by Dr.T.Sewell from this department.

CHAPTER 8

REFERENCES

Acampora,D., D'Esposito,M., Failla,A., Pannese,M., Migliaccio,E., Morelli,F., Stornaiuolo,A., Nigro,V., Simeone,A., Boncinelli,E. (1989) The human HOX gene family. Nucleic acids Res. 17, 10385-10402.

Affolter,M., Percival-Smith,A., Muller,M., Leupin,W., Gehring,W.J. (1990) DNA binding properties of the purified Antennapedia homeodomain. Proc.Natl.Acad.Sci.USA 87, 4093-4097.

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

Akam,M., Dawson,I., Tear,G. (1988) Homeotic genes and the control of segment diversity. Development 104, 123-133.

Akam.M. (1989b) HOX and HOM: Homologous gene clusters in insects and vertebrates. Cell 57, 347-349.

Angerer,L.M., Dolecki,G.J., Gagnon,M.L., Lum,R., Wang,G., Yang,Q., Humphreys,T., Angerer,R.C. (1989) Progressively restricted expression of a homeobox gene within the aboral ectoderm of developing sea urchin embryos. Genes & Development 3, 370-383.

Ausubel, F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A., Struhl,K. (1987) Current Protocols in Molecular Biology. John Wiley and Sons, New York.

Balling,R., Mutter,G., Gruss,P., Kessel,M. (1989) Craniofacial abnormalities induced by ectopic expression of the homeobox gene Hox-1.1 in transgenic mice. Cell 58, 337-347.

Bastian,H., Gruss,P. (1990) A murine even-skipped homologue, Evx 1, is expressed during early embryogenesis and neurogenesis in a biphasic manner. EMBO 9, 1839-1852.

Bateson,W. (1894) Materials for the study of variation. MacMillan and Co. London.

Beachy,P.A. (1990) A molecular view of the Ultrabithorax homeotic gene of Drosophila. TIG 6, 46-51.

Beachy,P.A., Krasnow,M.A., Gavis,E.R., Hogness,D.S. (1988) An Ultrabithorax protein binds sequences near its own and the Antennapedia P1 promotors. Cell 55, 1069-1081.

Beeman,R.W. (1987) A homeotic gene cluster in the red flour beetle. Nature 327, 247-249.

Beltz,G.A., Jacobs,K.A., Eickbush,T.H., Cherbas,P.T., Kafatos,F.T. (1983) Isolation of multigene families and determination of homologies by filter hybridization methods. Meth. Enzym. 100, 266-285.

Berg,O.G., von Hippel,P.H. (1988) Selection of DNA binding sites by regulatory proteins. TIBS 13, 207-211.

Berleth,T., Burri,M., Thoma,G., Bopp,T., Richstein,S., Frigerio,G., Noll,M., Nusslein-Volhard,C. (1988) The role of localization of bicoid mRNA in organizing the anterior pattern of the Drosophila embryo. EMBO 7, 1749-1756.

Bermingham,J.R., Scott,M.P. (1988) Developmentally regulated alternative splicing of transcripts from the Drosophila homeotic gene Antennapedia can produce four different proteins. EMBO 7, 3211-3222.

Bienz,M., Tremml,G. (1988) Domain of Ultrabithorax expression in Drosophila visceral mesoderm from autoregulation and exclusion. Nature 333, 576-578.

Biggin,M.D., Bickel,S., Benson,M., Pirrotta,V., Tjian,R. (1988b) Zeste encodes a sequence-specific transcription factor that activates the Ultrabithorax promotor in vitro. Cell 53, 713-722.

Biggin,M.D., Tjian,R. (1988a) Transcription factors that activate the Ultrabithorax promotor in developmentally staged extracts. Cell 53, 699-711.

Biggin,M.D., Tjian,R. (1989a) A purified Drosophila homeodomain protein represses transcription in vitro. Cell 58, 433-440.

Biggin,M.D., Tjian,R. (1989b) Transcription factors and the control of Drosophila development. TIG 5, 377-383.

Blatt,C., Aberdam,D., Schwartz,R., Sachs,L. (1988) DNA rearrangement of a homeobox gene in myeloid leukaemic cells. EMBO 7, 4283-4290.

Blochlinger,K., Bodmer,R., Jack,J., Jan,L.Y., Jan,Y.N. (1988) Primary structure and expression of a product from cut, a locus involved in specifying sensory organ identity in Drosophila. Nature 333, 629-635.

Bodner,M., Castrillo,J., Theill,L.E., Deerinck,D., Ellisman,M., Karin,M. (1988) The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. Cell 55, 505-518.

Boncinelli,E., Simeone,A., La Volpe,A., Faiella,A., Fidanza,V., Acampora,D., Scotto,L. (1985) Human cDNA clones containing homeobox sequences. Cold Spring Harb. Symp. Quant. Biol. 50, 301-306.

Bopp,D., Burri,M., Baumgartner,S., Frigerio,G., Noll,M. (1986) Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila. Cell 47, 1033-1040.

Botchan,M., McKenna,G., Sharp,P.A. (1973) Cleavage of mouse DNA by a restriction enzyme as a clue to the arrangement of genes. Cold Spring Har.Symp. 38, 383-395.

Brandt,W.F., Strickland,W.N., Strickland,M., Carlisle,L., Woods,D., von Holt,C. (1979) A histone programme during the life cycle of the sea urchin. Eur.J.Biochem. 94, 1-10.

Breier,G., Dressler,G.R., Gruss,P. (1988) Primary structure and developmental expression pattern of Hox-3.1, a member of the murine HOX 3 homeobox gene cluster. EMBO 7, 1329-1336.

Britten,R.J., Cetta,A., Davidson,E.H. (1978) The single-copy DNA sequence polymorphism of the sea urchin Strongylocentrotus purpuratus. Cell 15, 1175-1186.

Britten,R.J., Graham,D.E., Neufeld,B.R. (1974) Analysis of repeating DNA sequences by reassociation. Methods in Enzymology 29, 363.

Burglin,T.R., Finney,M., Coulson,A., Ruvkun,G. (1989) Caenorhabditis elegans has scores of homeobox-containing genes. Nature 341, 239-243.

Burri,M., Tromvoukis,Y., Bopp,D., Frigerio,G., Noll,M. (1989) Conservation of the paired domain in metazoans and its structure in three isolated human genes. EMBO 8, 1183-1190.

Cameron,R.A., Hough-Evans,B.R., Britten,R.J., Davidson,E.H. (1987) Lineage and fate of each blastomere of the eight-cell sea urchin embryo. Genes Dev. 1, 75-84.

Carroll,S.B., Scott,M.P. (1985) Localization of the fushi tarazu protein during Drosophila embryogenesis. Cell 43, 47-57.

Carroll,S.B., Scott,M.P. (1986) Zygotically active genes that affect the spatial expression of the ftz segmentation gene during early Drosophila embryogenesis. Cell 45, 113-126.

Casanova,J., Sanchez-Herrero,E., Morata,G. (1986) Identification and characterisation of a parasegment specific regulatory element of the Abdominal-B gene of Drosophila. Cell 47, 627-636.

Casey,J., Davidson,N. (1977) Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucl.Acids Res. 4, 1539-1552.

Chavrier,P., Vesque,C., Galliot,B., Vigneron,M., Dolle,P., Duboule,D., Charnay,P.(1990) The segment-specific gene Krox -20 encodes a transcription factor with binding sites in the promoter region of the Hox-1.4 gene. EMBO 9, 1209-1218.

Cho,K.W.Y., Goetz,J., Wright,C.V.E., Fritz,A., Hardwicke,J., DeRobertis,E.M. (1988) Differential utilisation of the same reading frame in a Xenopus homeobox gene encodes two related proteins sharing the same DNA-binding specificity. EMBO 7, 2139-2149.

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

Clarke,L., Carbon,J (1976) A colony bank containing synthetic ColE1 hybrids representative of the entire E.coli genome. Cell 9, 91-99.

Clerc,R.G., Corcoran,L.M., LeBowitz,J.H., Baltimore,D., Sharp,P.A. (1988) The B-cell specific Oct-2 protein contains POU box and homeobox-type domains. Genes Devel. 2, 1570-1581.

Cohen,S.M. (1990) Specification of limb development in the Drosophila embryo by positional cues from segmentation genes. Nature 343, 173-177.

Cohen,S.M., Brönner,G., Küttner,F., Jürgens,G., Jäckle,H. (1989) Distal-less encodes a homeodomain protein required for limb development in Drosophila. Nature 338, 432-434.

Cohen,S.M., Jurgens,G. (1990) Mediation of Drosophila head development by gap-like segmentation genes. Nature 346, 482-485.

Colberg-Poley,A.M., Voss,S.D., Chowdhury,K., Gruss,P. (1985) Structural analysis of murine genes containing homeobox sequences and their expression in embryonal carcinoma cells. Nature 314, 713-718.

Costa,M., Weir,M., Coulson,A., Sulston,J., Kenyon,C. (1988) Posterior pattern formation in C.elegans involves position-specific expression of a gene containing a homeobox. Cell 55, 747-756.

Coulson,A., Waterston,R., Kiff,J., Sulston,J., Kohara,Y. (1988) Genome linking with yeast artificial chromosomes. Nature 335, 184-186.

Crick,F.H.C., Brenner,S., Klug,A., Pieczenik,G. (1976) A speculation on the origin of protein synthesis. Origins Life 7, 389-397.

Dalton,D., Chadwick,R., McGinnis,W. (1989) Expression and embryonic function of empty spiracles: A Drosophila homeo box gene with two patterning functions on the anterior-posterior axis of the embryo. Devel. 3, 1940-1956.

Davidson,E.D. (1989) Lineage-specific gene expression and the regulative capacities of the sea urchin embryo: a proposed mechanism. Development 105, 421-445.

Dearolf,C.R., Topol,J., Parker,C.S. (1989) The caudal gene product is a direct activator of ftz transcription during Drosophila embryogenesis. Nature 341, 340-342.

Dearolf,C.R., Topol,J., Parker,C.S. (1990) Transcriptional regulation of the Drosophila segmentation gene fushi tarazu. BioEssays 12, 109-113.

DeRobertis,E.M., Oliver,G., Wright,C.V.E. (1989) Determination of axial polarity in the vertebrate embryo: Homeodomain proteins and homeogenetic induction. Cell 57, 189-191.

Deschamps,J., DeLaaf,R., Joosen,L., Meijlink,F., Destree,O. (1987) Abundant expression of homeobox genes in mouse embryonal carcinoma cells correlates with chemically induced differentiation. Proc.Natl.Acad.Sci.USA 84, 1304-1308.

Desplan,C., Theis,J., O'Farrell,P.H. (1985) The Drosophila developmental gene, engrailed ,encodes a sequence-specific DNA binding activity. Nature 318, 630-635.

Desplan,C., Theis,J., O'Farrell,P.H. (1988) The sequence specificity of homeodomain-DNA interaction. Cell 54, 1081-1090.

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

DiNardo,S., Kuner,J.M., Theis,J., O'Farrell,P.H. (1985) Development of embryonic pattern in D.melanogaster as revealed by accumulation of the nuclear engrailed protein. Cell 43, 59-69.

DiNardo,S., Sher,E., Heemskerk-Jongens,J., Kassis,J.A., O'Farrell,P.H. (1988) Two-tiered regulation of spatially patterned engrailed gene expression during Drosophila embryogenesis. Nature 332, 604-609.

Doe,C.Q., Hiromi,Y., Gehring,W.J., Goodman,C.S. (1988a) Expression and function of the segmentation gene fushi tarazu during Drosophila neurogenesis. Science 239, 170-175.

Doe,C.Q., Smouse,D., Goodman,C.S. (1988) Control of neuronal fate by the Drosophila segmentation gene even-skipped. Nature 333, 376-378.

Dolecki,G.J., Humphreys,T. (1988b) An engrailed class homeobox gene in sea urchins. Gene 64, 21-31.

Dolecki,G.J., Wang,G., Humphreys,T. (1988a) Stage- and tissue- specific expression of two homeobox genes in sea urchin embryos and adults. Nucleic Acids Res. 16, 1543-11558.

Dolecki,G.J., Wannakrairoj,S., Lum,R., Wang,G., Riley,H.D., Carlos,R., Wang,A., Humphreys,T. (1986) Stage-specific expression of a homeobox-containing gene in the non-segmented sea urchin embryo. EMBO 5, 925-930.

Dolle,P., Duboule,D. (1989a) Two members of the murine HOX-5 complex show regional and cell-type specific expression in developing limbs and gonads. EMBO 8, 1507-1515.

Dolle,P., Izpisua-Belmonte,J., Falkenstein,H., Renucci,A., Duboule,D. (1989b) Coordinate expression of the murine Hox-5 complex homeobox-containing genes during limb pattern formation. Nature 342, 767-772.

Doyle,H.J., Harding,K., Hoey,T., Levine,M. (1986) Transcripts encoded by a homeobox gene are restricted to dorsal tissues of Drosophil embryos. Nature 323, 76-79.

Dressler,G.R., Deutsch,U., Balling,R., Simon,D., Guenet,J., Gruss,P. (1988) Murine genes with homology to Drosophila segmentation genes. Development 104, 181-186.

Dressler,G.R., Gruss,P. (1988) Do multigene families regulate vertebrate development? TIG 4, 214-218.

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

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

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

Driever,W., Thoma,G., Nusslein-Volhard,C. (1989b) 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.

Duboule,D., Dolle,P. (1989) The structural and functional organization of the murine HOX gene family resembles that of Drosophila genes. EMBO 8, 1497-1505.

Dugaiczyk,A., Boyer,H.W., Goodman,H.M. (1975) Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. J.Mol.Biol. 96, 171.

Duprey,P., Chowdhury,K., Dressler,G.R., Balling,R., Simon,D., Guenet,J.L., Gruss,P. (1988) A mouse gene homologous to the Drosophila gene caudal is expressed in epithelial cells from the embryonic intestine. Genes Dev. 2, 1647-1654.

Edgar,B.A., Weir,M.P., Schubiger,G., Kornberg,T. (1986) Repression and turnover pattern fushi tarazu RNA in the early Drosophila embryo. Cell 47, 747-754.

Eichele,G. (1989) Retinoids and vertebrate limb pattern formation. TIG 5, 246-251.

Emmons,S. (1988) The nematode Caenorhabditis elegans. Editor Wood,W.B. Cold Spring Harbor Laboratory, New York.

Fasman,G.D. (1976) CRC Handbook of biochemistry and molecular biology: Nucleic acids, vol.II. CRC Press, Cleveland, Ohio.

Fienberg,A.A., Utsef,M.F., Bogarad,L.D., Hart,C.P., Awgulewitsch,A., Ferguson-Smith,A., Fainsod,A., Rabin,M., Ruddle,F.H. (1987) Homeobox genes in murine development. Current Topics in Dev.Biol. 23, 233-256.

Finkelstein,R., Perrimon,N. (1990) The orthodenticle gene is regulated by bicoid and torso and specifies Drosophila head development. Nature 346, 485-488.

Finney,M., Ruvkun,G., Horvitz,H.R. (1988) The C.elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. Cell 55, 757-770.

Fitzpatrick,V.D., Ingles,C.J. (1989) The Drosophila ftz polypeptide is a DNA-binding transcriptional activator in yeast cells. Nature 337, 666-668.

Frain,M., Swart,G., Monaci,P., Nicosia,A., Stampfli,S., Frank,R., Cortese,R. (1989) The liver-specific transcription factor LFB1 contains a highly diverged homeobox DNA binding domain. Cell 59, 145-157.

Fregien,N., Dolecki,G.J., Mandel,M., Humphreys,T. (1983) Molecular cloning of five individual stage- and tissue-specific mRNA sequences from sea urchin pluteus embryos. Mol.Cell.Biol. 3, 1021-1031.

Freyd,G., Kim,S.K., Horvitz,H.R. (1990) Novel cysteine-rich motif and homeodomain in the Caenorhabditis elegans cell lineage gene lin-II. Nature 344, 876-879.

Frigerio,G., Burri,M., Bopp,D., Baumgartner,S., 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.

Fritz.A., DeRobertis,E.M. (1988) Xenopus homeobox-containing cDNAs expressed in early development. Nucl.Acids Res. 16, 1453-1469.

Frohnhofer,H.G., Nusslein-Volhard,C. (1986) Organisation of anterior pattern in the Drosophila embryo by the maternal gene bicoid.

Galau,G.A.; Britten,R.J., Davidson,E.H. (1974) A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. Cell 2, 9-20.

Galau,G.A.; Britten,R.J., Davidson,E.H. (1976) Studies on nucleic acis reassociation kinetics: Rate of hybridization of excess RNA with DNA, compared to the rate of DNA renaturation. Proc.Natl.Acad.Sci.U.S.A. 74, 1020-1023.

Garcia-Blanco,M.A., Clerc,R.G., Sharp,P.A. (1989) The DNA-binding homeo domain of the Oct-2 protein. Genes Devel. 3, 739-745.

Gaunt,S.J., Miller,J.R., Powell,D.J., Duboule,D. (1986) Homeobox gene expression in mouse embryos varies with position by the primitive streak stage. Nature 324, 662-668.

Gaunt,S.J., Sharpe,P.T., Duboule,D. (1988) Spatially restricted domains of homeo-gene transcripts in mouse embryos : relation to a segmented body plan. Development 104, 167-179.]

Gaunt,S.J., Singh,P.B. (1990) Homeogene expression patterns and chromosomal imprinting. TIG 6, 208-212.

Gay,N.J., Poole,S., Kornberg,T. (1988) Association of the Drosophila melanogaster engrailed protein with specific soluble nuclear protein complexes. EMBO 7, 4291-4297.

Gehring,W.J. (1987) Homeoboxes in the study of development. Science 236, 1245-1252.

Gergen,J.P. (1986) Drosophila segmentation genes and blastoderm cell identities. BioEssays 6, 61-65.

Gergen,J.P., Wieschaus,E. (1986) Dosage requirements for runt in the segmentation of Drosophila embryos. Cell45, 289-299.

Giudice,G. (1973) Developmental biology of the sea urchin embryo. Academic Press, New York.

Glisin,V., Crkvenjakov,R., Byus,C. (1974) Rna isolated by cesium chloride centrifugation. Biochemistry 13, 2633-2637.

Gonzales-Reyes,A., Morata,G. (1990) The developmental effect of overexpressing a Ubx product in Drosophila embryos is dependent on its interaction with other homeotic products. Cell 61, 515-522.

Goto,T., Macdonald,P., Maniatis,T. (1989) Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell 57, 413-422.

Goustin,A.S., Wilt,F.H. (1981) Protein synthesis, polyribosomes, and peptide elongation in early development of Strongylocentrotus purpuratus. Dev. Biol. 82, 32-40.

Graham,A., Papalopulu,N., Krumlauf,R. (1989) The murine and Drosophila homeobox gene complexes have common features of organisation and expresssion. Cell 57, 367-378.

Gurdon,J.B. (1988) The origin of cell-type differences in early embryos. Reg.Mech.Develop.Proc., 1-6.

Hafen,E., Kuroiwa,A., Gehring,W.J. (1984a) Spatial distribution of transcripts from the segmentation gene fushi tarazu during Drosophila embryonic development. Cell 37, 833-841.

Hafen,E., Levine,M., Gehring,W.J. (1984b) Regulation of Antp transcript distribution by the Bithorax complex of Drosophila. Nature 307, 287-289.

Han,K., Levine,M.S., Manley,J.L. (1989) Synergistic activation and repression of transcription by Drosophila homeobox proteins. Cell 56, 573-583.

Hanahan,D. (1983) Studies on transformation of E.coli with plasmids. J.Mol.Biol. 166, 557-580.

Hanes,S.D., Brent,R. (1989) DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell 57, 1275-1283.

Harding,K., Hoey,T., Warrior,R., Levine,W. (1989) Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. EMBO 8, 1205-1212.

Harding,K., Levine,M. (1988) Gap genes define the limits of Antennapedia and Bithorax gene expression during early development in Drosophila. EMBO 7, 205-214.

Harrison,S.C., Aggarwal,A.K. (1990) DNA recognition by proteins with the helix-turn-helix motif. Ann.Rev.Biochem. 59, 933-969.

Harrison,S.D., Travers,A.A. (1988) Identification of the binding sites for potential regulatory proteins in the upstream enhancer element of the Drosophila fushi tarazu gene. Nucl.Acids Res. 16, 11403-11416.

Harvey,R.P., Melton,D.A. (1988) Microinjection of synthetic Xhox-1A homeobox mRNA disrupts somite formation in developing Xenopus embryos. Cell 53, 687-697.

Herr,W., Sturm,R.A., Clerc,R.G., Corcoran,L.M., Baltimore,D., Sharp,P.A., Ingraham,A.H., Rosenfeld,M.G., Finney,M.,Ruvkun,G., Horvitz,H.R. (1988) The POU domain: A large conserved region in the mammalian pit -1, oct -1,2 and C.elegans unc-86 gene products. Genes Dev. 2, 1513-1526.

Hill,R.E., Jones,P.F., Rees,A.R., Sime,C.M., Justice,M.J., Copeland,N.G., Jenkins,N.A., Graham,E., Davidson,D.R. (1989) A new family of mouse homeobox containing genes: molecular structure, chromosomal location, and developmental expression of Hox-7.1. Genes.Devel. 3, 26-37.

Hiromi,Y., Gehring,W.J. (1987) Regulation and function of the Drosophila segmentation gene fushi tarazu. Cell 50, 963-974.

Hoey,T., Levine,M. (1988) Divergent homeobox proteins recognize similar DNA sequences in Drosophila. Nature 332, 858-861.

Holland,P.W.H., Hogan,B.L.M. (1986) Phylogenetic distribution of Antp-like homeoboxes. Nature 321, 251-253.

Hörstadius,S. (1939) The mechanics of sea urchin development, studied by operative methods. Biol.Rev. 14, 132.

Hyman,L.H.(1955) The invertebrates: Echinodermata. Vol.IV, 413-588 McGraw-Hill, New York.

Ingham,P.W. (1988b) The molecular genetics of embryonic pattern formation in Drosophila. Nature 335, 25-34.

Ingham,P.W. (1990) The X,Y,Z of head development. Nature 346, 412-413.

Ingham,P.W., Baker,N.E., Martinez-Arias,A. (1988a) Regulation of segment polarity genes in the Drosophila blastoderm by ftz and eve. Nature 331, 73-75.

Ingraham,H.A., Chen,R., Mangalam,H.J., Eisholtz,H.P., Flynn,S.E., Lin,C.R., Simmons,D.M., Swanson,L., Rosenfeld,M.G. (1988) A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell 55, 519-529.

Irish,V.F., Martinez-Arias,A., Akam,M. (1989) Spatial regulation of the Antennapedia and Ultrabithorax homeotic genes during Drosophila early development. EMBO 8, 1527-1537.

Ish-Horowicz,D., Pinchin,S.M., Ingham,P.W., Gyurkovics,H.G. (1989) Autocatalytic ftz activation and metameric instability induced by ectopic ftz expression. Cell 57, 232-232.

Jack,T., McGinnis,W. (1990) Establishment of the Deformed expression stripe requires the combinatorial action of coordinate, gap and pair-rule proteins. EMBO 9, 1187-1198.

Jackle,H., Tautz,D., Schuh,R., Seifert,E., Lehmann,R. (1986) Cross-regulatory interactions among the gap genes of Drosophila. Nature 324, 668-670.

Jaynes,J.B., O'Farrell,P.H. (1988) Activation and repression of transcription by homeodomain-containing proteins that bind a common site. Nature 336, 744-749.

Johnson,R.A., Walseth,T.F.(1979) The enzymatic preparation of [α-^{32}P]ATP, [α-^{32}P]GTP, [^{32}P]cAMP and [^{32}P]cGMP, and their use in the assay of adenylate and guanylate cyclases and cyclic nucleotide phosphodiesterases. Adv. Cyclic Nucleotide Res. 19, 135-167.

Johnson,W.A., Hirsh,J. (1990) Binding of a Drosophila POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. Nature 343, 467-470.

Joyner,A.L., Skarnes,W.C., Rossant,J. (1989) Production of a mutation in mouse en-2 gene by nonhomologous recombination in embryological stem cells. Nature 338, 153-156.

Kamb,A., Weir,M., Rudy,B., Varmus,H., Kenyon,C. (1989) Identification of genes from pattern formation, tyrosine kinase, and potassium channel families by DNA amplification. Proc.Natl.Acad.Sci.USA 86, 4372-4376.

Kamps,M.P., Murre,C., Sun,X.H., Baltimore,D. (1990) A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein. Cell, in press (ex. EMBL Data Bank).

Kappen,C., Schughart,K., Ruddle,F.H. (1989) Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. Proc.Natl.Acad.Sci.USA 86, 5459-5463.

Karlsson,O., Thor,S., Norberg,T., Ohlsson,H., Edlund,T. (1990) Insulin gene enhancer binding protein isl-1 is a member of a novel class of proteins containing both a homeo- and a Cis-His- domain. Nature 344, 879-882.

Kedes,L.H., Cohn,R.H., Lowry,J.C., Chang,A.C.Y., Cohen,S.N (1975) The organization of sea urchin histone genes. Cell 6, 359-369.

Kessel,M., Gruss,P. (1990) Murine developmental control genes. Science 249, 374-379.

Kessel.M., Schulze,F., Fibi,M., Gruss,P. (1987) Primary structure and nuclear localisation of a murine homeodomain protein. Proc,Natl.Acad.Sci.USA 84, 5306-5310.

Kim,Y., Nirenberg,M. (1989) Drosophila NK-homeobox genes. Proc.Natl.Acad.Sci.USA 86, 7716-7720.

Ko,H., Fast,P., McBride,W., Staudt,L.M. (1988) A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeobox domain. Cell 55, 135-144.

Kongsuwan,K., Allen,J., Adams,J.M. (1989) Expression of Hox-2.4 homeobox gene directed by proviral insertion in a myeloid leukaemia. Nucleic Acids.Res. 17, 1881-1891.

Kongsuwan,K., Webb,E., Housiaux,P., Adams,J.M. (1988) Expression of multiple homeobox genes within diverse mammalian haemopoietic lineages. EMBO 7, 2131-2138.

Krasnow,M.A., Saffman.E.E., Kornfeld,K., Hogness,D.S. (1989) Transcriptional activation and repression by Ultrabithorax proteins in cultured Drosophila cells. Cell 57, 1031-1043.

Krause,H.M., Gehring,W.J. (1989) Stage-specific phosphorylation of the ftz protein during Drosophila development. EMBO 8, 1197-1204.

Kuziora,M.A., McGinnis,W. (1988) Autoregulation of a Drosophila homeotic selector gene. Cell 55, 477-485.

Kuziora,M.A., McGinnis,W. (1989) A homeodomain substitution changes the regulatory specificity of the Deformed protein in Drosophila embryos. Cell 59, 563-571.

Kuziora.M.A., McGinnis,W. (1988) Different transcripts of the Drosophila Abd-B gene correlate with distinct genetic subfunctions. EMBO 7, 3233-3244.

LaRosa,G.J., Gudas,L.J. (1988) Early retinoic acid-induced F9 teratocarcinoma stem cell gene ERA-1: Alternative splicing creates transcripts for a homeobox-containing protein and one lacking the homeobox. Mol.Cell.Biol. 8, 3906-3917.

Laskey,R.A. (1980) The use of intensifying screens or organic scintillators for visualising radioactive molecules resolved by gel electrophoresis. Methods.Enzymol. 65, 363.

Laughon,A., Carroll,S.B., Storfer,F.A., Riley,P.D., Scott,M.P. (1985) Common properties of proteins encoded by the Antennapedia complex genes of Drosophila melanogaster. Cold Spring Harb.Symp.Quant.Biol. 50, 253-262.

Laughon,A., Howell,W., Scott,M.P. (1988) The interaction of proteins encoded by Drosophila homeotic and segmentation genes with specific DNA sequences. Development 104, 75-83.

Laughon,A., Scott,M.P. (1984) Sequence of a Drosophila segmentation gene: Protein structure homology with DNA binding proteins. Nature 310, 25-30.

Lawrence,P.A. (1987) Pair-rule genes: do they paint stripes or draw lines? Cell 51, 879-880.

Lawrence,P.A. (1988) Background to bicoid. Cell 54, 1-2.

Lawrence,P.A. (1990) Compartments in vertebrates? Nature 344, 382-383.

Lee,J.J., Costlow,N.A. (1987) A molecular titration assay to measure transcript prevalence levels. Methods in Enzymology 152, 633-648.

LeMotte,P.K., Kuroiwa,K., Fessler,L.I. Gehring,W.J. (1989) The homeotic gene sex combs reduced of Drosophila: gene structure and embryonic expression. EMBO 8, 219-227.

Levine,M., Harding,K., Wedeen,C., Doyle,H., Radomska,H. (1985) Cold Spring Harb.Symp.Quant.Biol. 50, 209-222.

Levine,M., Manley,J.L. (1989) Transcriptional repression of eukaryotic promoters. Cell 59, 405-408.

Lewis,E.B. (1978) A gene complex controlling segmentation in Drosophila. Nature 276, 565-570.

Lewis,J. (1989) Genes and segmentation. Nature 341, 382-383.

Lewis.J., Martin,P. (1989) Limbs: a pattern emerges. Nature 342, 734-735.

Li,W., Luo,C., Wu,C. (1985) Evolution of DNA sequences in 'Molecular evolutionary genetics' 1-94. Edited by MacIntyre,R.J.; Plenum Press; New York.

Licht,J.D., Grossel,M.J., Figge,J., Hansen.U.M. (1990) Drosophila Kruppel protein is a transcriptional repressor. Nature 346, 76-79.

Lillie,J.W., Green,M.R. (1989) Activator's target in sight. Nature 341, 279-280.

Lumsden,A., Keynes,R. (1988) Segmental patterns of neuronal development in the chick hindbrain. Nature 337, 424-428.

Macdonald,P.M., Ingham,P., Struhl,G. (1986) Isolation, structure ands expression of even-skipped : A second pair-rule gene of Drosophila containinga homeobox. Cell 47, 721-734.

Manfioletti,G., Schneider,C. (1988) A new and fast method for preparing high quality lambda DNA suitable for sequencing. Nucl.Acids Res. 16, 2873-2884.

Maniatis,T., Fritsch,E.F., Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbour University Press, Cold Spring Harbour.

Mavilio,F., Simeone,A., Giampolo,A., Faiella,A., Poiana,G., Russo,G., Peschle,C., Boncinelli,E. (1986) Differential and stage-related expression in embryonic tissue of a new human homeobox gene. Nature 324, 664-668.

McGinnis,W., Garber,P.L., Wirz,J., Kuroiwa,A., Gehring,W.J. (1984) A homologous protein coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell 37, 403-408.

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

McMaster,G.K., Carmichael,G.G. (1977) Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc.Natl.Acad.Sci.USA 74, 4835-4838.

Mihara,H., Kaiser,E.T. (1988) A chemically synthesized Antennapedia homeodomain binds to a specific DNA sequence. Science 242, 925-927.

Mitchell,P.J., Tjian,R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245, 371-378.

Mlodzik,M., Fjose,A., Gehring,W.J. (1988) Molecular structure and spatial expression of a homeobox gene from the labial region of the ANT-C. EMBO 7, 2569-2578.

Mount,S.M. (1982) A catalogue of splice junction sequences. Nucl.Acids Res. 10, 459-472.

Muller,M., Affolter,M., Leupin,W., Otting,G., Wuthrich,K., Gehring,W.J. (1988) Isolation and sequence specific binding of the Antennapedia homeodomain. EMBO 7, 4299-4304.

Muller-Immergluck,M.M., Schaffner,W., Matthias,P. (1990) Transcription factor Oct-2A contains functionally redundant activation domains and works selectively from a promoter but not from a remote enhancer position in non-lymphoid (HeLa) cells. EMBO 9, 1625-1634.

Murphy,P., Davidson,D.R., Hill,R.E. (1989) Segment-specific expression of a homeobox-containing gene in the mouse hindbrain. Nature 341, 156-159.

Nei,M., Li,W. (1979) Mathematical model for studying genetic variation in terms of restriction nucleases. Proc.Natl.Acad.Sci.USA 76, 5269-5273.

Nicosia,A., Monaci,P., Tomei,L., DeFrancesco,R., Nuzzo,M., Stunnenberg,H., Cortese,R. (1990) A myosin-like dimerization helix and an extra-large homeodomain are essential elements of the tripartite DNA binding structure of LFB1. Cell 61, 1225-1236.

Nourse,J., Mellentin,J.D., Galili,N., Wilkinson,S., Stanbridge,E., Smith,S.D., Cleary,M. (1990) Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell 60, 535-545.

Ohkuma,Y., Horikoshi,M., Roeder,R.G., Desplan,C. (1990a) Engrailed, a homeodomain protein, can repress in vitro transcription by competition with the TATA box-binding protein transcription factor IID. Proc.Natl.Acad.Sci.USA 87, 2289-2293.

Ohkuma,Y., Horikoshi,M., Roeder,R.G., Desplan,C. (1990b) Binding site-dependent direct activation and repression of in vitro transcription by Drosophila homeodomain proteins. Cell 61, 475-484.

Ohlendorf,D.H., Anderson,W.F., Matthews,B.M. (1983) Many gene-regulatory proteins appear to have a similar α-helical fold that binds DNA and evolved from a common precursor. J.Mol.Evol. 109-114.

Oliver,G., Wright,C.V.E., Hardwicke,J., DeRobertis,E.M. (1988a) A gradient of homeodomain protein indeveloping forelimbs of Xenopus and mouse embryos. Cell 55, 1017-1024.

Oliver,G., Wright,C.V.E., Hardwicke,J., DeRobertis,E.M. (1988b) Differential antero-posterior expression of two proteins encoded by a homeobox gene in Xenopus and mouse embryos. EMBO 7, 3199-3209.

Otting,G., Qian,Y., Muller,M., Affolter,M., Gehring,W. Wuthrich,K. (1988) Secondary structure determination for the Antennapedia homeodomain by NMR and evidence for a helix-turn-helix motif. EMBO 7, 4305-4309.

Peifer,M., Karch,F., Bender,W. (1987) The BX-C: Control of segment identity. Genes Dev. 1, 891-898.

Pfeffer,P.L. (1985) Honnours project. University of Capetown.

Posakony,J.W., Flytzanis,C.N., Britten,R.J., Davidson,E.H. (1983) Interspersed sequence organisation and developmental representation of cloned poly(A) RNAs from sea urchin eggs. J.Mol.Biol. 167, 361-389.

Ptashne,M. (1988) How eukaryotic transcriptional activators work. Nature 335, 683-689.

Qian,Y.Q., Billeter,M., Otting,G., Muller,M., Gehring,W.J. Wuthrich,K. (1989) The structure of the Antp homeodomain determined by NMR spectroscopy in solution : Comparison with prokaryotic precursors. Cell 59, 573-580.

Rebagliati,M. (1989) An RNA recognition motif in the bicoid protein. Cell 58, 231-232.

Reeves,R. (1984) Transcriptionally active chromatin. BBA 782, 343-393.

Regulski,M., McGinnis,N., Chadwick,R., McGinnis,W. (1987) Develomental and molecular analysis of Deformed; a homeotic gene controlling Drosophila head development. EMBO 6, 767-777.

Renkawitz,R. (1990) Transcriptional repression in eukaryotes. TIG 6, 192-196.

Rieber,M.S., Rieber,M. (1990) A Hoechst H33258 agarose plate assay for the estimation of nanogram DNA levels without RNA interference: applications in PCR and in estimations of plasmid and cytoplasmic DNA. Nucl.Acids Res. 18, 1918.

Robert,B., Sassoon,D., Jacq,B., Gehring,W., Buckingham,M. (1989) Hox-7, a mouse homeobox gene with a novel pattern of expression during embryogenesis. EMBO 8, 91-100.

Rosa,F.M. (1989) Mix.1, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of Xenopus embryos. Cell 57, 965-974.

Rosenberg,S.M., Stahl,M.M., Kobayashi,I., Stahl.F.W. (1985) Improved in vitro packaging of coliphage lambda DNA: A one-strain system from endogenous phage. Gene 38, 165-175.

Rosenberg,U.B., Schroder,C., Preiss,A., Kienlin,A., Cote,S., Riede,I., Jackle,H. (1986) Structural homology of the product of the Drosophila Kruppel gene with Xenopus transcription factor IIIA. Nature 319, 336-339.

Rosner,M.H., Vigano,M.A., Ozato,K., Timmons,P.M., Poirier,F., Rigby,P.W.J., Staudt,L.M. (1990) A POU domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature 345, 686-692.

Ruiz i Altaba,A., Melton,D.A. (1989a) Interaction between peptide growth factors and homeobox genes in the establishment of anterior-posterior polarity in frog embryos. Nature 341, 33-38.

Ruiz i Altaba,A., Melton,D.A. (1989b) Involvement of the Xenopus homeobox gene Xhox3 in pattern formation along the anterior-posterior axis. Cell 57, 317-326.

Rushlow,C., Frasch,M., Doyle,H., Levine,M. (1987) Maternal regulation of zerknullt: a homeobox gene controlling differentiation of dorsal tissues in Drosophila. Nature 330, 583-586.

Saint,R., Kalionis,B., Lockett,T.J., Elizur,A. (1988) Pattern formation in the developing eye of Drosophila melanogaster is regulated by the homeobox gene, rough. Nature 334, 151-154.

Sambrook,J., Fritsch,E.F., Maniatis,T. (1989 Molecular Cloning: A Laboratory Manual, 2nd Edition. Cold Spring Harbor University Press, Cold Spring Harbor.

Samson,M., Jackson-Grusby,L., Brent,R. (1989) Gene activation and DNA binding by Drosophila Ubx and abd-A proteins. Cell 57, 1045-1052.

Sanger,F., Nicklen,S., Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc.Natl.Acad.Sci.USA. 74, 5463-5467.

Savard,P., Gates,P.B., Brockes,J.P. (1988) Position dependent expression of a homeobox gene transcript in relation to amphibian limb regeneration. EMBO 7, 4275-4282.

Sawadogo,M., Sentenac,A. (1990) RNA polymerase B (II) and general transcription factors. Annual Review Biochem. 59, 711-754.

Schaffner,W. (1989) How do different transcription factors binding the same DNA sequence sort out their jobs? TIG 5, 37-39.

Schaller,D., Wittmann,C., Spicher,A., Müller,F., Tobler,H. (1990) Cloning and analysis of three new homeobox genes from the nematode C.elegans. Nucleic Acids Res. 18, 2033-2036.

Schneider,W.C. (1957) Determination of nucleic acids in tissues by pentose analysis. Methods in Enzymology 3, 680-684.

Schneuwly,S., Klemenz,R., Gehring,W.J. (1987) Redesigning the body plan of Drosophila by ectopic expression of the homeotic gene Antennapedia. Nature 325, 816-818.

Scholer,H.R., Ruppert,S., Suzuki,N., Chowdhury,K., Gruss,P. (1990) New type of POU domain on germ line-specific protein Oct-4. Nature 344, 435-438.

Schughart,K., Kappen,C., Ruddle,F.H. (1989) Duplication of large genomic regions during the evolution of vertebrate homeobox genes. Proc.Natl.Acad.Sci.USA 86, 7067-7071.

Scott,M.P., Tamkun,J.W., Hartzell 111,G.W. (1989) The structure and function of the homeodomain. BBA 989, 25-48.

Scott,M.P., Weiner,A.J., (1984) Structural relationships among genes that control development: Sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc.Natl.Acad.Sci.USA 81, 4115-4119.

Seed,B., Parker,R.C., Davidson,N. (1982) Representation of DNA sequences in recombinant libraries prepared by restriction enzyme partial digestion. Gene 19, 201-209.

Shen,W., Largman,C., Lowney,P., Corral,J.C., Detmer,K., Hauser,C.A., Simonitch,T.A., Hack,F.M., Lawrence,H.J. (1989) Lineage-restricted expression of homeobox-containing genes in human haematopoietic cell lines. Proc.Natl.Acad.Sci.USA 86, 8536-8540.

Shepherd,J.C.W., McGinnis,W., Carrasco,A.E., DeRobertis,E.M., Gehring,W.J. (1984) Fly and frog homeodomains show homologies with yeast mating type regulatory proteins. Nature 310, 70-74.

Shott,R.J., Lee,J.J., Britten,R.J., Davidson,E.H. (1984) Differential expression of the actin gene family of Strongylocentrotus purpuratus. Developmental Biol. 101, 295-306.

Simeone,A., Acampora,D., Arcioni,L., Andrews,P.W., Boncinelli,E., Mavilio,F. (1990) Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. Nature 346, 763-766.

Simeone,A., Pannese,M., Acampora,D., D'Esposito,M., Boncinelli,E. (1988) At least three human homeoboxes on chromosome 12 belong to the same transcription unit. Nucleic Acids Res. 16, 5379-5389.

Stern,S., Tanaka,M., Herr,W. (1989) The Oct-1 homeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature 341, 624-630.

Struhl,G. (1982) Genes controlling segmental specification in the Drosophila thorax. Proc.Natl.Acad.Sci.USA 79, 7380-7384.

Struhl,G. (1989b) Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. TIBS 14, 137-140.

Struhl,G., Struhl,K., Macdonald,P.M. (1989a) The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell 57, 1259-1273.

Stryer,L. (1981) Biochemistry. 2nd edition. W.H.Freeman and Co., San Francisco.

Stubblefield,E. (1986) A theory for developmental control by a program encoded in the genome. J.Theor.Biol. 118, 129-143.

Sturm,R.A., Das,G., Herr,W. (1988) The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeobox subdomain. Genes Devel. 2, 1582-1599.

Sullivan,W. (1987) Independence of ftz expression with respect to cellular density in Drosophila embryos. Nature 327, 164-167.

Tabin,C.J. (1988) Isolation of potential vertebrate limb-identity genes. Development 105, 813-820.

Tabor,S., Richardson,C.C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767-4771.

Tanaka,M., Herr,W. (1990) Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. Cell 60, 375-386.

Tautz,D., Lehman,R., Schnurch,H., Schuh,R., Seifert,E., Kienlin,A., Jones,K., Jackle,H. (1987) Finger protein of novel structure encoded by hunchback, a second member of the gap class of Drosophila segmentation genes. Nature 327, 383-389.

Thomas,P.S. (1983) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods in Enzymology 100, 255-266.

Treisman,J., Gonczy,P., Vashishtha,M., Harris,E., Desplan,C. (1989) A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell 59, 553-562.

Tremml,G., Bienz,M. (1989a) Homeotic gene expression in the visceral mesoderm of Drosophila embryos. EMBO 8, 2677-2685.

Tremml,G., Bienz,M. (1989b) An essential role of even-skipped for homeotic gene expression in the Drosophila visceral mesoderm. EMBO 8, 2687-2693.

Trifonov,E.N., Brendel,V. (1986) Gnomic: A Dictionary of Genetic Codes, VCH Publishers, Weinheim.

Upholt,W.B. (1977) Estimation of DNA sequence divergence from comparison of restriction nuclease digests. Nucleic Acids Res. 4, 1257-1265.

USBC - United States Biochemical Corporation (1988) Sequenase : Step-by-step protocols for DNA sequencing with Sequenase. 4th Edition.

Vincent,J., Kassis,J.A., O'Farrell,P.H. (1990) A synthetic homeodomain binding site acts as a celltype specific, promoter specific enhancer in Drosophila embryos. EMBO 9, 2573-2578.

Way,J.C., Chalfie,M. (1988) Mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C.elegans. Cell 54, 5-16.

Wilkinson,D.G., Bhatt,S., Chavrier,P., Bravo,R., Charnay,P. (1989a) Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. Nature 337, 461-464.

Wilkinson,D.G., Bhatt,S., Cook,M., Boncinelli,E., Krumlauf,R. (1989b) Segmental expression of HOX-2 homeobox-containing genes in the developing mouse hindbrain. Nature 341, 405-409.

Winslow,G.M., Hayashi,S., Krasnow,M., Hogness,D.S., Scott,M.P. (1989) Transcriptional activation by the Antennapedia and fushi tarazu proteins in cultured Drosophila cells. Cell 57, 1017-1030.

Wolgemuth,D.J., Behringer,R.R., Mostoller,M.P., Brinster,R.L., Palmiter,R.D. (1989) Transgenic mice overexpressing the mouse homeobox-containing gene Hox-1.4 exhibit abnormal gut development. Nature 337, 464-467.

Wolpert,L. (1969) Positional information and the spatial pattern of cellular differentiation. J.Theor.Biol. 25, 1-47.

Wright,C.V.E., Cho,K.W.Y., Hardwicke,J., Collins,R.H., DeRobertis,E.M. (1989) Interference with function of a homeobox gene in Xenopus embryos produces malformations of the anterior spinal chord. Cell 59, 81-93.

Young,B.D., Anderson,M.L.M. (1985) Quantitative analysis of solution hybridisation. Nucleic acid hybridisation, a practical approach. Hames,B.D. Higgins.S.J. eds., IRL Press, Oxford, p47-71.

Zimmer,A., Gruss,P. (1989) Production of chimaeric mice containing embryonic stem cells carrying a homeobox Hox-1.1 allele mutated by homologous recombination. Nature 338, 150-153.