



This thesis has been digitised as part of a Library project to commemorate Warwick's 50th anniversary. The project closely connects with [Warwick 50's](#) key themes of celebrating our international research, individual achievement, connecting alumni and leaving a lasting legacy.

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/67748/>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

"The thing can be done" said the butcher "I think:
The thing must be done I am sure
The thing will be done. Bring me paper and ink
The best there is time to procure."

The Beaver brought paper, portfolios, pens
And ink in unfailing supplies
While strange creepy creatures came out of their dens
And looked on with wondering eyes.

So engrossed was the Butcher he heeded them not
As he wrote with a pen in each hand,
Explaining the while in a popular style
Which the Beaver could well understand...

The Structure of cloned Histone genes of *Xenopus borealis*

A thesis submitted for the degree of PhD at the University of
Warwick by

William Bains BA

Developmental Biology Group
Department of Biological Sciences
University of Warwick
October 1982

Table of contents

Contents	ii
Table of Figures	ix
Abbreviations	xiii
Acknowledgements	xv
Declaration	xv
Summary	1
Introduction	
1. Histones	3
1.1 Histone proteins	3
1.2 Nucleosome structure and chromatin	6
1.3 Histone variants	9
1.4 Histone modification	11
1.5 Histones and the cell cycle	13
1.6 What should we expect from histone genes?	16
2. Genes	18
2.1 The eukaryotic genome	18
2.2 Copy number of histone genes	18
2.3 Structure of cloned histone genes	21
2.3.1 The sea urchin pattern	21
2.3.2 The chick non-pattern	24
2.3.3 Conserved features of cloned histone genes	25
2.3.4 The chick model for vertebrate histone genes	27
2.3.5 The Amphibian semi-pattern	28
2.3.6 Cluster variation and conservation	33
2.3.7 Cloned vertebrate genes: summary	37
2.4 Chromosomal location of histone genes	38
2.5 Histone genes as a multigene family	39
2.5.1 Gene clustering	40

2.5.2 Intra-cluster variation and inter-cluster conservation	42
2.6 Histone gene clusters: summary	44
2.7 Repetitive DNA and genome plasticity	45
2.7.1 C-value and highly repeated DNA	45
2.7.2 Dispersed repetitive elements	47
2.7.3 Function of intermediate repetitive DNA	50
2.7.4 RATS and the fluid genome	54
2.7.5 Repeated DNA and histone genes	62
3. <i>Xenopus borealis</i>	64
4. Aims and approaches of this project	68
4.1 Original aims	68
4.2 Approaches	69
4.2.1 Molecular cloning	69
4.2.2 Characterisation of a cloned DNA	71
4.2.3 Sequence analysis and other explorations	72
4.2.4 Functional dissection	73
4.3 Subsequent aims	73
Materials and Methods	
5. Materials and Methods	76
5.1 Bacterial, plasmid and phage strains	76
5.2 Microbiological culture media and conditions	79
5.3 DNA preparation procedures	82
5.4 DNA labelling procedures	87
5.5 Other techniques for recombinant manipulation	88
5.6 Electrophoretic gel methods	91
5.7 DNA-DNA hybridisation	94
5.8 Preparation of DNA from agarose gels	96
5.9 DNA sequence determination	96
5.10 Standard buffers	96
Results	

6. Construction of a gene library from <i>X.borealis</i>	100
6.1 Introduction	100
6.2 Methods	101
6.3 Results	102
6.3.1 Construction of a genomic gene library	102
6.3.2 Amplification of the library	105
6.3.3 Demonstration that the library is recombinant	106
6.3.4 Size of the inserted DNAs	108
6.4 Conclusion	
7. Selection of histone H4 gene-containing recombinants	110
7.1 Introduction	110
7.2 Methods	111
7.3 Results	111
7.3.1 Screening the gene library	111
7.3.2 Only one type of clone, Xbhl, is a true +ve	112
7.3.3 The Xbhl to 5 clones are all identical	114
7.3.4 Abundance of Xbhl in the gene library	114
7.3.5 Re-cloning the Xbhl insert into pBR325	115
7.4 Conclusion	116
8. Restriction map of Xbhl	118
8.1 Introduction	118
8.2 Methods	120
8.3 Results	120
8.3.1 XbaI, KpnI, SstI	120
8.3.2 Other six-cutters	122
8.3.3 4-cutters	124
8.4 conclusions	125
9. Gene mapping	127
9.1 Introduction	127
9.2 Methods	130
9.3 Results	133

9.3.1 What genes are on Xbhl?	133
9.3.2 What is their orientation?	134
9.3.3 What are the locations of the genes?	135
9.4 Conclusion	138
10. Repetitive DNA in Xbhl	140
10.1 Introduction	140
10.2 Methods	142
10.3 Results	143
10.3.1 Locating repetitive DNA in Xbhl	143
10.3.2 Copy number of the repetitive elements	143
10.3.3 Sequence studies on IR2	147
10.3.4 Phylogeny of IR2	149
10.3.5 Isolating lambda-XIR clones	150
10.4 Conclusion - the repetitive DNA families of Xbhl	153
11. Genomic distribution of Xbhl sequences	156
11.1 Introduction	156
11.2 Methods	156
11.3 Results	157
11.3.1 Genomic distribution of histone genes	157
11.3.2 Genomic distribution of IR2	160
11.3.3 Extra-chromosomal distribution of IR2	160
11.4 Conclusion	161
12. In vivo transcription of IR2	162
12.1 Introduction	162
12.2 Methods	162
12.3 Results	163
12.3.1 Hybridisation of Xbhl to cDNA probes	163
12.3.1 Hybrid release translation from IR2	163
12.4 Conclusions	164
13. 'In his Image...': Xbhl and Xlh8	166
13.1 Introduction	166

13.2 Methods	166
13.3 Results	167
13.3.1 Xlh8 and Xbhl are very similar	
13.3.2 Xbhl is very similar to but different from Xlh2 and Xlh202	168
13.4 Conclusions	168
14. Functional studies on Xbhl	171
14.1 Introduction	171
14.2 Methods	171
14.3 Results of microinjection experiments	172
14.4 Conclusions	174
15. DNA sequence studies on the coding regions of Xbhl	178
15.1 Partial sequence of the H1 gene of Xbhl	178
15.2 Partial sequence of the H3 gene of Xbhl	179
15.3 Analysis of the H3 genes from several species	179
15.3.1 Coincident codons	179
15.3.2 Codon usage and sequence conservation in H3	180
15.3.3 CpG and sequence conservation in H3 genes	183
16. Heteroduplex analysis of Xbhl and Xlh8	187
16.1 Introduction	187
16.2 Methods	188
16.3 Results	191
16.3.1 Heteroduplex analysis of Xlh8 and Xbhl	191
16.3.1 'Snapback structures' in Xbhl	192
16.4 Physical structures in Xbhl	194
Discussion	
17. Discussion	197
17.1 Structure of Xbhl	197
17.1.1 H3 and H4 genes	197
17.1.2 The H1 gene	199
17.1.3 The H2A and H2B genes	201

17.1.4 Snapback elements in Xbh1	204
17.1.5 Intermediate repetitive sequences in Xbh1	205
17.2 Xbh1 and other histone gene clusters	208
17.2.1 Xbh1 and the X.borealis major gene cluster	208
17.2.2 The X.laevis 'major cluster'	210
17.2.3 The circular histone gene map of X.laevis	211
17.2.4 Xlh8 and Xbh1 - chance or conservation?	216
17.2.5 Eclectic evolution and horizontal gene transfer	217
17.3 The 50kb gene	223
17.3.1 Multigene families and multigene clusters	223
17.3.2 How many genes in a eukaryote	228
17.3.3 The size of a 'gene'	231
17.3.4 The 50kb gene and repetitive DNA	235
17.4 The 50kb histone gene	236
Appendices	
18. Appendix 1. Cosmid cloning: an attempt and its failure	240
18.1 Introduction	240
18.2 Methods	241
18.3 Results	242
18.3.1 Initial cloning attempt	242
18.3.2 A successful E.coli gene library	243
18.3.3 Non-recombinants are concatomer derivatives	244
18.3.4 Failure to clone is due to X.borealis DNA	245
18.4 Conclusions and a speculation	246
19. Appendix 2. Critical variables in lambda-gtWES cloning	248
19.1 Introduction	248
19.2 Method	248
19.3 Results	248
19.3.1 The approach	248
19.3.2 The model results	249
20. Appendix 3. Genomic representation of IR1 and IR2	251

20.1 Introduction	251
20.2 Results	252
21. Appendix 4. Sequence differences determined by restriction	254
21.1 Introduction	254
21.2 Band differences caused by a new site	254
21.3 Site differences caused by base differences	255
References	
References	258

The Bellman, perceiving their spirits were low
 Recited in musical tone
 Some Jokes he had kept for a season of woe
 But the crew would do nothing but groan.

Quotations are from "The hunting of the Snark" by Lewis Carroll

Figures

<u>Figure</u>	<u>Following Page</u>
1.1 The nucleosome	6
2.1 Histone gene numbers	19
2.2 Invertebrate histone genes	21
2.3 Higher vertebrate histone genes	24
2.5 Yeast histone genes	25
2.6 Consensus gene structures	26
2.7 Amphibian histone genes	29
2.8 <i>Xenopus</i> major histone gene clusters	32
2.9 H2A and H2B genes	34
3.1 <i>Xenopus laevis</i> and <i>borealis</i>	64
3.2 Phylogeny of selected higher animals	66
6.1 Lambda gtWES cloning strategy	101
6.2 Isolation of lambda 'arms'	102
6.3 EcoRI digestion and ligation	103
6.4 Library phage preparations	107
6.5 XIR Phage preparations	108
6.6 Recombinant phage insert sizes	108
7.1 First screen for histone genes	111
7.2 Recombinants isolated in first screen	113
7.3 Third screen for histone genes	113
7.4 Fourth screen for histone genes	113
7.5 Restrictions of Xbhl to 5	114
7.6 pBR325 subcloning strategy	116
7.7 Recloning Xbhl into pBR325	117
8.1 Restriction mapping by partial digestion	120
8.2 6-cutters : 1	121
8.3 6-cutters : 2	123

8.4	6-cutters : 3	123
8.5	Initial six - cutter map	123
8.6	6-cutters : 4	124
8.7	4-cutters : 1	124
8.8	4-cutters : 2	125
8.9	4-cutter data from Dr. P.C.Turner	125
8.10	Restriction map of Xbhl	125
8.11	Restriction sites near H3	126
9.1	Sea urchin histone gene clones	130
9.2	HRT from Xbhl	133
9.3	HRT from separate strands of Xbhl	134
9.4	HRT from regions of Xbhl	135
9.5	Gene mapping blots : 1	136
9.6	Gene mapping blots : 2	137
9.7	The gene map of Xbhl	138
9.8	Xbhl and the 'major cluster'	139
10.1	Repetitive DNA in Xbhl	143
10.2	Genomic representation of IR2	144
10.3	Structure of IR2	146
10.4	DNA sequences from IR2	147
10.5	Analysis of IR2 sequence	148
10.6	Phylogeny of IR2	149
10.7	First screen for XIR clones	150
10.8	XIR clones	151
10.9	Rescreen of XIR clones	152
10.10	Homologies between XIR and histones	152
11.1	RPC-5 column eluant	157
11.2	Genomic DNA probed with Xbhl	157
11.3	Histone genes in RPC-5 fractions	158
11.4	IR2 in genomic DNA	159
11.5	Genomic distribution of IR2	160

11.6 IR2 on circular DNA	161
12.1 Expression of IR2	163
13.1 Restriction of Xbhl and Xlh8	167
13.2 Differences between Xbhl and Xlh8	167
13.3 The Xlh8 like clones	168
13.4 Sequence differences between clones	168
14.1 Microinjecting oocytes	172
14.2 Microinjection of Xbhl clones	173
14.3 H2A expression in Xbhl and Xlh8	174
15.1 H1 gene sequences	178
15.2 H3 gene sequences	179
15.3 Coincident codons in H3 genes	180
15.4 Codon usage in 10 H3 genes	181
15.5 Dinucleotide frequencies	182
16.1 Heteroduplex of pWB1 and pRW8a	191
16.2 Snapbacks in pWB1 : 1	192
16.3 Snapbacks : 2	193
16.4 Snapbacks : 3	193
16.5 Analysis of snapbacks	194
17.1 Xbhl	197
17.2 IR2	205
17.3 Xbhl and the major cluster	208
17.4 Formation of Xbhl	209
17.5 The X.laevis 'circular map'	212
18.1 Cosmid 'recombinants'	243
18.2 Cosmid recombinants	243
18.3 Analysis of cosmid methodology	244
18.4 Plasmid instability	244
18.5 Cloning several DNAs in cosmids	245
18.6 Cloning several DNAs in lambda-gtWES	245
19.1 Output from cloning model	249

20.1 Data to measure IR2's representation	252
20.2 Data for genomic representation	253
20.3 Hybrid melting curves	253

Abbreviations used in this thesis

Ap	Ampicillin
ATP	Adenosine triphosphate
BAP	Bacterial Alkaline Phosphatase
Cm	Chloramphenicol
DMSO	Dimethylsulphoxide
EDTA	Sodium Ethylenediaminetetracetate
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HPFH	Hereditary persistence of fetal haemoglobin
HPRT	Hypoxanthine phosphoribosyltransferase
HRT	Hybrid release translation
IR	Intermediate repetitive (of DNA)
kb	kilobases (or kilobase pairs of double stranded DNA)
Kn	Kanamycin
Mb	Megabase (or Mega-base pairs)
mg	milligrams
mRNA	Messenger RNA
rRNA	Ribosomal RNA
rDNA	DNA coding ribosomal RNA
tRNA	transfer RNA
My	Megayear
ng	nanogram
NTP, dNTP	: Nucleoside triphosphate , deoxynucleoside triphosphate
PEG	Polyethyleneglycol
PPO	2,5-diphenyloxazole
RATS	Retroviral and transposable sequences
RPC-5	Reverse phase chromatography medium 5
RPM	revolutions per minute
SSC	Standard saline citrate (see section 5.10)
Tc	Tetracycline
TE	Tris/EDTA

TEMED N,N,N',N'-Tetramethyl-ethylenediamine
Tris Tris(hydroxymethyl)aminomethane, hydrochloride salt buffer
µg microgram
© circa or approximately equals
3MM 3MM chromatography paper

Acknowledgements

I am pleased to record my thanks to Dr. Robert Old, my supervisor , for his unstinting help, guidance and encouragement during this work. I also thank Dr. P.C.Turner and Professor H.R.Woodland for frequent help and suggestions, Dr Lesley Coggins and Mrs Alison Slater for help with heteroduplex and electron microscopic techniques, Ms Elizabeth Balantine, Mr. Tom Aldridge and Drs Ian Jones, Andy Cummings, John Adair , John Norton, Alan Colman and John Clewley for donations of materials and guidance on their use as recorded, Professor Larry Kedes for making available the pSp clones and giving me a job, and many others for their helpful criticisms. I am also indebted to Deena Blundell, Neil Coveney , Rachel Countryman, Chris Mayers and especially to Alister Wilson for help producing this manuscript, and to Jan for keeping me on the rails for three years.

Declaration

The work described in sections 14 (Microinjection studies), 15 (histone gene sequencing) and 16 (Heteroduplex and snapback analysis) were performed in collaboration with Dr. P.C.Turner, Professor Hugh Woodland and Dr. Lesley Coggins respectively. Mr. Tom Aldridge helped to confirm some of the results presented in section 13, and Dr. P. C. Turner contributed some results to section 8. Apart from this, all the results described in this thesis are entirely my own.

WBains

Summary

I have constructed a genomic gene library of EcoRI fragments from Xenopus borealis DNA cloned in lambda-gtWES. The library contains 425,000 phage with a mean insert size of 6.1kb. I have screened this library for histone genes and for a family of repetitive DNA sequences.

I have isolated one histone gene clone from the library (Xbhl). It is overabundant in the library, so no other histone gene clones are isolatable. The cloned DNA fragment is 8.55kb long and contains one copy of each histone gene in the order H4-H2A-H2B-H1-H3. H2A is transcribed in the opposite direction to the other four genes. There is a short repetitive element between H2A and H2B present in 5000 copies in the genome, and an inverted pair of larger elements 3' to H2B present in 10000 copies. I have isolated clones containing homologues to the latter and characterised them and homologous genomic sequences: it is a member of a dispersed, diverse family of elements with some sequence homology to histone mRNA and present in anurans and urodeles.

Xbhl is a minor species in the genome, but is homologous to the major histone gene cluster in the H4-H2A-H2B gene region. Its overall sequence is only 1-2% divergent from three histone gene clones isolated from X.laevis, which shared an ancestor with X.borealis 8 million years ago.

The H4, H2A, H1 and H3 genes of Xbhl direct the production of normal proteins when Xbhl is microinjected into oocytes. The H3 gene codes a protein 96% homologous to Calf H3, and contains codons conserved in all H3 genes sequenced. The H1 gene codes for a protein identical in 16 out of 21 central amino acids to an H1 from X.laevis and homologous to calf and salmon H1 proteins. It is possible that the H2B gene is a pseudogene.

Introduction

1. Histones

Come listen my men, while I tell you again
Of the five, unmistakable marks
By which you might know, wheresoever you go,
The warrented, genuine Snarks.

1.1 Histone proteins.

"Histone" is the name for a class of basic proteins found in all eukaryotes. As well as being literally central to the eukaryotic cell - they make up 30% of the dry weight of the nucleus - their role in nuclear structure makes them of special interest to molecular geneticists.

There are five classes of histone protein (Isenberg 1979), the four 'nucleosomal core' proteins H2A, H2B, H3 and H4, and the fifth 'linker' protein H1. They were initially distinguished solely on chemical and electrophoretic grounds, but it is now clear that these chemical differences reflect different roles in the structure of chromatin (Kossel 1928, Darlington 1942, Bustin and Cole 1968, Kornberg 1977, Isenberg 1979). The proteins which play the central role in the formation of chromatin must be of interest to the study of gene control. Recent advances in two areas have suggested that they may be more than just the physical support of the genetic material. These are the discovery of the 'nucleoskeleton' and its relation to gene action, and the re-appraisal of the relation of histone variants and histone modification to differentiation and differential gene activity. As the latter is important to the theme of this thesis - the genes for these proteins - I will start by describing briefly what the

1. Histones

histones are, and how they are believed to interact to form chromatin. In particular the intra-species variation of protein sequence within histone classes, and its possible role in the large-scale structure of chromatin, will be of importance to this work.

The outline structure of the nucleosome has been known for some time. The four 'core' histones interact strongly in low salt conditions to form oligomers of the formula $(H2A.H2B.H3.H4)_n$ (Elgin and Weintraub 1975, Bradbury 1976, Kornberg 1977, Isenberg 1979), where n had been identified as ≈ 2 in native chromatin and ionic conditions which mimic it (Finch and Klug 1976). H3 and H4 are arginine-rich (18/135 amino acids in calf H3, 14/102 in calf H4 (Isenberg 1979)), and are very highly conserved between species (Panyim, Bilek and Chalkley 1971). There are only two conservative substitutions between calf and pea H4, a staggering degree of homology which is repeated in many other sequence comparisons between higher eukaryotes, although Saccharomyces H3 and H4 vary from the mammalian protein sequence by 7% (Brandt, Patterson and van Holt 1980, Isenberg 1979). H3 is nearly as strongly conserved, but differences between gel mobilities of Yeast and Tetrahymena H3s and the calf protein (Brandt and van Holt 1982) suggest that a greater degree of variability is tolerated in this protein.

H3 and H4 bind tightly to each other in solution in a double dimer, and once associated become incapable of being incorporated into a normal nucleosome without being biochemically separated again. This attests to the strength of the forces binding them together, which must be sufficient to cause major changes in the protein's secondary structure (Daban and Cantor 1982). This is believed to reflect their physically central position in the nucleosome in a 'horse-shoe shaped' tetramer (Felsenfeld 1978, Solner-Webb, Camerini-Otero and Felsenfeld

1. Histones

1976 , McGhee and Felsenfeld 1980). The evolutionary constraints reflected by the need to conserve these and other strong physiochemical interactions in such a vital pair of proteins, the physical and physiochemical centre of the nucleosome, can reasonably be assumed to be sufficient to account for the extraordinary sequence conservation of H3 and H4.

H2A and H2B are highly basic with an arginine/lysine ratio intermediate between H3 and H4, and H1 (Bradbury 1976) . The sequences of both proteins are more variable than H3 or H4, but still are highly conserved throughout eukaryotic evolution (Isenberg 1979). Notably the majority of the conserved residues fall in the C-terminal 60% of the molecule, and over half the basic amino acids in the N-terminal 30%, in both H2A and H2B. As these latter regions are probably involved with binding DNA, and also possibly other 'acidic' nuclear proteins (McGhee and Felsenfeld 1980, Bohn et al 1982, Allan et al 1982) , this paints an interesting picture of the nucleosome in which an evolutionarily constant core has the capacity for variable interactions with other macromolecules conferred upon it by the superficial regions of H2A and H2B. This is supported by findings which implicate the 'tails' of H2A and H2B in the higher order folding of chromatin through protease digestion studies (Bohm et al 1982), specifically in stabilizing the solenoid structure (Allan et al 1982), while the C-termini interact within the nucleosome's centre with each other and with H3 and H4. Further evidence for this view is discussed below. H2A and H2B associate with the (H3.H4)₂ tetramer as two heterodimers, the C-terminal part of the H2B component of each binding strongly to the C-terminal region of H4 (McGhee and Felsenfeld 1980).

The fifth histone class, H1, is the most variable both between and within organisms. These lysine-rich histones have a central region

1. Histones

conserved in evolution : amino acids 93-110 in particular are found to be highly conserved between highly divergent organisms (Isenberg 1979). NMR and spectroscopic studies of the H1 proteins in solution show this central region to have a static structure typical of a globular protein, while the N- and C-termini have little rigid structure, rather resembling random polypeptide chains (Cary et al 1981). The flexible 'tails' are assumed to bind to DNA (Bradbury 1976), but could equally interact with 'acidic' non-histone proteins now known to be integral to chromatin (Elgin and Weintraub 1975). H1 is only loosely associated with the core nucleosome, but it does assist in the compaction of nucleosomal particles into higher order structure in oligonucleosomes and probably in intact chromatin (Felsenfeld 1978, Lia and Cole 1981, Brasch 1976, Rao, Rao and Ganguly 1982). In this regard it is notable that the highly variant histone H5 comprises up to 50% of the lysine-rich histones in avian erythrocytes (Brasch 1976), and a lower fraction of amphibian (Flynn and Woodland 1980) and fish (Miki and Nedin 1975) erythrocyte H1-like protein, where the chromatin is condensed and transcriptionally inert.

1.2 Nucleosome structure and chromatin.

The consensus structure of the nucleosome derived from hydrodynamic, spectral, cross-linking, stoichiometric and electron microscopic studies (reviewed in McGhee and Felsenfeld 1980) is presented in fig 1.1 . This shows also where the other major component of the nucleosome, the DNA, probably fits onto the particle. About 140 bp of DNA are wrapped 1.7 times around the 'core' nucleosome, and another 20 - 50 bp are probably associated with H1 at each end to make the complete particle (Finch and Klug 1976, Felsenfeld 1980, McGhee and Felsenfeld 1980). DNA interacts with all four core nucleosomes. It

Fig 1.1

The Nucleosome

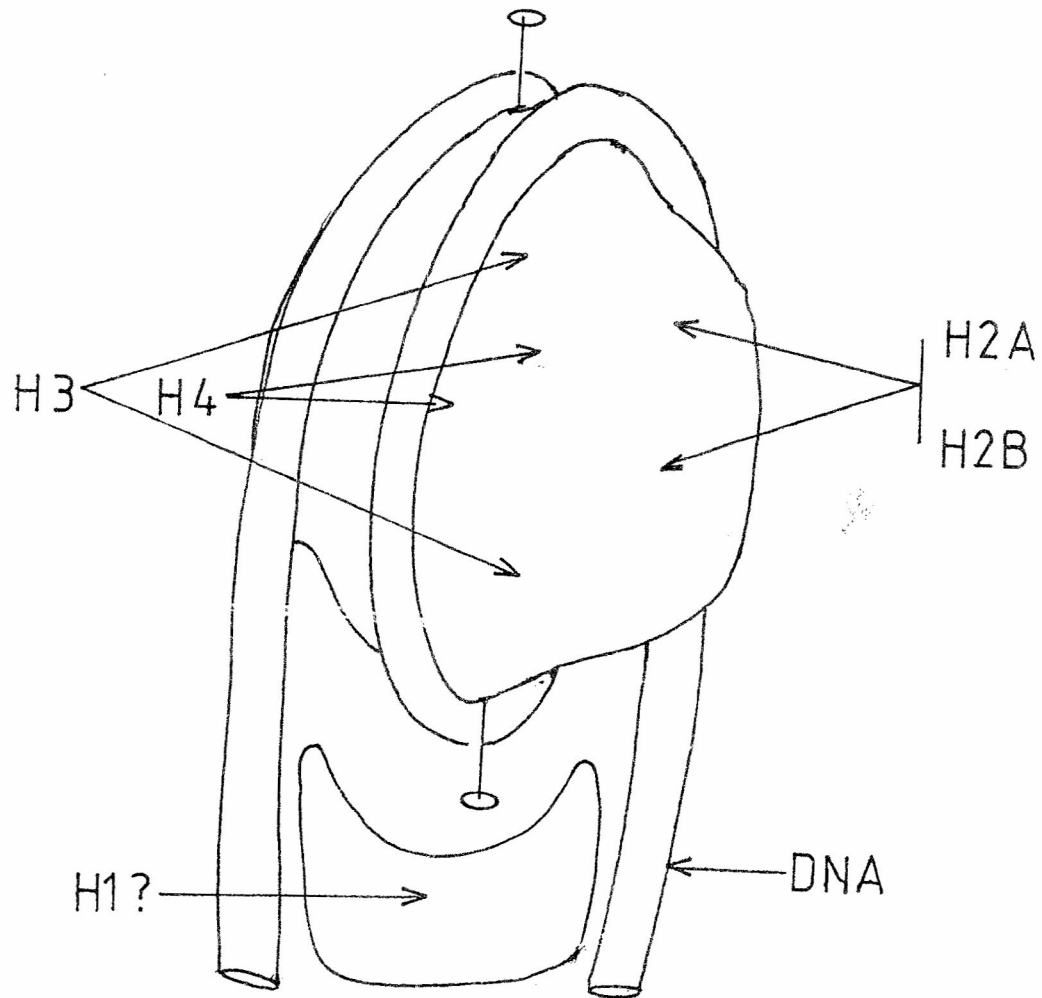


Fig 1.1 Consensus structure of the calf thymus chromatin nucleosome.
After Kornberg 1977, McGhee and Felsenfeld 1980, Felsenfeld 1980. See
text for details.

1. Histones

appears to bind to a 'ridge' which runs round the core structure and is composed of sequences from all the protein molecules in the core.

Chemical cross-linking studies have shown which part of the 140 bp of DNA interacts with which protein, with interesting results (reviewed in McGhee and Felsenfeld 1980). The central turn, ≈ 90 bp, interacts mainly with H3 and H4. The rest of the DNA which is associated with the 'core' interacts primarily with H2A and H2B, and the extra-nucleosomal 'tail' interacts with H1. This gives a glimpse of how H2A, H2B and H1 condense chromatin through interactions between their highly basic 'tails' and either DNA or other proteins bound to it. Increasing the strength of these interactions would compress the chromatin from a loose coil of 100bp stretches of (H3+H4)-bound DNA interspersed with 100bp of loosely bound DNA into a structure in which essentially all the DNA within and between the 'cores' is bound into rigid nucleoprotein. Binding of H1 molecules to each other, either by natural covalent cross-linking or non-covalently through binding to an intermediary protein, would have a similar condensing effect, a point we will come to later.

The nucleosome is only the basic unit of nuclear structure: nucleosomes are built into higher order helices with 5 - 6 nucleosomes per 'supercoil' (Finch and Klug 1976). These supercoils are relatively unstable in solution, but larger structures appear, paradoxically, not to share this evanescence. 50 - 60 nucleosomes on a single DNA strand make a more stable structure than 5 - 6, suggesting an even higher order structure with 10 solenoidal turns per unit.

The specificity of the arrangement of proteins relative to each other and to the DNA strand in the nucleosome is not mirrored by a corresponding DNA-sequence specificity for the position of nucleosome

1. Histones

placement. Discussions of the 'phasing' of nucleosomes, the position of the particles on defined sequences of DNA in chromatin, has been plagued by artefacts generated during in situ nuclease digestion of chromatin. However it is clear that, in at least some repeated DNA sequences (5S rDNA and satellite DNA) nucleosomes are positioned preferentially at specific sites on the DNA (Zachau and Igo-Kemenes 1981). The significance of this is uncertain.

As well as histones, chromatin contains an equal mass of non-histone proteins, often called 'acidic' although many are actually quite basic (Elgin and Weintraub 1975). Non-histone proteins are a general feature of chromatin, being bound to over 90% of nucleosomes and not just to a restricted set of them (Bustin, Goldblatt and Spalding 1976). Although this study will not investigate any facet of non-histone nuclear proteins, it is important to remember that it is they, and not the histones, which are probably responsible for overall nuclear and chromosomal shape (Cook and Brazell 1975, Adolph, Cheng and Laemmli 1977, Paulson and Laemmli 1977, Lebkowski and Laemmli 1982a,b) by formation of a nucleoskeleton, and that DNA replication (Pardoll, Vogelstein and Coffey 1980, McCready et al 1980) and RNA synthesis (Miller, Huang and Pogo 1978, Herman, Weymouth and Panman 1978, Jackson, McCready and Cook 1981) occur in association with this nucleoskeleton. Furthermore two non-histone proteins, HMG 14 and HMG 17, are responsible for conferring on chromatin the 'DNaseI-sensitive' conformation characteristic of transcribed genes (Weintraub and Groudine 1975, Weisbrod, Groudine and Weintraub 1978, Weisbrod 1982), believed to reflect an open, accessible form of the nucleosome in which RNA polymerase molecules may approach the DNA and move along it. DNA which is tightly wound around nucleosomes is clearly inaccessible as a template for RNA synthesis. The nature of the 'open' configuration detected by these experiments must therefore be

1. Histones

different from condensed nucleosomes, although its nature is still unclear.

1.3 Histone variants.

Variants or subtypes of each of the histone classes except H4 are known (Isenberg 1979). The most dramatically variant histone is the H1-like histone H5 which comprises up to 50% of the lysine- , rich histones of the erythrocytes of some non-mammalian vertebrates (vide supra) . Many other H1 variants are known. Multiple H1 types are known in rabbits (Bustin and Cole 1968, Langan, Rall and Cole 1971), calf (Kincaid 1969), three rodents (Seyedin and Kistler 1979a,b, Gurley, Walters and Tobey 1979), man (Sherod, Johnson and Chalkley 1974, Borun et al 1977), chick (Sotirov and Jones 1972), newt (Imoh 1978, Asao 1969,1970), Xenopus (Flynn and Woodland 1980), trout (Miki and Neslin 1975), Buffalo fish (Panyin, Bilek and Chalkley 1971), several species of sea urchin (Easton and chalkley 1972, Cohen, Newrock and Zweidler 1975, Newrock et al 1978, Heiter et al 1979) and peanut worms (Mæær and Champagne 1976). Few of the species investigated have less than three H1 proteins, and the rat, calf and rabbit, the most extensively characterised, have at least six in the tissues examined (Bustin and Cole 1968, Kincaid 1969, Kincaid and Cole 1966, Sotirov and Jones 1972, Seyedin and Kistler 1979a,b). Stedman and Stedman (1943, 1950) suggested that histones might be associated with the control of gene action, and since they made this suggestion the unequal distribution of histone variants between tissues has lent weight to this idea. The ratio of the amounts of variant H1s to each other varies between tissues in mammals (Seyedin and Kistler 1979a,b , Kincaid 1969, Bustin and Cole 1968, Stellwagen and Cole 1968) and other species (Mæær and Champagne 1976, Brasch 1976, Asao 1969,1970). In particular H1 variants peculiar to

erythrocytes (vide supra) and sperm (Seyedin and Kistler 1979a,b, Easton and Chalkley 1972, Meistrich, Reid and Barcellona 1976, Maruchige and Maruchige 1975) have been found in several species. H1 variants have also been found to have differing affinities for non-histone proteins (Smerdon and Isenberg 1976).

H2A and H2B are also present in a wide spectrum of species as variants (Borun et al 1977, Stephens et al 1977, Shires, Carpenter and Chalkley 1975, Brandt, Patterson and van Holt 1980, Brandt et al 1979, Poccia and Hinegardner 1975). These variants are also found in differing amounts in different tissues in primates and rodents (Shires, Carpenter and Chalkley 1975, Franklin and Zweidler 1977, Cohen, Newrock and Zweidler 1975, Benezra et al 1981 Rao, Rao and Ganguly 1982), and between different developmental stages in sea urchins (Heiter et al 1979, Ruderman, Baglioni and Gross 1974, Cohen, Newrock and Zweidler 1975). H2A and H2B variants have also been correlated with the presence of HMG-17 on the same nucleosomes (Benezra et al 1981), itself correlated with 'active' chromatin (Weisbrod 1982), and also directly with a chromatin fraction bearing nascent RNA chains (Gabrielle, Hancock and Faber 1981).

Notably, the variant H1 protein types in erythrocytes and sperm have both been strongly implicated in the high degree of chromatin condensation in these inactive nuclei (Rao, Rao and Ganguly 1982, Brasch 1976), and are found selectively in the highly condensed regions of the chromatin of spermatids. Chromosomal de-condensation has been directly associated with transcription by Pol-I and Pol-II by electron microscopic studies (Scheer 1978, Foe, Wilkinson and Laird 1976, McKnight and Miller 1976).

A variety of experiments suggest that these variant histones are

1. Histones

primary structure variants, and not the product of varying post-translational modification. As well as partial or complete sequence data on the proteins (Franklin and Zweidler 1977, Brandt, Patterson and van Holt 1980, Kincaid 1969, Langan, Rall and Cole 1971, Kincaid and Cole 1966), studies showing the heterogeneity of the histone mRNAs (vide infra) suggest that variant histones differ in amino acid sequence. They may also differ in modification as well, of course (Langan, Rall and Cole 1971, Gurley, Walters and Tobey 1974, 1979)

Thus we picture the nucleosome which interacts with other nucleosomes and non-nucleosomal structures through H1, H2A and H2B, whose primary structure is altered to accomodate different developmental requirements for chromatin condensation, and add the note that these altered interactions are almost certainly associated directly with the transcriptional activity of the DNA in the chromatin.

1.4 Histone modification.

Supporting this model are observations on histone modification. Acetylation and phosphorylation occur on most histones, and might be associated with their assembly into chromatin (Ruiz-Carrillo, Waugh and Allfrey 1975, Jackson et al 1976, Kranse and Stein 1976, Gurley, Walters and Tobey 1974, 1975). Poly(ADP-ribosyl)ation is, however, more selective: H1 is modified, and even cross-linked into dimers (Hayaishi and Ueda 1977), and H2A and H2B are also extensively ADP-ribosylated in the nucleus. The cross-linking of H1 reminds us of our suppositions about the forces between nucleosomes, and the effect that enhancing these forces would have on chromatin condensation. Early assumptions that Poly(ADP-ribosyl)ation was connected solely with DNA repair have been cast into doubt by the tissue-specific and embryonic

stage-specific pattern of both poly(ADP-ribose) chain initiation and elongation (Hayaishi and Ueda 1977, Bredehorst et al 1981), and the involvement of its precursor (NADH) in the events committing a cell to DNA replication and division (Whitaker and Steinhardt 1981). Interestingly in view of the effects of variant Hls on chromatin condensation noted above, poly(ADP-ribosyl)ation of chromatin has been shown to cause relaxation of chromatin structure (Poirier et al 1982). There is no information on whether particular histone sub-types are preferentially poly(ADP-ribosyl)ated.

Poly(ADP-ribose) is turned over extremely fast in the nucleus by the simultaneous action of degradative glycohydrolases and of the synthetic polymerase (Hilz and Stone 1976). It is notable that, while an energetically inefficient way of maintaining a static system, such cycling has been shown to confer extreme sensitivity and rapidity of response of the cycling metabolite's concentration to external control in a dynamic system (Hers 1976, Newsholme 1976). Thus an alteration of the activity of either enzyme by 10% in such a cycle can at once alter the net flow through the cycle by 50% or more (Crabtree 1976). Note that any enzyme in the cycle, synthetic or degradative, is an equally effective a point of control as any other: the cross-linking of H1 proteins by poly(ADP-ribose) is controlled at the level of glycohydrolase activity (Lorimer, Stone and Kidwell 1977). Such rapidity and completeness of response is clearly seen in a number of genetic systems in response to quite slight stimuli. The already-known metabolic behaviour of nuclear poly(ADP-ribose) would well account for the speed of such dramatic 'gene switches', were it connected with the control of gene activity, something that the slower turnover of other ligands and of nuclear proteins, both histone and non-histone, could not do.

1. Histones

Another interesting modification of chromatin is by the protein ubiquitin (Goldstein et al 1975, Schlessinger, Goldstein and Niall 1975). This enormously widely distributed protein has been found covalently linked to H2A (Searle 1980), and the conjugate is found selectively on nucleosomes associated with transcribed genes (Levinger and Varshavsky 1982). It is highly significant that Ubiquitin has been found to be a cofactor required for an ATP-dependant protease (Ciechanover et al 1980, Wilkinson, Urban and Haas 1980) which is extremely peptide-specific (Sous et al 1982), and that Eichbush, Watson and Moudriankis (1976) have described a protease activity which is extremely tightly bound to chromatin and is specific for H2A (Watson and Moudriankis 1982). Whether Ubiquitin is involved in this H2A-specific protease has not been tested. However it is inviting to suggest that the ubiquitin modification of H2A is concerned with the specific proteolytic removal of H2A from chromatin, possibly for its later replacement by different variants of H2A to modulate the structure of chromatin.

1.5 Histones and the cell cycle.

As well as appearing at different times and places in ontogeny, different histone subtypes are synthesised at different times in the cell cycle. The normal view of histone synthesis was that it occurred only during S-phase when new histone was needed to package newly replicated DNA into chromatin (Spalding, Kajawara and Mueller 1966, Ruderman and Gross 1974, Stephens et al 1975, Deleage and Lee 1982, Lewin 1980, Woodland 1980). Groppi and Coffino (1980) have challenged this view, finding histone synthesis throughout the cell cycle, and also setting off a spate of rebuttals, some of which failed to duplicate their observations (although using different cell lines) (Delegeane and Lee 1982). Wu and Bonner (1981) studied histone

1. Histones

synthesis by 2-D gel electrophoresis of total cell and nuclear protein, and found the majority of histone is made in S-phase. This synthesis is closely coupled to DNA synthesis as judged by the inhibition of its synthesis by hydroxy-urea. However 10% of the synthesis occurs outside S-phase, is not coupled to DNA replication and consists of different histone variants than those synthesised in the S-phase. In particular the highly variant H2A protein H2A-Z was noted to be a 'Basal synthesis' (as opposed to S-phase synthesis) histone. H2A-Z is more highly conserved between sea urchins and mammals than other H2As, confirming different functional constraints upon its evolution (Wu, Nishioka and Bonner 1982). These authors suggested on the basis of these findings that the genome contains two quite distinct sets of histone genes: one of which produces the 'normal' histones required during S-phase, and one of which is responsible for the 'Basal synthesis' of histones which occurs throughout the cell cycle and had some other, possibly regulatory role in the cell.

Two would appear to be a minimum estimate for the number of sets of histone genes experiencing different control during the cell cycle. In rabbit eye lens epithelia a sub-population of histones is synthesised during S-phase, but is not coupled to DNA synthesis (Briggs, Rothstein and Wainwright 1976); several sub-populations of hamster H1s are defined by their phosphorylation behaviour during the cell cycle (Gurley, Walters and Tobey 1974, 1979); Stephens et al (1977) found that sub-populations of histone mRNAs defined by the sensitivity of their synthesis to inhibitors of DNA synthesis, are themselves internally heterogenous.

Interestingly Roberts et al (1975) have noted alterations in the Poly(ADP-ribose)synthetase activity during the cell cycle. Histones

1. Histones

are the main nuclear acceptors for Poly(ADP-ribose) and H2A and H2B are the most poly(ADP-ribos)ylated histones (Isenberg 1979).

It is not possible to draw a detailed picture of chromatin structure in relation to gene activity from these isolated data. However we may adduct some general conclusions.

1) H3 and H4 form a central core to the nucleosome, being highly conserved within and between species.

2) H1 interacts with the 'outside' of the nucleosome. It is present in a number of variants which are correlated with different genetic programs. This correlation may be related to the protein's role in condensing chromatin.

3) H2A and H2B are functionally (although not chemically) intermediate between H3 and H4, and H1. Their C-termini interact with each other to form strongly bound dimers, so it is reasonable to think of them as protein pairs rather than as individual species. Their N-terminal 'tails' interact outside the nucleosome, which interaction is connected with chromatin condensation. Variant H2A and H2B proteins are found differently distributed between cells expressing different genetic programs, and may be directly concerned with differential gene activity.

4) Post-translational modification, especially by poly(ADP-ribose) and ubiquitin, adds to the variability in the chemistry of the histones found between different differentiated states, and some may also be associated with altered gene activity. Again H1, H2A and H2B are the major proteins altered.

1. Histones

1.6 What should we expect from histone genes?

The picture of the histone proteins drawn above gives some strong predictions about the organisation of their genes, features whose discovery was paradoxically treated with some surprise when they were actually found (Kedes and Maxson 1981). The heterogeneity of histones described above would immediately suggest a similar diversity of genes, a minimum of 4 for each of H2A and H2B, and 6 for H1 to account for the observed protein diversity in the more studied species. Other studies above suggest that a similar diversity applies to a wide variety of species, including the amphibia.

In a few taxa one set of these - those transcribed during oogenesis - would be present in higher copy number than the others. The demand for histones in amphibian oogenesis and sea urchin embryogenesis is so high that a single set of histone genes could not synthesise sufficient mRNA to meet it. Hence more than one set is needed (Woodland and Adamson 1977, Adamson and Woodland 1977, Davidson 1976, Davidson, Hough-Evans and Britten 1982, Woodland 1980). For Xenopus laevis at least 40 genes are needed during oogenesis to synthesise the huge pool of maternal histone mRNA and protein from only two diploid chromosome sets in the time available (Woodland and Wilt 1980a,b). As this histone synthesis is uncoupled from DNA synthesis in Xenopus (Adamson and Woodland 1974), we can reasonably expect these reiterated genes to be different from those histone genes which act mainly during S-phase in sequence, chromosomal location or both. Sea Urchins require a similarly amplified set of histone genes to provide chromatin for the cleavages in early development (Davidson 1976). Indeed, the reiteration in sea urchins is so dramatic that it led to the idea that the amplified genes were essentially the only histone genes in these organisms, and so the

1. Histones

organisation of all histone genes in both this and other taxa would follow the simple, tandem-repeated structure in which these embryonic genes are arranged (Lewin 1980). This has subsequently turned out to be untrue.

As well as variant proteins, subsequent work using cloned probes has found that identical histones may be coded by a number of mRNA species of different lengths, even within an 'homogenous' cell or embryo population in mouse (Sittman et al 1981), man (Sierra et al 1982, Stephens et al 1977, Borun et al 1977, Lichtler et al 1982), sea urchin (Heiter et al 1979, Maxson and Wilt 1981, Grunstein 1978, Childs, Maxson and Kedes 1979) and Xenopus (Old per com, Zernik et al 1981). So in the family of genes coding histone proteins we can expect to find a number of different gene types for each histone class, and at least one type which is reiterated to cater for the high transcriptional demands of oogenesis and embryogenesis in organisms where these processes require a large amount of mRNA to be produced in a short time. The number of variant proteins found will give a minimum expectation of the number of genes, which is raised by further heterogeneity among mRNAs coding for the same protein. Some of these genes will be differently regulated from others, requiring transcription at different times and in different tissues, and so could be different in sequence structure, chromosomal location or both. With these expectations in mind we can now move on to discuss the subject of this thesis, the histone genes themselves.

2. Genes.

2.1 The eukaryotic genome.

The DNA of eukaryotes may be conveniently divided into classes according to its self-renaturation kinetics (Lewin 1980). 'Snapback' DNA renatures essentially instantly, and consists of inverted repeat sequences. The fast-renaturing or 'highly repeated' DNA class contains all sequences present in the genome at 20,000-fold reiteration or more (the cutoff value is arbitrary), including satellite DNAs. Intermediate repetitive DNA falls between this class and 'unique', single copy genes: it is a highly heterogenous class with respect to its complexity and repetition and, as we shall see, to its actual or possible 'function'. The definition of 'unique' in this context depends on hybridisation stringency and the divergence between low copy-number genes.

Histone genes fall between intermediate repetitive and unique classes, being present in 10-1000 copies each in different species. We will therefore be concerned mainly with a family of genes which are Intermediate Repetitive. However we shall see later that they bear many affinities to unique genes.

2.2 Copy number of histone genes.

2. Genes

Some effort has been expended on showing how many histone genes different organisms possess. As a result we have a good idea not only of how histone gene numbers vary between species but also how the genes' sequence varies within an organism.

The sea urchin histone genes may be separated from most of the rest of their genomic DNA by centrifugation on equilibrium gradients, and so comprise a formal satellite (Kedes and Birnsteil 1971). They were the subject of most pre-cloning studies on histone genes. The genes in this satellite are reiterated @400 times and, as is required of a satellite, are closely clustered. Subsequent work has shown that @50 copies of at least one histone gene (H3) exist outside this cluster, and some of these have been shown to be pseudogenes (Childs et al 1981). The total histone gene number in these organisms is about 500 copies of each gene type. The sea urchins

Strongylocentrotus purpuratus, Psammechinus miliaris, Echinus esculentus and Lytechinus pictus have very similar gene numbers and 'satellite' organisations (Kedes and Birnsteil 1971, Portmann, Schaffner and Birnsteil 1976, Kedes et al 1975, Sures et al 1978, Kedes 1979), although Strongylocentrotus and Lytechinus belong to different orders and live on opposite sides of the American continent (Barnes 1974). This degree of conservation of structure has been seen many times since in histone and other gene families.

Drosophila melanogaster contains 110 copies of each histone gene (Lifton et al 1977, Kedes 1979). Other organisms contain numbers around 100-1000 for lower vertebrates, 10-30 for some mammals. Some available histone gene numbers are given in Fig 2.1 . There is no clear pattern to this distribution, apart from that organisms with high C-values have large numbers of histone genes. The converse is not true: the sea urchin has nearly as many genes as the newts

Fig 2.1

Histone gene numbers

Species		ref
Sea Urchin	500	(1)
<u>Drosophila</u>	110	(2)
<u>Xenopus laevis</u>	90	(3,4)
<u>Xenopus borealis</u>	90	(4)
<u>Triturus</u>	640	(3)
<u>Ambystoma</u>	2700	(3)
Mouse	10-20	(6)
Man	30-40	(7)
Chick	10	(8)
Yeast	2	(9)

Fig 2.1 Histone gene numbers . Haploid genome compliments of each type of histone gene. references: 1) Kedes and Birnstiel 1971, Kedes et al 1975a,b. 2) Lifton et al 1977, Pardue et al 1977a,b; 3) Hilder et al 1981, 4) Turner and Woodland 1982b), 6) Jacob 1976, 7) Wilson and Melli 1977, 8) Crawford et al 1979, 9) Wallis, Hereford and Grunstein 1980.

2. Genes

Triturus and Notophthalmus, but only one 25th of their C-value (Kedes and Birnstiel 1971, Hilder et al 1981). Although the histone gene numbers of the Amphibian correlate with C-value rather better, as would be expected if we assume the critical factor in determining histone gene number is the number of genes required to produce stored histone message during oogenesis for use in the rapid divisions of embryogenesis (Woodland and Wilt 1980a,b) (this point is discussed further below), and that this depends on the c-value, gene numbers are still not proportional to C-value (Hilder et al 1981).

Another problem encountered when trying to compare these figures is that different methods have been used to obtain them: direct quantification of DNA purified on CsCl gradients (Kedes and Birnstiel 1971); hybridisation with poly-A⁺ (Jacob 1976) or size-fractionated polysomal (Wilson and Melli 1977) RNA, and hybridisation with a variety of cloned probes under widely differing conditions (Stephenson, Erba and Gall 1981b, van Dongen et al 1981, Turner and Woodland 1982b). Turner and Woodland (1982b) have demonstrated that there is considerable sequence heterogeneity among histone genes in X.laevis, so previous estimates of histone gene numbers in this species, based on hybridisation with homologous probes under stringent conditions which produced the figure 20-50 of each gene type (Jacob, Malacinski and Birnstiel 1976, van Dongen et al 1981) were underestimated by a factor of two. Despite the known conservation of protein sequence it is clear that even H4 genes vary by up to 20% between species at the level of nucleotide sequence, almost entirely due to variation in the third bases of codons (Busslinger et al 1982, Turner and Woodland 1982b). Thus hybridisation in stringent conditions will fail to detect a substantial proportion of H4 genes.

2. Genes

Whatever their uncertainties we should bear these figures in mind so that we may relate the cloned genes discussed below to the total genome from which they are derived.

2.3 Structure of cloned histone genes.

2.3.1 The Sea urchin pattern.

The sea urchin contains 300-500 histone genes which are clustered into a 'satellite' DNA. This proved to be a tandem repeat of a cluster containing one of each of the five histone gene classes (Kedes et al 1975a,b, Portmann, Schaffner and Birnstiel 1976, Sures et al 1978, Kedes 1979) (see Fig 2.2) . The structure of this repeat has been highly conserved between species with respect to gene order and polarity, although wide divergence is seen between the spacer sequences of different species. Within a species the clusters are homogenous with a few specific exceptions. Lytechinus pictus has two variants on the basic repeat structure which are non-allelic, not intermingled, and present in approximately equal numbers (Cohn and Kedes 1979). Their overall structure is the same, but there is considerable sequence heterogeneity between them. By contrast the two classes are internally homogenous with respect to sequence as revealed by restriction mapping. Echinus esculentus has two allelic types of cluster, differing by a 1kb 'insert' near the H4 gene (Hentschel and Birnstiel 1981), and Strongylocentrotus purpuratus has two cluster variants which are also probably allelic (Overton and Weinberg 1978).

The different orders of sea urchin from which histone genes have been cloned diverged from a common ancestor in the mid Mesozoic

Fig 2.2

Invertebrate Histone genes

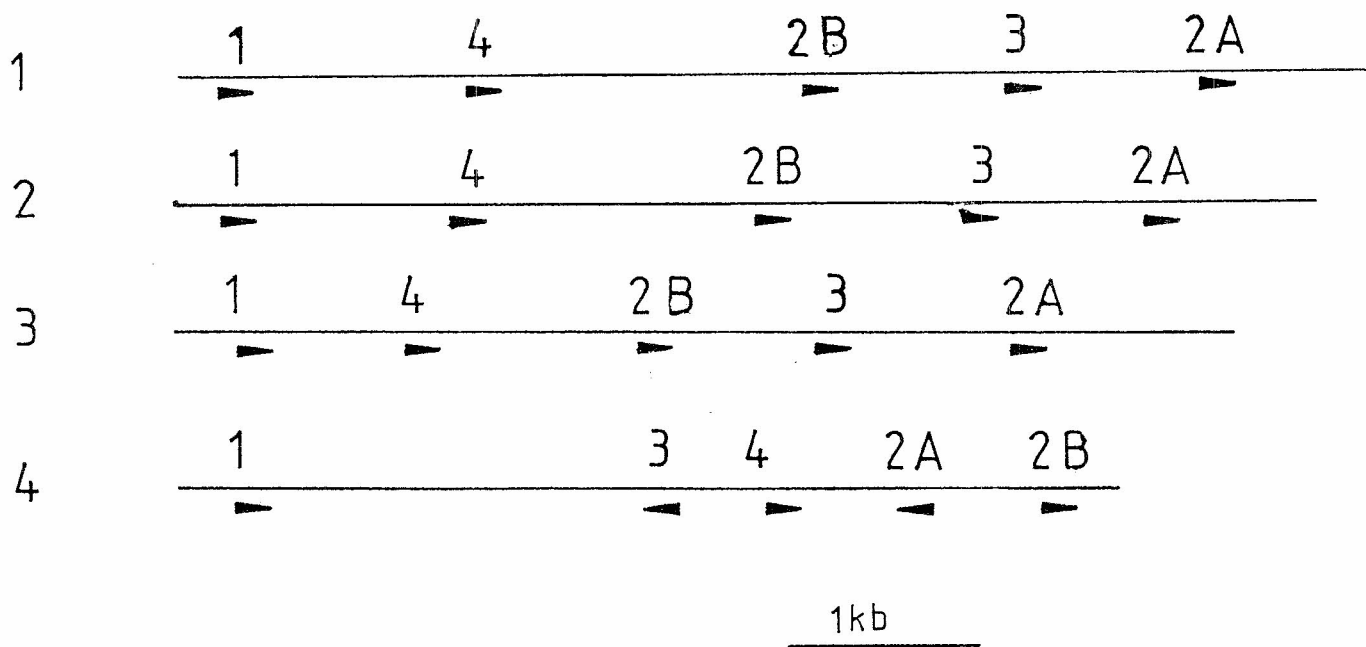


Fig 2.2 Embryonic histone gene repeats of invertebrates. 1) Lytechinus pictus, 2) Strongylocentrotus purpuratus, 3) Psammechinus miliaris, 4) Drosophila melanogaster. Arrows indicate the polarity of transcription of the genes. (1=H1, 2A=H2A, 2B=H2B, 3=H3, 4=H4) After Kedes et al 1975a,b, Portmann, Schafner and Birnstiel 1976, Sures et al 1978, Cohn and Kedes 1979, Lifton et al 1977.

2. Genes

(Barnes 1974), and presumably the consensus sea urchin histone gene cluster has been preserved since then. This points to considerable evolutionary stability. Busslinger, Rusconi and Birnsteil (1981) note that within coding regions of H3 and H4 genes of a number of sea urchin species the 'silent' bases - bases which are redundant in the genetic code - evolve as fast as other unique genes (with a notable exception), even though the overall gene cluster structure is constant. The main mode of genome fluidity here is therefore base substitution. Turner and Woodland (1982) have pointed out that all silent bases do not evolve at equal rates, some being conserved in a wide range of species from sea urchin to man, suggesting selection pressures acting on non-coding as well as coding bases in these genes. A notable example of such conservation was discovered by Busslinger et al (1982), who discovered a minor gene cluster in Psammechinus miliaris which was essentially identical to one from Strongylocentrotus drobachensis in the H3 gene and 300bp of adjacent, non-transcribed spacer. As these species shared a common ancestor no more recently than the Cretaceous this represents an extraordinary degree of conservation of DNA sequence. Some time prior to their publication we had found an analogous situation in the Xenopinae, and similar severe restraints on non-coding gene sequence have been found in other gene families (vide infra).

The sea urchin genes are all transcribed in the same direction on separate messages (Mauron et al 1981). Sequence studies have shown the tandemly repeated genes to code for embryonic histones (Hieter et al 1979, Childs, Maxson and Kedes 1979, Spinelli et al 1980, Grunstein et al 1981). This predicts that other histones should be coded by other genes not in this repeating structure. We noticed above that we can expect at least 6 of each histone gene type per haploid genome to account for the observed diversity of

2. Genes

proteins. Childs et al (1981) have cloned representatives of these minor sea urchin genes, which are unlinked to the embryonic genes (cited in Davidson, Hough-Evans and Britten 1982). They also found a number of copies of pseudogenes derived from the embryonic repeat.

The Drosophila histone genes are organised in a tandem repeat similar in size to the sea urchin's with 110 copies per haploid genome (Lifton et al 1977). The gene order is different from that in the sea urchin cluster, however, and unlike the sea urchin transcription occurs off both DNA strands (Fig 2.2). Minor Drosophila histone gene clones have been isolated, but seem to be variants on the major embryonic gene cluster like the sea urchin variants rather than the genes coding for variant proteins (Kedes 1979). Again, this might represent the difficulty of finding one gene which is not homologous to the probe among over 100 which are homologous.

The tandemly repeated structure of the sea urchin and Drosophila embryonic histone genes is now regarded as an oddity of the requirements of embryogenesis, and not as a model for all histone gene clusters. It is supposed that one copy of a proto-embryonic histone gene has been amplified in the past in response to the transcriptional requirements of embryogenesis, and that selection pressure, and/or unequal crossover and gene correction, have conserved this structure ever since. In this respect the variation between Lytechinus clusters is interesting in that it could represent a secondary amplification of another gene cluster after the initial gene expansion event. By contrast the variant alleles found in Drosophila, Echinus esculentus and Strongylocentrotus purpuratus must be derivatives of a single expansion event. These could have diverged from each other through

2. Genes

selection though heterozygous advantage, or by random drift of neutral polymorphisms in transiently isolated sub-populations of the species. No reasonable mechanism of gene amplification and selection for the products could explain how two different gene clusters could be amplified at allelic positions on sister chromosomes. Thus variation among the highly conserved embryonic histone gene clusters have been generated by a number of mechanisms.

2.3.2 The Chicken non-pattern.

Initially it was thought that the sea urchin pattern of homogenous, tandem repeats of gene clusters was going to be repeated in all organisms (Kedes 1979, Lewin 1980), and claims to find such a pattern in chick (Crawford et al 1979) and Xenopus (van Dongen et al 1981) supported this idea. However it is now clear that such a simple pattern is not prevalent in vertebrates. Given our predictions about the organisation of histone genes made above, and the known copy-number of these genes in vertebrates, this is not surprising: if there must be at least 6 types of each histone gene in chick, and there are only ten of each histone gene altogether, the potential for satellite-like repetition of one cluster is clearly limited.

The available information on the chick, mouse and human genes is summarised in Fig 2.4 . There is no clear pattern here. Some clones may be seen as being overlapping representatives of the same genomic sequence. The number of overlapping groups of clones may be taken as an estimate of the number of histone gene clusters: in this sense chick must have at least three clusters, man 6 and mouse 2. However the cloned clusters may be linked by greater lengths of DNA than have been examined, so that the clones represent isolated sections

Fig 2.4

Higher vertebrate Histone genes

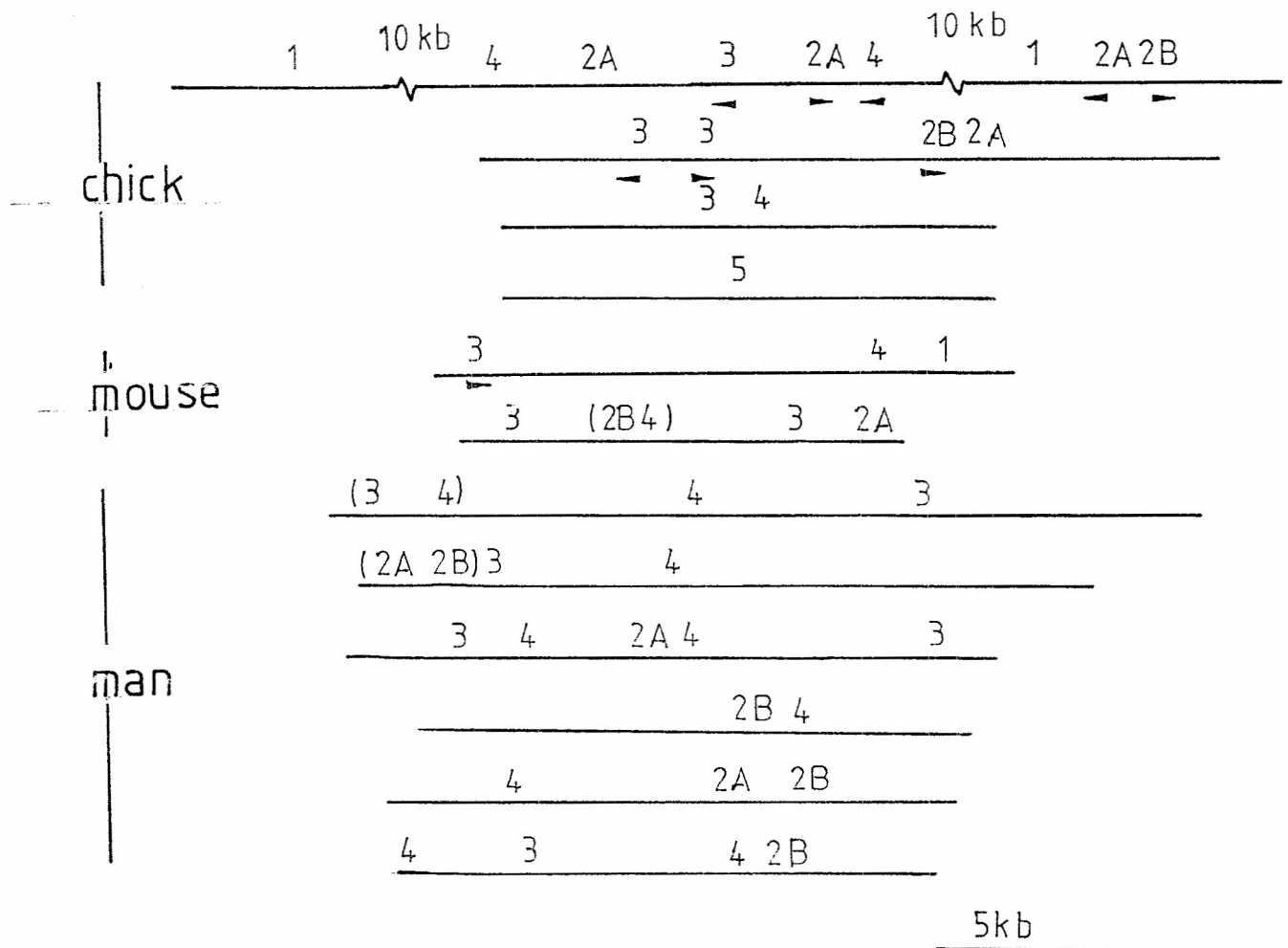


Fig 2.4 Structure of cloned histone gene clusters from higher vertebrates. The chick clusters are the combined map derived from several overlapping clones. Symbols as in Fig 2.2. (chick: Harvey et al 1981, R. D'Andrea and P. Krieg per com, Engel and Dodgson 1981, Harvey et al 1981, Krieg et al 1982. mouse: Sittman et al 1981. man, Heintz, Zernik and Roeder 1981, Sierra et al 1982).

2. Genes

of one cluster, or may be diverse representatives of a cluster which is extremely polymorphic. Clearly for all species except the chick these clones represent only a small minority of the histone genes present in the genome, and even chick has yet to yield a complete library of histone sequences.

Separate genes for H5 (Krieg et al 1982) and two variant core histones designated H2A-F and H2B-F (Krieg per com) have been cloned, and do not appear to be linked to other histone genes. H2A-F has been tentatively identified with H2A-Z, the H2A variant characteristic of non - S phase, DNA synthesis - uncoupled histone synthesis in growing cells (Wu and Bonner 1981). A fourth variant gene, coding a fairly 'normal' H3 protein, has been found to contain two introns (Engel, Sugarman and Dodgson 1982), the first in nearly 120 cloned histone genes isolated to be split by introns.

It is clear from the chick, mouse and human histone gene clones that in the higher vertebrates there is no five gene: one cluster rule seen so frequently in invertebrate embryonic histone genes.

Several histone genes have been cloned from Saccharomyces cerevisiae (Wallis, Hereford and Grunstein 1980, Choe, Kolodrubetz and Grunstein 1982). Their structure is shown in Fig 2.5 . These both contain a pair of divergently transcribed H2A and H2B genes which code for similar but different pairs of proteins.

2.3.3 Conserved features of cloned histone genes.

Histone genes from many diverse species have been found to have some features in common. They were originally noticed in invertebrate histone genes, and have since been confirmed in

Fig 2.5

Yeast histone genes

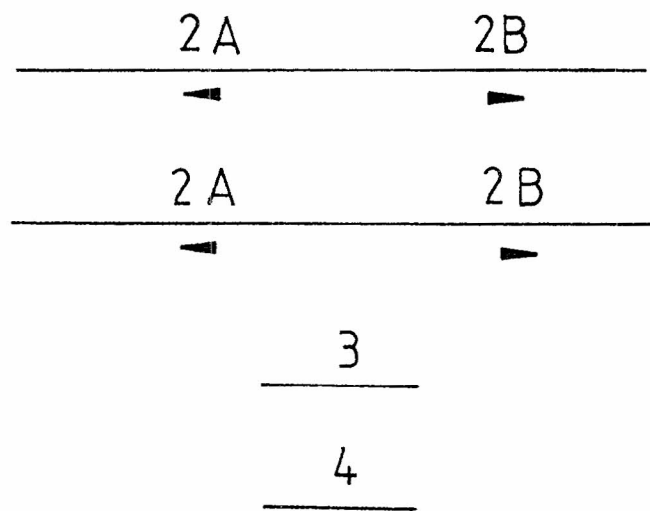


Fig 2.5 The identified histone gene clusters of yeast. Symbols as in Fig 2.2. These represent the yeast genome's total complement of histone genes. (Wallis, Hereford and Grunstein 1980, Choe, Koldrubetz and Grunstein 1982).

2. Genes

but no 3' homology block, although it does possess a hyphenated repeat of quite different sequence starting at the equivalent position downstream of the coding sequence. The H3 gene variant does not have a polyadenylation signal, but contains the genomic sequence AAAAAAAAAA.G.AAAAAAAAAA at its 3' end which, if it is transcribed, would produce an oligo-A tract in the message which could bind to oligo-dT in message purification and cDNA synthesis. H2A-F and H2B-F histones might also be polyadenylated (Krieg per com).

2.3.4 The chick model for vertebrate histone genes.

As the chick histone genes have been the most thoroughly characterised of the vertebrate histone genes, it is worth considering them as a new archetype at this point. The genes are clearly clustered: the occurrence of 9 out of 30 genes in a 40kb segment of a 1900 Mb genome shows a strong tendency for histone genes to stay together. Genes on both clusters examined have similar sequences throughout the coding region, but vary completely outside these regions. They are conserved between individuals of an inbred strain of chickens: in four haploid chromosome sets examined (2 animals) only one polymorphic EcoRI site has been found in the 40kb segment (D'Andrea and Krieg per com). Separate from these clusters are highly variant genes which have features not common in histone genes, but usual for other protein-coding genes. As there is no highly-reiterated class of histone genes in chicken, these variants are relatively more abundant than in sea urchin or Drosophila, where variants are numerically very minor and hence hard to clone. In short, the chicken histone genes, and presumably those of the mouse and man as well although data are more scanty here, support the suggestions we made regarding gene structure on the basis of

2. Genes

protein structure and mRNA heterogeneity.

Two unsolved problems remain (apart from the incompleteness of the data). Firstly, what necessitates the extensive conservation of some histone gene clusters within and between species, as exemplified by the sea urchin embryonic genes, while not constraining, say, all H3 or all H4 genes to have remotely similar sequences outside their coding region, with the exception of the short conserved features mentioned above. This is equivalent to asking: what are the various histone genes' and their variant protein products' function? As yet isolation, characterisation and application of the clones to biological problems is at too embryonic a stage to answer such a question, let alone discuss how their function might constrain their structure.

Secondly how did the histone genes in chick or man arrive in their current configuration? A descriptive answer to this latter, simpler question might shed some light on the former, as well as giving a classificatory structure to a bewildering diversity of genes. Such questions cannot be easily approached in the mammals or birds so far, as too few species have been studied and their histone genes' organisation appears too complex to reveal much evolutionary detail. However the Amphibia offer more hope.

2.3.5 The Amphibian semi-pattern.

The histone genes of the Amphibian, specifically the frogs Xenopus laevis and Xenopus borealis with histone gene numbers of ©90 each, are intermediate in number between the highly repetitive sea urchin pattern and the disordered mammalian one. It is

reasonable to suppose that the Xenopus histone genes expressed in oocytes will be reiterated in response to the demand for histone protein and mRNA in the oocyte during oogenesis (Woodland and Wilt 1980a,b), which suggests that there will be a 'major' type of histone gene cluster coding for these histones. However it is no longer thought sufficient to study these genes alone to characterise 'The' histone gene structure in Xenopus, any more than a car-spotter would be content to record only Fords because they comprise 30% of cars. A Ferrari or a Rolls, while less numerically significant, is just as important to a student of car design. Similarly we need to study not just the numerically major clusters of genes in the Amphibian, but the structures of all histone gene clusters or representatives thereof.

The available information on Amphibiann histone genes is summarised in Fig 2.7 . This figure also includes the results from the clone Xbhl described in this study. One feature immediately notable is that the majority of the clones from X.laevis fall into a single pattern of gene order, although spacing between the genes and restriction maps vary somewhat from clone to clone, and that the clones of Zernik et al (1980) and van Dongen (1982) contain a pair of H3 and H4 genes at each end, apparently derived from a tandem repeat of this structure. Zernik et al (1981) also report that clusters of this type contain different H1s homologous to one or other of the two oocyte H1 proteins, H1a and H1b, although whether this means that the genes actually code oocyte proteins is not proven, merely that they are strongly homologous to their mRNAs as assayed by hybridisation.

In one respect Fig 2.7 is misleading in that it shows types of clones and not individual isolates. In some cases multiple isolates

Fig 2.7

Amphibian histone genes

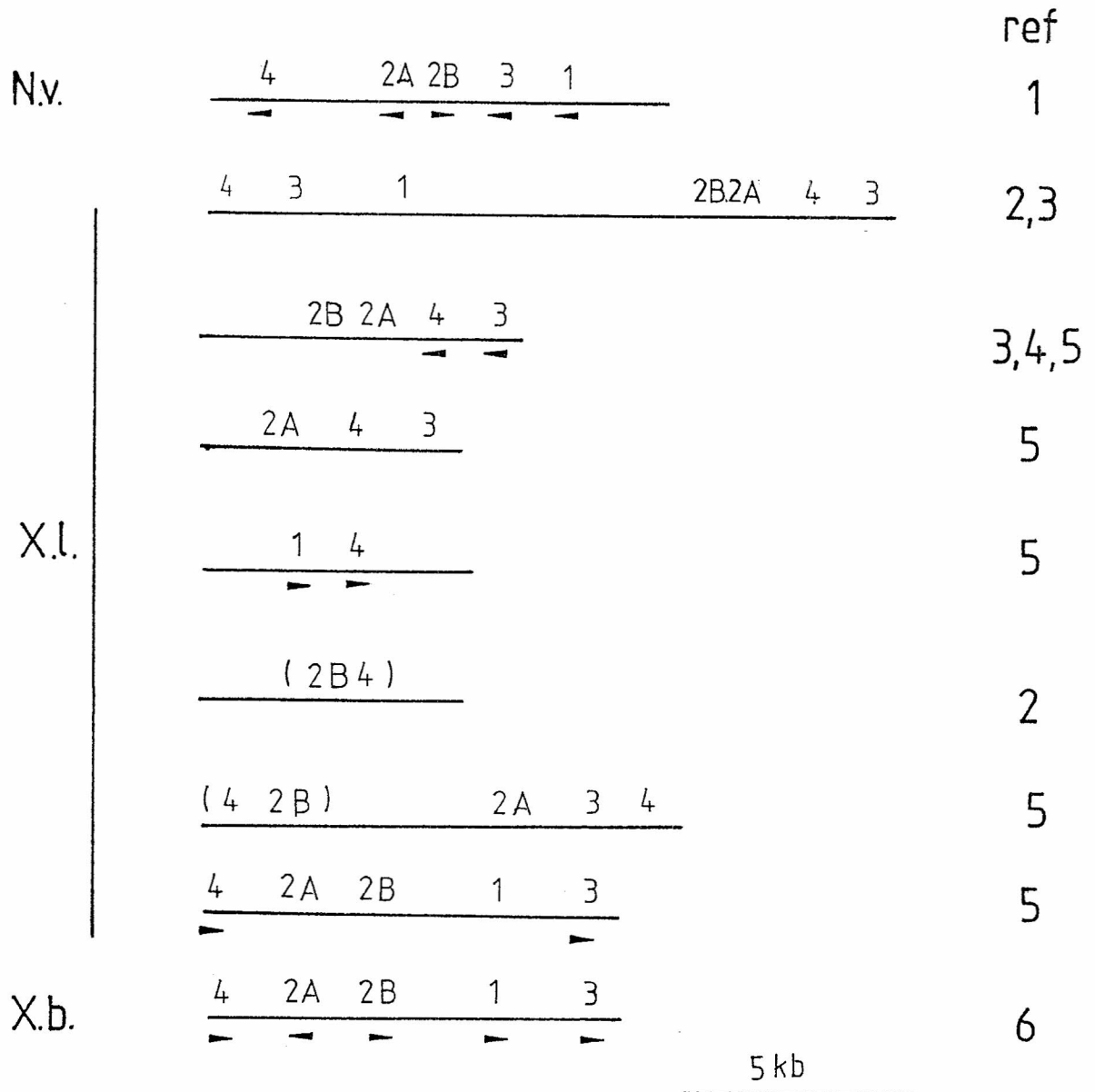


Fig 2.7 Structure of cloned amphibian histone genes of Notophthalmus viridescens (N.v.) , Xenopus laevis (X.l.) and Xenopus borealis (X.b.) Note that several workers have isolated clones of identical gene order and similar restriction map to the first two clones from X.laevis. Symbols as in Fig 2.2. Refs: 1) Stephenson, Erba and Gall 1981, 2) Zernik et al 1981, 3) van Dongen 1982, 4) Moorman et al 1980, 5) T.C.Aldridge and R.W.Old per com, 6) this work.

2. Genes

of very similar clones are reported as such (Harvey et al 1981, van Dongen 1982), especially when this reveals heterogeneity between similar clusters. However workers usually choose to concentrate on a 'typical' example of a cloned sequence rather than publish characterisations of all independent isolates of apparently identical clones. This reflects the idea that different gene clusters will vary widely, so if cloned clusters are similar or identical as judged by a cursory restriction map then they may be grouped together and a single representative analysed. This practice also reflects problems in estimating genomic copy-numbers from the frequency of a gene in a gene library. A gene present in one or two copies per haploid genome in the original DNA from which the library was made may not be represented at all in the final library, or may be over-represented dozens of times. (Possible causes and effects of this are discussed in Section 7.3.4 and Appendix 2, as they apply to this work.) Workers who find a gene at high frequency in a gene library will assume that this is due to cloning artefacts unless compelling physical mapping data from experiments on the original DNA shows otherwise. An example of the loss of interesting information which failure to examine similar clones in detail can cause is shown by the phenomena associated with R.W.Old's clone Xlh8, discussed below (section 13).

The distortions of relative gene numbers introduced during the cloning process also makes estimating the copy-number of rare sequences from cloning data impossible. Because a number of workers have reported independent isolates of clones from X.laevis with the same gene order and approximate gene spacing as those of Zernik et al (1981), it is tempting to assign a copy-number to this 'major' gene cluster type. While, because of its abundance in at least three gene libraries and because of other data mentioned

2. Genes

below, we can say that this 'major cluster' of X.laevis is major in that it is present in more copies than any other type of gene cluster, we cannot assign an accurate copy number to it, and can only say of the 'minor' clusters that they are less abundant than the 'major' one.

The vast majority of the histone genes in the newt Notophthalmus viridescens lie in tandemly repeated clusters consisting of 9kb gene clusters separated by 20kb to more than 50kb of satellite DNA (Stephenson, Erba and Gall 1981a,b). Digestion of genomic DNA with restriction enzymes which cut the gene cluster but not the satellite and probing with cloned histone probes shows there to be no fragments detectable smaller than ≈ 20 kb which are not consistent with the existence of some restriction site polymorphism in this cluster, and hybridisation occurs up to the limit of resolution of the gels, at least 50kb. Notophthalmus has a very high C-value (45pg) (Stephenson, Erba and Gall 1981b), one of the highest vertebrate C-values known. Hilder et al (1981) pointed out that this would imply the need for a very large number of histone genes to provide histone to 'package' DNA into chromatin during the early stages of embryogenesis, when histone is synthesised off stored maternal message. Hilder et al (1981) suggested that the Notophthalmus histone genes were reiterated more than could be accounted for by this line of argument. However, if we assume that ≈ 50 of the X.laevis histone genes are reiterated oocytic genes, then we would expect Notophthalmus to have $50 \times 45/3.2$ reiterated histone genes, ≈ 700 , i.e. exactly what we observe (Xenopus laevis C-value ≈ 3.2 pg). A similar calculation for Triturus cristatus (a related newt) with a C-value of 23pg (Hilder et al 1981) gives an expectation of $23/3.2 \times 50$ reiterated histone genes, ≈ 350 . The actual number is 640 (Hilder et al 1981), suggesting that these genes are

2. Genes

'over-amplified'. An even greater over-amplification occurs in the sea urchin genes as we noted above. This may reflect other evolutionary pressures at work than we have accounted for here.

We can note that the order of the core histone genes is the same in Notophthalmus as in many X.laevis histone gene clones.

Several studies have looked at the histone genes in genomic DNA, rather than in clones, by probing Southern blots of genomic DNA with cloned histone gene probes. Here we can make definite statements about the relative copy-numbers of genes, as the studies are examining directly the genomic DNA. However detection of variation between individual gene clusters and fine mapping of genes is more difficult than with cloned DNAs. The early study of van Dongen et al (1981) on X.laevis has subsequently been found to be severely flawed (van Dongen 1982, Turner and Woodland 1982b, Turner per com). Xenopus laevis has a complex pattern of divergent gene clusters which eludes rigorous mapping by Southern blotting of genomic DNA. What can be deduced is a partial gene order which is given in Fig 2.8, and applies to the most strongly-hybridising fragments in total genomic DNA restriction digests. The restriction map of this presumably major cluster type is heterogenous between and within individuals (Turner per com), and is often obfuscated by a number of 'minor' hybridising species whose total hybridisation is at least equal to the intensity of this 'major' gene cluster. It is notable, however, that the gene order in this partial map is consistent with that in many of the clones isolated from this species.

The situation in X.borealis is much simpler. The majority of the 90 histone gene sets in this species are in a single, relatively

Fig 2.8

Xenopus major histone gene clusters

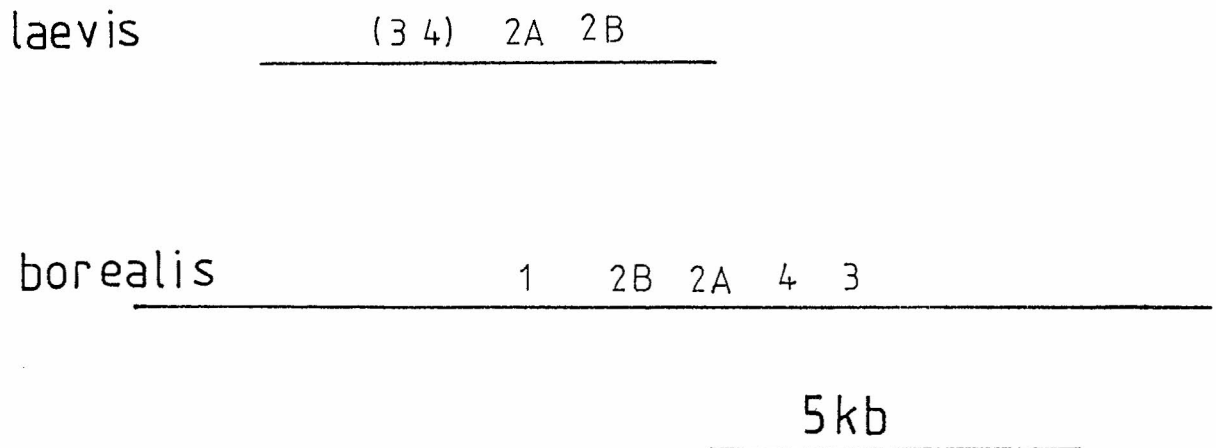


Fig 2.8 The Xenopus 'major' histone gene clusters: the structure of histone gene clusters deduced from probing Southern blots of restricted genomic DNAs with cloned histone gene probes. Note there is some uncertainty about the validity of the X. (laevis) result (see text). Xenopus laevis ('laevis') van Dongen et al 1981, van Dongen 1982. Xenopus borealis ('borealis') Turner and Woodland 1982b).

2. Genes

homogenous type of cluster which has been mapped by Turner and Woodland (1982b). The map of this cluster is presented in Fig 2.8 . There is a possibility that this is a tandemly repeating structure of repeat length 16kb (Turner and Woodland 1982b). X.borealis also contains two 'semi-minor' clusters (which might be the halves of a single cluster not shown to be linked in this experiment) present in 5-20 copies per haploid genome, and a number of minor clusters. Unlike those of X.laevis , the histone gene clusters of X.borealis vary little between individuals as judged by Southern blots of genomic DNA. In contrast to our expectation that a reiterated gene cluster be present in Xenopus to provide genes from which histone mRNA can be made during oogenesis, there is a possibility that the numerically major cluster type in Xenopus borealis does not code for the oocyte H3. A cDNA clone derived from the extreme 3' end of an ovarian H3 message does not hybridise well to the major cluster, instead hybridising preferentially to one of the semi-minor gene clusters present in 5-20 copies (Turner and Woodland 1982b). In the absence of a clone of the major cluster, it is not possible to say whether it codes for oocyte H1, H2A or H2B proteins, which would be more useful as indicators of whether the major cluster is active during oogenesis or not.

2.3.6 Cluster variation and conservation.

These various but related patterns of genes conjure up several interpretations of the function and the evolution of the genes represented in the Amphibiann histone gene clones. However as some of them have been prompted by the findings of this study, I will leave them until the Discussion section 17.2. Here one interesting point can be made by examination of the mass of data accumulated so far. In most cases where H2A and H2B are present on a clone together

2. Genes

they are adjacent and, if transcriptional polarity has been determined, divergently transcribed. This point is illustrated in Fig 2.9 . While statistics are not conclusive of significance in themselves, it is notable that this correlation is seen in all species examined except the sea urchin: even yeast follows this rule. (Wallis, Hereford and Grunstein 1980) . For the vertebrates, one may calculate the probability that, if we took 14 different vertebrate histone gene clones (each containing on average one H2A, one H2B and two other genes randomly arranged) the probability that we find the H2A and H2B would be adjacent. (I assume here, and for reasons stated above this may be invalid, that similar or overlapping clones are representatives of just one gene cluster type.) For each clone the chance is 50% : for 11 or more out of 14 it is 2%. The probability that 4 clones of the five in which transcriptional polarity has been determined should have H2A and H2B in divergent orientation is 1.2% . Thus we can say with reasonable confidence that, although the order of genes in vertebrate histone gene clusters appears random, it actually is not.

The histone gene clusters of X.laevis, and to a lesser extent of X.borealis, are quite variable between individuals. with respect to sequence and length of intergenic spacer as revealed by restriction mapping. Of the organisms examined, only the Xenopinae show this variation. The Notophthalmus tandemly repeated genes are slightly heterogenous in the intergene spacer, and the length of the satellite spacer between clusters varies widely (Stephenson, Erba and Gall 1981b). Allelic and non-allelic differences have been seen between sea urchin histone gene clusters as we noted above (Kedes et al 1975, Cohn and Kedes 1979, Overton and Weinberg 1978), and within inbred Drosophila (Lifton et al 1977), but appear restrained to a few specific variations on a basic pattern in each

Fig 2.9

H2A and H2B genes

	H2A/H2B	mean	adjacent	divergent
sea urchin	7	5	0	0
<u>Drosophila</u>	3	5	3	1
Chick	3	4	2	0.5
Man	3	3.6	2	nd
Mouse	1	5	0	nd
Yeast	2	2	2	1
<u>Notophthalmus</u>	1	5	1	1
<u>X.laevis</u>	4	4	4	nd
<u>X.borealis</u>	2	5	2	1
Total	16			

Fig 2.9 Divergent transcription of H2A and H2B. The H2A and H2B genes of several cloned clusters are summarised. Columns: 1) Species 2) (H2A/H2B) number of clones on which H2A or H2B and adjacent genes on both sides have been found. 3) (Mean genes) mean number of genes on those clusters 4) (Adjacent) number of clones in which H2A and H2B are adjacent. 5) (Divergent) fraction of those clones where transcriptional polarities have been assigned where H2A and H2B are found to be divergently transcribed. Refs as for Figs 2.2, 2.4 and 2.7

species.

However the histone-gene containing restriction fragments of X.laevis genomic DNA, identified in Southern blots by hybridisation to cloned probes, rarely occur in all individuals in equal numbers (van Dongen et al 1981, van Dongen 1982, Turner and Woodland 1982b). A minimum of about 30 restriction fragments are required in the Xenopus population to explain the observed diversity of bands obtained in several individual frogs, and this number probably represents the limit of resolution of the methods involved as much as that of the diversity of genes. In all likelihood there are at least 40 different gene clusters in X.laevis distinguished by this method. Only a few of these will be variants of the major repeat presumed to be required to code for oocyte histone. This variation may represent true heterozygosity at histone loci or non-allelic variation between different populations. It is notable that X.laevis and X.borealis are evolutionary tetraploids. It is possible that this fluidity of genome organisation is a result of the redundancy of, and hence reduced conservative selection pressure on, the doubled-up gene copies generated by tetraploidisation, (Widmer et al 1981, Walli et al 1979, May, Weber and Westley 1982, Bozzoni et al 1981). It is notable that, of the nine histone gene clusters isolated by R.W.Old (per com), four (Xlh24 & Xlh25, and Xlh11 & Xlh23) are pairs of clusters with very similar gene and restriction maps, and another three (Xlh8, Xlh2 and Xlh202; see also section 13) are almost identical to each other. Thus this 'doubling up' also appears in the histone gene family in Xenopus. Skirbinski and Ward (1981,1982) note a positive correlation between heterozygosity and the rate of protein evolution at a number of other loci in various species, again probably resulting from selection pressure acting on

2. Genes

'redundant' loci or alleles. Minor restriction site differences could be responsible for much of the variation seen between X.laevis individuals in the Southern blot experiments on genomic DNA.

The genetic variability in X.borealis' histone genes is less obvious than that in X.laevis (Turner and Woodland 1982b, Turner per com). The majority of the histone genes are in the conserved major cluster, which, with the exception of one specific insertion of 200bp at one end of the mapped region, does not seem to vary significantly between individuals. Those minor genes not part of this structure or of the equally conserved 'semi-minor' clusters are likely to be present in only 1 or 2 copies per haploid set. The variation within these clusters (while hard to detect due to the problem of detecting such a minor species in a Southern blot) appears to be small (Turner per com). If these minor genes are present in only 1 copy/genome, unlike analogous X.laevis clusters, the selective forces on them, and the consequent suppression of variation in them, would be correspondingly enhanced.

2.3.7 Cloned vertebrate histone genes : summary.

In summary, the histone genes fall into three classes;

a) 'normal' histone genes possessing no introns or polyadenylation signal and having a 3' homology block,

b) reiterated histone genes coding for embryonic or oogenic stored mRNAs and proteins,

c) divergent histone genes coding for variant histone proteins. At least some of class b) clearly fall into class a), and we shall see later that many of the 'normal' histone genes studied to date are actually likely to be members of class b). The consensus of what a 'normal' histone gene is like has probably been coloured by this, and future studies on genes coding variant histones might show that this 'normal' pattern is actually an exceptional feature of oocytic or embryonic reiterated histone genes.

The embryonic genes are usually clustered in a tandem repeat, which often is present as variants which have accumulated different spacer mutations. The non-embryonic genes are also clustered, the case of the chick suggesting a cluster occupying not more than 0.05% of the genome.

Gene numbers would appear to be constant between species at 10-30, these probably coding for the spectrum of variant and variably controlled histone proteins we noted in section 1.6, plus however many copies of a conserved, repeated gene cluster are

2. Genes

needed for embryogenesis or oogenesis. In Triturus and sea urchin this reiteration is even greater than the demands of embryogenesis apparently require.

This gives little information other than simply to describe gene structure. This is necessarily so; the expression of histone genes has been extensively studied in the oocyte and early embryo of Xenopus and sea urchin in terms of both RNA and protein synthesis, but which genes are active is unknown except in sea urchin, and no other tissues or developmental stages have been rigorously analysed. The complex area of histone gene expression is not the target of this thesis, although the cloned genes described here are key components of such studies as S1 nuclease mapping, cDNA primer extension and hybridisation studies, and so their isolation and characterisation are useful first steps to studies of gene expression.

In the majority of the histone gene clusters analysed, H2A and H2B are adjacent and divergently transcribed.

2.4 Chromosomal location of histone genes.

The observed clustering of some cloned histone genes suggests that all histone genes might be clustered into only a few, maybe only one, gene cluster, which will effectively be a point locus on a chromosomal scale. The divergence of histone genes noted above sheds doubt on whether this would be true of H5 or of other highly variant histone genes, and, even if they were located with other histone genes, whether in situ hybridisation procedures would detect them. Given the limits of the available probes, however, histone genes have been found to be clustered at a few loci in the four species

2. Genes

examined.

In two newts, Triturus cristatus with a C-value of 23pg and Notophthalmus viridiscens with 45pg, the lampbrush chromosomes of the oocyte are sufficiently large to allow in situ hybridisation to locate a gene to within 1% of the length of a chromosome. Old, Callan and Gross (1977) found 8 or more loci transcribing histone genes in Triturus on four chromosomes. However it is still unclear how many of these loci represent true histone gene clusters, and how many spurious hybridisation of the probe used to alternating copolymers in satellite DNA (Callan and Old 1980) known to be transcribed in newts (Diaz et al 1981). In Notophthalmus histone genes are present at only two loci, which have the genes fairly evenly divided between them (Gall et al 1981). The satellite sequences associated with the clustered histone genes are also found elsewhere, notably at the centromeres (Diaz et al 1981).

In man histone sequences have been located by in situ hybridisation to the distal tip of chromosome 7 (Yu et al 1977, Chandler et al 1979, Hentschel and Birnsteil 1981). With the proviso about sensitivity noted above, this would suggest that a significant proportion, perhaps the majority, of human histone genes are in one cluster.

In Drosophila histone genes show similar clustering, suggesting that most if not all the genes are at one locus (Pardue et al 1977a,b). However that locus, which covers several chromomeres, is not a continuous tandem array of histone genes (Lifton et al 1977).

2.5 Histone genes as a multigene family.

2. Genes

In this last section on histone genes I wish to consider them against the background of what we know of other multigene families. If, as the chicken gene model suggests, histone genes are not a class of intermediate repetitive, satellite-like DNAs which happen to code for protein, but instead are a diverged family of genes some of whose members are reiterated to meet a need for high levels of expression, then we should find illuminating parallels between them and other multigene families. The existence of this section makes it unsurprising that such parallels exist.

2.5.1 Gene clustering.

A wide range of eukaryotic genes have been studied by recombinant DNA techniques, and the majority are members of multigene families (see also Discussion section 17.3.1). The most reiterated transcribed genes in the Xenopus genome are ribosomal RNA genes (but see below as well). There are 24,000 5SRNA genes in the Xenopus genome present in tandemly repeated clusters of fairly uniform structure (Ford 1978). Sequence determination shows these to be oocyte 5SrDNA rather than somatic type, so we can class them with the Notophthalmus and sea urchin highly reiterated histone genes as being reiterated to accomodate the extraordinary transcriptional loads of oogenesis rather than with the histone genes of other species. They are also transcribed by a quite different route to protein-coding genes, using PolIII and a specific activator (Engelke et al 1980, Kay, Schmidt and Gall 1981), as are tRNA genes (Galli, Hofstetter and Birnstiel 1981). Ribosomal 18+28S genes in Xenopus are also tandemly repeated and separated by non-transcribed spacer (Wellauer et al 1976, Boseley et al 1979). It is clear therefore that these genes are models for embryonic histone genes in their evolutionary response to transcriptional demands, but in no other

way.

It is likely that the extensively reiterated heat-shock protein genes of Drosophila are also duplicated to increase their transcriptional capacity, and are probably better models for embryonic histone genes in this respect. They are present in densely packed clusters of duplicated genes whose transcription is synchronously initiated on heat shock to synthesise very large amounts of mRNA (Compton and McCarthy 1978, Mirault et al 1979, Ingolia and McCarthy 1980, Mason et al 1982). The chorion protein locus of the silk moth is another densely populated locus whose reiteration is probably an ancient evolutionary response to transcriptional demands (Rodakis, Moschonas and Kafatos 1982, Eichbach and Kafatos 1982).

The most studied gene family after the histones (or maybe even before them) is the globin family, and they are illustrative of several features found in other multigene families and relevant to histone genes. The beta globin cluster in mammals is particularly relevant (Efstratiadis et al 1980). The cluster consists of four or five beta-like genes with an essentially identical pattern of introns in their transcribed regions closely linked in 40 - 60kb of DNA, sometimes with one or two pseudogenes (the exact pattern depends on the species). The spacing of five 2kb genes in 50-odd kb of genome (in man) is similar to the chick histone spacing of nine 600-800bp genes in 40kb, a ratio of 0.2 and 0.14 transcribed/non-transcribed respectively. This tight clustering is a feature of other genes: human interferon (Ullrich et al 1982), Drosophila heat shock (Ingolia, Craig and McCarthy 1980, Mason et al 1982), chick ovalbumin (Royal et al) and many other gene families show clusters of related genes separated by one kb to tens of kb of

non-coding DNA. (I document this point more fully in Discussion section 17.3.1).

2.5.2 Intra-cluster variation and inter-species conservation.

Two other features link globin and histone gene clusters. First is the interspersion of intermediate repetitive elements in both globin (Coggins et al 1980, van Dommelen et al 1980) and histone (Krieg per com, van Dongen 1982, Lifton et al 1977 and this work) gene clusters. Rather than a satellite or copia-like element these are short 'Alu-like' sequences, or their presumed avian or amphibian equivalent. This is a common feature in protein-coding multigene families but not of other reiterated sequences such as rDNA or satellites, which have their own distinctive repetitive DNA patterns.

Secondly, despite the presence of these repetitive elements, which are usually associated with genome plasticity (see section 2.7 below), the globin locus gene clusters of primates (Jeffreys et al 1982) and anurans (Patient et al 1980) are highly conserved, both in the gross structure and sequence (of both the coding and the non-coding and non-transcribed regions) between equivalent positions in different species. By contrast, equivalent gene segments within a species are quite divergent: the alpha-1 gene 3' end in man is easily distinguishable from the alpha-2 gene 3' end, despite the genes' identical coding capacity, close genetic linkage and recent evolutionary divergence (Michelson and Orkin 1980). The same pattern of inter-species conservation coupled with intra-cluster variation has been seen in interferon (Owerbach et al 1981, Ullrich et al 1982) genes in mammals, heat shock genes in

2. Genes

Drosophila (Mason et al 1982) and elsewhere (see section 17.3.1). This evolutionary conservation is moderated by slight polymorphism of DNA sequence, as detected by restriction site differences, caused by single base-changes rather than by larger substitutions in globin (Jeffreys 1979, Efstratiadis et al 1980, Orkin et al 1982) and actin (Engel, Gunning and Kedes 1982b) gene families. Variation in the globin locus may be contrasted to the large-scale deletions found in the human population which invariably degrade the function of the whole beta-globin cluster (Fritsch, Lawn and Maniatis 1979, Weatherall and Clegg 1982) as do some smaller lesions (Bernards and Flavell 1980), despite occurring in the same DNA as point mutations which have no visible phenotypic effect (Jeffreys 1979, Orkin et al 1982). Intra-cluster variation of duplicate genes has also been found in the chick ovalbumin (Royal et al 1979, Hughes et al 1979), Drosophila heat shock (Ingolia, Criag and McCarthy 1980) and tubulin (Mische and Pardue 1982) genes, and mouse alpha-amylase (Schibler et al 1982) and H2 (Steinmetz et al 1982) loci; this last contains a complex pattern of duplication and reduplication with overlying divergence very reminiscent of that found in many mammalian globin clusters. If this is normality for eukaryote multigene families, then the histones are normal indeed: they are clustered, divergent within a cluster so that, say, H3 genes can be quite different within an organism, but conserved between organisms, for example between different sea urchins and between X.laevis and X.borealis (see also Discussion section 17.2).

We may look ahead a little to contrast this pattern with that seen in other genome components. Ribosomal RNA genes are conserved between species, but also within species to at least the same extent. The intermediate repetitive DNA elements discussed below generally show the reverse pattern to multigene families: they are

2. Genes

conserved within a species but divergent between them (Gillespie, Pequinot and Strayer 1981, Schmid and Jelinek 1982). (This is not true of all such elements - some are highly conserved between species as well (Jelinek et al 1980)). Thus this feature of the families of DNA sequences we call multigene families must be an effect of their function, and not a concomitant of their duplication or their distribution in the genome.

An interestingly predictive side-light of this is that, in several of the gene clusters investigated, the variant cluster members are not under strict coordinate control but rather are expressed at different times in their specific tissues' history . This is true of mammalian globins (Efstradiatis et al 1980), chick ovalbumin (Heilig, Muraskowsky and Mandel 1982) and mouse H2 (Klein 1978) gene clusters. Analagous data are not available for the histone gene clusters with the exception of the Drosophila and sea urchin embryonic gene clusters (Lifton et al 1977, Heiter et al 1979, Childs, Maxson and Kedes 1979, Kedes 1979). However if histone genes resemble other gene families in this respect too, we should not expect closely linked members of gene clusters necessarily to be expressed at the same time as each other just because they are linked.

2.6 Histone gene clusters - summary.

Contrary to previous assumptions it is now clear that histone gene clusters are essentially similar to other gene clusters in eukaryotes in many features. . As a result we can draw useful analogies between histone and other gene clusters to suggest how histone genes might be distributed, might function and might have evolved. The first of these points is also directly eluctable by gene

2. Genes

cloning. The second has revealed such a diversity of controls and patterns of expression that there is as yet no unifying data base on which to found deduction or speculation. The third I will take up again in the Discussion.

2.7 Repetitive DNA and genome plasticity.

The genome of eukaryotes does not consist solely of single-copy, protein - coding DNA. Early studies showed that 'unique' DNA , defined as being present at less than five copies per haploid set, was interspersed with 'repetitive' DNA (Davidson et al 1973). More recent work has shown that these sequences are indeed interspersed among functional genes in gene clusters, inter alia among the histone genes. I will therefore review this complex area very briefly, with no regard for historicity, mentioning only the areas of information on the structure and role of repetitive DNA which are relevant to our later discussion of histone genes.

2.7.1 C-value and highly repeated DNA.

Eukaryotic cells contain a large quantity of nuclear DNA - up to 10,000 Mb per haploid genome in some amphibia (Lewin 1980). There is no evidence to suggest that they have a correspondingly enormous number of 'genes' in the prokaryotic sense that the whole genome is composed of contiguous lengths of DNA a few kb long each coding for a protein. The question "What is the DNA for?" is therefore a vital one to a comprehensive theory of molecular genetics, and has prompted extensive study of the structure and genetic activity of this overgenerous complement of DNA. From this a number of points have escaped, the most striking being that the genome of animals is much simpler in terms of sequence complexity than its size would

2. Genes

indicate, and that much of it is composed of simple sequences which are reiterated many times to make up the observed bulk of DNA (reviewed by Lewin 1980).

Highly repeated DNAs (the name is arbitrary: 100,000 copies per genome is highly repeated, 1000 copies is just repeated) are present in two forms. Many animals are found to have satellite DNAs, sequences which band apart from the bulk of the DNA on various CsCl density gradients. They are simple, highly tandemly repeated sequences (Endow, Polan and Gall 1975, Southern 1975, Biro et al 1975, Hsieh and Brutlag 1979, Calson and Brutlag 1979, Plucienniczak, Skowronski and Jaworski 1982), often present at chromosomal centromeres (Rosenberg, Singer and Rosenberg 1978, Diaz et al 1981). They may be detected in genomic DNA as very long lengths of DNA uncut by certain restriction enzymes which, failing to cut the unit repeat, will be unable to cut the entire array (Botchan, McKenna and Sharp 1973). This is a quick diagnostic for such tandem arrays, and as a result the definition of a satellite has moved away from the CsCl banding and towards a definition based on extensive tandem repetition of a short, simple sequence. In this sense the sea urchin embryonic histone genes, although present as a CsCl satellite (Kedes and Birnstiel 1971, Portmann, Schaffner and Birnstiel 1976), are not a satellite DNA. This also excludes rRNA genes; although tandemly repeated their sequence is not simple (Brown and Weber 1968, Ford 1978).

Satellite DNAs are not usually transcribed (although homologous sequences outside the tandem array may be), and appear to have no coding function. Smith (1976) has pointed out that any repeated heterogenous cluster of short sequences could tend to homogenise and amplify its unit sequence by unequal crossover, in the absence of

2. Genes

constraints on its sequence imposed by informational requirements forming a satellite-like array, Dover (cited in Robertson 1981) suggests that gene correction could have a similar effect on non-adjacent sequences. Whether this is the real method of generation and maintenance of satellite DNAs, which therefore have no function, or whether they do have a non-coding function is not known. The association of such sequences with chromosomal centromeres is suggestive of a role in chromosomal or nuclear structure.

Satellite DNAs have been found in a number of species. Their distribution cannot account for the variation in C-value between species. It is notable that the histone genes of Notophthalmus are embedded in satellite DNA (Stephenson, Erba and Gall 1981b), where they must be able to resist any sequence homogenisation driving them to uniformity with the satellite around them. The histones are not located on the centromere, although the satellite is found there as well. This suggests that the satellite has a different function in the histone gene clusters of this species than it has at the centromeres. What this might be is discussed by Cavalier-Smith (1978) among others.

2.7.2 Dispersed repetitive elements.

The majority of highly repeated sequences in most eukaryotes are not present in tandemly repeated satellites but as small blocks of sequence interspersed with less repeated DNAs (Davidson et al 1973). The pattern of interspersion derived for Drosophila (Manning, Schmid and Davidson 1975) is quite different from that of a wide range of other species (Davidson et al 1973, Graham et al 1974, Firtel and Kindle 1975, Schmid and Deininger 1975, Angerer, Davidson

2. Genes

and Britten 1975), but recent studies have indicated that this apparent division may be an artefact (Scheller et al 1981, Krumlauf, Jeanpierre and Young 1982, Moysis et al 1982a,b). This interpretation, which is in fact inherent in earlier work, shows that mammalian families of repetitive sequences intersperse unique DNA once every few 100s to 1000s of bp as previously deduced, but that this is not the 'tandem repeat' length. Different families of repeats have members interspersed in the same region of DNA, so that unique DNA can be flanked and interrupted by single members of a variety of gene families. This association is random with respect to the repetitive sequence and the length of unique DNA between repetitive elements. Consequently one could have to 'walk' tens of kb along the genome from one member of a specific family of repeated DNA elements, coming across several members of other families, before finding another member of the first family from which you started. This scrambled interspersion pattern is hidden in reassociation experiments which are performed at low hybridisation stringencies, as some families can cross-react with others, yielding an unrealistic measure of the inter-member distance. This cross-reaction can be reduced either by increasing the stringency or by choosing an organism in which the sequence divergence between families of repetitive elements is large. The former was performed by Moysis et al (1982a,b), the latter was found in Drosophila (Manning, Schmid and Davidson 1975). The suggestion is, therefore, that Drosophila and vertebrate interspersion patterns are essentially similar, differing only in the degree to which different families of repeated sequences differ from each other.

This pattern of scrambled families of interspersed, highly repeated DNAs has been confirmed by two routes. Several members of highly repeated, dispersed families of sequences have been cloned

2. Genes

and characterised (Houck, Rinehart and Schmid 1979, Jelinek et al 1980, Spohr, Reith and Sures 1981, Meunier-Rotival 1982, Singer 1982). The Alu family in man is archetypical (Houck, Rinehart and Schmid 1979, Jelinek et al 1980, Schmid and Jelinek 1982). It is a family of 300bp units cut once by AluI (hence the name) present in at least 100,000 copies in the human genome, and distributed over at least 60% of it (Houck, Rinehart and Schmid 1979). Alu sequences are usually flanked by short direct repeats in the genome. They are related to the b1 sequences in mouse, which are similarly dispersed in the murine genome, and show a 30bp homology with many mammalian repetitive DNAs, part of which is also homologous to some small nuclear RNAs, and to the replication origin of Papovaviruses (Jelinek et al 1980, Schmid and Jelinek 1982).

Similarly repeated sequences from other organisms show different structures. A highly repeated element from Xenopus laevis consists of 8 repeats of a 77bp element, itself showing considerable conservation of sequence within and between elements (Spohr, Reith and Sures 1981). This sequence is present in ~100,000 copies per genome. All these short interspersed sequences are heterogenous within a genome, so the number of elements any hybridisation experiment detects will depend on the stringency demanded of the hybridisation.

Longer sequences have also been cloned. The human Kpn family is a repeat of 6.4kb, not obviously flanked by repeated sequences (Singer 1982). Analogous sequences have been found in mice (Meunier-Rotival 1982). The murine sequence, present in 30,000 copies, is not distributed uniformly in the genome, unlike unique genes, suggesting both a different location to, and hence a lack of functional correlation with, true 'genes',

2. Genes

That these sequences are interspersed with each other and with unique, coding DNA has been shown in a number of cloning experiments where clones selected for their content of unique DNA, and hence taken essentially at random as far as their repetitive sequences are concerned, contain a variety of repeated sequences intermixed with each other and with the unique DNA in complex patterns. Such situations are known from the human globin (Coggins et al 1980, van Dommelen et al 1980), interferon (Ullrich et al 1982) and insulin (Ullrich et al 1982b) genes, vitellogenin (Ryffel et al 1981), histone (van Dongen et al 1982 and this work) genes and a cluster of solely repetitive elements (Spohr, Reita and Sures 1981) from Xenopus, murine alpha-amylase (Schibler et al 1982), avian keratin (Kemp 1975) and chick lysozyme (Baldacci et al 1981) inter alia. In this last case a complex series of overlapping duplications was claimed spanning the gene itself, although whether these represent duplications or simply the proximity of members of closely related families of sequences is unclear.

2.7.3 Function of Intermediate repetitive DNA.

Intermediate repetitive DNA makes up 50% or more of most vertebrate genomes, and up to 75% of some larger amphibian genomes (Lewin 1980). Yet it does not code for recognisable protein, and although some of it is transcribed the function of such transcripts is highly controversial. The convolutions of the arguments for and against various theories about Intermediate repetitive DNA are too much for this thesis. In summary:

2.7.3.1 They are regulatory elements.

This, the oldest and still one of the most widely held

theories, suggests that they are control elements analagous to promoter and operator elements in prokaryotes, although operating through different mechanisms (Britten and Davidson 1969, Davidson and Britten 1979). Finding intermediate repetitive DNA both transcribed on mRNA and hydrogen bonded to it has lead support to this theory (Davidson and Posakony 1982), although other interpretations of these findings are possible. Further elucidation of the interspersion pattern of IR sequences has not aided this model, in which regard it is notable that about 20% of some eukaryotic DNA is not within 3kb of a repetitive element (Krumlauf, Jeanpierre and Young 1982). However as we do not know the mechanism of control of eukaryotic gene expression, this theory is not contradicted by fact.

2.7.3.2 They are structural elements.

This maintains the role of these sequences in gene expression, but removes sequence and position constraints from their evolution. In particular the model of Cavalier-Smith (1978) proposes that selection operates on the nuclear surface area through the need for rapid transport of RNA from the nucleus, and hence on the mass of DNA which forms a 'skeleton' determining the nucleus' size. This theory has several logical holes in it, but is attractive anyway in that it ties in with the empirical control models invoking a 'domain' structure in the nucleoskeleton as a unit of transcription and replication (McCready et al 1980, Jackson, McCready and Cook 1981, Pardoll, Vogelstein and Coffey 1980, Buongionno et al 1982). (see also section 17.3.3)

2.7.3.3 They are Selfish DNA.

2. Genes

This, the modal group of theories, suggests that IR DNA has no phenotypic function, although not necessarily no effect, instead acting 'selfishly' by containing replication origins and/or transposition mechanisms (vide infra) which allow its spread through and perpetuation in the genome (Doolittle and Sapienza 1980, Orgel and Crick 1980, Robertson 1981). While such a position is tenable for relatively rare elements, it makes two strong predictions about more common ones which are contradicted by fact. If these elements are selfish, why does the genome not expand due to their continued replication until all organisms had C-values of 10,000 Mb or more? If this is selected against by a mechanism like that proposed by Cavalier-Smith (1978), competition between the selfish sequences would result in their attaining an 'optimum' sequence instead of diverging in different species, whereas in fact the sequences vary drastically between organisms (vide supra), and in primates different members of related families have been amplified to different extents, suggesting that their sequence and their copy-number are only weakly connected (Gillespie, Pequignot and Strayer 1980). Truly selfish transposable DNA would also result in a genome far more fluid than is observed in any animal. However the Selfish DNA hypothesis does explain why these sequences are so widespread, apparently capable of self-transposition and arranged with no apparent regard for their genomic surroundings.

2.7.3.4 They are accidentally amplified 'junk'.

As I mentioned above, random unequal crossover is believed to be able to cause sequences to converge on a tandemly repeated structure like a satellite DNA (Smith 1976). Analogous mechanisms involving gene conversion have been suggested as having the same effect on dispersed sequences not in tandem repeats (cited in

2. Genes

Robertson 1981). These would be entirely random in sequence, and position, and would in effect be 'spinoffs' of the normal mechanisms of recombination and error-correction. Gene conversion causing co-evolution of related genes has been suggested in immunoglobulin (Baltimore 1981) and Drosophila rRNA (Coen, Thoday and Dover 1982) genes, and Chakravarti and Beutow (1982) estimate that the rate of base 'correction' in the primate globin locus is 40 times that of mutation due to base substitution. However studies on Xenopus 5S and 18+28S rRNA genes suggest that unequal crossover is of at least as much importance here (Glover 1976). The rapid divergence of homologous repeated families of DNAs on the X and Y chromosomes of man, which do not cross over, also suggests a gene conversion/unequal crossover mechanism for the homogeneity of other sequences (Cooke and McKay 1978). The idea of 'concerted evolution' in an attractive one in that it explains the sequence divergence between species with conservation within a species (Dover, cited in Robertson 1981, Chakravarti and Beutow 1982), but one wonders why such 'spinoffs' are generated to such a variable extent by presumably identical recombination and correction mechanisms in closely related species which are observed to have very different C-values.

These cover all the possibilities of sequence selection or non-selection and function or non-function. None are contradicted directly by any experimental data, but much circumstantial evidence opposes all four classes of theory. One possible reason for this is mentioned in the Discussion (section 17.3.4). This debate is kept alive by the above - mentioned close association of functional genes and repetitive elements, and the other facet of IR DNA, the transposing and gene-moving activities of some repetitive elements.

2.7.4 RATS and the fluid genome.

Eukaryotic transposable elements have been known in maize for some time, and more recently have been identified in Drosophila and possibly some other Diptera (Calos and Miller 1980, Green 1980). They were initially detected genetically by their effect on mutation and cross-over rates, but have since been identified biochemically by cloning their DNA and identifying its transcripts and insertion sites. No transposable elements have been detected by genetic studies in vertebrates, although DNA structures which are biochemically related to the Drosophila transposable elements have been found through other routes.

However, many vertebrates (man is an almost unique exception) are subject to infection by Retroviruses, RNA viruses whose replicative cycle is through a double-stranded DNA circle copied by Reverse Transcriptase from the RNA genome. These DNAs integrate into the host cell chromosome, and may be stably transmitted to progeny cells (Dunsberg et al 1979, Hughes et al 1979). They may excise, and in doing so occasionally excise cellular sequences with them forming transducing viruses (Bishop 1981, Rigby 1982). These may re-integrate into other cells where, if suitable host genes have been acquired by the virus, the viral genome's expression will lead to transformation of the cells (Donoghue et al 1979, Rigby 1982). It is this transforming property that has led to much of the research into retroviruses.

In the past two years it has become increasingly apparent that retroviruses are just transposable elements that carry genes for

2. Genes

viral coat proteins - or conversely that Drosophila transposable elements are endogenous retroviruses.

(This is not a purely semantic difference. An endogenous origin suggests a host function for these sequences, whether past or present. Exogenous origin postulates that the original pro-retrovirus need have had no function in the hosts presently afflicted by retroviruses and transposable elements. Instead these elements evolved integration and transposition mechanisms as a method of replication much as phage Mu has in E.coli (Couturier 1976). This latter view is popularly supported by the idea that an 'ideal' parasite is one which does its host no harm at all, but peacefully coexists with it, and thus that retroviruses are ideal parasites. This argument, while true in some contexts, cannot be applied to all cases (May 1982). It is clear that a virus which is capable of producing more than one infected individual per host generation would be at a selective advantage over one whose replication was tied to that of their host, and that therefore retroviruses are hardly optimal as parasites from their own point of view. Rather they resemble unruly endogenous elements. Temin (1980) also favours an endogenous origin for retroviruses, and transposable sequences in general, for other reasons concerning the biochemistry of their transposition.)

The copia and 412 families of Drosophila transposable elements (TEs) have been most studied at the DNA level. There are about 20 families like these per Drosophila genome, making up as much as 10% of the fly's DNA (Muræbekov 1981). They are 5-6kb long and bounded by direct repeats a few 100bp long, themselves bounded by short inverted repeats. A short duplication of 'host', IE pre-transposition, DNA brackets the integrated unit. They may excise

2. Genes

themselves from the genome to form circles of two distinct size classes (Stanfield and Lengyel 1980). In these respects they much resemble prokaryotic transposons (Calos and Miller 1980). Recent evidence suggests that they are transcribed into full-length RNA copies, and that these are present in cells as RNase-resistant RNP particles similar in density to viral particles (Flavell and Ish-Horowitz 1982 unpub).

Retroviral DNA genomes are also present as a twin-sized circular population, having essentially the same integrated structure as copia and 412, are of similar size to full-length transcripts of copia, and of course are found in full-length copies in viral particles (Finnegan 1981). This similarity of biochemistry is backed by comparative sequence studies of Copia, 412 and retroviral DNA (Will, Bayev and Finnegan 1981) which show strong sequence homologies between their Long Terminal Repeats.

The properties of these elements (Retroviral and Transposable Sequences : RATS) have implications for the eukaryotic genome structure and evolution in general and this study in particular: effects on intra- and inter-organism gene transfer, on genome fluidity and on the origin and nature of shorter intermediate repetitive elements. It has been shown that RATS have been exchanged between primates and cats in the fairly recent past (Weinberg 1981); in the latter species primate-related retroviruses have been found in the genomic DNA. A similar conclusion has been arrived at with regards to chick endogenous retroviruses (Hughes et al 1979), so germline insertion of retroviruses could be quite common on an evolutionary timescale. It has been demonstrated to occur spontaneously in mice (Rowe and Kozak 1980). There is no proof that a functional gene has been successfully transferred between higher

2. Genes

eukaryotes by such a process, but this might be extremely difficult to detect. A number of possible candidate events are discussed in Section 17.2.5. Certainly the use of synthetic retroviral vectors points to the physical possibility of this idea (Gluzman 1982, Tabin et al 1982). In principle this could lead to genetic homogenisation within a species or introduction of novel genes from one species to another and hence, in a limited sense, to genomic homogenisation across species boundaries.

Copia is present in the Drosophila genome in @20 copies - each of 20 or so other RATS families are also present at a similar level (Mirzabekov 1981), so 5 - 10% of the genome of Drosophila is recognisable as RATS. They are therefore a subclass of intermediate repetitive DNA. To have such a load of transposable elements implies that they would play a major part in mutation both by insertional inactivation of genes and by mediating deletions, inversions and transpositions of host DNA (Calos and Miller 1980). This is borne out by the observed effects of the transposable elements causing hybrid dysgenesis (Berg 1982). In this system mobilization of a transposable element present in only @20 copies per genome causes an extremely high rate of mutation due to its insertion into expressed DNA. This burst of mutagenic activity is explicable in terms of a RATS-coded repressor for the 'transposase' which accumulates in the oocyte over time (Thompson and Woodruff 1981, Berg 1982). The failure of both this and the better studied RATS to break up the Drosophila genome in less restricted circumstances is not explained by this model, however, given the large background of RATS in the fly's DNA (Hickey 1982). Observations that the hybrid dysgenesis RATS is capable of chromosome fragmentation (Thompson and Woodruff 1981) suggests that either other factors are limiting the effect of these elements or that they are not as 'selfish' as we

2. Genes

think. Foldback DNA elements in Drosophila are also known to be able to generate heterozygosity in adjacent DNA and to transpose large stretches of DNA between two elements (unpub, cited in Potter 1982). Their structure is different from Copia, notably their terminal repeats are almost satellite-like in their internal repetition, but their transposing properties and mechanisms are believed to be similar (Schmid, Manning and Davidson 1975, Potter 1982). They are therefore probably another class of RATS.

If RATS transpose through RNA intermediates as the retroviral example would suggest, it is possible that mis-incorporation of mRNA into viral particles or random reverse-transcription of mRNA would generate novel message-like DNA sequences, pseudogenes, which would subsequently be inserted into the genome distant from the gene which originally coded the message. Such sequences have been found (Nishioka, Leder and Leder 1980, Vanin et al 1980, Hollis et al 1981, Wilde, Crowther and Cowan 1982) complete with spliced-together exons, poly-A tail and, in the latter case, flanking direct repeats. In the case of the immunoglobulin 'processed' pseudogene found by Hollis et al (1982) we can be fairly certain that the mRNA from which this sequence was derived was packaged into a viral particle which was infectious. Immunoglobulin genes are not found in a functional form in the germ line. They are rearranged in lymphoid stem cells to generate the final, functional Ig-coding gene (Seidman et al 1978, Brack et al 1978, Bernard, Nobumichi and Tonegawa 1978, Rabbits and Forster 1978). Therefore the pseudogene must be derived from somatic mRNA, but is now present in the germ-line DNA, bespeaking the transfer from one to the other by some agent able to leave and re-enter cells. It is not impossible that functional genes could be generated in this way: the histone and beta-interferon gene families, both almost entirely without

introns, spring to mind.

Lastly there is still some doubt about just how much of the genome is related to RATS. Endogenous RATS must be lost from the germline to counter the continuous observed influx from 'viral' particles (Hughes et al 1979, Rowe and Kozak 1980), but the mechanism by which this occurs is unclear (Weinberg 1981). Genetic drift seems the most likely - the number of RATS-related sequences detected in mammalian genomes far outweighs the numbers of actual RATS recoverable by induction or infection with helper virus (in man's case no virus is recoverable this way). Keshnet et al (1980) cloned the DNAs of a number of RATS-like sequences coding large RNAs and found a range of sequence divergence between them suggesting that they were the products of random drift from an originator sequence.

This reappraisal of endogenous retroviruses, copia-like elements and similar sequences has encouraged a second look at other intermediate repetitive DNA of greater repetition frequency. In particular the Alu sequences, flanked by short direct repeats of 'host' DNA (IE DNA sequence not shared by all Alu repeats) were looked at as possible eukaryotic insertion elements (Schmid and Jelinek 1982). The similarity between the two classes of sequence is by no means striking, but some facets are notable. Alu sequences are flanked by short direct repeats of non-Alu DNA (Jelinek et al 1980), and are found in human cells in circular DNA which lack these repeats (Krolewski et al 1982). Polydisperse circular DNA (1.5 - 3kb long) from green monkey cells has also been shown to contain sequences related to two characterised IR families, Alu and Alpha, and also a third not known from a genomic clone (Bertelsen et al 1982). Integration to produce direct repeats and presence on

2. Genes

extrachromosomal circles are only two phenomena associated with prokaryotic insertion elements: the latter can also be shown to integrate, excise and to 'mobilise' the DNA between them as a transposon (Calos and Miller 1980). In this regard short intermediate repetitive DNA elements have been implicated in polymorphism in the human genome, where Alu-rich regions of the genome are found to be more polymorphic between individuals than Alu-poor regions (Calbretta et al 1982). The authors state that transposition via circular DNA is involved in this variation. Repetitive DNA is also involved in the variation in the polymorphic region near the human insulin gene (Bell, Selby and Rutter 1982), and has been implicated in a beta-globin locus deletion which ends in one of a pair of opposed Alu sequences (Kagadeeswaran et al 1982). As well as the known effects of Copia and similar elements in Drosophila, other middle-repetitive DNA elements have been associated with heightened polymorphism in adjacent sequences (Mirault et al 1979, Young 1979). Abundantly distributed pseudogenes from small nuclear RNAs of man are also flanked by inverted repeats (van Arsdell et al 1981): the reader will recall that these RNAs had some homology with viral replication origins and with Alu sequences (Jelinek et al 1980). We might also recall that one of the Histone H3 gene-related 'Orphans' of Childs et al (1981) is adjacent to DNA which is moderately repetitive in the genome and which contains a short inverted repeat hyphenated by 480bp of other sequence.

So we can say that Alu and similar short, intermediate repetitive DNA families show all the static characteristics of insertion sequences, and reveal that they might be involved in the dynamic aspects through gene cloning experiments on polymorphic loci. One possibility that ties these various lines of evidence together is that the short intermediate repetitive elements are

2. Genes

derived from RATS, possibly from the Long Terminal Repeats (Donoghue et al 1979) by analogy with prokaryotic transposons, these sequences having subsequently diverged from the original RATS family. Support for this view comes from the recent work of Streeck (1982), who has found two related intermediate repetitive families in cow with structural homology to retroviral long terminal repeats. Subsequent 'concerted evolution' (Robertson 1981, Cohen, Thoday and Dover 1982) of these scattered fragments would ensure that they evolved as families of sequences and not as separate elements. Excision and/or insertion as an insertion-like element would then depend on their chance similarity to the RATS from which they derived. This is a very tentative hypothesis, of course, supported largely by a few features of the sequences concerned, their known association with and possible role in genome alteration and the lack of any other role for them.

The finding that intermediate repetitive DNAs in general (Davidson and Posakony 1982) and the Alu family in particular (Ullu and Melli 1982) are transcribed does not contradict this model: 'integration' of these sequences next to a functional, transcribed gene could lead to the sequences being transcribed without requiring that they have any function relevant to the adjacent gene. Depending on the element's site of integration and its subsequent evolution, these transcripts may be differentially expressed in development, differentiation or through the cell cycle (Davidson and Posakony 1982).

Whatever the origin of the short interspersed repetitive elements, it is clear that they and the RATS are major agents in genome alteration, and hence potential mediators of evolutionary change. This would appear to be a major support for the 'selfish

2. Genes

DNA' view of the genome, with the problems that entails. It does suggest a solution to one of these problems: if only complete RATS or their Long Terminal Repeats can transpose, and the majority of RATS detected by hybridisation, and all the short intermediate repetitive DNA elements are actually degenerate RATS, then the reason the genome remains heterogeneous between species and intact within a species is that the majority of the 'selfish' DNA is actually just the relics of such DNA, and is not capable of transposition or self-replication (apart from as part of the chromosome, of course). However this begs the question as to why those RATS should become diverged and degenerate, and what prevented their transposition in the past and limits that of RATS today. I will return to this point in the discussion.

2.7.5 Repeated DNA and Histone genes.

If, as I have argued, repeated DNA is not significant for normal gene control or for nucleosome structure, but only in its own propagation and the resulting genome plasticity, then its relevance to histone genes is limited to evolutionary considerations. These are of some importance, however. The dispersal of gene clusters, either to produce 'scrambled' gene clusters of the chick type where there is no clear gene order (Engel and Dodgson 1981, Harvey et al 1981, Krieg et al 1982, D'Andrea and Krieg per com) or Orphans (Childs et al 1981) will significantly increase the diversity of histone gene structure we have to analyse. Conversely finding histone gene clusters containing such elements will point to possible evolutionary origins for the clusters concerned, the selective pressures on the non-coding regions and possible affinities (in sequence, proteins coded, or time and place of expression) with other gene clusters. Studies of the phylogeny of gene clusters have

2. Genes

already given us insights into functional constraints upon them (the globin cluster especially - Jeffreys 1981), and it is not unreasonable to expect useful results along the same lines from histone gene clusters.

3. Xenopus borealis.

The fifth is Ambition. It next would be right
To describe each particular batch;
Distinguishing those that have feathers, and bite
From those that have whiskers, and scratch.

The organisms we chose for this work were the primitive anuran Xenopus laevis and the closely related Xenopus borealis. Xenopus laevis, the South African Clawed Toad, is found in much of continental Africa and is considered the most primitive of the sub-family Xenopinae on morphological grounds. Xenopus borealis, the Kenyan Clawed Toad, is a central African relative. Both are actually frogs (Noble 1954, Nieukoop and Faber 1975) (Fig 3.1).

Xenopus laevis is widely used for molecular biological research for a number of reasons. It is easy to keep, cheap to feed and may be induced to breed with indecent efficiency by hormone injections (Nieukoop and Faber 1975). Its ovaries, liver and nucleated erythrocytes may all be prepared easily for biochemical analysis. As well as these general points, it was suitable for this study for two reasons.

Firstly, unless it is intended to make a complete de novo survey of an organism, it is always useful to use an experimental species which has been characterised by previous workers. This is particularly true of genetic and molecular genetic research, where generalisations and 'rules' about gene structure and function may often only be drawn by comparisons of a result with a background of other knowledge. This point will be illustrated later when the results from this thesis

Fig 3.1

Xenopus laevis and *borealis*

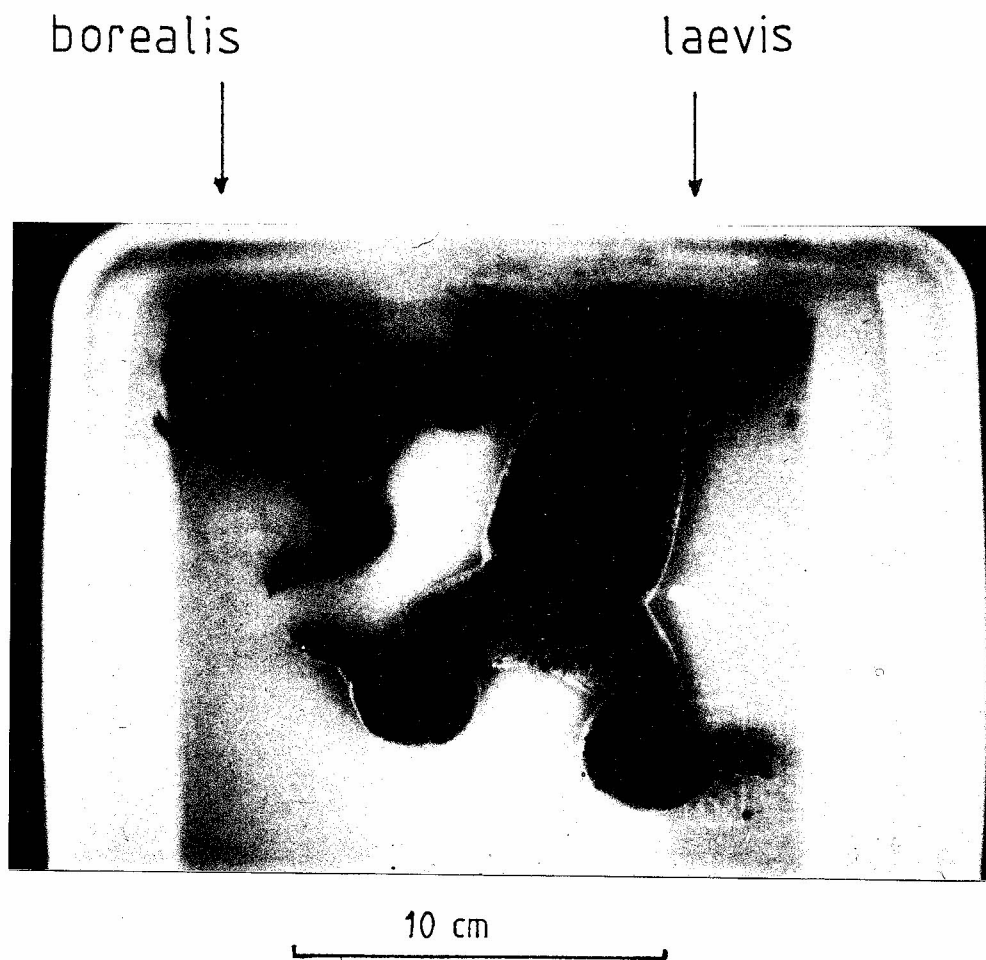


Fig 3.1 Female adult *Xenopus laevis* and *Xenopus borealis*. The frogs are in water and photographed from above.

3. Xenopus borealis

will be compared with those from other Xenopus histone gene cloning projects.

Secondly the Amphibia possess relatively enormous eggs and oocytes with correspondingly enlarged nuclei. They are sufficiently large to micromanipulate and microinject with exogenous DNAs and RNAs using relatively simple equipment (Merz and Gurdon 1977, Wickens et al 1980, de Robertis and Merz 1977). These DNAs are transcribed, and the RNA produced is processed, capped, poly-adenylated and translated. There is no effective competition with endogenous message (Richter and Smith 1981) and transcription (Hentschel, Probst and Birsteil 1980) and translation are faithful to the normal processes in the cells from which the injected DNA or RNA was derived as far as the sequence of the product is concerned. This is true even if relatively enormous amounts of DNA are introduced into the oocyte: 1 ng of cloned DNA will function in an oocyte containing 4 haploid chromosome sets (C-value = 3.2 pg (Theibaud and Fischberg 1977)) and a similar amount of amplified extrachromosomal rDNA.

This is a powerful system for introducing cloned DNAs back into a living cell to study their function, and in particular for studying the function of genes altered by in vitro mutagenesis. This system has already been extensively exploited in studying the function of sequences to the 5' of genes (Grosschedl and Birnstiel 1980, Grosschedl et al 1981, Wasylyk and Chambon 1981, Grosschedl and Birnstiel 1982). These studies have all been carried out with heterologous cloned genes, however, and some aberrant transcription seems to result because of this. We therefore decided that, as well as cloning histone genes from X.laevis we would clone those from X.borealis as well, both for structural comparisons and in order to

3. Xenopus borealis

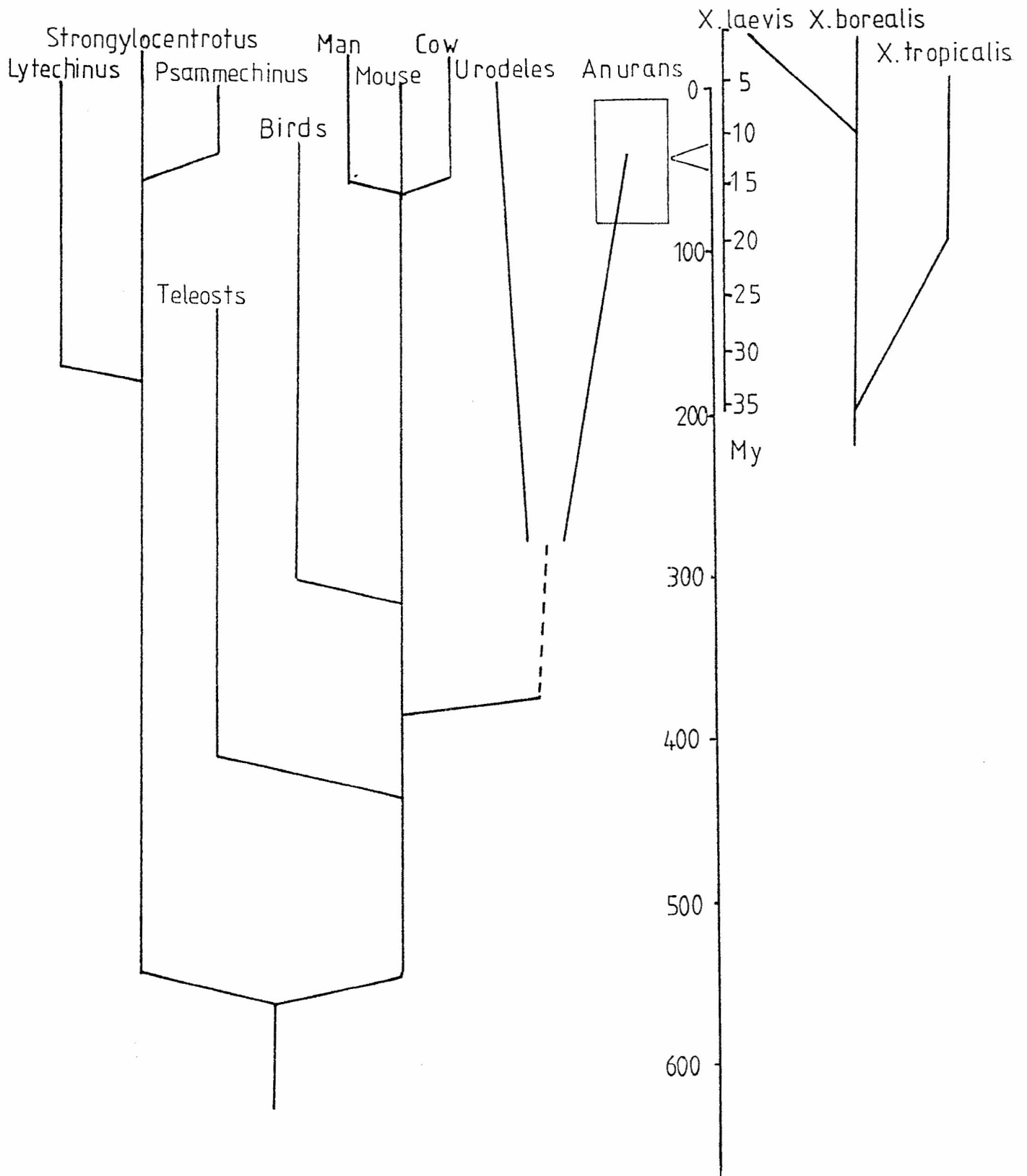
obtain an homologous set of genes to micro-inject into X.laevis oocytes for functional dissection. The X.borealis major and minor oocyte H1 proteins may be distinguished on 1-dimensional SDS-PAGE from the major X.laevis H1 (Woodland per com), and other gel systems were to be investigated to detect differences between the other histones of the two species. In this way the expression of X.borealis genes microinjected into X.laevis (and visa versa, although X.laevis are easier to obtain for dissection) could be readily monitored, while the oocyte received genes from a closely related species rather than genes from a species that diverged from the Amphibia at least 500My ago.

As some parts of this thesis ponder on evolutionary relationships, a brief phylogenetic tree showing the species mentioned at various points in this thesis is presented in Fig 3.2 . In particular we should note that immunological data on the albumins (Maxson, Sarrich and Wilson 1974) suggest that X.laevis and X.borealis diverged from a common ancestor about 8My ago (Bisbee et al 1977)

The Xenopinae have undergone tetraploidization several times since separated from the other Pipids . Relative to X.tropicalis, which contains the proposed ancestral karyotype, both X.laevis and X.borealis are tetraploid (Tymowska and Fischberg 1973, Theibaud and Fischberg 1977). Tetraploidization is common on an evolutionary timescale in the Amphibia (Noble 1954, Bogart 1980) - relative to X.laevis, X.vestitus is tetraploid and X.ruvenæriensis is hexaploid (Theibaud and Fischberg 1977). By contrast the sub-species of X.laevis have identical C-values and karyotypes (Theibaud and Fischberg 1977). Some chromosomal rearrangements have accompanied the tetraploidization events (Tymowska and Fischberg 1975). Thus it is not valid to say that X.borealis must have two copies of any given gene on two separate chromosomes, although this was presumably true

Fig 3.2

Phylogeny of selected higher animals



3. Xenopus borealis

immediately after the tetraploidization event. Subsequent divergence, large-scale chromosomal alteration and smaller-scale insertions and deletions could have destroyed any trace of the tetraploidization event at the molecular level. However several families genes have been cloned from X.laevis, and their members have been found in duplicated pairs: this is true of ribosomal protein (Bozzoni et al 1981), globin (Patient et al 1980, Widmer et al 1981), vitellogenin (Wahli et al 1979) and albumin (May, Weber and Westley 1982) genes, where a duplication of whatever other pattern of diversity each gene family possesses is seen, suggesting all the genes in the frog have been duplicated and have not diverged extensively since. The complexity of the histone gene organisation found in this animal is also more easily explained by the assumption that these genes have been doubled up by an ancient tetraploidization event and have not diverged extensively since (Section 2.3.6).

4. Aims and Approaches of this project.

"For the Snarks a peculiar beast that won't
Be caught in a commonplace way.
So do all that you know, try all that you don't
Not a chance must be wasted today."

4.1 Original aims.

Initially this project was to answer the following questions.

What is the structure of the X.borealis histone gene cluster?

What is the function of the different regions of the cluster?

Question 1 has been shown to be partly meaningless since this project was conceived. The notion of The histone gene cluster, typified by the embryonic histone genes of the sea urchin and Drosophila, has given way to that of A histone gene cluster. The stochastics of cloning mean that in this study this equated with 'a histone gene cluster present as a minor class of histone genes in X.borealis'. While this project was underway, the structure of the majority cluster type in X.borealis was determined (Turner and Woodland 1982b).

Question 2 has only been briefly addressed. The coding regions of the clone have been identified and some regions characterised by DNA sequencing. The functional studies using micro-injection assays for function have been begun by asking what genes in the clone are transcribed and what effect vector sequences have on this

4. Aims and Approaches.

transcription. These assays can also show whether the genes code for variant proteins, and, if so, relate these to variants previously identified by electrophoresis. This cannot be done by hybridisation techniques or DNA sequencing in the absence of protein sequence information by which variant protein sequences may be identified. The micro-injection studies will be continued in this laboratory after the work described here.

4.2 Approaches.

4.2.1 Molecular cloning.

The cornerstone of modern molecular genetics is 'molecular cloning'. This is the generic name for any technique which used biological systems to produce a large number of identical copies of a molecule (almost invariably DNA) from an original stock by clonal growth of an organism in which the molecule is situated. In particular the ability to produce recombinant molecules of DNA from any source in vitro has meant that small fragments of eukaryotic genomes have been 'cloned' (a word I will use hereafter without apostrophisation or amplification), and hence made available for biochemical analysis free of the rest of the bulk of DNA in the eukaryotic nucleus (Glover 1977, Dahl, Flavell and Grosveld 1981). Thus a specific gene 1 to 10kb long can be studied free of the background of 1,000 Mb of DNA in the smaller eukaryotic genomes.

Gene cloning strategies almost all use autonomously replicating prokaryotic DNAs as vectors to provide a replication origin and selectable markers for the DNA to be cloned (Dahl, Flavell and Grosveld 1981). Viral and plasmid vectors are both common. With the

4. Aims and Approaches.

exception of Cosmids (Collins and Bruning 1978, Collins and Hohn 1978, see Appendix 1) the latter suffer from the drawback that, being maintainable in their host cell in any size, they possess no positive mechanism for selecting in vitro recombinants containing large segments of eukaryotic DNA. By contrast bacteriophage vectors, notably lambda-derived vectors (M13 is like a plasmid in this as in other respects) require a certain minimum size of inserted DNA before they can be packaged by their own virion proteins into viable phage particles, and hence before they can be propagated (Hohn 1975, Hohn and Hohn 1974, Hohn and Murray 1977, Thomas, Cameron and Davies 1974). Thus inserting random DNA fragments into phage vector molecules results in a spectrum of molecular species, including many which contain no inserted DNA at all (see Appendix 2), but of which only those with inserts can propagate as phage (Thomas, Cameron and Davis 1974). This eliminates non-recombinant clones from a gene library, which clones would contribute an unacceptable background of potential recombinants which were actually useless were they left in with the true recombinants.

In particular I used the lambda vector Lambda-gtWES (Thomas Cameron and Davis 1974, Enquist et al 1974, Tiemeier, Enquist and Leder 1974) to clone EcoRI fragments of total genomic DNA of X.borealis. The use of restriction enzymes to cut DNA at defined, reproduceable sites and of DNA ligase to rejoin it makes such operations technically simple (in theory). Lambda -gtWES has an optimum insert size of about 8kb and a minimum insert size of 1kb (Tiemeier, Enquist and Leder 1974; see Appendix 2). As histone genes are rarely interrupted and code for small proteins this clearly gives a good chance of obtaining an intact histone gene with its flanking sequences in a single DNA fragment. If the genes are clustered (as we suspected they would be and now know they are), 8kb

4. Aims and Approaches.

of insert should give a reasonable probability of obtaining two or more genes from a cluster, especially if overlapping clones are obtained.

The cloning of specific eukaryotic genes depends on the availability of hybridisation probes which can be used to identify plaques or colonies carrying the recombinant molecules containing the required sequence (Benton and Davis 1977). Initially we had clones and sub-clones of the Strongylocentrotus purpuratus embryonic gene cluster (Sures et al 1978), but these do not hybridise well with Xenopus histone genes due to the sequence divergence between the two species, and contain spacer sequences which are known to cross-hybridise with satellite in Triturus (Callan and Old 1980) and possibly would with Xenopus as well. During this study P.C.Turner produced an H4 cDNA clone from X.laevis (Turner and Woodland 1982). This clone was therefore used as a probe to isolate a histone-gene containing clone from the gene library. If this clone had not been available there is little doubt that the S.purpuratus clone would have been sufficient. However extra screening procedures would have had to be employed to eliminate possible hybridisation artefacts.

4.2.2 Characterisation of a cloned DNA.

To locate the genes on a clone hybridisation probes are required again, but as the uncharacterised cloned DNA is available in large quantities the conditions of hybridisation are not so demanding as those for screening a gene library or studying uncloned genomic DNA, and so less homologous or less productive hybridisations may be used. Two types of hybridisation analysis can be made. The clone can be fragmented, usually with restriction enzymes, the fragments

4. Aims and Approaches.

isolated, labelled and hybridised to DNA or RNA of known (or suspected) sequence to determine if these sequences are homologous to the unknown. Alternatively the clone may be fragmented and separated, and the probe, which has been labelled, hybridised to it. The latter can be expanded to use an unlabelled mRNA probe which is subsequently identified by its ability to direct an in vitro translation system to make its protein product. This last is Hybrid Release Translation.

It should be noted that using the clone as a labelled probe to detect homology in restricted DNA is a sufficiently sensitive technique to be used to detect single-copy genes in genomic DNA. This is a major technique for localising genes within the uncloned genome, and has already been referred to in the studies of the histone gene structure in the genomes of X.laevis and X.borealis (Turner and Woodland 1982b, van Dongen 1982).

4.2.3 Sequence analysis and other explorations.

DNA sequence analysis now plays a major part in any investigation of detailed DNA structure. It was intended to use DNA sequencing techniques to characterise regions of the clone prior to in vitro mutagenesis for functional dissection of the clone. However the project did not proceed as far in this direction as had been planned, so this was not done. Some coding region sequence was obtained to characterise the H3 and H1 genes, and a repetitive element in the clone. As we have seen above, sequence analysis of intermediate repetitive elements can give strong indications as to their affinities with other, better characterised genomic sequences. In particular I wished to see if its sequence, position or distribution in the genome gave any clues as to its role in the

4. Aims and Approaches.

function or evolution of the histone gene cluster it was in. The relevance of section 2.7 above to these points will be appreciated. The location of repetitive elements in the clone and their repetition in the genome may be determined by filter hybridisation techniques described in more detail below. These do not guarantee that they are detecting exactly homologous sequences, of course, merely that they are detecting a family of sequences of defined divergence from the cloned sequence.

4.2.4 Functional dissection.

This section, which was to have been the main thrust of the project, was only started in this work. The initial experiments must examine the micro-injection system, determine that the genes in the clone are transcribed correctly in the system and that this is not affected by any contaminating vector sequences or other potential artefacts. A short series of experiments can show 1) what proteins (if any) are produced from mRNA synthesised from the clone and 2) that they are coded by the cloned DNA. This is of considerable interest in characterising the clones, as it shows that the coding and adjacent regions of the transcribed genes must be intact (Grosschedl and Birnstiel 1980, 1982, Grosschedl et al 1981), and also indicates whether the proteins produced are variants.

4.3 Subsequent aims.

With the discovery of the repetitive DNA segment in the clone, and the interest in where the clone reported here as Xbhl had come from, two further aims were added which partially superseded the second aim of the original project.

4. Aims and Approaches.

What is the genomic frequency and distribution of a) the whole cloned cluster and b) the repetitive section of it?

Where did Xbhl come from, both in the evolutionary and the laboratory sense?

The approach to the description of repetitive DNA sequences has been mentioned above. Quantitative hybridisation techniques are available in a range of precisions and speeds, and are well characterised (Kafatos, Jones and Efstradiatis 1979, Britten, Graham and Neufeld 1981). I will leave detailed discussion of the dot-hybridisation method used and its analysis for Appendix 3.

The approach to the second aim derives from the reason for asking the question. The clone described in this work, although superabundant in the genomic library from which it was isolated, is barely detectable in X.borealis genomic DNA. It is also extremely similar to a clone isolated from X.laevis by Dr. R.W.Old (vide infra section 13). This gave rise to the suspicion that the former could be a contamination of the gene library by the latter. To remove this suspicion and to characterise the degree and nature of the similarity between these two clones, I needed a probe of primary DNA structure which is not dependant on DNA function. Restriction enzyme digests are such a probe (Roberts 1978), and can be used to characterise the similarity between DNA molecules as well as fragment them for other analysis. This was the approach used here. The results have implications for histone gene evolution in Xenopus in particular, and for the structure and function of multigene clusters in general.

Methods

"The Method employed I would gladly explain
While I had it so clear in my head -
If I had but the time, and you had but the brain -
But much remains to be said."

5. Methods

This section describes the methods used generally throughout the rest of this thesis. Methods which have been used to produce results described in only one section are described in that section.

5.1 Materials.

Restriction enzymes were obtained from Boehringer Mannheim (London) Ltd. (AluI and TaqI) and BRL (Bethesda Research Laboratory Inc, Maryland, USA) (all others), and were used exactly according to the instructions supplied by the manufacturers. E.coli DNA polymerase I was obtained from Boehringer Mannheim. E.coli DNA polymerase Klenow fragment was obtained from Boehringer Mannheim or from CBL (Uniscience Ltd, England).

All radio-isotope containing reagents were obtained from Amersham International Ltd (England) at the following specific activities: alpha-p(32)-dNTP 2000-3000 Ci/mM at 10 mCi/ml in water. Tritiated lysine: 75-100 Ci/mM at 10mCi/ml in water. S(35)-Methionine: 600 Ci/mM at 10mCi/ml in water.

Nitrocellulose sheets were obtained from Schleicher and Schull GmbH (W.Germany) (their 'membranfilter BA-85'). Type II Agarose for electrophoresis was bought from the Sigma Chemical Company (England). Acrylamide for electrophoresis was obtained from Eastman Kodak

Company (N.Y., USA). Materials for bacteriological media were obtained from Difco Laboratories (Michigan, USA) and Oxoid Ltd (England). X-ray film and autoradiographic intensifying screens were from Ilford Ltd (England): all other photographic materials were obtained from Ilford or Eastman Kodak. All other chemicals and reagents were obtained from BDH Chemicals Ltd (England) (this supplier's 'Analar' grade reagents were used throughout), Fisons Scientific Apparatus (England), Sigma and Eastman Kodak.

5.1 Bacterial, plasmid and phage genotypes.

Only the genotype features relevant to this study are listed here. References are provided for the other genetic features of these strains.

5.1.1 E.coli strains

HB101. SupE RecA- mK- rK- (Bolivar and Beckman 1979)

LE392. SupE SupF RecA- rK- (Hohn and Murray 1977, Hohn 1979, Enquist et al 1979)

BHB2688. N205 RecA- (Lambda imm434 cIts b2 red3 Eam4 Sam7) (Hohn 1979)

BHB2690. N205 RecA- (Lambda imm 434 cIts b2 red3 Dam15 Sam7) (Hohn 1979)

5.1.2 Plasmid vectors.

pBR322. ColE1 derivative Ap-r, Tc-r 4.3kb long (Bolivar et al

1977)

pAT153. pBR322 deletion derivative. Ap-r. 3.9kb long (Twigg and Sherrat 1980)

pBR325. pBR322 derived vector for EcoRI cloning. Ap-r, Tc-r, Cm-r 5.4kb long (Bolivar 1978)

pCR1. ColE1 derivative Kn-r. 13.1 kb long (Bolivar and Beckman 1979)

pHC79. pBR322 derived cosmid vector. Ap-r, Tc-r 6.4kb long (Hohn and Collins 1980)

Homer-1. pAT153 derived cosmid vector. Ap-r 5.8kb long (Chia, Scott and Rigby 1982).

5.1.3 Plasmid and M13 recombinants.

pX1H4W1. pAT153 cDNA clone of ovarian H4 from X.laevis (Turner and Woodland 1982)

pX1H3W1. pAT153 cDNA clone of ovarian H3 from X.laevis (Turner and Woodland 1982)

pXbH4W1. pAT153 cDNA clone of ovarian H4 from X.borealis (Turner and Woodland 1982)

M13-H2a. M13 phage subclone of H2A gene region of a genomic gene clone from X.laevis (R.W.Old per com)

M13-H1. M13 phage subclone of central region of an H1 histone gene from a genomic gene clone from X.laevis (R.W.Old per com)

pSp series. pSC101 derivatives containing fragments and subfragments of the Strongylocentrotus purpuratus embryonic gene repeat. (Kedes et al 1975a,b , Sures et al 1978)

pSp102 H1, H4 and H2B genes and intervening spacer.

pSp117 H3 and H2A genes and intervening spacer.

pSp2b-5. 3' end of H2B gene.

pSp1-3. central region of H1 gene coding region.

pSp3-1. Most of the central coding region of the H3 gene.

pXlr101. pCR1 recombinant containing part of the X.laevis rDNA repeat (ETS, 18S and part of 28S) (Boseley et al 1979).

5.1.4 Lambda phage.

lambda cI857S7. 'wild type' lambda - cI ts857, Sam7, kept as a lysogen (Hedgpeth, Goodman and Boyer 1972, Enquist et al 1974)

lambda-gtWES. Eco RI cloning vector. See section 6.1

5.2 Microbiological culture media and conditions.

5.2.1 Bacteriophage plating.

This is essentially as described (Murray and Murray 1974, Enquist et al 1974, Enquist and Sternberg 1979) . Bacteria and phage were plated on 9cm diameter plates containing autoclaved L.Agar (10g NaCl, 10g Bacto Tryptone, 5g Difco Yeast Extract, 15g Agar in 1 litre, autoclaved for 15mins, 120°C then poured into sterile plates and allowed to cool and set). Phage were diluted to suitable concentration in Phage Buffer (1.5g potassium dihydrogen phosphate, 3.5g disodium hydrogen phosphate, 2.5g NaCl, 0.125g magnesium sulphate, 0.1mM Calcium chloride, 50mg gelatin in 500 mls, autoclaved). Indicator E.coli (LE392) were grown in 10 mls L.M.Broth (vide infra) overnight, centrifuged for 5 mins in an MSE benchtop centrifuge (5,000rpm) and re-suspended in 1/10 original volume of 10mM magnesium sulphate. 0.1ml of phage was added to 0.2ml indicator bacteria. The phage were incubated with the bacteria for 10 minutes at room temperature to allow them to adsorb onto the bacterial cells, and then 3.5mls of Top Agar (as above for L.Agar but containing only 7g/litre agar and supplemented with 0.4% Maltose, 10mM magnesium sulphate) melted and then cooled to 45°C were added, and the mix immediately poured onto an L.Agar plate. Plates were incubated overnight at 37°C.

5.2.2 Plating E.coli.

E.coli were spread on L.Agar plates using a glass rod pre-sterilised by washing in ethanol and burning off the surplus ethanol. Antibiotics were included in the L.Agar after autoclaving but before pouring plates to the following concentrations: ampicillin 100ug/ml , chloramphenicol 50ug/ml tetracycline 10ug/ml , kanamycin 30ug/ml final concentration.

5.2.3 Storage of E.coli strains

E.coli strains were kept for up to 1 month as colonies on plates. For longer storage they were kept in glycerol thus. 0.5mls of an overnight culture of the bacteria in L.Broth with suitable antibiotic supplement was added to a 4-ml 'bijou' bottle in which 0.5 mls of glycerol and 3mls packed volume of glass beads had been previously autoclaved. The bottles were stored at -20°C, and one or two beads taken out into L.Broth whenever growing stocks were required.

5.2.4 Bulk growth of lambda bacteriophage.

This was modified from Murray and Murray (1974, Hohn and Murray 1977). Lambda phage isolates were plated out at densities of 100 to 1000 on one plate, and the top agar scraped into 8mls Phage Buffer. This was allowed to stand for 1 hour at 4°C, and then bacterial and agar debris was centrifuged out at 8000 rpm, 15mins. This was the starting inoculum, and typically had a volume of 3-5mls. Indicator bacteria were grown overnight in L.M.Broth (L.Broth supplemented with 0.4% Maltose, 10mM magnesium sulphate). 1 ml of phage inoculum was added to 0.3mls of indicator bacterial culture and incubated at 37°C for 15 minutes: all components of this and subsequent stages were prewarmed to 37°C. The bacteria were then added to 500ml or 1 litre volumes of prewarmed L.Broth supplemented with 10mM magnesium sulphate in a 2 litre flask, and shaken vigorously (120-150rpm) for at least 16 hours. Successfully completed growth is indicated by the collection of fine strands and blobs of bacterial debris on the bottom of the flask. At this stage the yield of phage is maximal, and is typically about 10 to the 14 pfu per litre.

5.2.5 Purification of lambda phage from bulk cultures.

Lambda phage were concentrated from bulk culture by Polyethylene glycol 6000 (PEG) precipitation (Yamamoto et al 1970). 500ml or 1 litre cultures, on completion of growth, were lysed with 1ml of chloroform (which both kills the bacteria and lyses the cells releasing additional phage particles), which was added to the flasks 10 minutes before centrifugation. The cultures were centrifuged at 5000 rpm in an MSE 6x250ml rotor at 4°C to pellet out bacterial debris. The supernatant was made 0.5M in NaCl, 10% in PEG. It was then stirred at 4°C for 2 hours plus however long the PEG took to completely dissolve. This suspension was then centrifuged at 3500 rpm in an MSE 6x250ml rotor at 4°C for 15mins, and the supernatant carefully poured off. The pots were inverted for ten minutes to allow supernatant to drain from them. The pellet, including the material on the sides as well as that on the bottom of the centrifuge pot, was resuspended in Phage buffer by swirling and gently stirring with a pipette. It was then shaken gently for 1 hour to disperse lumps, centrifuged for 5 minutes at 10,000rpm in an MSE 8x50ml rotor, and the phage-containing supernatant banded on CsCl (Murray and Murray 1974).

To each ml of the phage concentrate prepared above was added 0.709g of CsCl solid. This was dissolved, and the volume made up to a convenient one for the rotor to be used with phage buffer made 0.709gCsCl per 1ml. This was then centrifuged in an MSE 10x10ml or an MSE 8x25ml rotor at 40,000rpm for 48hours at 18°C. The resulting gradient had just one band in its centre: this is the phage band. This was removed with a pasteur pipette, made up to a convenient volume with more phage buffer + CsCl (1ml+0.709g) and recentrifuged as before. The phage band was removed. This was the standard phage

stock, for use in DNA preparations or for storage.

5.3 DNA preparation procedures.

5.3.1 Preparation of Phage DNA.

This was after Murray and Murray (1974) Phage prepared as above were dialysed against 10mM EDTA (pH 8.0) for 2 hours or overnight. They were then extracted with an equal volume of phenol until the phenol:aqueous interface was clear: typically this took four successive extractions. The volume of the phage preparation frequently had to be increased to overcome problems with the viscosity of the DNA: a volume of 10mls for 1 litres' worth of phage culture was a typical volume to phenol extract. The phage DNA prep was then extracted twice with chloroform to remove the phenol, once with ether to remove the chloroform and was then dialysed for at least 48hours at 4°C against TE to remove the ether. At no stage was the phage DNA precipitated from ethanol. It was found to be extremely hard to redissolve after such a precipitation.

5.3.2 Preparation of plasmid DNAs.

Plasmid DNAs were prepared by the 'cleared lysate' procedure (Clewell and Helinski 1970) as modified by Coleman et al (1978). A 10ml inoculum of E.coli was grown in L.broth overnight. This was added to 1 litre of L.broth supplemented with suitable antibiotic (to the same concentration as was used in preparing plates), and shaken vigorously at 37°C until the OD at 600nm of the suspension reached 0.8 to 1.2 (typically 6-9 hours) . 150mg of solid chloramphenicol was added and the flasks shaken for a further 14-18

hours (Clewell 1974).

The bacteria were harvested by centrifugation for 15 minutes at 5000rpm in an MSE6x250ml rotor. The centrifuge pots were drained thoroughly, and then the bacteria from each pot (1E equivalent to 250mls of the original culture) were suspended in 1ml of 10% sucrose (w/v), 2mM magnesium chloride. The pots were pooled: the following figures refer to the pooled bacteria from one 1 litre culture. The pooled bacteria were chilled on ice, and then 0.8mls of 20mg/ml freshly made lysozyme solution in 10mM Tris (pH8.35) were added. This was incubated on ice for 5 minutes, then 1.6mls of 0.25M EDTA (pH 8.0) were added and mixed in well. Another five minutes on ice, and then 6.4mls of 0.5% Triton X-100 (NP-40 was interchangeably used), 0.0625MEDTA, 0.05M Tris pH 8.35 were added, mixed in well. The mix was incubated on ice for 20minutes, during which time the bacteria lysed. The lysate was cleared by centrifugation for 30 minutes in an MSE 8x50 rotor at 25,000rpm at 4°C in an MSE '65' .

Note that here 'Tris pH8.35' refers to a Tris/HCl solution made up to pH8.35 at a concentration of 1M, and then diluted with water as required. The actual pH of the diluted buffer will differ slightly from 8.35.

5.3.3 CsCl/ethidium bromide equilibrium centrifugation.

Plasmids were purified by banding on a CsCl/ethidium bromide equilibrium gradient (Vinograd 1963, Radloff, Bauer and Vinograd 1967). To each ml of a cleared lysate plasmid sample prepared as above was added 0.1ml of 7mg/ml ethidium bromide solution in water, and the solution centrifuged briefly to remove precipitated protein. To the supernatant was added 1g CsCl solid for each ml of original

cleared lysate. This was dissolved, and the solution centrifuged in an MSE 10x10ml or an MSE 8x25ml rotor for 48hours at 40,000rpm at 18°C. Two bands appeared on the resulting gradient: the lower of the two ethidium-stained bands was the supercoiled plasmid band. This was removed by pumping from the top with a peristaltic pump, carefully avoiding inclusion of any of the upper, chromosomal DNA band or any particles of precipitated protein floating in the gradient. The plasmid preparation was then extracted with isoamyl alcohol previously equilibrated with water saturated with CsCl until it no longer showed a pink colouration: typically four extractions each with four times the sample volume of isoamyl alcohol sufficed. The plasmid preparation was then diluted to twice its volume with water, phenol extracted once, made 0.1M in sodium acetate, and two times the total sample volume of ethanol were added to precipitate the DNA. Plasmid DNA was stored as an ethanol pellet at -20°C, and resuspended in TE as required for use.

5.3.4 Preparation of Genomic DNA from Xenopus.

A Xenopus, preferably a female (as females are larger), was anaesthetised in 0.2% 'MS222' (aka 'Tricaine', Ethyl m-amino benzoate solid dissolved in the tank water to 0.2% final concentration). The thorax was dissected open from the ventral side, the heart exposed and a hypodermic syringe with 23-G needle attached was inserted into the ventricle. The heart still beats when Xenopus are anaesthetised for 10-20 minutes with this concentration of MS222, and so pumps the blood into the syringe from the circulation. The syringe, the receiving glass tube and the frog's thorax was rinsed in heparin solution in NMT (vide infra) to prevent blood clotting. Typically two or three Xenopus individuals were used in one DNA preparation.

The blood was collected in NMT (0.1M NaCl, 0.01M Tris pH7.4, 3mM magnesium chloride), 10 mls per X.laevis, or per three or four X.borealis. They cells were pelleted by centrifugation at 2000rpm for 10mins at 4°C, and then washed in fresh NMT and re-pelleted until the supernatant was no longer pink. To the final 10mls of washed, suspended cells in NMT was added a further 10mls of NMT made 1% in NP-40 : this was left on ice for 10 mins. This lyses the cells releasing the chromatin as dense, 'nuclear' bodies. The chromatin was pelleted by centrifugation at 4000rpm for 5mins at 0°C. This pellet should be white.

The chromatin pellet was resuspended in 10mM NaCl, 10mM Tris pH7.4, 10mM EDTA. This is difficult to do, and it is usually necessary to break up the pellet with a few strokes of a fairly loose-fitting teflon homogeniser before it is suspended. To the suspended chromatin was added pancreatic RNase A to 10ug/ml. This was incubated for 1 hour at 37°C. Then I added SDS to 0.2%, and proteinase K to 40ug/ml, again homogenising briefly to mix in the components. This was incubated for 1.5hours at 37°C. The digested chromatin material, now extremely viscous, was then extracted with an equal volume of phenol until no phenol:aqueous interface precipitate was seen. This required quite vigorous extraction and separation of the phases by centrifugation at 8000rpm for 5mins, 20°C due to the viscosity of the DNA. However it is essential to ensure that the DNA is thoroughly purified of protein, even if this means additional mechanical shearing during the extraction procedure. Once clean of protein, the DNA was extracted twice with chloroform and dialysed for 24 hours against TE at room temperature and then for a further 48hours at 4°C. As with lambda DNA, ethanol precipitation was not attempted on undigested genomic DNA.

5.3.5 Preparation of E.coli genomic DNA.

This is a combination of the Cleared lysate procedure and the X.borealis genomic DNA preparation procedure. 1.5litres of L.Broth supplemented with 0.5% glucose were inoculated with a 10ml overnight culture of the chosen strain of E.coli, and shaken vigorously overnight at 37°C. The bacteria were pelleted by centrifugation at 5000rpm, 15mins, 4°C and resuspended in 8mls of 10% sucrose, 2mM magnesium chloride. To this was added 1.5ml of freshly made up 100mg/ml lysozyme in 10mM Tris pH8.0. This was incubated for 60mins at 0°C, and then 1.5mls of 0.5M EDTA (pH 8.0) and 9mls of 1% Triton X-100 in water were added. This lyses the bacteria. This lysate was phenol extracted twice, which did not remove all the protein present, and then extracted twice with chloroform. To it was added Pancreatic RNase A to 20ug/ml final concentration, and the lysate was then incubated for 1 hour at 37°C. Proteinase K was added to 10ug/ml final concentration, and SDS to 0.2%, and the lysate was incubated at 37°C for a further hour. The lysate was then extracted with equal volumes of phenol until the phenol:aqueous interface was clear, extracted once with chloroform and dialysed for 24 hours at room temperature against TE, and then a further 48hours against TE at 4°C.

5.4 DNA labelling procedures.

5.4.1 Nick translation.

This is after Rigby et al (1977). To 100ng - 2ug of DNA to be labelled in not more than 8ul of TE or water were added the following: 4ul of 5xNT buffer (250mM Tris pH7.9, 50mM magnesium chloride, 50mM Dithiothreitol, 250ug/ml bovine serum albumin), 1ul each of three unlabelled dNTPs (1mM each), 1-5ul of alpha-32-P dNTP (10mCi/ml in aqueous buffer), water to a total volume of 19ul. This reaction was assembled on ice, then started by adding 1ul of E.coli DNA polymerase I (2-5u/ul). The reaction was allowed to continue for 2-3hours at room temperature, and then terminated by heating it to 70°C for 5mins. 10ug of E.coli tRNA was added as carrier. The labelled DNA was separated from unincorporated label by passing it down a Sephadex G-50 column equilibrated in TE, collecting the excluded fraction (monitored by detecting the radioactivity travelling down the column with a bench monitor), making it 0.5M in sodium acetate and adding two volumes of ethanol to precipitate it.

For labelling DNA for electrophoresis or restriction, a slightly different protocol is used which results in more intact double-stranded molecules being produced. The reaction is incubated for 45mins in a water bath at 12°C only, and at least 1ug of DNA is used as starting material. Typically this was not separated from unincorporated label on a column, instead being diluted with TE 10- to 50-fold and being used in this form.

5.4.2 End labelling using Klenow polymerase.

For some restriction mapping and for DNA sequence determination fragments were end-labelled by filling in the cohesive restriction cleavage site ends with the Klenow fragment of E.coli DNA Polymerase (Jacobsen, Klenow and Overgaard-Hansen 1974, Sanger, Nicklen and Coulson 1977). To 1-5ug of DNA in TE were added 4ul 5xEL buffer (250mM Tris pH8.0, 250mM sodium acetate, 25mM magnesium chloride, 10mM 2-mercaptoethanol), 2ul of alpha-labelled dNTP (the actual NTP used depends on the sequence of the ends to be filled, of course) 10mCi/ml in aqueous buffer, water to 18ul. This was assembled on ice, and to it was added 2ul Klenow fragment (12u). This was incubated for 2 hours at room temperature, then was heated to 70°C for 5mins and treated as a nick-translation.

5.5 Other techniques for recombinant manipulation.5.5.1 Transforming E.coli with DNA.

E.coli were transformed by the technique of Mandel and Higa (1970). Recipient E.coli were grown overnight in 10mls L.Broth at 37°C. 2 10ml aliquots of L.Broth were inoculated with 0.2 mls of this overnight culture, and were grown at 37°C until their OD (600nm) was 0.7-1.0. They were then chilled briefly on ice, centrifuged at 3000rpm in an MSE bench-top centrifuge and resuspended in 5mls each of chilled 0.1M calcium chloride. The bacteria were kept on ice for 20mins, and then centrifuged again and resuspended in a total volume of 0.1ml. To this competent cell suspension was added the DNA in 0.1ml of SSC. The mixture was kept on ice for 15mins, then incubated at 37°C for 20mins. It was then added to 10mls of prewarmed L.broth

and incubated for a further 30mins to allow expression of resistance markers, harvested by centrifugation at 4000rpm, 5mins and plated on suitable antibiotic- containing agar.

5.5.2 In vitro packaging of lambda DNAs.

This is the system of Enquist and Sternberg (1979), slightly modified by Ish-Horowitz (per com).

The system consists of two bacterial lysates, the Sonicated Extract (SE) and the Freeze-Thawed lysate (FTL). The FTL is prepared thus. 3 10ml cultures of BHB2688 are grown overnight at 30°C. These were used to inoculate 3 500ml cultures which were shaken in 1 litre flasks at 30°C until their OD(600) has reached 0.3. They were then warmed over a bunsen until their temperature reached 45°C, and then were vigorously shaken at 37°C for 1.5hours. The cultures were cooled on ice, loaded into an MSE 6x250 rotor and the bacteria harvested by centrifugation at 6000rpm for 10mins at 0°C. The centrifuge pots were drained thoroughly to remove all the supernatant, and then each of the 6 pellets was suspended in 0.5mls of cold 10% sucrose, 50mM Tris pH8.5. The suspensions were pooled, and placed into a 10ml centrifuge tube, warmed at 37°C for 4 mins, and then rapidly frozen in liquid nitrogen. They were then kept at -70°C overnight, thawed out at room temperature the following day and then cooled on ice. 75ul of buffer M1 (made up by mixture of: 110ul water, 6ul 0.5M Tris pH 7.5, 1ul 2ME, 300ul 0.0M Spermidine, 0.1M Putrescine, 9ul 1M magnesium chloride, 75ul 0.1M ATP) was added and mixed in with the gelatinous lysate, which was then centrifuged at 35,000rpm for 35mins in an MSE 10x10ml rotor. The supernatant was removed, aliquotted into 50ul aliquots and frozen in liquid nitrogen for storage at -70°C until needed.

SE was prepared thus. 3 10ml cultures of BHB2690 were grown overnight at 30°C, and used to inoculate 3 500ml L.broth cultures in 1 litre flasks, which were shaken at 37°C until their OD(600) reached 0.3. They were heated over a bunsen to 45°C and then shaken for 1.5hours at 37°C. The bacteria were harvested by centrifugation at 6000rpm, 0°C. The pellets were drained and taken up in a total of 3mls of Buffer A (20mM Tris pH8.0, 3mM magnesium chloride, 0.05% v/v 2-mercaptoethanol, 1mM EDTA). The suspended bacteria were pooled, another 2.6mls of buffer A added, and then the suspension was sonicated on ice with brief blasts once every 20-30 seconds being careful to avoid foaming. Sonication was continued until the suspension was no longer stringy and viscous. The extract was centrifuged for 6mins at 8000rpm, 0°C, and the supernatant was pipetted into 50ul aliquots for rapid freezing in liquid nitrogen and storage at -70°C.

5.5.3 In vitro packaging of DNA.

A packaging reaction was made up thus. 1 or 2 ul of DNA in water or TE were added to 7ul of buffer A, 2ul of buffer M1, 6ul of SE and 10ul of FTL. The reaction was assembled on ice, then incubated at 37°C for 1 hour. 100ul of Phage buffer are then added, and the phage plated out as above. The standard reaction cannot efficiently package more than about 1ug of lambda DNA, and operates most efficiently on amounts of 0.5ug or less.

5.5.4 The Benton and Davis plaque-lift hybridisation procedure.

This is the bulk phage-screening procedure of Benton and Davis (1977). Phage are plated out on L.Agar in the usual fashion, noting that it is wise to extensively dry the plates first in an oven under

UV light before using them, as otherwise the Top Agar does not bond efficiently to the L.Agar, and sticks instead to the nitrocellulose. After overnight incubation the plates were cooled to 4°C, and then were used as soon as convenient. Nitrocellulose filters, previously autoclaved, were layed on the surface Top Agar and left until the moisture seaped through, typically about 1 minute. Their orientation on the plates was marked with a pin-prick before they were removed and washed in 0.5M NaOH, 1.5MNaCl for 20 seconds, 0.5MTris pH7.0, 3MNaCl for 20 seconds, rinsed in 2xSSC and then dried in air and baked for 2 hours at 85°C.

5.6 Electrophoretic gel methods.

5.6.1 Agarose gel electrophoresis of DNA.

100 ml 15x20cm agarose slab gels were run in Tris/Phosphate buffer (4.35g Trizma base, 4.68g Sodium dihydrogen phosphate, 0.37gEDTA in 1litre) for all purposes except preparation of restriction fragments. In the latter case Tris/Acetate (4.84gTrizma base, 2.72g Sodium acetate, 0.37g EDTA, adjusted to pH8.0 with acetic acid) buffer was used (Sharp, Sugden and Sambrook 1973, Murray and Murray 1974) . Low percentage gels (0.5% - 1.2%) were run overnight at 1volt/cm (40mA). High percentage gels (1.2% - 2%) were run for 4 to 8 hours (depending on the fragments to be examined) at 2 volts/cm (80mA). Gels of DNA samples to be visualised by ethidium bromide fluorescence were made up containing 1µg/ml ethidium bromide, and run with running buffer with the same stain concentration in it. Samples were diluted with 1/2 volume of loading buffer (0.1x running buffer, 10% sucrose, 0.01% Orange G or Bromophenol blue) before runing.

Sizes of fragments detected in agarose gels were deduced by interpolation from marker DNAs of known size after Southern (1979). Size markers were: for 0.6 - 24kb lambda cI857S7 EcoRI or EcoRI+HindIII digests (Hedgepeth, Goodman and Boyer 1972, Glover 1977), for 150bp - 1.6kb a HinfI digest of pAT153 (Sutcliffe 1978, Twigg and Sherrat 1980).

5.6.2 Acrylamide gels for electrophoresis of double stranded DNA.

These were made up as for denaturing gels, but without the urea (vide infra) (Maniatis, Jeffrey and van de Sande 1975, Southern 1979). They were run for 8-16 hours (depending on fragment sizes), at 6 volts/cm, 10mA. Samples were freeze-dried and redissolved in 5ul of Loading buffer (50% glycerol, 25mM EDTA pH8.0, 0.1% each Xylene cyanol and bromophenol blue) before loading.

5.6.3 Acrylamide gels for single stranded DNA.

These were gels for use mainly with DNA sequencing samples. Gels of 20% and 12% were occasionally used, but the majority of samples (including double stranded ones as above) were analysed on 8% gels (Maxam and Gilbert 1978, Sanger, Nicklen and Coulson 1977). The acrylamide concentration in the polymerisation mixture was the only difference between these gels. 4g of acrylamide, 0.1g N.N bis-acrylamide, 2lg urea and 5mls of 10xTBE buffer (TBE buffer = 10.9g Trizma base, 5.5g Boric acid, 0.95g EDTA in 1 litre) were dissolved in water to a total of 49.5 mls. Two 40cmx15cm electrophoresis plates were taped together with 0.4mm spacers between them. To the acrylamide solution was added 300ul of 10% (w/v) Ammonium persulphate, 40ul TEMED: this was mixed and then quickly poured through the plates onto the bench. A slot-former was inserted

between the top of the gel plates immediately afterwards, and the gel left to set for 1-2 hours. Acrylamide DNA gels were run vertically in TBE buffer: sequencing gels were run at constant 25mA (800-1000volts) Sequencing gels were autoradiographed frozen: a sheet of "Clingfilm" was placed over the gel, a strip of X-ray film over this and a glass plate over the film. The whole was clipped together, wrapped in foil and frozen at -70°C for the required exposure time.

5.6.4 SDS-PAGE of histones.

This is the gel system of Laemmli (1970). Histones were separated on an SDS 4mm thick, 20x15cm gel. The separating gel was made up thus. 52.5mls of 30% acrylamide stock in water were mixed with 1.5mls of 2.5% NNbisacrylamide, 0.45mls of 20%SDS, 33.6mls of 1M Tris pH 8.8, and then 300ul of 10% ammonium persulphate and 30ul of TEMED were added immediately before the solution was poured between the gel plates. When the separating gel was set, a stacking gel, made up of 3.3mls of 30% acrylamide, 1.3mls of 2.5% NNbis-acrylamide, 0.2mls 20% SDS, 2.5mls of 1M Tris pH 6.8, 12.8mls of water, 100ul 10% ammonium persulphate and 10ul TEMED, was poured on top, and the slot-former inserted into it. Gels were run in Tris/glycine running buffer (25mM Trizma base, 192mM glycine, 0.1% SDS) at 100-120volts constant voltage overnight. Samples were loaded in SDS loading buffer: 2% SDS, 10% glycerol, 0.1M dithiothreitol, 62.5mM Tris pH 6.8, 0.002% tryptan blue. When the dye had reached the bottom of the gel, the electrophoresis was complete.

Tritium and S35 label was detected in gels by fluorography (Bonner and Laskey 1974). Gels were stained in Coomassie blue (lg coomassie brilliant blue in 227mls Methanol, 227mls water, 46mls acetic acid for 1 hour, followed by detaining in 875mls water, 50mls

methanol, 75mls acetic acid for 2-3 hours) to locate size markers, and soaked in two changes of DMSO for 30mins each with gentle shaking, then in 22% PPO in DMSO (500mls) for 3 hours. The gel was then rinsed under running water for at least 60mins, dried onto blotting paper and autoradiographed.

5.7 DNA-DNA hybridisation.

5.7.1 Southern blotting.

This is the procedure of Southern (1975b) to transfer DNA from agarose gels to nitrocellulose. Agarose gels are run with 0.1µg/ml ethidium bromide stain in or less, preferably none. The DNA was denatured n situ in the gel by soaking it for twenty minutes in 0.5M NaOH, 1.5M NaCl followed by neutralisation for 20 mins in 0.5M Tris pH 6.9, 3.0M NaCl. Typically 0.5l of each buffer was used for each 100ml agarose slab gel. The gel was laid on a sheet of 3MM chromatographic paper ('3MM') previously moistened in 20xSSC on a glass platform itself mounted in a tray. A sheet of nitrocellulose paper was laid on top of the gel, and marked so that tracks could be identified on it later. A sheet of polythene plastic was cut so as to have a window in it a few mm smaller than the nitrocellulose filter: this sheet was laid over the filter so that the filter's edges were covered. Two sheets of 3MM, previously moistened in 20xSSC were laid on top of the assembly, a pile of tissues about 5cm thick laid on top of that and a weight of ^1kg put on the top. The tray in which this assembly was mounted was filled with 20xSSC, and the whole left overnight at room temperature for the SSC to transfer the DNA from the gel to the nitrocellulose. The following day the nitrocellulose was removed, rinsed briefly in 2xSSC, dried thoroughly in air and

then baked for 150mins at 80°C.

5.7.2 Filter hybridisation.

These conditions are after Denhart (1966). These are the conditions used for all nitrocellulose-filter hybridisations. Filters were moistened in 2xSSC, then put into a plastic bag. Hybridisation buffer: 1x Denhart solution (0.02% Ficoll, 0.02%polyvinylpyrrolidone, 0.02% bovine serum albumin) was made 2xSSC: typically a 10x10cm filter required about 20mls of buffer. The radioactive probe was diluted to 0.5mls in water and boiled for 5mins. It was then added to the hybridisation buffer, and this was poured into the plastic bag, which was sealed shut and immersed in a water bath at the required temperature for hybridisation overnight. For some experiments the filters were pre-hybridised first, by incubating them under these conditions for 6-16 hours without any probe being present. After hybridisation the filters were removed from the bag, rinsed in cold 2xSSC and then washed in pre-warmed 2xSSC in a beaker in a water bath. Filters were usually extensively air-dried before autoradiography. Unless otherwise stated, hybridisation and washing temperatures throughout this study were 60°C.

5.8 Preparation of DNA from agarose gels.

DNA was electrophoresed in agarose containing 0.1µg/ml ethidium bromide to visualise unlabelled fragments, no ethidium bromide if labelled fragments were to be prepared. The DNA-containing gel slices were passed through a 21-G syringe needle to fragment them, mixed with 10x volume of Buffer A (150mM NaCl, 10mM Tris pH8.0, 1mM EDTA) and shaken at 4°C overnight to allow the DNA to diffuse out of the agarose. The agarose solids were then pelleted by centrifugation at

35000rpm for 30mins at 4°C, and the supernatant loaded onto a column of DE-52 diethylaminoethyl-cellulose equilibrated with buffer A. The column was washed with 30mls of buffer A, and the DNA then eluted from it with 2 2ml aliquots of 1M NaCl and concentrated by adding 10mls of ethanol to the eluted material in a siliconised glass tube.

5.9 DNA sequence determination.

5.9.1 Preparation of labelled fragments for sequencing.

DNA restriction fragments were prepared, labelled with Klenow polymerase, restricted so as to separate the ends and electrophoresed on a suitable agarose gel without ethidium. Bands were located by autoradiography of the wet gel. An abbreviated procedure was followed for preparing DNA from this gel. The bands were excised, mashed up in an eppendorf tube in 600ul of TE and left at 4°C overnight. The suspension was centrifuged through glass wool, the filtrates pooled and extracted with phenol once and isoamyl alcohol (not equilibrated with water) twice, made 0.2M in sodium acetate and precipitated with 2 volumes of ethanol.

5.9.2 Maxam and Gilbert sequence determination.

This was exactly as described by Maxam and Gilbert (1978). Labelled DNA was redissolved in water, aliquotted into five tubes, 10ul of sonicated calf thymus DNA was added (100ug/ml), 10ul of 0.4M sodium acetate and 50ul of ethanol. The tubes were frozen for 2hours at -70°C, centrifuged and the pellets dried and resuspended for sequencing. The chemistry was performed exactly as described. Cleaved DNA was concentrated by freeze-drying, taken up in 4ul of loading

buffer (vide supra) and run on a denaturing (urea-containing) polyacrylamide gel.

5.10 Standard buffers.

These buffers have been described elsewhere, but are assembled here for convenience.

5.10.1 Phenol

The phenol used in several procedures in this thesis was redistilled phenol equilibrated with 0.1M Tris pH 7.4.

5.10.2 Tris

The buffer referred to throughout as 'Tris' is Trizma base neutralised to an appropriate pH with HCl.

5.10.3 SSC

SSC is 0.15M NaCl, 0.015M Sodium citrate. 20xSSC, twenty times this concentration, is 175g NaCl, 88.25g sodium citrate per litre.

5.10.4 TE

TE (Tris/EDTA) is Trizma base (10mM), EDTA (1mM) pH adjusted to pH7.5 with HCl.

5.10.5 Phage Buffer

The neutral phosphate phage suspension buffer is 1.5g Potassium

dihydrogen phosphate, 3.5g sodium dihydrogen phosphate, 2.5g NaCl, 0.125g magnesium sulphate, 50ul of 1M Calcium chloride and 5mg gelatin per 500mls.

5.10.6 Denhart

Denhartz solution is 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% Bovine serum albumin. These are final concentrations to be attained in hybridisation reactions.

5.10.7 L.Broth, L.M.Broth, L.Agar and Top Agar.

L. broth is 10gNaCl, 10g Bacto tryptone and 5g Difco yeast extract per 1 litre. L.M.Broth is L.broth supplemented with 0.4% Maltose, 10mM magnesium sulphate. L.Agar is an agar gel made by autoclaving 15g/litre of agar with L.Broth. Top Agar is an agar gel made by autoclaving 7g/litre of agar with L.M.Broth.

Results

6. Construction of a gene library from X. borealis

"Just the place for a Snark" the Bellman cried
As he landed his crew with care
Supporting each man on the top of the tide
By a finger entwined in their hair.

6.1 Introduction

A 'gene library' is a collection of small, clonally derived DNA fragment derived from a single species' DNA, each of which may be isolated and perpetuated free of the rest of the genome of the donor organism. In this form they can be studied and manipulated in vitro as individual segments of the genome and not simply as part of bulk DNA. In usual practice this means a collection of plasmid-containing E.coli colonies or E.coli bacteriophage plaques which contain restriction fragments of the donor's DNA in the plasmid or phage DNAs respectively (reviewed in Dahl, Flavell and Grosveld 1981).

The cloning system I used is based on lambda-gtWES (Tiemeier, Enquist and Leder 1974), a lambda bacteriophage vector designed for cloning EcoRI restriction fragments. This lambda has had one of the five EcoRI restriction sites in 'wild-type' lambda removed by passage through an EcoRI - restricting E.coli (Murray and Murray 1974). It also carries a deletion (Nin5 - see Davidson and Szybalski 1971) which removes another EcoRI site. Thomas, Cameron and Davis (1974) deleted a small central EcoRI fragment from this phage to produce a lambda which was incapable of lysogenic growth and only just large enough to be packaged into lambda phage particles. As mentioned above (section 4.2.1), lambda DNA above or below certain size limits cannot be

6. Gene library

accommodated in infectious particles by the virion assembly proteins, and so cannot form the chromosome of a viable phage strain (Hohn 1975, Hohn and Murray 1977). If the remaining small EcoRI fragment is deleted, the size of the DNA drops below this packaging limit. Thus the cloning strategy for use of this vector is to biochemically delete the central EcoRI fragment leaving the two outer parts of the molecule, the 'arms', ligate these 'arms' with EcoRI-digested donor DNA, package the result into infectious phage particles and plate it on a suitably permissive E.coli (Thomas, Cameron and Davis 1974). Only phage containing sufficient DNA to counteract the deletion from the original chromosome will be able to produce viable phage.

Enquist et al (1974) subsequently introduced amber mutations into the E, S and W genes of this phage, making it non-viable when plated on any but laboratory suppressor strains, and Tiemeier, Enquist and Leder (1974) altered this phage to produce the final lambda-gtWES.lamB Δ B by replacing the original EcoRI-C fragment in the phage by EcoRI-B.

The general procedure for using lambda vectors for isolating eukaryotic genes has been described by Maniatis et al (1978). The strategy used in this work, a simplified version of the Maniatis procedure, is summarised in Fig 6.1. This section deals with the construction and characterisation of a gene library from Xenopus borealis DNA, and the next deals with the isolation from that library of histone gene clones.

6.2 Methods

6.2.1 Preparation of lambda gtWES arms

Fig 6.1

Lambda-gtWES cloning strategy

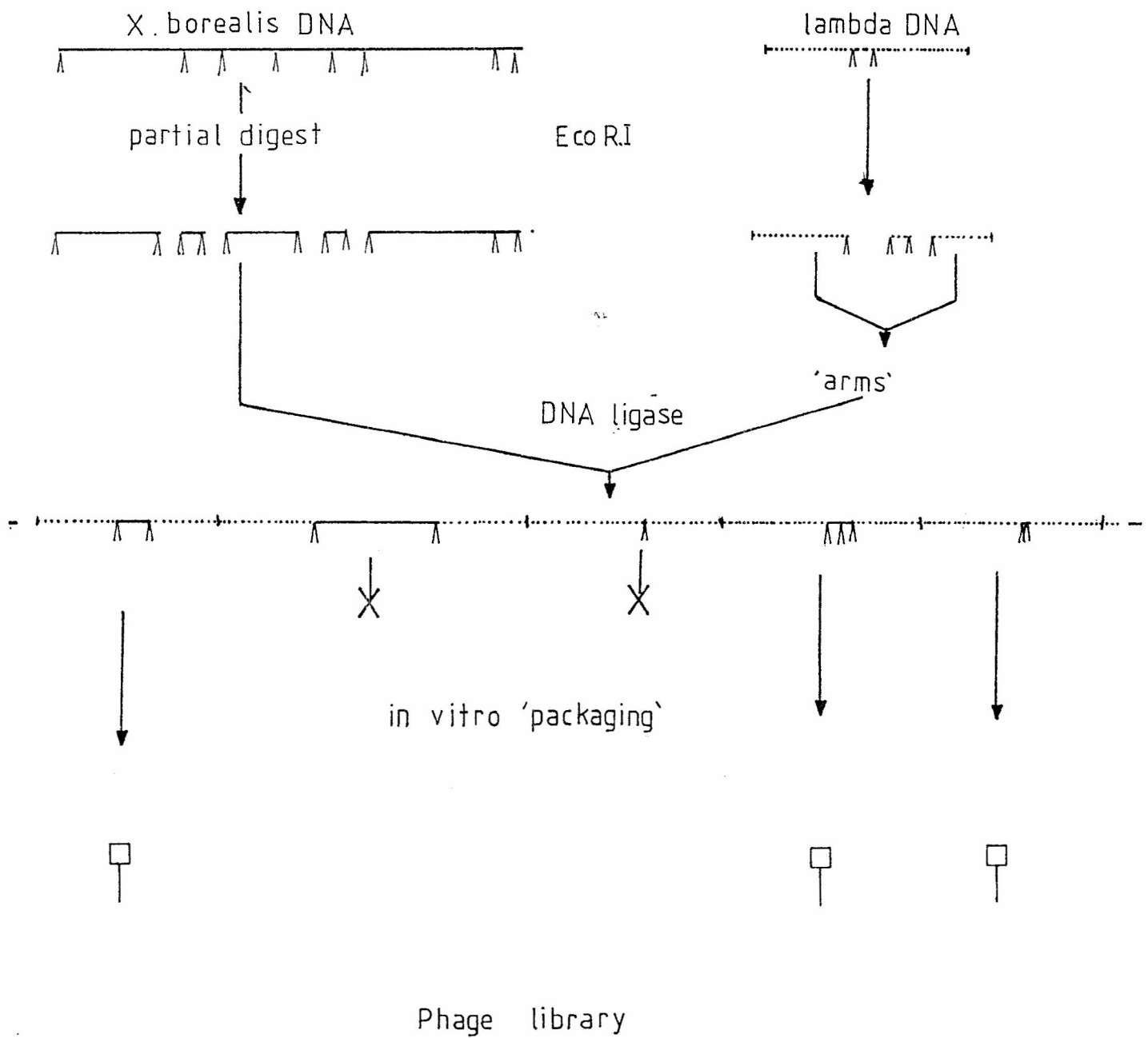


Fig 6.1 The cloning strategy used in this study. Solid lines - X.borealis genomic DNA. Dotted lines - lambda DNA. Arrows - EcoRI sites. 'In vitro packaging' is the incorporation of DNA into phage particles using a bacterial lysate and exogenously added DNA. See text for details and references.

6. Gene library

100 µg of lambda-gtWES DNA was digested with EcoRI. A sucrose step gradient was constructed by layering 3ml aliquots of 20, 15, 10 and 5% sucrose (w/v made up in 0.1M NaCl, 10mM Tris pH7.4, 1mM EDTA) into an MSE 14-ml swing-out rotor centrifuge tube in that order and leaving them to diffuse overnight at room temperature. The digested DNA was layered onto the top of the gradient and the ensemble was centrifuged for 16 hours at 22,000 rpm in an MSE 6x14s.o. rotor. The gradient was fractionated into 700µl samples and an aliquot of each was electrophoresed on a 0.7% agarose-ethidium gel to locate the 'arms'. Fig 6.2 shows a typical result. To the pooled fractions were added 1/10 volume of 2M sodium acetate, 2 volumes of ethanol to precipitate the DNA.

6.3 results

6.3.1 Construction of genomic gene library from *X.borealis*.

Three initial attempts to make a gene library from *Xenopus borealis* were not successful.

My first attempt was to use Cosmid vectors to produce a plasmid gene library. This was unsuccessful, despite several attempts and the success of parallel 'control' experiments. These results, and a possible interpretation of them, are presented in Appendix 1. All subsequent attempts used the lambda-gtWES cloning system outlined above.

The second attempted to fractionate EcoRI-restricted *X.borealis* DNA on an RPC-5 column, and then size-fractionate the result on an agarose gel prior to cloning in lambda-gtWES. The RPC-5 column

Fig 6.2

Isolation of lambda 'arms'

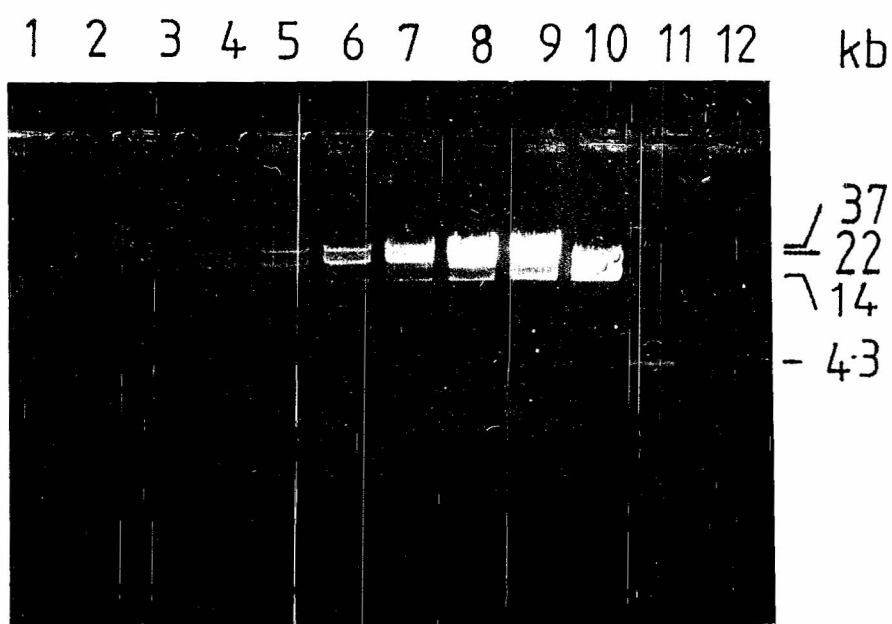


Fig 6.2 0.7% agarose gel of 30ul samples of gradient fractions from an 'arms' preparation to prepare the large terminal EcoRI 'arms' from lambda-gtWES as described in section 6.2. The gel is stained with ethidium bromide. The track numbers are the same as the fraction numbers : 1=bottom of tube, 12=top. Fractions 7,8 and 9 were pooled as an 'arms' preparation in this experiment.

6. Gene library

chromatographic method is described in section 11.2. We will just note here that this is the reason that the RPC-5 chromatography was carried out.

Further studies suggested that trace impurities in the X.borealis DNA could severely hamper the construction of a gene library, so new DNA was prepared and used in the construction of the genomic library, and all other experiments in this thesis. I would like to stress this point in view of the variability found between individuals of X.laevis and to a lesser extent X.borealis (Turner and Woodland 1982b, van Dongen et al 1981) : the genomic DNA used in all the experiments in this thesis with the exception of the RPC-5 analysis described in section 11 was made from in one batch from three female X.borealis.

26µg of X.borealis DNA was digested with 50u of EcoRI for 1 hour. Another 26µg was digested with 5u. These digests were pooled and used as the EcoRI-digested DNA to be cloned. The reasons for using partial restriction digests of genomic DNA for gene cloning are given by Maniatis et al(1978). Briefly, it is never certain whether any particular gene or gene cluster is going to be cut by a restriction enzyme. So, in order to be sure that the intact gene or gene cluster exists in the gene library, a partial digest is performed. Thus, for all restriction sites, there exist in the digest fragments that site still uncut within them. This poses some problems in interpretation of results: if two EcoRI fragments are found in the same recombinant lambda, did they originate from adjacent genomic EcoRI fragments in the genome or did they fortuitously come together in the ligation? Mapping the genes in genomic DNA by Southern blotting can answer this question but this was not needed in this study. Fig 6.3 shows the electrophoresis of a sample of these

Fig 6.3

EcoRI digestion and ligation

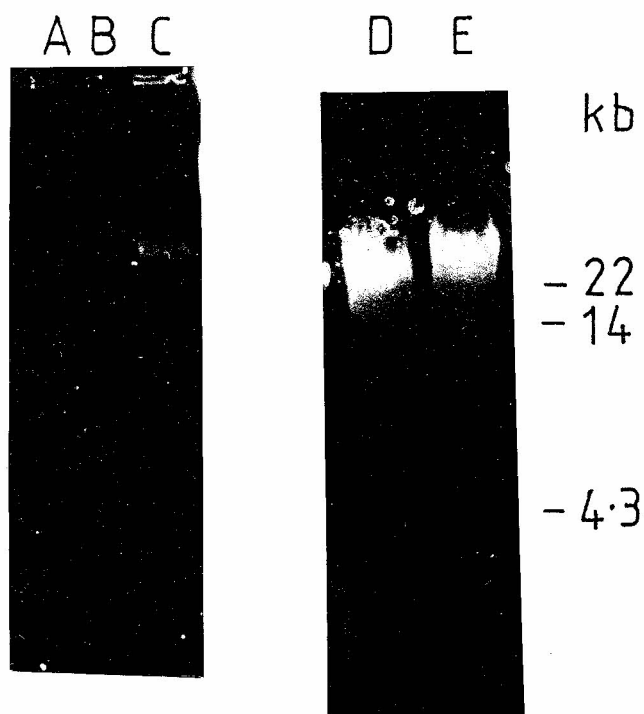


Fig 6.3 Restriction and ligation of DNA for construction of a genomic library. Tracks: A,B and C EcoRI digests of genomic X.borealis DNA A)4u/μg, B) 1u/μg, C) 0.2u/μg, all incubated for 1 hour (one unit digests one microgram to 95% completion in 1 hour). D)and E) - ligation of partial digest of X.borealis DNA with lambda-gtWES 'arms' as described. D=before reaction, E=after. 0.7% agarose/ B) Autoradiograph of 0.7% agarose ethidium gel.

restriction digests.

3.1µg of this EcoRI-digested DNA was mixed with 11.0ug of lambda-gtWES 'arms' in a total volume of 84µl in 10mM magnesium chloride, 66mM Tris pH7.6, 10mM Dithiothreitol, and heated to 70°C for 5' and then incubated at 42°C for 30' to anneal the lambda arms together via their cohesive ends (Hedgpeth, Goodman and Boyer 1972, Maniatis et al 1978). The in vivo substrate from phage DNA packaging is concatomeric lambda molecules (Hohn and Hohn 1974, Hohn 1975), and it is unclear whether monomeric lambda or cosmid molecules can be packaged into phage particles by the in vitro system used here (Hohn and Murray 1977, Maniatis et al 1978, Enquist and Sternberg 1979). So the lambda-gtWES 'arms' were joined not only to EcoRI fragments of donor DNA but also to other pairs of 'arms' via their cohesive ends. A ratio of 3 donor:11 lambda DNA was used so that EcoRI-donor DNA fragments (mean length @6kb) and lambda-gtWES 'arms' (total unit length @40kb) should be in a near to equimolar ratio. Too low a ratio of donor:lambda DNA can cause an excessive background of non-recombinant lambda-gtWES.B regenerated from traces of the B-fragment present as a contaminant in the 'arms' preparation, so exact molarity was not sought.

The ligation mix was then made 1mM in ATP and incubated with 12u of DNA ligase at 10°C for 36 hours. The effect of this ligation is also shown in Fig 6.3

The ligated DNA was packaged in 20 separate reactions using the packageing system of Enquist and Sternberg (1979) described in section 5.5.3 (each reaction used 5 times the volume of reagents recommended in the recipe in section 5.5.3, and 2.2µl = 0.3µg of ligated DNA per reaction). This produced a total of 425,000 phage, an

6. Gene library

efficiency of $\approx 140,000$ pfu/ μ g donor DNA. This is an average efficiency for such a procedure (Ish-Horowitz per com, Maniatis et al 1978, Patient et al 1980).

6.3.2 Amplification of the library.

The primary library contains a good representation of the number and size of the EcoRI fragments found in the starting DNA. However each cloned genomic fragment is only there in one copy, so in order to screen for a low copy-number gene a substantial amount of this library needs to be plated out, in which form it will last only a few months and can be screened only a few times (Benton and Davis 1977). So I amplified the library by plating it out and harvesting the phage from the plates, thus achieving a multiplication of each original phage by 1000 to 10000 times (Benton and Davis 1977, Maniatis et al 1978). This procedure has its drawbacks, as we shall see in the next section. The primary library of 425,000 phage was diluted with 30mls of Phage Buffer and plated onto 400 9cm L.Agar plates on LE392. This was a rather excessive number of plates: preliminary platings had suggested that the library contained more phage than actually appeared as plaques in this amplification step. The plaque size on these plates was quite variable, betokening less than optimal plating conditions. On average the plates held ≈ 1000 phage each.

The library was amplified by growing the phage on plates rather than in a bulk liquid culture to reduce the distortion of the representation of some sequences in the library during amplification. In liquid culture phage are in direct competition for host bacteria and nutrients throughout the growth period during which several rounds of replication take place, and so phage with selective

6. Gene library

disadvantages or advantages due to properties of their insert DNA will be severely under- or over-represented in the resulting amplified library. In plaque growth on plates there is no direct competition between phage, so this sort of distortion should not occur. Severe cases of distortion can still occur in libraries amplified by this method, however (May, Weber and Westley 1982, Patient et al 1980, and also *vide infra*).

The Top Agar from these 400 plates was scraped into 1.2 litres of Phage Buffer containing 2mls of chloroform and stirred at room temperature for 30mins. The agar and bacterial debris were centrifuged out at 6000 rpm for 20' at 4°C in a Sorval 6L centrifuge and re-extracted with 400 mls of Phage Buffer. The pooled supernatants were made 10% w/w in Polyethylene glycol 6000, 0.5M NaCl and left at 4°C overnight. This mixture was centrifuged at 5500rpm in an MSE 6x250 rotor 4°C, 10' and resuspended in 70mls of Phage buffer. This concentrated stock was the amplified library used in subsequent experiments.

6.3.3 Demonstration that the library is recombinant.

Before proceeding to screen the library for specific recombinants, I checked that it did contain recombinant phage. This can be done three ways.

- 1) The presence of non-vector phage can be demonstrated by isolating and identifying a random selection of clonally derived phage.

- 2) The heterogenous nature of the entire library can be demonstrated either by characterising the phage or their DNA.

3) The library may be screened for sequences present in the donor DNA but not in the vector.

In the first library attempt I tried approaches 2 and 3. In particular the failure of rDNA clones to detect any ribosomal RNA genes in the library was a strong indication that it contained little if any X.borealis DNA. Approaches 1 and 2 were used on the successful library initially.

Several procedural problems were encountered when growing up bulk phage in liquid culture: the procedure given in section 5.2.4 is the result of considerable adjustment of the conditions of Phage growth and harvest made during this phase of the work.

An aliquot of the gene library was plated out at low phage density and eight plaques picked at random. These were designated BPX1 to 8. Another aliquot of @100 phage was taken, plated out and the phage from the whole plate harvested and designated BPL. These nine inocula were cultured in 500ml liquid cultures and the phage preparatively banded on CsCl as described in section 5.2.5. A photograph of the gradients on which BPX2, BPX3 and BPL were loaded is shown in Fig 6.4 . Note that the phage band in the BPL tube is significantly broader than that in BPX3, denoting that it is a collection of many different phage bands. BPX2 is in fact three phage species, again showing the heterogeneity of the phage in the library. The physical size of the phage particles is roughly fixed (Hohn, Flick and Hohn 1975), so the density of different phage is dependant on their mass alone. As the protein content of the phage is constant, their mass is determined by the length of DNA in them (Davidson and Szybalski 1971). Thus heterogeneity of banding position seen in Fig 6.4 directly reflects heterogeneity of insert sizes in the

Fig 6.4

Library phage preparations

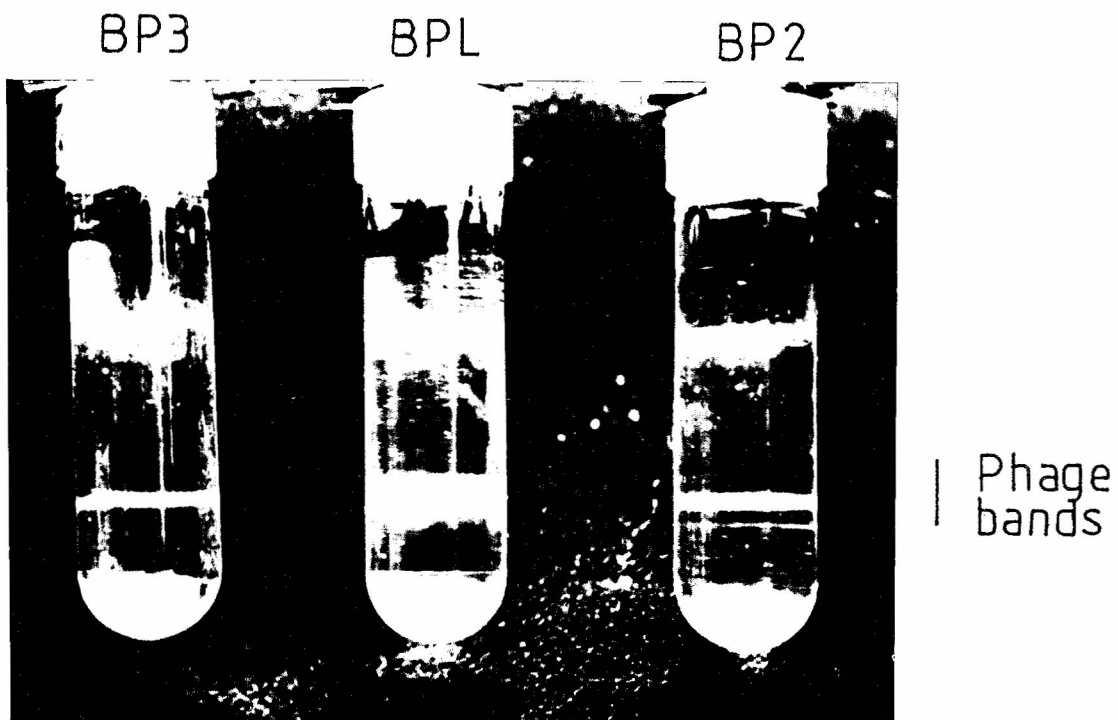


Fig 6.4 1 litre cultures of several phage were grown, concentrated, and banded on CsCl gradients as described in the text. These are the 10ml capacity MSE centrifuge tubes after 48 hours centrifugation at 40,000rpm. Photographed under side illumination. The prominent bands near the bottom of the tube are phage bands. Details of phage samples (BP3, BPL and BP2 as labelled) given in text.

recombinant phage.

In a subsequent experiment (see section 10.3.5) ten recombinant lambda phages were isolated from the library, grown and banded on CsCl at the same time: these gradients are shown in Fig 6.5 . This shows more clearly the heterogeneity of between individual clones. Note that one of these 'recombinants' is the vector lambda-gtWES.

From this it follows that it is unnecessary to show that molecules other than lambda-gtWES vector are present in the gene library. However we can note that I have directly demonstrated that these non-vector molecules have X.borealis DNA in them by showing the presence of X.borealis histone (section 7.3) and intermediate repetitive (section 10.3.5) DNA in recombinant phage from this library, and Mr. G.Cross has also isolated X.borealis Actin gene clones from it (per com). The recombinants therefore contain X.borealis DNA and not E.coli DNA or rearranged vector sequences.

6.3.4 Size of the inserted DNAs in the library

The most comprehensive way to quantify the size of the inserts in a gene library is to enumerate the phage of different densities as separated on CsCl gradients, and relate this to the size of the insert by simultaneously banding several control phage of known genome size (Maniatis et al 1978). This experiment was not performed on this library. However a number of recombinants have been isolated from the library, at random as far as their insert sizes are concerned, and an estimate of the mean insert size may be obtained from these. Their sizes are listed in Fig 6.6 . The mean size of 6.1 for the insert lengths is midway between the limits of lambda-gtWES' cloning capacity (1kb-12kb Thomas, Cameron and Davis 1974, Tiemeier

Fig 6.5

XIR phage preparations

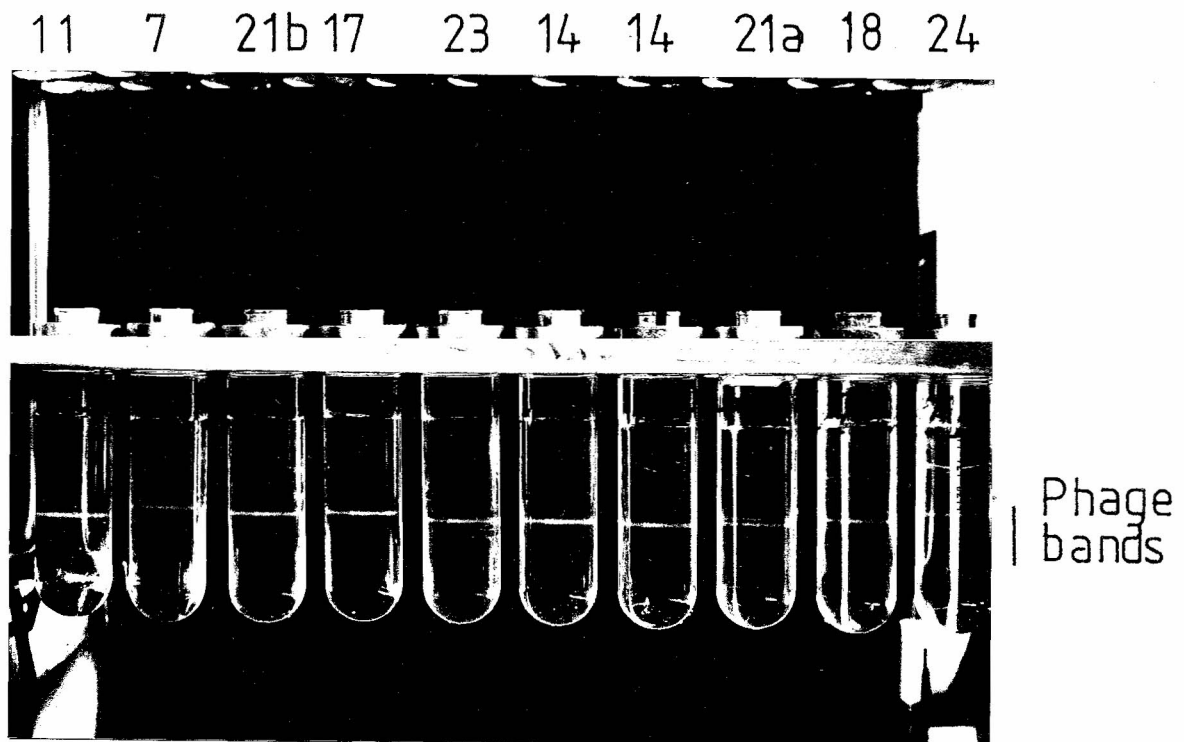


Fig 6.5 500ml cultures of several phage were grown, concentrated and banded on CsCl gradient as described. These are 10ml capacity MSE centrifuge tubes after 48hours centrifugation at 40,000 rpm. Photographed under side-illumination. The bands appear more blue here than in life due to the different colour responsivity of the human eye and photographic emulsion. Note the variable banding positions of these different phage. These are XIR clones (number as marked in the figure) whose isolation is described in section 10.

Fig 6.6

		Recombinant phage inserts sizes (kb)				
phage	section	R1				insert
51	7	5.6	1.7			7.3
161		3.5	1.7	1.0	0.5	6.7
171		9.0				9.0
191		4.9				4.9
991		3.3	1.9			5.2
921		5.5				5.5
9912		4.8				4.8
Xbh1		8.5				8.5
XIR7	10	2.8				2.8
11		9.5				9.5
17		lambda gtWES				
18		6.0				6.0
21a		6.8	1.4			8.2
actin	35A	G Cross		2.7		
	5AP	per com		3.5	1.7	5.2
means		4.2				6.1

Fig 6.6 Characteristics of recombinant phage isolated from X.borealis gene library. The first seven are recombinants isolated during the screen for histone gene clones described in section 7, and are all not histone gene clones. Xbhl is the single type of histone gene clone isolated in section 7. The XIR clones are various repetitive DNA recombinants described in section 10: in both cases the multiple re-isolates of Xbhl have not been included. 35A and 5AP are actin gene clones isolated from this gene library by Mr. G.C.Cross. 'RI' - EcoRI fragment size. 'Insert' -total insert length.

6. Gene library

and Leder 1974). Note that this is not the mean of the EcoRI fragment sizes in the clones: this latter figure is 4.1kb. The number-mean size in the original digest is 6kb, showing that the original digest has not been cloned representatively. As all EcoRI fragments over 12kb long are inelligible for cloning in lambda-gtWES, this is not surprising. This size distribution is similar to what I would expect of the reagents used (see Appendix 2).

6.4 Conclusion.

In this section I have described the construction of a gene library from X.borealis genomic DNA. The library is derived from 425,000 original phage with a mean X.borealis DNA content of 6.1kb. This gives a 60% chance of finding a unique sequence in this library (Maniatis et al 1978). Of 10 phage which have not been selected for their content of a particular sequence, I.E. which were negatives accidentally isolated in subsequent screens, one is probably a non-recombinant vector molecule. Together with the lack of any visible vector band in the CsCl preparation of the whole library (Fig 6.4) , this suggests that the library is at least 90% recombinant.

7. Selection of Histone H4 gene-containing recombinants.

They sought it with thimbles, they sought it with care
They pursued it with forks and hope
They threatened its life with a railway share
They charmed it with smiles and soap

7.1 Introduction.

In order to study unique genes by cloning recombinants from a gene library, it is essential to be able to screen a large number of recombinants rapidly for the gene concerned. In this section I describe how I used the mass screening procedure of Benton and Davis (1977) to isolate histone gene clones from the gene library whose construction I described in section 6.

There are ©90 H4 histone genes in X.borealis (Hilder et al 1981, Turner and Woodland 1982b), most of which are in the uniform major repeat (Turner and Woodland 1982b). The C-value is 3200 Mb in this species (Theibaud and Fischberg 1977). So we would expect about 1 in 5000 plaques to contain an H4 gene (assuming an average insert size of 6kb). The H4-coding cDNA clone pcX1H4W1 was available at this time (Turner and Woodland 1982). Both hybridisation studies using Southern blots (Turner and Woodland 1982b) and the known sequence conservation of H4 gene- and protein- sequences (Isenberg 1979, Turner and Woodland 1982, Busslinger et al 1982) made it most probable that this clone would be closely homologous to a number of X.borealis H4 genes, probably to all of them. This cDNA clone was therefore to be used as a probe. The various clones isolated would be placed into groups depending on whether they shared a common restriction pattern

or not, ie whether they were likely to be derived from the same genomic sequence.

7.2 Methods

All the methods used in this section have been described in section 5.

7.3 Results

7.3.1 Screening the gene library.

27 square L.Agar plates (9cmx9cm) were plated with 4200 plaques from the amplified gene library each. The plating conditions varied slightly from standard: 9mls of Top Agar was used per plate to give a thicker Top Agar layer. This was found to give more reproduceable plaque sizes. The plates were all dried extensively in an oven under U.V. light before plating to reduce surface moisture and to ensure the Top Agar bonded effectively to the L.Agar during the plaque lift procedure of Benton and Davis (1977) . A 28th plate was plated with 3000 plaques from the gene library and 200 of lambda Xlh7, a recombinant from X.laevis isolated by Dr. R.W.Old and known to contain an H4 gene. This latter was a control experiment to ascertain that the plaque-lift and hybridisation steps had worked.

Nitrocellulose filters were prepared from these plates according to Benton and Davis (1977). These were hybridised with nick-translated pcX1H4W1 in 3xSSC, 1xDenhardt's at 68°C, and washed in 2xSSC at 68°C before exposure. A sample of the result is shown in Fig 7.1

7. Screening for H4

The alert reader will notice that each of the filters shown in Fig 7.1 shows @200 spots, all of which correspond in morphology to what is expected of hybridising phage DNA, and most of which could be aligned with plaques on the original plates. This means that about 1 in 100 phage in the original library contain H4 sequences, and not 1 in 5000 as we expected. This suggests contamination of the gene library by an exogenous phage which contains H4 sequences, for example the Xlh7 used as a positive hybridisation control. The way to establish that this is so is to isolate the hybridising species and show that it is identical to the species suspected to be the source of the contaminant. Showing that it is not a contaminant is harder, as it requires that all possible contaminants be identified and eliminated. I will return to this point later.

I selected a number of these positive clones and re-screened them at a density of 200-500 phage/plate with the X.borealis H4 cDNA clone pcXbH4W1 (Turner and Woodland 1982). Hybridisation conditions were as above, except that synthetic poly-A was added to prevent the hybridisation of the oligo-A:T tracts in the cDNA clones to spacer or satellite DNA: this addition made no difference. I selected 13 clones, several isolated twice from the same plate to ensure that the hybridising species had been isolated. They were picked in a plug of agar with the blunt end of a Pasteur pipette, replated at 20-60 phage/plate, re-picked to ensure clonal purity and kept in 1ml of Phage Buffer + 1 drop of chloroform as a stock. Aliquots from the phage buffer were used for subsequent experiments. These stocks are quite stable for several months if kept at 4°C.

7.3.2 Only one type of clone, Xbhl, is a true positive.

These potential H4-containing clones were grown in 500ml liquid

Fig 7.1

1st screen for histone genes

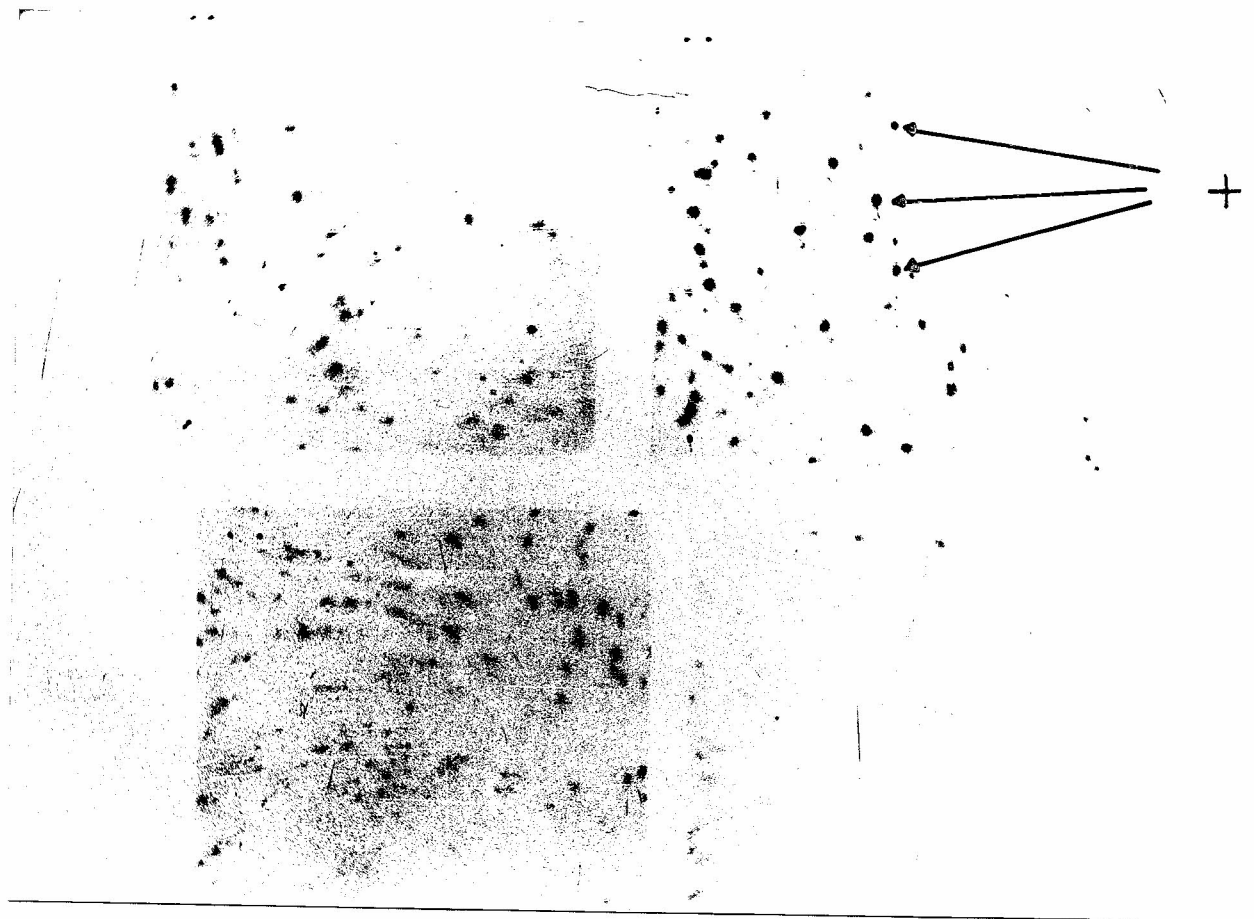


Fig 7.1 Primary screen of the gene library for H4-containing gene clones. Filters prepared after Benton and Davis (1977) were hybridised with nick-translated pcX1H4W1 at 60°C in 3xSSC and washed at 60°C in 2xSSC. This figure shows the autoradiograph of four typical filters '+' = spots marking hybridising phage plaques.

7. Screening for H4

cultures for initial analysis. The EcoRI digestion patterns of these phage are shown in Fig 7.2 . It can be seen that 7 out of the 12 clones have very similar insert sizes, while the others are different from them and from each other.

To confirm that these are histone H4-containing clones, these single DNAs were used in two series of experiments. Firstly, pcX1H4W1 was Southern blotted onto strips of nitrocellulose, together with lambda cI857S7 ('wild type lambda') digested with EcoRI as a size marker and hybridisation control. These strips were hybridised with nick-translated DNA from these recombinants. The results, shown in Fig 7.3 show that a number of these phage do not hybridise to pcX1H4W1 DNA, and hence cannot be H4 recombinants. An example is clone 921. Clone 2252 is an example of a DNA which does hybridise to pcX1H4W1, and hence is an H4 positive. On the basis of this screen the phage can be grouped into three groups. 1) Phage which hybridise to pcX1H4W1 and are similar to clone 2522 on the basis of EcoRI restriction pattern: these five were designated Xbh1 to 5. 2) Phage which hybridise to pcX1H4W1 but are not similar to 921: these were designated Xbh6 to 9. 3) Phage which do not hybridise to pcX1H4W1. These negatives were put aside.

Subsequently it was shown that all of group 2 were really negatives as well. The clones in this group were digested with BamHI, XbaI and PstI electrophoresed on 0.7% agarose, Southern Blotted and probed with pcX1H4W1. Only Xbh1, included as a control, showed significant hybridisation (see Fig 7.4). The apparent positive results shown by Xbh6 to 9 in the previous screens must therefore be spurious, as they do not contain sequences which hybridise significantly to an H4 cDNA. It is curious that they came up positive in the three screens prior to this one. This is probably an effect of

Fig 7.2

Recombinants isolated in 1st screen

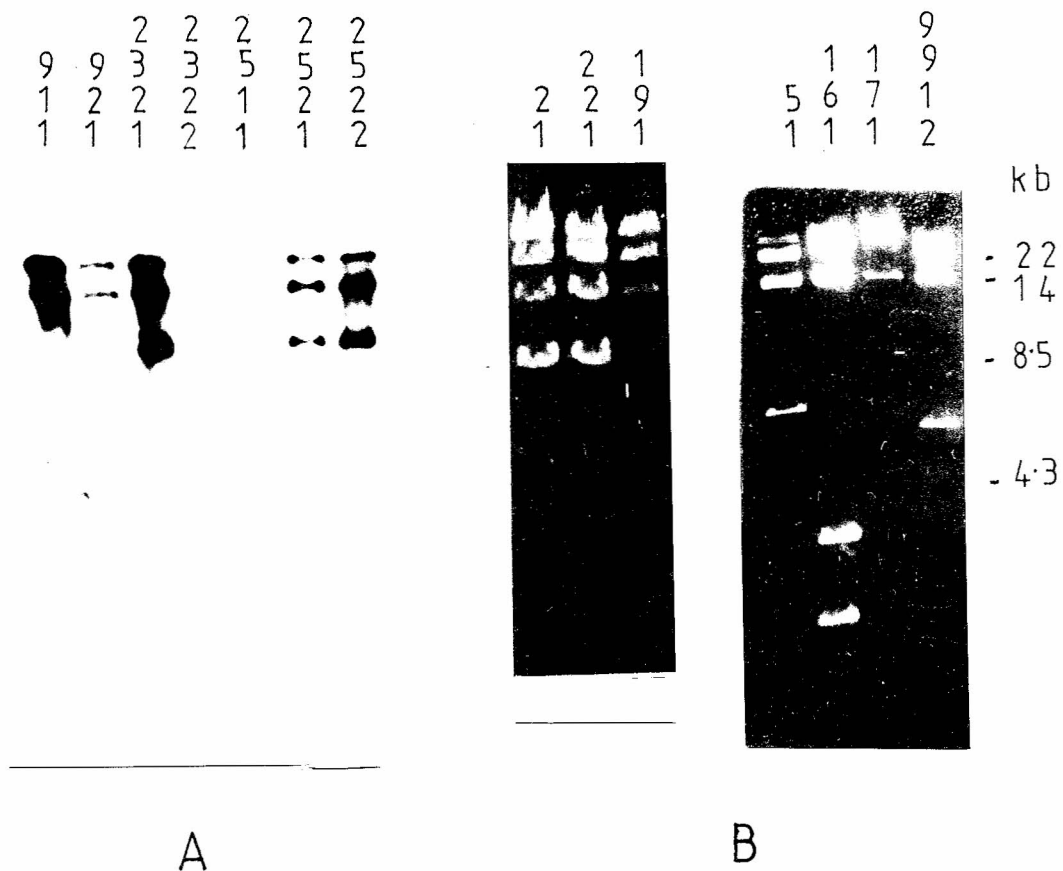


Fig 7.2 EcoRI digests of the DNA from recombinants isolated in primary screen. A) Recombinants were nick translated, digested , separated on 0.7% agarose and the gel autoradiographed . B) More recombinants were EcoRI digested, separated on 0.7% agarose gels and stained with ethidium bromide. Figures at the head of each track are the numbers used to designate the original isolates: the same numbering system is used in Figs 6.6 and 7.3. Note that 2321 and 2322, and 2521 and 2522, are duplicate pairs of isolates from one initial plaque in the primary screen.

Fig 7.3

3rd screen for histone genes

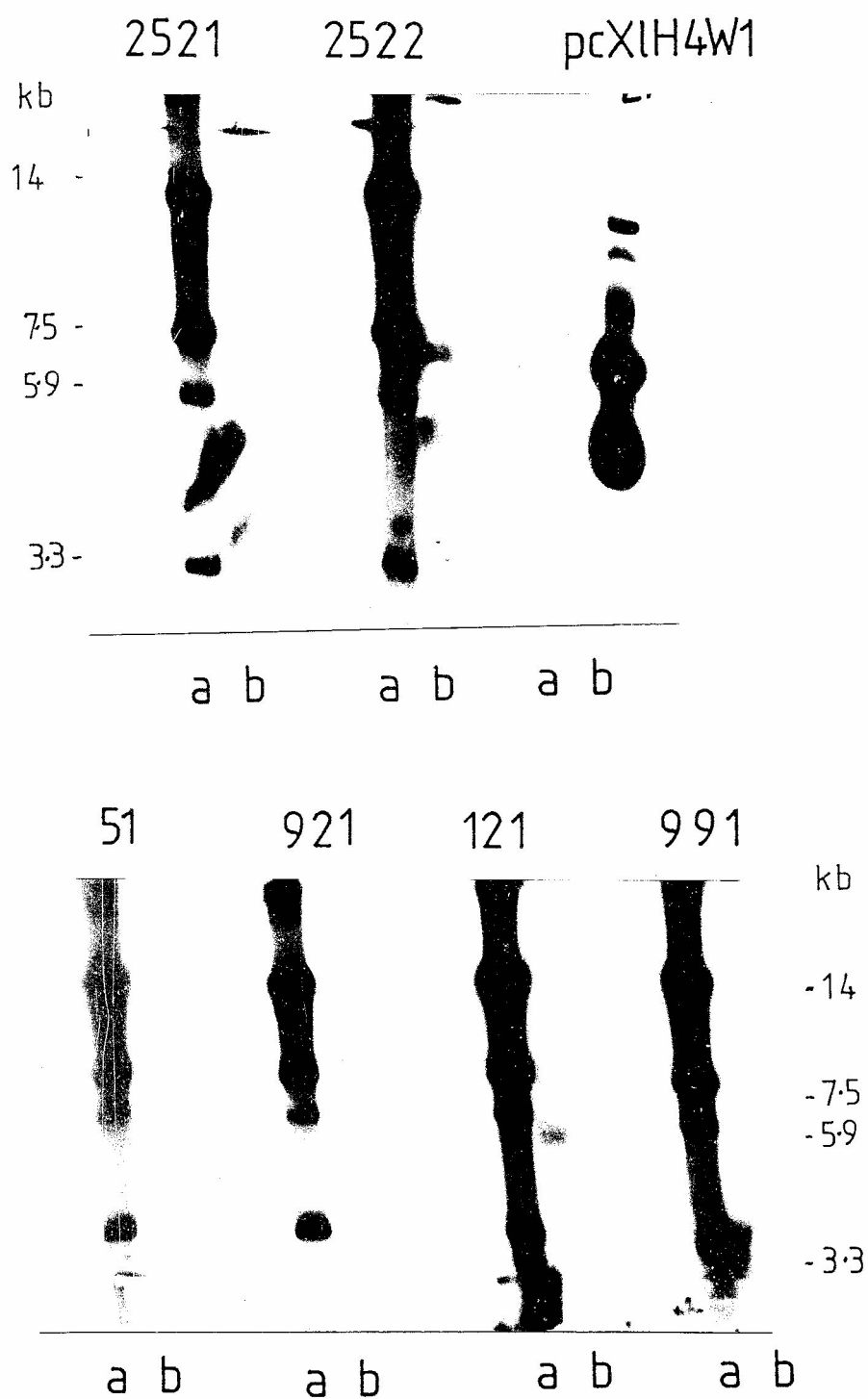


Fig 7.3 3rd screen of recombinant clones. The results for six lambda clones (2521, 2522, 51, 191 and 991) and pcX1H4W1 are shown. Lambda cI857S7 was EcoRI digested and run in a series of 0.7% agarose gels in track 'a'. pcX1H4W1 was run in the parallel track 'b'. The gels were Southern blotted and probed with nick-translated recombinant lambda, or pcX1H4W1 as a 'positive control'. Hybridisation and washing were at 60°C, 2xSSC. The object was to see if the recombinants hybridised to pcX1H4W1, i.e. if they contained H4-homologous sequences.

Fig 7.4

4th screen for histone genes



Fig 7.4 Recombinant clones were digested with A) BamHI, B)SstI, C)PstI and electrophoresed on a 0.7% agarose gel, Southern blotted and probed with pcX1H4W1. Hybridisation and wash were at 60°C, 2xSSC. Clones are Xbh1 and Xbh6 to 9 as indicated at the top of each track.

non-specific binding of the hybridisation probe to any DNA bound to nitrocellulose (Britten, Graham and Neufeld 1981, Kafatos et al 1980), and shows the need for careful screening of any recombinant to eliminate hybridisation artefacts. It is also possible that these represent highly diverged or extensively variant H4 genes: the "hybridisation" of LF53 to every large restriction fragment in these clones to the same minimal level makes this improbable, however.

7.3.3 The Xbhl to 5 clones are all identical.

The clones Xbhl to 5 have very similar EcoRI digestion patterns. Careful gel techniques allowed me to show that their insert sizes were essentially identical. I digested these clones with a number of other enzymes: the pattern produced in every case was also identical (Fig 7.5). These restriction enzymes can therefore show no sequence differences between these five clones. Thus, to within 0.3% sequence difference (see Appendix 4) they are the same. In view of their anomalous abundance in the gene library, I am sure that this means that they are identical. I designated them all Xbhl, and the rest of the work in this thesis is based around one of these Xbhl isolates.

7.3.4 Abundance of Xbhl in the gene library.

Why is Xbhl present in such enormous numbers in the gene library? To look ahead to section 9, it is not a clone of the major histone gene cluster: the gene spacing and restriction maps are quite different. Nor is it one of the semi-minor clusters detected by Turner and Woodland (1982b) in X.borealis. Hence it must either represent a minor histone gene cluster present in 1 to 5 copies per

haploid set, or a contaminant not present in X.borealis at all. Which is clearly of some importance to the interpretation of this work, so I tried to identify the source of any potential contamination and any possible route by which a phage could become disproportionately numerous in the gene library.

To check whether this is an effect of a selective amplification acting on phage with 8 - 9kb inserts during the gene library construction due to their relatively greater efficiency of ligation or packaging, I have attempted to model the DNA ligation and packaging stages of the shotgun cloning in a semi-empirical computer model, presented briefly in Appendix 2. The results of this model accord with the known characteristics of the library described in section 6.3.5, but only suggest that an 8.5kb EcoRI fragment should be at a selective advantage over the rest of the genome by 1 to 3 times during the construction of the library. Xbh1 and the very similar Xlh8 have been anecdotally observed to grow reproducibly well in liquid bulk culture, but it is doubtful, even if this does represent a phage genuinely well adapted to growth on LE392, whether this could explain an amplification over other phage of more than 5 to 10 times during the library amplification step. Thus all identifiable factors cannot combine to explain more than a 30-fold amplification of this clone over its sisters. If the original sequence of which Xbh1 is a representative was present in the genome in 1 to 3 copies, its presence in 1 in 100 clones in the gene library represents an amplification of 2000 times. Thus identifiable factors fail to explain the observed amplification by at least two orders of magnitude.

This suggests that Xbh1 is a contaminant from another cloning experiment, most probably from those of Dr. R.W.Old, who used the

Fig 7.5

Restrictions of Xbh 1-5

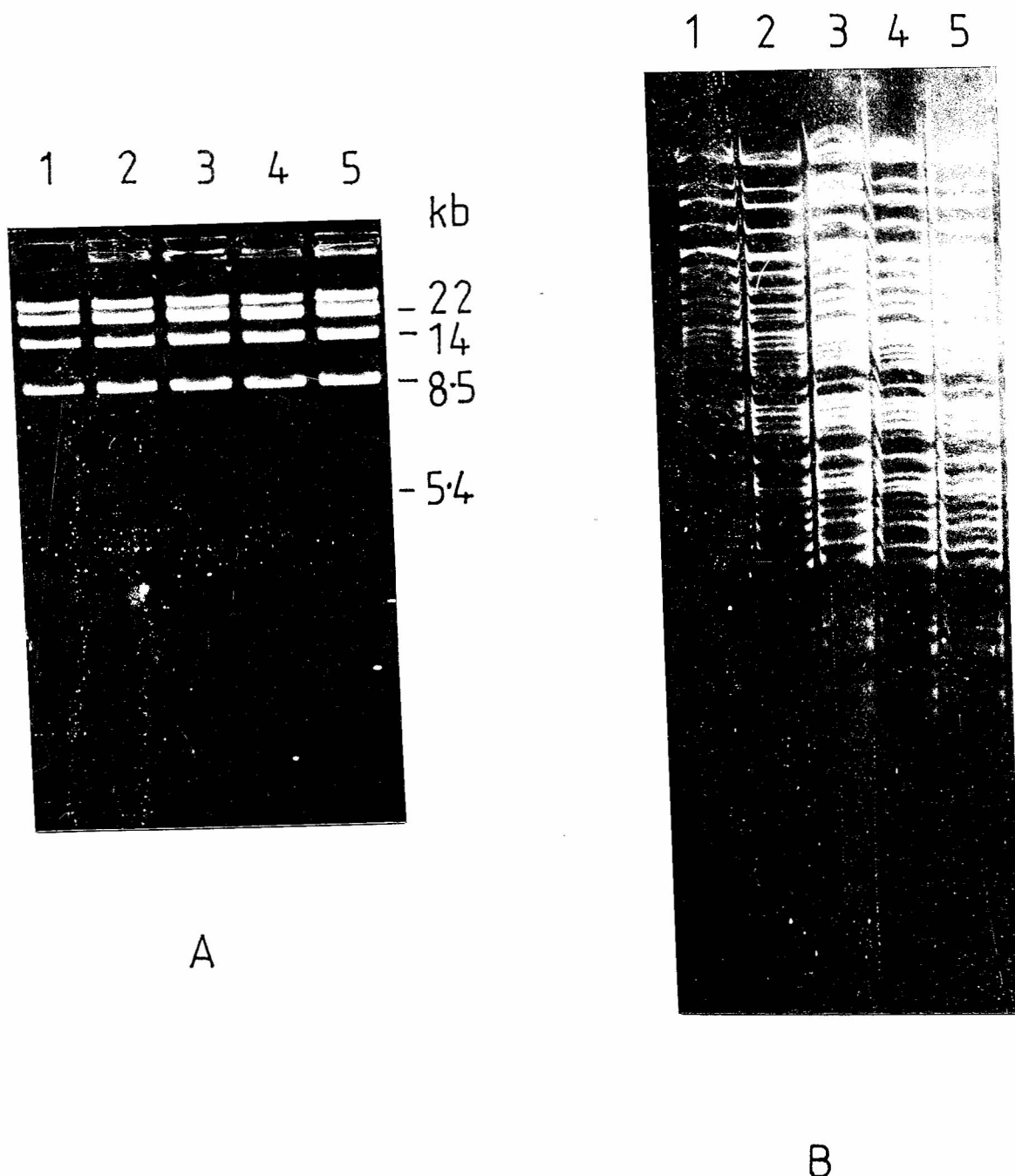


Fig 7.5 A) 0.8% Agarose gel of EcoRI digests of Xbh1 - 5 ('1' to '5' as indicated). B) 8% polyacrylamide non-denaturing gel of HinFI digests of Xbh1 to 5. Both gels stained with ethidium.

7. Screening for H4

same vector and cloning strategy to clone the histone genes of X.laevis. However it is not similar to the positive control used in the initial screen, Xlh7, and, although very similar to several other X.laevis clones which Dr. Old has isolated from his gene library, is not identical to any of them (Section 13 below).

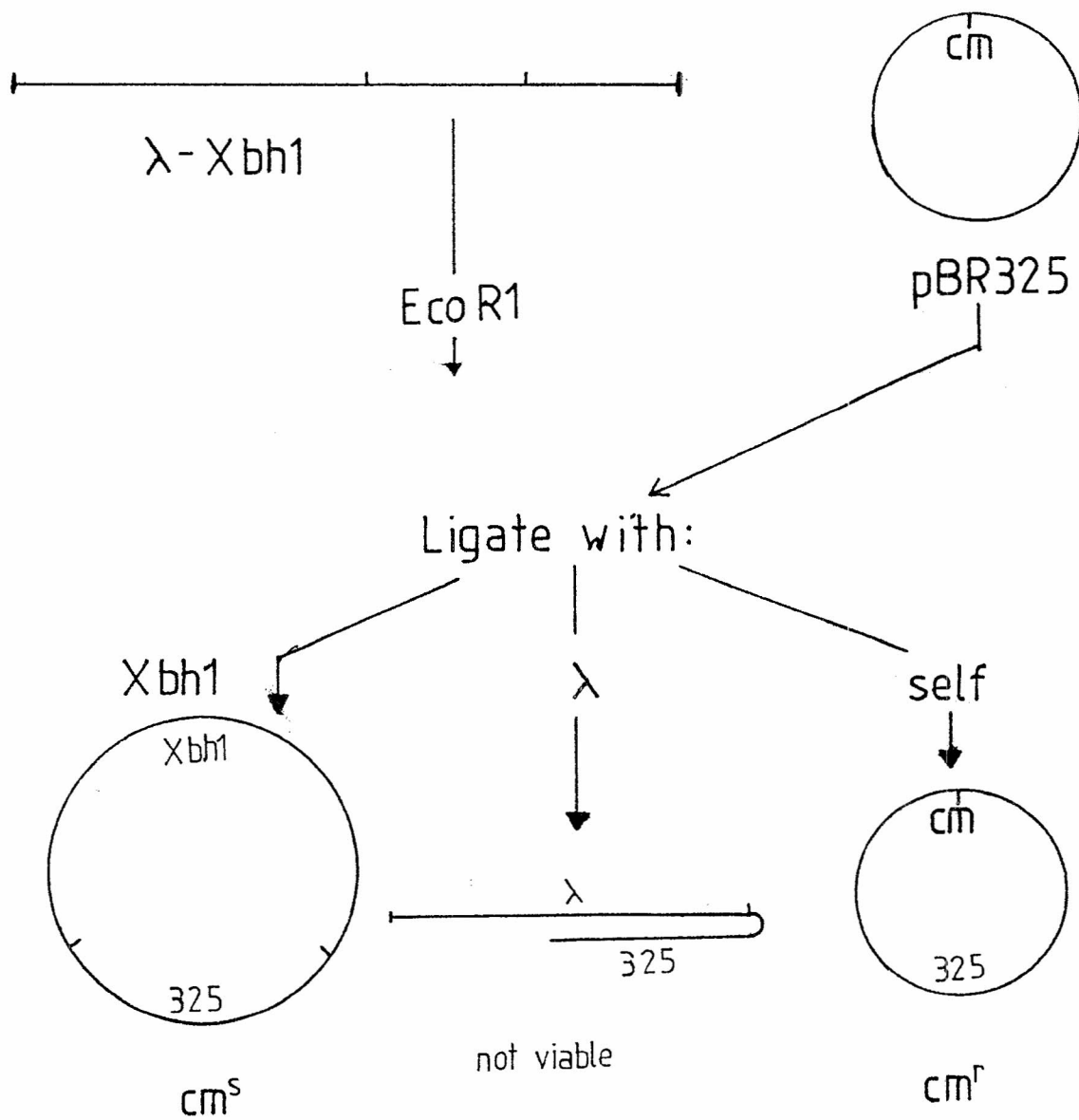
Thus all I can say is that I have eliminated all known sources of contamination and all possible reasons for distorted representation in the library as explanations for why Xbhl is present at such high copy-number in my X.borealis gene library.

7.3.5 Re-cloning the Xbhl insert into pBR325

For convenience of DNA preparation and manipulation I recloned the EcoRI insert of Xbhl into the ColEI-derived vector pBR325 (Bolivar 1978). Xbhl DNA was digested with EcoRI, and the mixed fragments ligated with EcoRI-linearised pBR325. As only the insert of the lambda recombinant has two EcoRI ends, only this will be able to form inserted closed circles with pBR325 (Fig 7.6). 3µg of Xbhl DNA was ligated with 4µg of pBR325 in 100µl: in view of the number of non-recombinant plasmids obtained it would clearly have been more productive to greatly increase the ratio of Xbhl:pBR325 DNAs. The ligation product was transformed into E.coli strain HB101 to give a total of 458 colonies. The HB101 were spread on plates containing Ampicillin to select for pBR325. The unique EcoRI site in pBR325 is in the middle of the chloramphenicol resistance gene, and insertion into this site inactivates the gene (Bolivar 1978). To detect recombinant plasmids I picked colonies and plated them in duplicates on chloramphenicol - containing and ampicillin - containing plates: colonies growing on Amp but not on Cm contained inserts. Four were isolated from 224 examined, the plasmids grown up and digested with

Fig 7.6

pBR 325 sub-cloning strategy



EcoR1 site = —|—

λ end = —————

7. Screening for H4

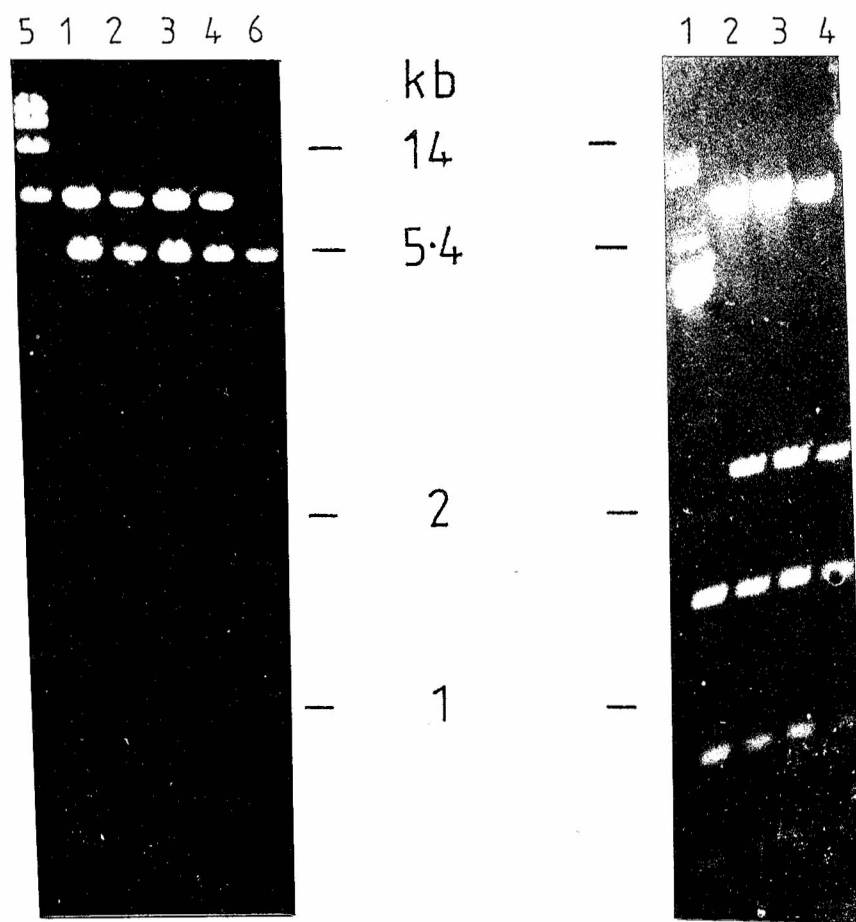
several restriction enzymes. The gel shown in Fig 7.7 demonstrates that they are re-clones of the Xbhl insert. Knowing the restriction data presented in section 8 below we can say that the insert in one clone is the opposite orientation to those in the other three. These two orientations were designated pWB1 and pWB2.

7.4 Conclusions.

I have described the isolation of a lambda-gtWES recombinant containing a sequence complementary to an H4 cDNA clone, and the re-cloning of its central EcoRI fragment in pBR325. The fragment is 8.55kb long. The recombinant carrying this fragment is super-abundant in the X.borealis gene library, a fact not presently explicable by any testable hypothesis. It was isolated five times in a single primary screen, together with a number of phage which proved to be negative for H4 sequences. For practical purposes, therefore, this is the only histone gene clone isolable from this gene library.

Fig 7.7

Recloning Xbh1 into pBR325



A

B

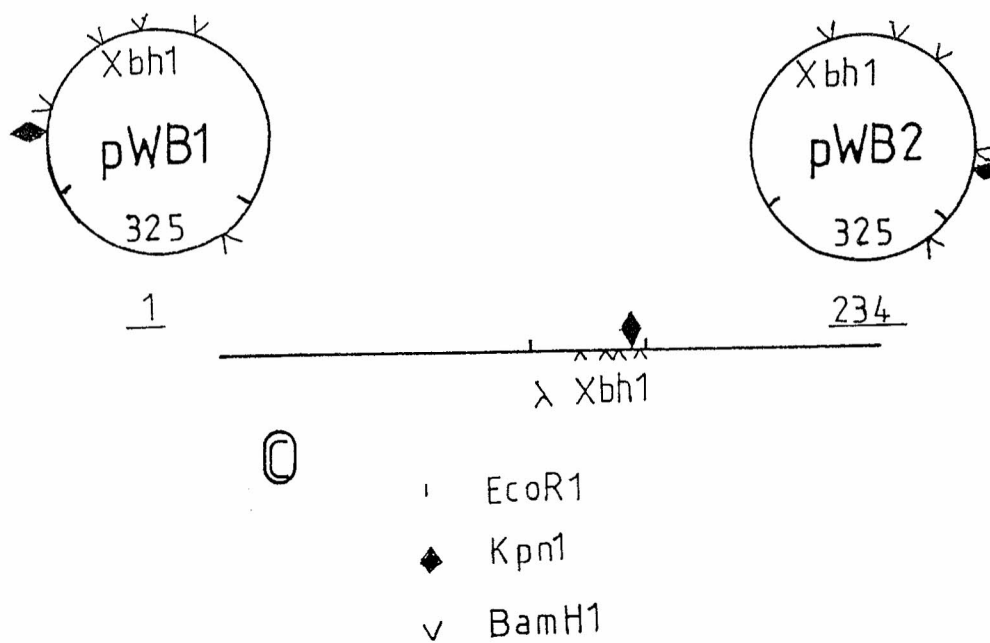


Fig 7.7 A) BamHI and B) EcoRI digests of four pBR325-Xbhl recombinants, numbered 1 to 4, lambda Xbhl DNA (5) and pBR325 DNA (6). 0.8% agarose gel stained with ethidium bromide. C) Brief restriction maps of lambda Xbhl, recombinant 1 (designated pWB1) and recombinants 2,3 and 4 (designated pWB2). Note that no BamHI sites in the lambda vector have been shown.

8. Restriction mapping Xbhl.

He had brought a large map representing the Sea
Without the least vestige of land.
And the crew were much pleased when they found it to be
A map they could all understand

8.1 Introduction.

I wished to find out where the H4 gene was on Xbhl, what other genes were contained in it, in what orientation these genes are and what other sequences of potential interest, such as repetitive elements or direct or inverted repeats, are present. To do this I needed to be able to dissect the cloned DNA into defined sub-fragments, and the most convenient way of doing this is by digestion with restriction enzymes.

Like 'satellite DNA', 'Restriction Enzyme' is a term which has undergone passive re-definition in the last decade. It is now commonly used to denote any endonuclease with an absolute requirement for a specific DNA sequence for cleavage of the molecule, usually a palindromic sequence 4 to 6bp long (Roberts 1978). The link with a host-range restriction system has been demonstrated for only a few of these enzymes. The enzymes and their specificities are listed by Roberts (1978), from which paper the information on them used in this work has been culled.

Restriction enzymes may be conveniently classified as '4-cutters' or '6-cutters' depending on their recognition site length. Thus 6-cutters recognise a 6bp sequence in the DNA, and so cut relatively

8. Restriction mapping

infrequently (once every 4096 bp on average in DNA of 1:1:1:1 A:G:T:C composition in each strand, or about once every 3340bp for EcoRI in average X.laevis DNA which is rather higher in A+T than this ideal (Roberts 1978)). Thus they make useful tools for dissecting DNA 2 - 20kb in length. 4-cutters cut at 4bp recognition sequences, IE once every 256bp on average, so they generate a complex pattern of bands from an 8.55kb length of DNA (see Fig 7.5 for example). Thus it is relatively easy to map the sites at which a six-cutter cuts, but more useful to map those at which a 4-cutter cuts, the latter in all probability being able to dissect the DNA more precisely.

There are two approaches to restriction mapping (location of the cutting sites which a restriction enzyme recognises on a DNA molecule, IE making a 'restriction map') which I used in this study. Firstly, the DNA may be digested with an enzyme, the fragments sized by electrophoresis, then digested with another enzyme, the double-digest sized and so on. By building up a grid of the products of single, double and triple digest the positions on the molecule at which the various enzymes cut can be deduced. The more enzymes that are used in this procedure the more rigorous it becomes, as a greater number of double digests become available against which to check the potential restriction map. However doing every pairwise combination of, say, eight restriction enzyme digests leads to a profusion of data whose analysis is not easy, as well as consuming a lot of DNA and enzyme.

A modification of this procedure is to perform the first digest, then to isolate the fragments produced and re-digest them as new DNA fragment for mapping in their own right. If the second enzyme is one which cuts the DNA frequently, IE generates a large number of bands, this second procedure aids identification of where in the total molecule each of these fragments comes from.

8. Restriction mapping

In the second procedure the DNA is prepared labelled at one end. It is then partially digested with the test enzyme to generate a mixture of products, which are separated by electrophoresis and detected by autoradiography (Smith and Birnstiel 1976). Only those with the labelled end attached will be detected, so those bands seen will correspond to the terminal fragment, the terminal plus sub-terminal fragments, the terminal, sub-terminal and third fragment and so on. An irregular ladder is seen, in which the spacing between the rungs corresponds to the size of the fragments in the order they are in the DNA. This is illustrated in Fig 8.1. I had some trouble with this second technique due to failure to generate end-labelled material. Some of the information below is produced from this second approach, but the large majority from the first.

8.2 Methods.

All the methods used in this section have been previously described.

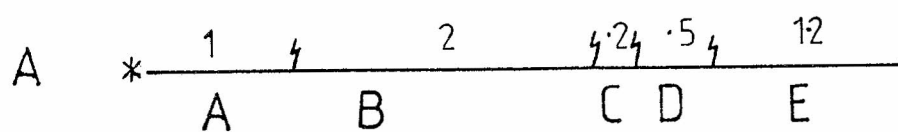
8.3 Results.

8.3.1 XbaI, KpnI, SstI.

To provide a framework around which other results could be analysed, I performed some initial digests of XbaI with EcoRI, XbaI, SstI and KpnI. EcoRI defines the ends of the X.borealis DNA, and, as we have seen, does not cut within it. The other three enzymes were chosen for their effect on lambda-gtWES DNA: XbaI and SstI do not cut the vector 'arms' and KpnI has only two sites close together in the left, 22kb 'arm'. Thus the pattern of bands these enzymes generate

Fig 8.1

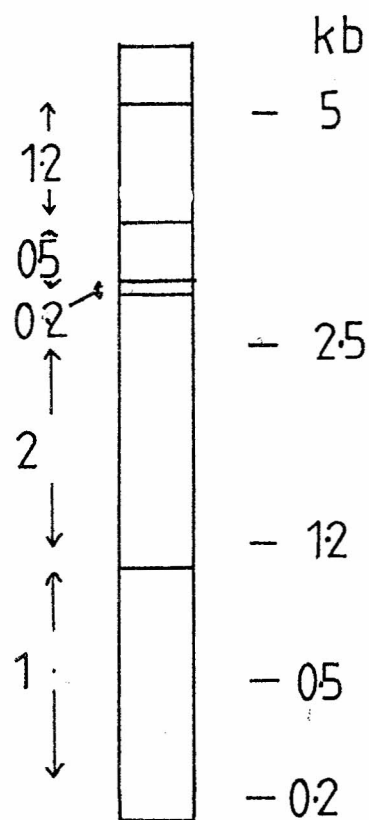
Restriction mapping by partial digestion



B



C



D

Fig 8: a demonstration.1 Restriction mapping by partial digests. A) The location of four sites for a restriction enzyme on a length of DNA. Numbers are lengths in kb. B) Total digestion product, as it would be seen on a 1% agarose gel. This shows the size of the fragments, but not their order. C) Partial digestion product. This could be used to deduce the fragment order, but is too complex to be easily analysable. D) Autoradiograph of track C in a sample where only the left-hand end of the DNA has been labelled (Fragment A). Only five bands are seen, and the fragment order may be directly determined from the distance between them.

8. Restriction mapping

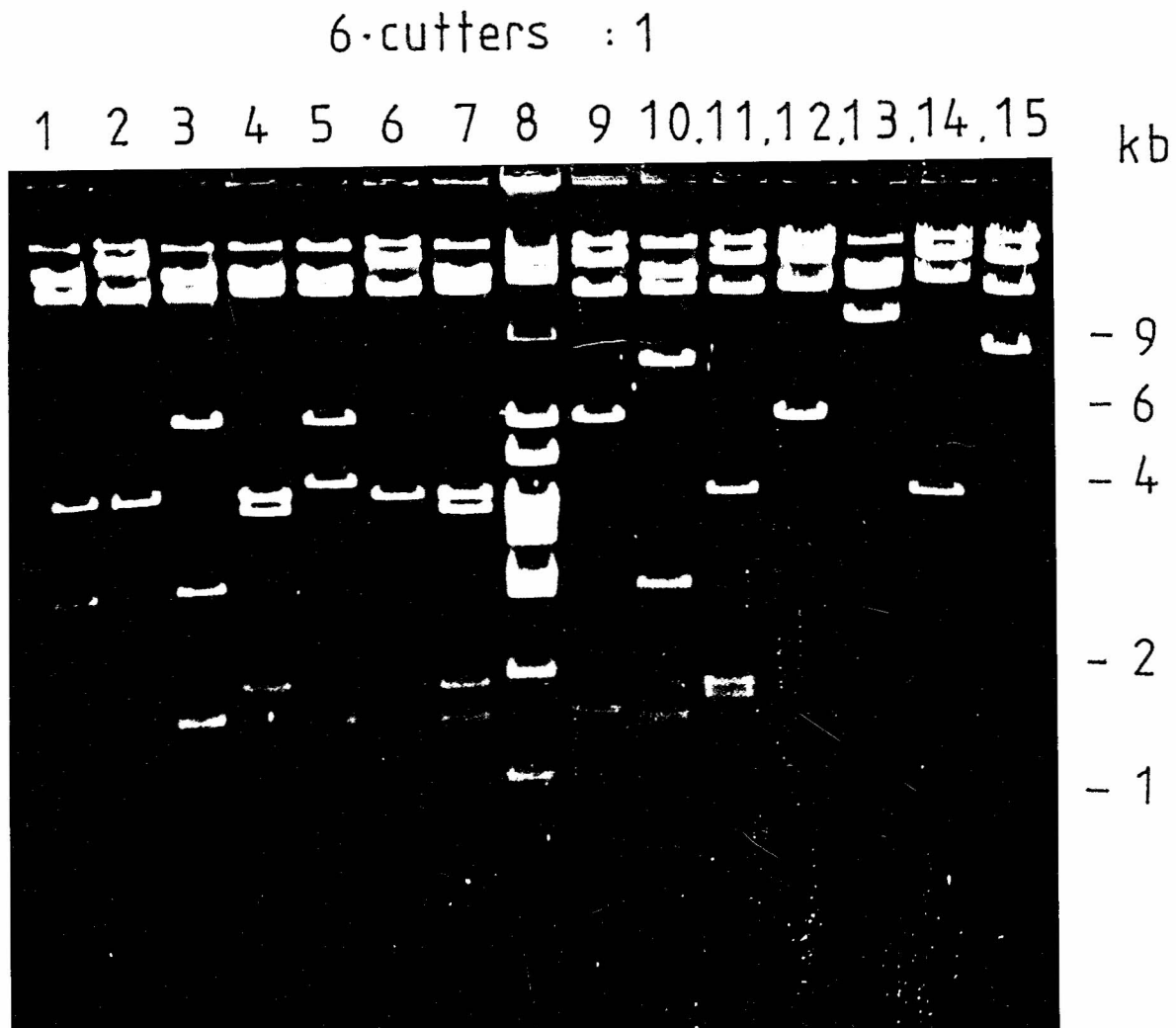
from any lambda-gtWES recombinant will not be complicated by a large number of vector bands.

The results of these digests are shown in Fig 8.2 . A number of general points are illustrated here. Firstly, digestion with and without EcoRI allows the terminal fragments to be easily identified. Considering tracks 12 and 9 in Fig 8.2 (SstI and SstI + EcoRI), SstI and EcoRI together produce four fragments beside the two vector 'arms' (5.97kb, 1.37kb, 670bp and 470bp). Two of these are produced by digestion with SstI alone (5.97kb and 670bp). Hence these latter two must be central, the other two terminal. The same conclusion could have been drawn if I had digested the whole recombinant and the isolated insert and compared the two.

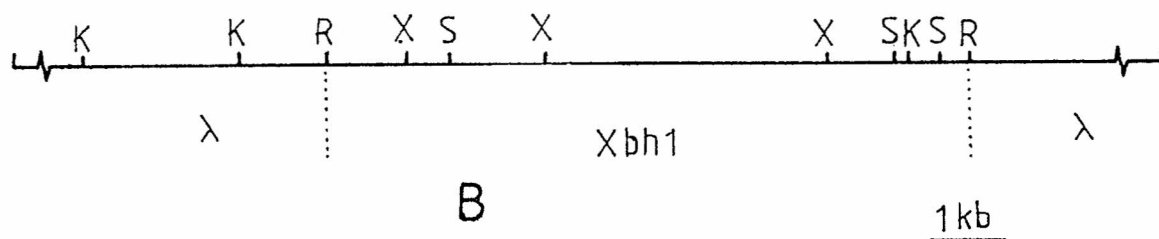
Secondly, simple logic can show the relative position of the sites for the different enzymes. . KpnI cuts the insert only once, near one end. In the KpnI + SstI double digest, the larger SstI fragment remains unaltered from its appearance in the SstI-only digest, but the 670bp SstI fragment is missing, and instead we see two extremely small bands. Hence KpnI cuts the 670bp SstI fragment and so, as the KpnI site is near the right-hand end of the clone (IE adjacent to the EcoRI site separating the clone from the 14.5kb lambda 'arm'), the order of internal SstI fragments is 5.95kb-670bp.

This reasoning is significant because it does not rely on accurate sizing of the fragments. This is important for three reasons. Firstly if the size determination is not very accurate, any maps which depend on measuring relative lengths will be correspondingly inaccurate. The determination of DNA fragment sizes by agarose gel electrophoresis have been improved by Southern (1979) who devised the plot of Reciprocal Distance Migrated versus Fragment

Fig 8.2



A



B

Fig 8.2 A) Single, double and triple digests of lambda-Xbh1 DNA. Tracks: 1) KpnI+ XbaI+ EcoRI, 2)EcoRI+ SstI+ XbaI, 3) EcoRI+ SstI+ KpnI, 4) KpnI+ SstI+ XbaI, 5)KpnI+SstI, 6) SstI+XbaI, 7)KpnI+ XbaI 8)size marker, 9) EcoRI+SstI, 10 EcoRI+KpnI, 11) EcoRI+XbaI, 12) SstI, 13) KpnI, 14)XbaI, 15 EcoRI. 0.7% agarose-ethidium bromide gel.

8. Restriction mapping

length to determine the length of double-stranded DNAs from known standards. However this is only accurate in a limited region of any gel and, when dealing with large DNA fragments cut several times by an enzyme the range of product sizes might well exceed this accuracy limit. Secondly, by not using the size data to create the restriction map it becomes valid either to use the size data to confirm the map, or to use the map to confirm the size data. Thirdly, on gels designed to separate large DNA fragments, small fragments may be easily overlooked. Subsequent studies showed that an SstI digest will generate a fragment 160-200bp long, but this is not evident on Fig 8.2. If we were relying on the fragment sizes for map data, the fact that the sizes (as determined from Fig 8.2) for the fragments generated by SstI digestions of the insert did not add up to the known size of the insert would cause some confusion.

The XbaI, KpnI and SstI maps of XbhI are shown in Fig 8.2.

8.3.2 Other six-cutters.

As we noted above, a 4-cutter will in general be able to dissect a clone more finely than a 6-cutter, and hence a 4-cutter map will be of more use for the functional dissection of a clone. However it is rare for a 4-cutter restriction pattern to be interpretable in the absence of either an extensive 6-cutter map to which double digests can be correlated or accurate data from end-labelled, partial digest restriction mapping (Smith and Birstiel 1976). As I mentioned above, the latter was not available at the time these experiments were being performed, so I made as extensive a collection of 6-cutter maps as I could to enable me to perform useful 4-cutter mapping later.

I will not go through all the data in detail here. They are

8. Restriction mapping

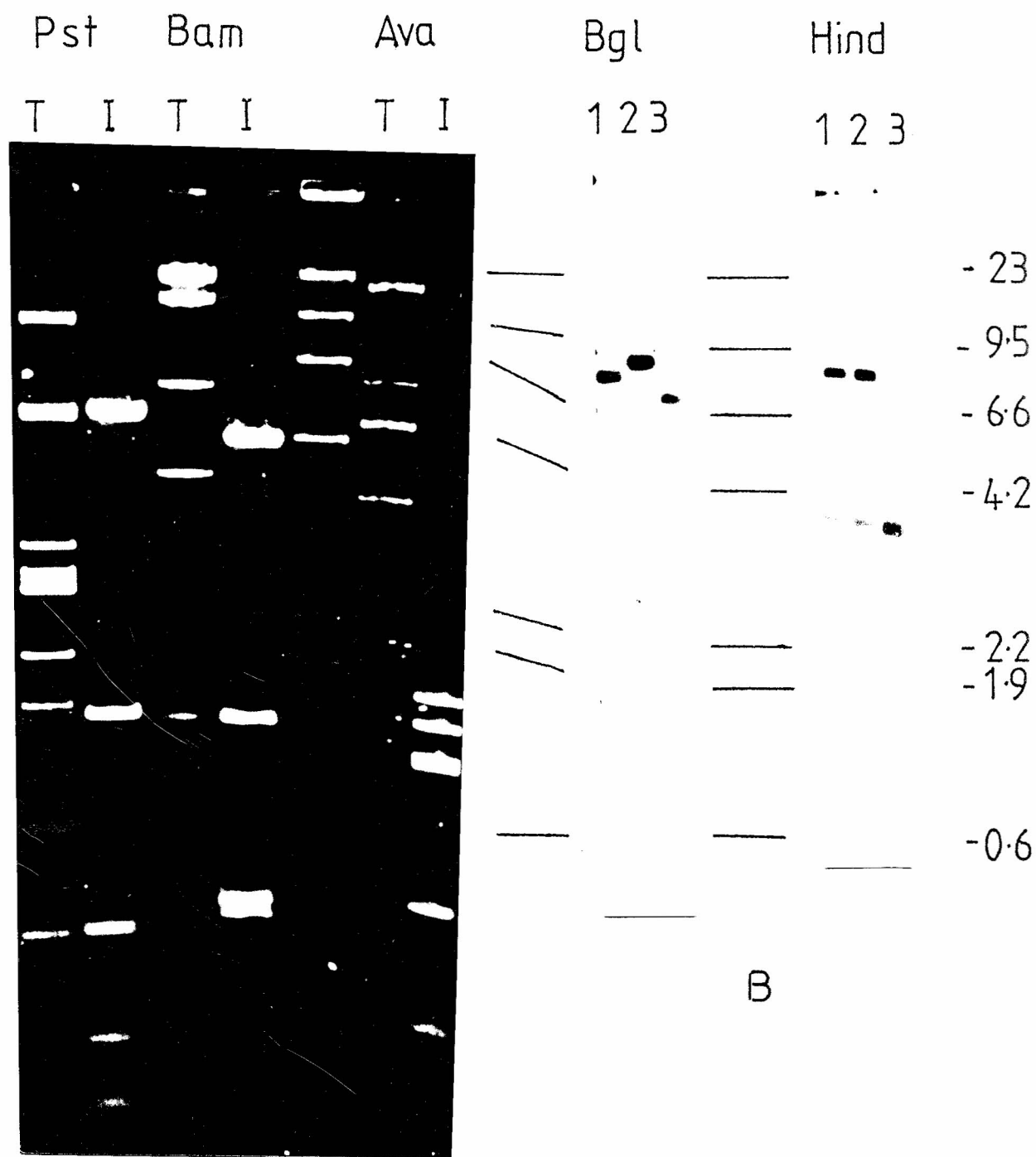
presented in Figs 8.3, 8.4 and 8.5, and notes on the digests used, the DNA digested and the region of the restriction map this helps to illuminate are given in the legends to the figures. Some notes are essential, however.

Fig 8.3 shows the identification of the end fragments by two methods. Digestion of whole lambda recombinant and of isolated insert with an enzyme distinguishes the terminal from the internal fragments, as we discussed above. The end fragments of the Hind III and BamHI maps were determined by the method of Smith and Birnstiel (1976). The insert was end-labelled, partially digested and then electrophoresed, the bands being detected by autoradiography. The terminal fragments are detected by this method. Note that only one band appears on the autoradiograph of the HindIII total digest : this is because the two terminal HindIII fragments are of nearly identical size, unresolvably different on this gel system.

Fig 8.4 shows a variety of single and double digests performed with nick-translated, isolated insert. This procedure labels DNA uniformly, as does Ethidium bromide staining, and so extremely small fragments will be poorly labelled, poorly resolved and hence virtually invisible on this sort of gel. The map deduced from this data and presented in Fig 8.5 has several ambiguities. The SstI, Aval and BamHI total fragment sizes do not add up to the known insert size, bespeaking of small fragments not detected by these digests. These fragments were detected later:- the BamHI and SstI fragments as small, unexpected 6-cutter bands seen on gels run to analyse 4-cutter digestion products, and the 200bp Aval fragment by Dr. P.C.Turner as a result of preliminary work for DNA sequence analysis. Other ambiguities in the central region of the clone were cleared up by mapping the XbaI 3.95kb, XbaI 1.75kb, HindIII 1.54kb and HpaII

Fig 8.3

6 cutters : 2



A

B

Fig 8.3 A) 0.7% agarose-ethidium gel. Pst=PstI digest, Bam= BamHI digest, Ava=AvaI digest. T=whole recombinant lambda-Xbhl, I= isolated insert of Xbhl.

Autoradiograph of 0.7% agarose gel of labelled material. Both experiments are on end-labelled Xbhl insert. Bgl=BglII digest. 2)= undigested, 1)=partial digest, 3)= total digest. Hind= HindIII digests 1) and 2)= partial digests, 3)=total digest. These gels allow us to identify the terminal fragments for these five restriction enzymes. In the case of BglII and HindIII this only leaves one fragment, which must therefore be central. This is confirmed by the partial digest products.

Fig 8.4

6 cutters : 3

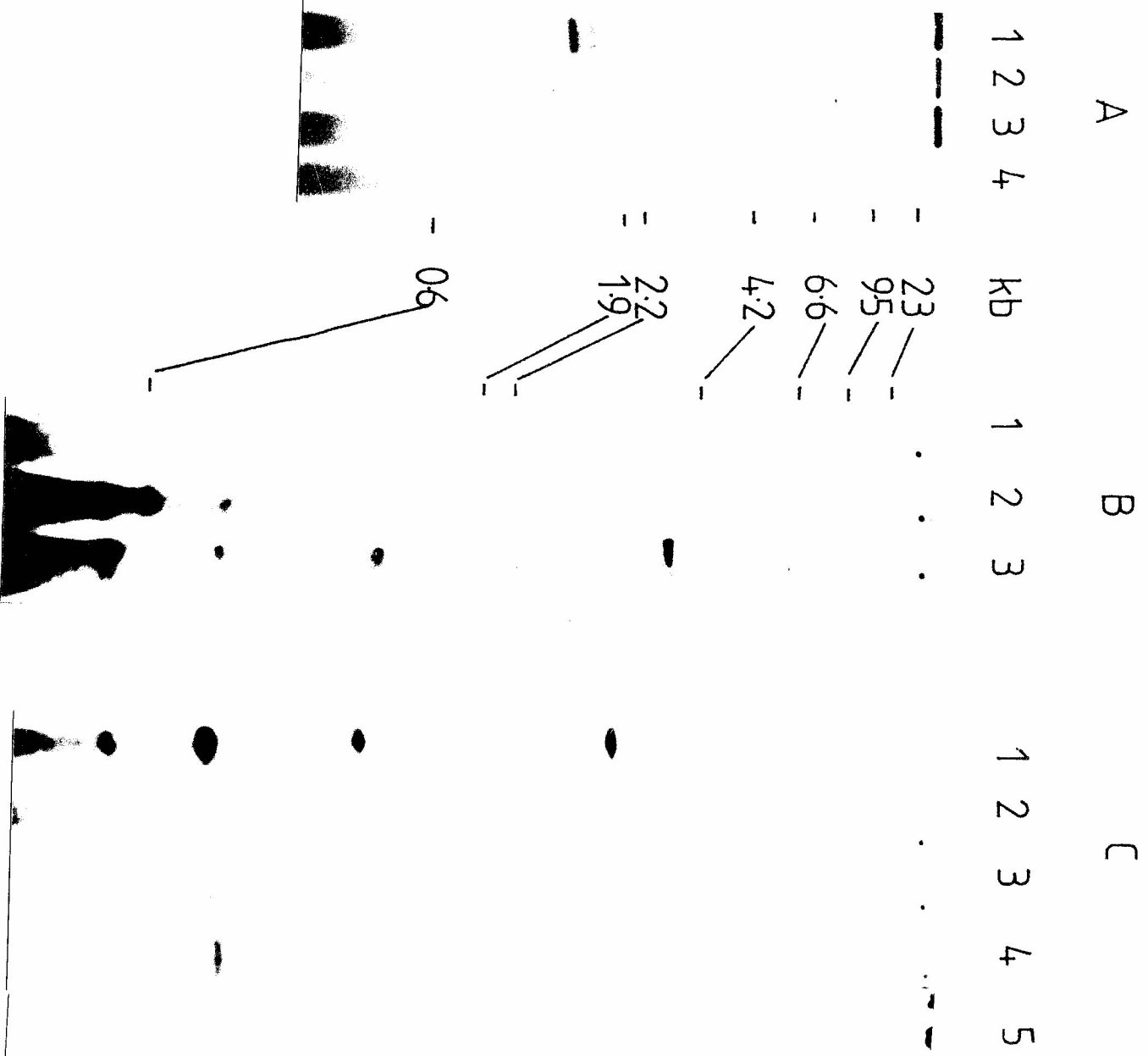


Fig 8.4 All gels: autoradiographs of digests of nick-translated insert of Xbhl, separated on 0.7% agarose gel. A) HindIII digests with 1) AvaI, 2) HindIII alone, 3) BamHI, 4) BamHI digest alone. B) PstI digests with 1) KpnI, 2) alone, 3) SstI. This allows most of the PstI map to be built up, together with the results from C) BamHI digests 1) alone, 2) with HindIII, 3)PstI, 4) XbaI, 5) KpnI. These digests provide the PstI map, and the BamHI and AvaI maps with a few ambiguities remaining.

Fig 8.5

Initial 6 cutter map

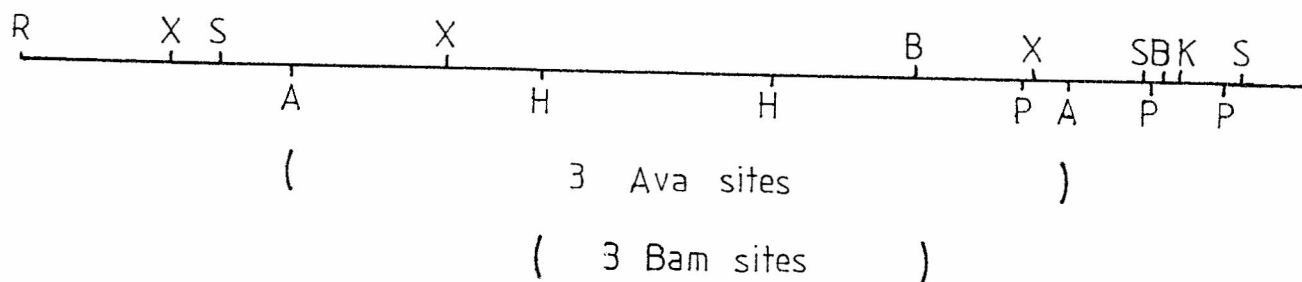


Fig 8.5 6-cutter restriction map. Note that there are some limits on the position of the un-mapped BamHI and AvaI sites (not shown here) which allow these maps to be finalised in conjunction with later data. R=EcoRI, X=XbaI, S=SstI, A=AvaI, H=HindIII, B=BamHI, P=PstI.

1.75kb fragments separately. The data for these are presented in Fig 8.6. This also gives some data for a HincII map of the central region of the clone.

8.3.3 4-cutters.

Mapping 4-cutter restriction sites has proved to be harder than I anticipated. Although I have collected fragmentary data on a number of 4-cutter digests of Xbhl, only for two enzymes can I present reasonably complete maps. These are HaeIII (chosen for its physical robustness and hence the ease with which it may be used in double digests) and HpaII (chosen for its use in preparing end-labelled fragments for DNA sequencing and its possible use in studying genomic DNA methylation (Bird and Southern 1978, Bird 1978)). Some data for HinfI is also included, as it is relevant to later sections.

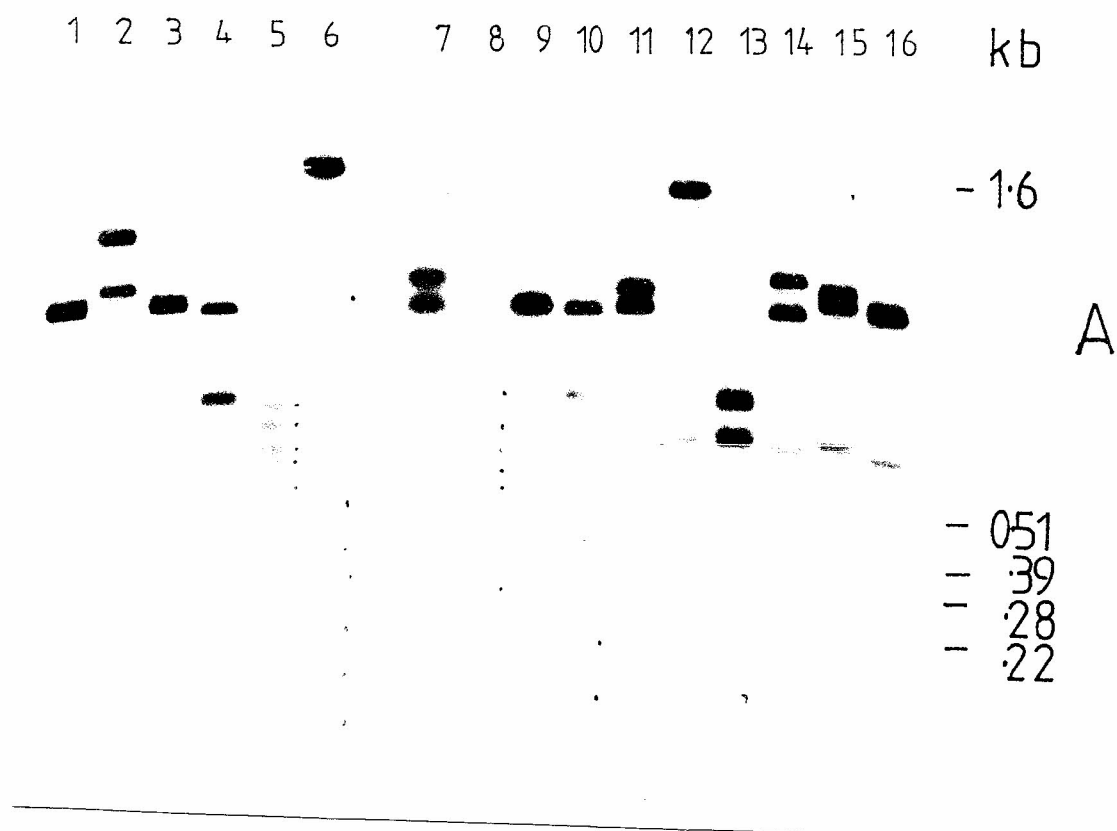
All 4-cutter data come from three series of studies: 1) Digests of uniformly labelled, isolated fragments of Xbhl insert, 2) Digests of end-labelled fragments of Xbhl and 3) Information gleaned by Dr. P.C.Turner in the course of sequence determination studies.

8.3.3.1 Studies on uniformly labelled material.

Gels of digests of nick-translated, isolated fragments of Xbhl are presented in Fig 8.7. I will not go through the deduction of the 4-cutter restriction maps from this data: the principle is the same as outlined above for the 6-cutters, but the amount of data involved is too great for concise written exposition. Again, more detail is given in the figure legends.

Fig 8.6

6 cutters : 4



1 2 3 4 5 6 7 8 kb

- 1.6



Fig 8.6 cont

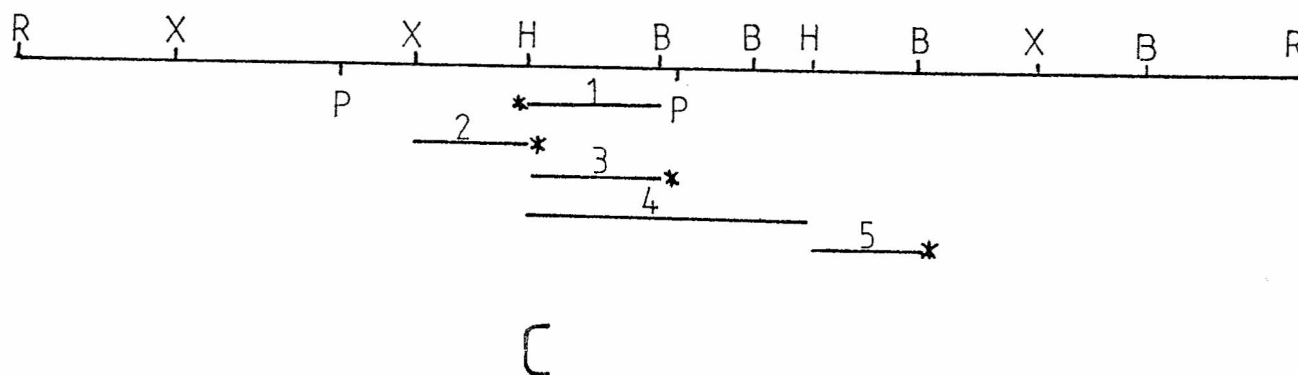


Fig 8.6 A) Digests of XbaI 3.98kb fragment, isolated and nick translated before digestion, separated on 1.5% agarose gel. 1) AvaI, 2) SmaI, 3) AvaI and SmaI, 4) BamHI, 5) AvaI and BamHI, 6) undigested fragment, 7) HindIII, 8) BamHI and HindIII, 9) AvaI, 10) SmaI+ BamHI, 11) HincII, 12) HincII and BamHI, 13) HincII and HindIII 14) HincII and SmaI, 15) HincII and AvaI. This gel locates the BamHI and AvaI sites within the central XbaI fragment, and hence within the whole clone. It also adds data on SmaI and HincII for this region.

B) End-labelled material digested thus: 1) Frag 2, 2) Frag 1, 3) Frag 5, 4) Frag 4/BamHI. 5) Frag 3/HpaII, 6) Frag 4/HpaII, 7) Frag 3, 8) Frag 4.

C) The location and labelling of fragments 1 - 5 alluded to above. R=EcoRI, X=XbaI, P=HpaII, H=HindIII, B=BamHI. These digests cover the junction between the XbaI 1.82kb and 3.98kb fragments, and locate the HpaII and HindIII sites relative to their ends.

Fig 8.7

4 cutters : 1

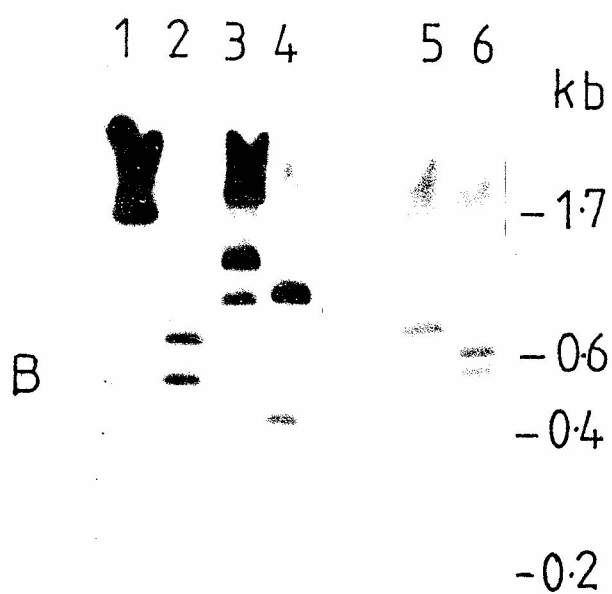
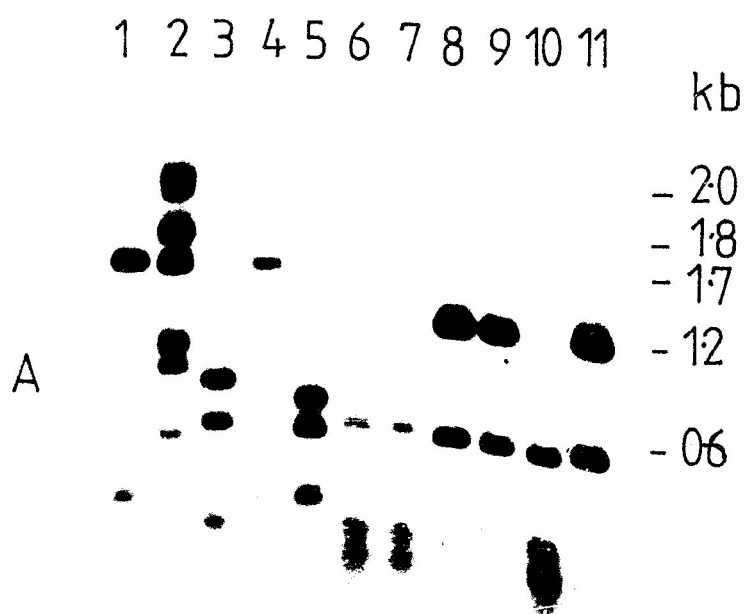


Fig 8.7 cont.

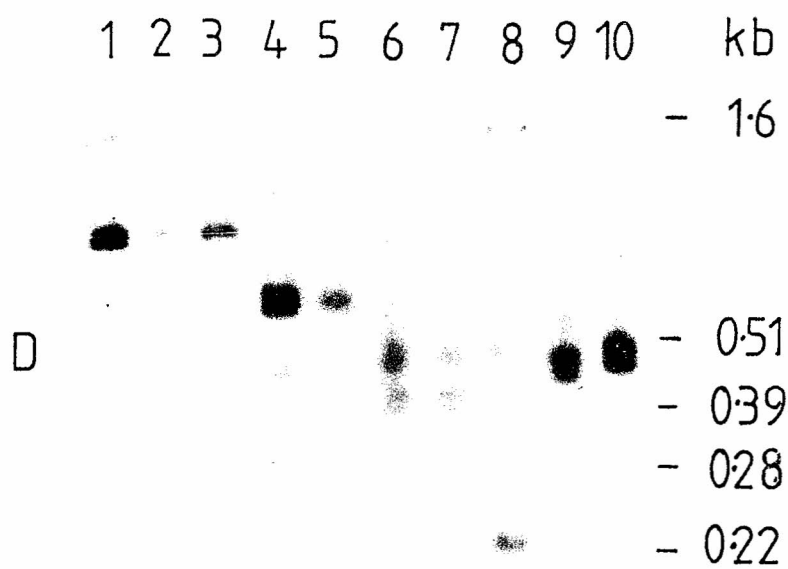
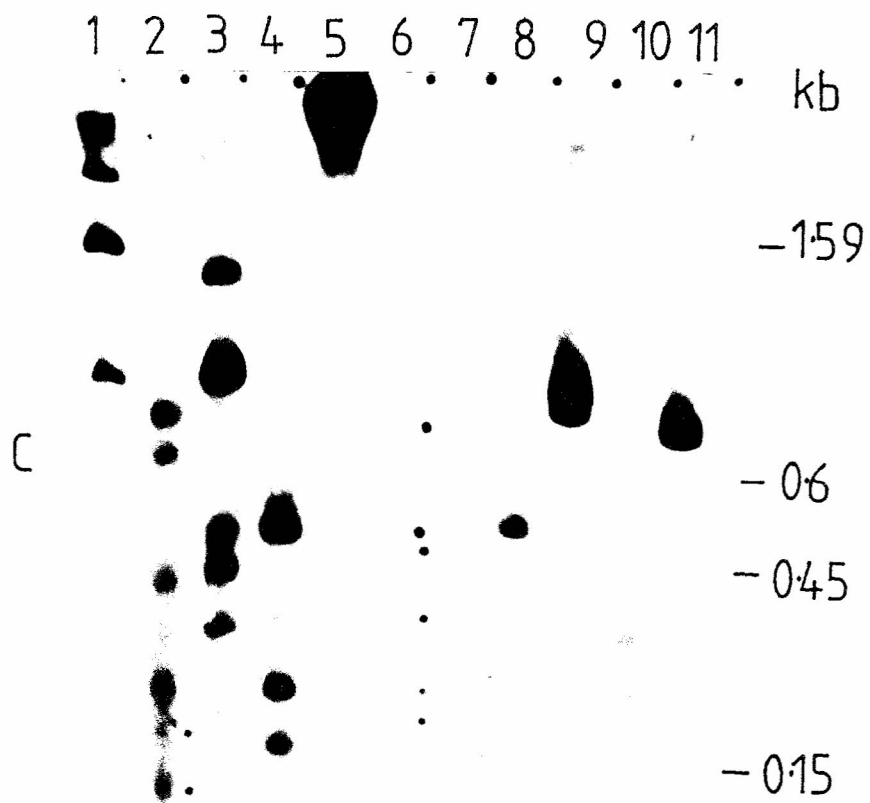


Fig 8.7 A) Restriction of Xba 1.82kb fragment - autoradiograph of nick-translated material digested and separated on 1.2% agarose gel. 1) HincII + HinfI, 2) marker, 3) HincII + AvaI, 4) HincII, 5) HincII+PstI, 6) HincII+HpaII, 7) HpaII, 8) PstI, 9) PstI+SstI, 10) PstI+HpaII, 11) PstI+HinfI. This gel locates the major HinfI and HpaII fragments in this region, allows the PstI, AvaI and SstI sites to be accurately located and the HincII map to be extended into this region.

B) Restriction of the HpaII 1.75kb fragment. Autoradiograph of nick-translated, digested material separated on 2% agarose gel. 1) HhaI, 2) HaeIII+HhaI, 3)HindII, 4) HinfI+HindIII, 5) HinfI, 6) HaeIII. This gel allows us to order the HaeIII, HpaII and HinfI fragments around the left HindIII site.

C) HpaII digests - 2% agarose gel of nick-translated material. Tracks are HpaII digests of: 1) XbaI 3.98kb fragment, 2) XbaI 1.82kb fragment, 3) Xba 1.75kb fragment, 4) XbaI 1.05kb fragment, 5) arker track, 6) SstI 1.59kb fragment, 7) SstI 0.69kb fragment, 8) SstI 0.49kb fragemnt, 9) HindIII 1.54kb fragment, 10) BamHI 1.62kb fragment, 11) BamHI 0.9kb fragment. This data, in conjunction with other above, can lead to a fairly certain HpaII map of most of the clone. In particular this gel assigns the larger fragemnts detected elsewhere to regions of the clone, and explores right of the left-most HindIII site.

D) HaeIII digests of 1) XbaI 1.82kb fragment, 2) ditto digested with PstI, 3) as 1) digested with SstI, 4) HindIII 1.54 fragment, 5) as 4) digested with BamHI, 6) SstI 1.59 kb fragment, 7) as 6) digested with XbaI 8) marker track, 9) SstI 0.69kb fragment, 10) SstI 0.49kb fragment. This gel locates most of the HaeIII sites bewteen the leftmost XbaI site and the left-most BamHI site, and gives supportive data for other regions. Note that several pairs of these tracks are

comparisons of digests with and without a third enzyme, which tells whether the third enzyme cuts within a given HaeIII fragment, and hence locates that fragment.

8. Restriction mapping

8.3.3.2 Mapping by end-labelling.

This data was mostly a 'spinoff' from DNA sequence studies, as end-labelled material is generated for this technique and needs to be restricted subsequently. Results are presented in Fig 8.8 . Studies have concentrated on the region between the left-most BamHI and SstI sites. In particular this region contains the 1.75kb HpaII fragment, itself the subject of sub-mapping.

8.3.3.3 Other information.

Dr. P.C.Turner has kindly made available to me a considerable body of information on 4-cutter restriction sites in the left-hand half of the clone, which he generated by digestion of end-labelled fragments and by searching for restriction sites in DNA sequence. The restriction map which this information gives us is presented in Fig 8.9

Together these studies enable a fairly complete HpaII and HaeIII map to be drawn, and some HinfI data to be presented too. The complete restriction map is presented in Fig 8.10 .

8.4 Conclusion. What can the srestriction map tell us?

I generated a restriction map of Xbhl for a purely functional reason - to be able to dissect Xbhl into convenient sized chunks for gene mapping experiments. As restriction enzymes are prokaryotic it is improbable that they will have any functional relationship with Xbhl DNA. But they can point to structures within that DNA which distort the distribution of their recognition sites. One such potential structure is seen in the right hand lkb of the clone.

Fig 8.8

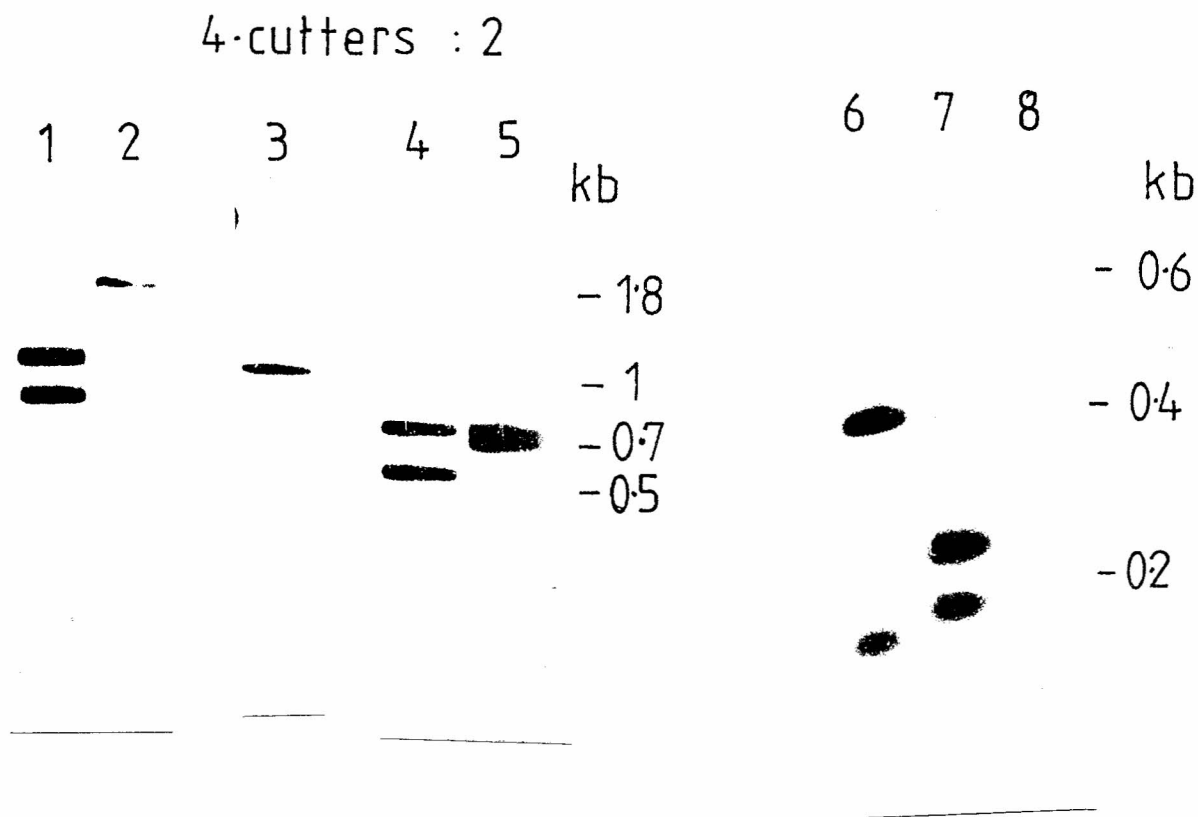


Fig 8.8 Digests of end-labelled material. All are autoradiographs of 2% agarose gels. 1) HpaII 1.75kb fragment end-labelled and digested with HindIII. 2) BamHI 4.2kb fragment labelled at the BamHI site only and HpaII digested, 3) HindIII 3.2(left)kb fragment labelled at the HindIII site only and HpaII digested, 4) HindIII 1.54 kb fragment end-labelled and BamHI digested. 5) as 4) but HpaII digested. 6-8) BamHI 0.6kb fragment end labelled and 6) HpaII, 7) HaeIII, 8) HinfI digested. This gives confirmatory data on the BamHI and HindIII maps in the central region, and locates the cutting sites of HpaII in this area.

Fig 8.9

4 cutter data from Dr. P.C.Turner

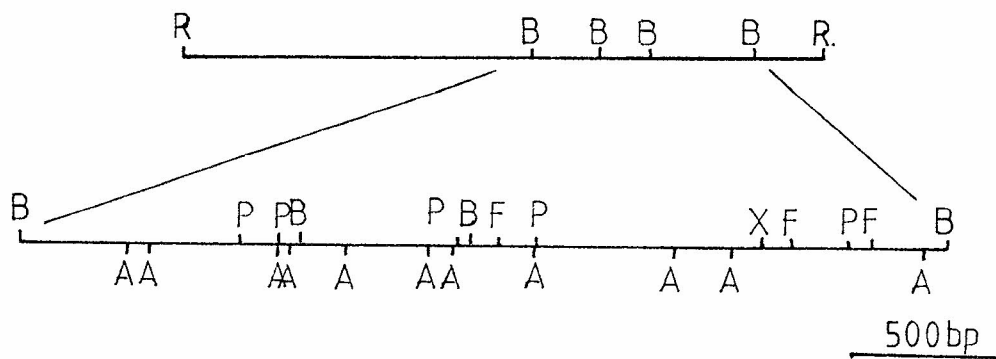
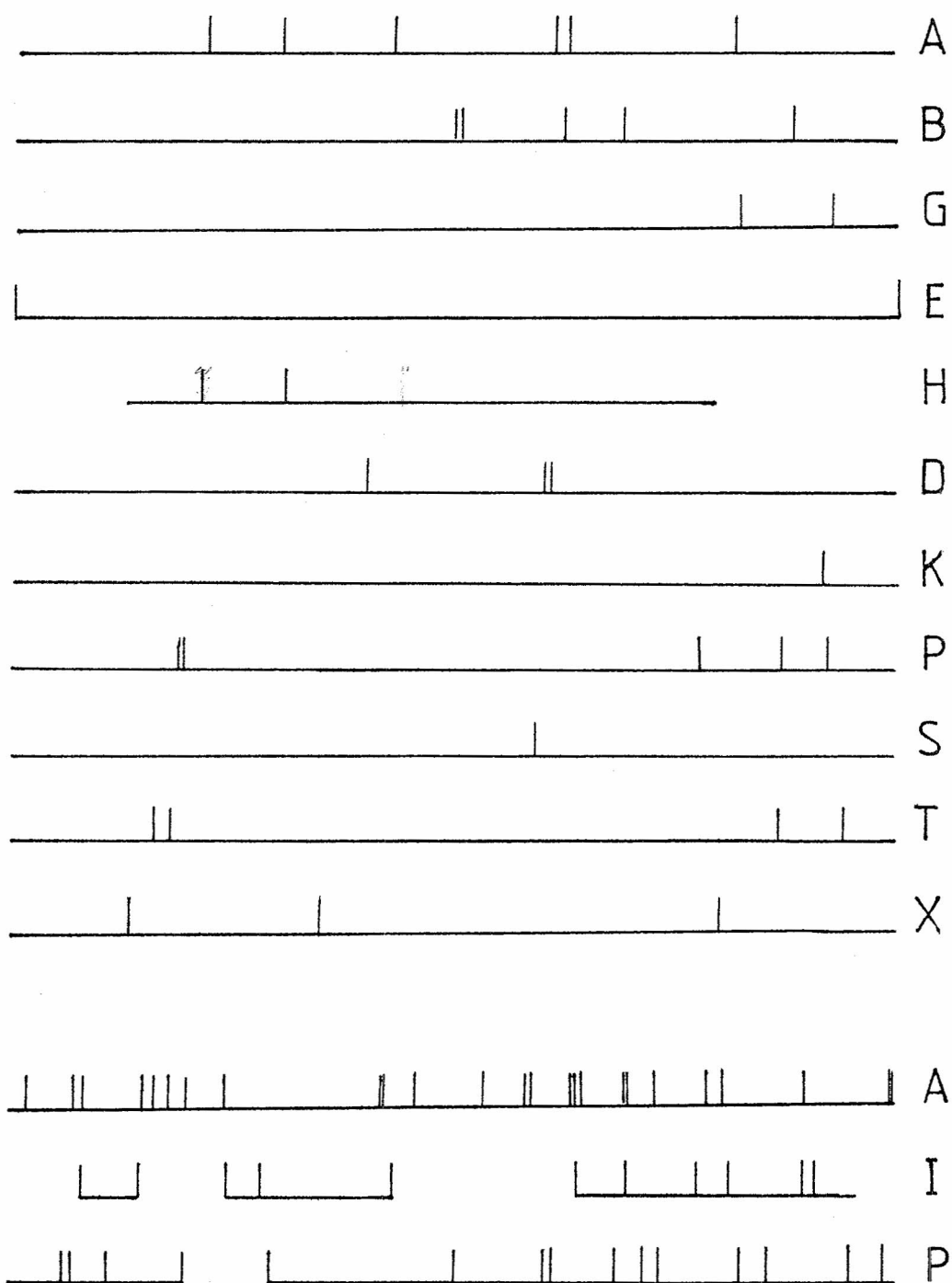


Fig 8.9 Restriction map data on HpaII (P), HaeIII (A) and HinfI (F) provided by Dr. P.C.Turner. This information is derived from studies of material end-labelled for sequence determination. Some of this data may also be derived from Figs 8.6 and 8.7

Fig 8-10

Restriction map of Xbh1



1kb

Fig 8.10. Restriction map of Xbhl. 6-cutters: A=AvaI, B=BamHI, G=BglII, E=EcoRI, H=HincII, D=HindIII, K=KpnI, P=PstI, S=SmaI, T=SstI, X=XbaI. 4-cutters: A=HaeIII, F=HinfI, P=HpaII. Fig 8.11 The 'inverted repeat' at the right-hand end of Xbhl. The right-hand end of Xbhl is drawn, and then re-drawn underneath in the opposite orientation. H=HpaII, F=HinfI, S=SstI, P=PstI, B=BamHI, G=BglII, X=aligned restriction site. H3= H3 gene region (see next section).

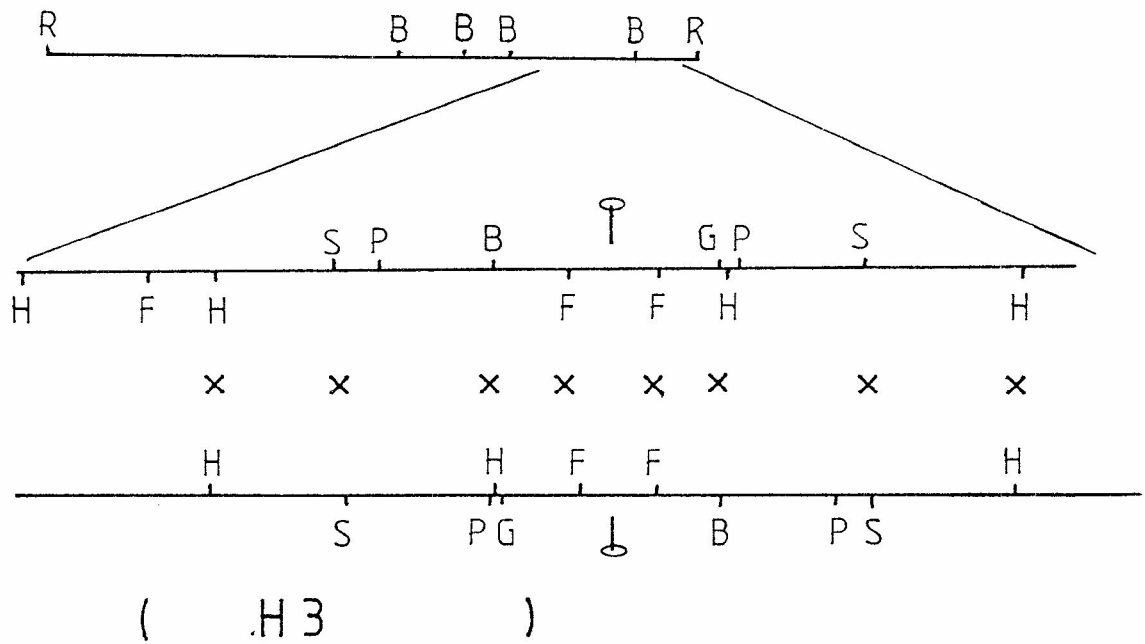
8. Restriction mapping

In this region, from 300bp from the right-hand end to 1300bp, there are 7 6-cutter sites for 10 6-cutters mapped, giving a 70% chance that any given 6-cutter will cut here. For the rest of the clone the figure for any 1kb segment is 3 sites, i.e. a 30% chance for any one enzyme. 6-cutters cut on average once every 4000bp, hence a chance of 30% that any enzyme will cut in a given 1kb is within our expectation. Why is this value higher at the right-hand end of XbaI? The chance that 7 out of 29 restriction sites should occur in 1kb out of 8.5 is 5.5%, low enough to suggest a significance without proving it. This region contains a suggestion of an inverted symmetric structure in the restriction map which involves many of the supernumary sites (fig 8.11). Notice that a BamHI and a BglII site coincide in this transformation: these two restriction enzymes recognise the same central 4bp sequence in their 6bp recognition sites (BamHI= GGATCC, BglII= AGATCT). The significance of this is uncertain as nearly half the supposedly duplicated block is a part of the protein-coding region of a functional H3 gene (vide infra). However if this really does represent an inverted duplication of DNA, the abundance of 6-cutter sites in this region can be explained by postulating that the duplicated region had more 6-cutter sites than average within it by statistical fluke, and that this preponderance was perpetuated by the duplication.

Apart from this region, the clone shows no obvious duplication or other structure in its restriction map. Similarities between this map and that of other histone gene clusters will be discussed later.

Fig 8.11

Restriction sites near H3



100bp

9. Gene mapping.

..."Other maps are such shapes with their Islands and capes,
But we have our Captain to thank"
So the crew would protest. "He has brought us the best
A perfect and absolute blank."

9.1 Introduction.

In this section I will describe the genes on Xbhl, their location and transcriptional orientation. This study is especially important when a gene clone is likely to contain more than one gene, as is so of the clustered Xenopus histone genes (Hentschel and Birnstiel 1981, Moorman et al 1980, Zernik et al 1981, Turner and Woodland 1982b, van Dongen 1982). Of the various procedures available for finding out where genes are located, I used a selection based on nucleic acid hybridisation techniques.

The first section uses DNA-RNA hybridisation to select mRNA complementary to sequences in the clone from a total cell RNA pool, and then identifies this mRNA by identifying what proteins it codes for - this is the elegant technique of Hybrid Release Translation (HRT). The cloned DNA is bound to a nitrocellulose filter, and then this is used in a hybridisation with a preparation of total cellular RNA. Only those RNAs complementary to some DNA sequence in the clone will bind to the DNA, and hence to the filter. The rest of the RNA is then washed off the filter, and the bound RNA eluted by melting its hybrids with the DNA. This bound RNA fraction is then used as a template for protein synthesis in a cell-free protein synthesis

system. Thus the proteins made by any RNA homologous to a region of the cloned DNA is identified (McGrogan et al 1979, Ricciardi, Miller and Roberts 1979, Zernik et al 1981, Patterson and Roberts 1981) This directly identifies gene products rather than homologous DNA sequences. However if there are no mRNAs homologous to the cloned DNA in the RNA population then the assay will give a negative result regardless of the sequence of the clone. Thus only positive HRT results can be reliably interpreted.

One purely technical problem in the technique as used here is that much of the mRNA in Xenopus oocytes, the source of RNA used in these experiments, codes for histones anyway (Adamson and Woodland 1977, Woodland and Adamson 1977, Woodland 1980), and so it can be hard to tell whether the presence of a histone protein in the translation products is due to the hybridisation of its RNA to the clone, or to contamination of the hybridising RNA by the starting RNA population. This problem is especially acute if the clone codes for all five histones. To surmount this the clone may be dissected to show that different sections of, or the two different strands of, the DNA hybrid release different mRNAs.

The second approach is DNA-DNA hybridisation. These tests are straightforward and conclusive when homologous probes are available to hybridise to the uncharacterised clone, but can give ambiguous results when heterologous probes from evolutionarily distant species or diverged genes are used. This is more troublesome when identifying genes which diverge rapidly through evolution such as H1 (Isenberg 1979) than with ones that are highly conserved like H3 and H4 (Turner and Woodland 1982, Busslinger et al 1982). If better cloned probes are not available, the only way round this problem is to alter the conditions of hybridisation until the heterologous probes hybridise to

9. Gene mapping

the uncharacterised clone, and to try to control for the non-specific DNA-DNA binding and possible misleading cross reactions between partly homologous sequences that can result.

This second approach can take two forms, as mentioned in the Introduction (Section 4.2.2). Either the uncharacterised, cloned DNA can be used as a labelled probe to detect sequences in the characterised DNA, or visa versa. Both techniques use Southern blots to transfer restriction fragments from agarose gels onto nitrocellulose paper. The nitrocellulose filters are then used in hybridisations with the labelled probe, so that the regions of binding of the radioactive probe to them may be related directly to the bands on the original gel.

The former has the advantage that the effects of the hybridisation of non-coding sequences in either clone to each other are much easier to detect. If the sequence composition and gene location of the standard clone are accurately known, then the expected pattern of hybridisation which would be caused by any particular gene in the uncharacterised clone may be predicted, and this pattern compared with what actually occurs. Deviations from the prediction point to spurious cross-reaction of inter-genic spacer or of vector sequences. However a separate hybridisation needs to be performed for each region of the uncharacterised clone to be used as a probe, which can lead to a lot of work if an accurate gene map of an 8kb DNA segment is needed.

The converse approach, hybridisation of characterised, labelled probe to uncharacterised DNA which has been suitably separated and Southern Blotted allows a dozen restriction digests to be screened at once for a 'gene'. However there is no indication of which part of the

9. Gene mapping

labelled, characterised probe is hybridising to the uncharacterised clone. So in the absence of probes containing solely the coding regions of known genes, this approach can also lead to artefacts.

The cloned DNA probes used have already been described (Section 5.1). Briefly, pcXlH4W1 and pcXlH3W1 are H4 and H3 cDNA clones respectively from Xenopus, pSpl02, pSpl17, pSp3-1, pSpl-4 and pSp2b-5 are genomic clones from Strongylocentrotus containing H1+H2B+H4, H3+H2A, H3, H1(coding only) and H2B genes respectively, and M13-H2A and M13-H1 are H2A- and H1- containing M13 subclones of Xenopus laevis genomic clones. References and details about these clones are given in section 5.1 (see Fig 9.1)

9.2 Methods.

9.2.1 Preparation of Total ovary RNA from Xenopus.

This was after Turner and Woodland (1982). The ovaries were dissected out of one X.borealis and placed in 200mls of Turner buffer (10mM EDTA, 2%SDS, 6%4-aminosalicylate, 1%LiCl, 0.5MTris pH9.0 : 470 mls autoclaved and then brought to 500mls with redistilled phenol). 100µl of Diethylpyrocarbonate was added to reduce RNase activity. The ovary was homogenised in a blender with 200 mls of phenol, and the homogenate centrifuged in a Sorval 6L for 15' at 20°C. The supernatant was taken and re-extracted until the phenol-water interface was clear: this typically took a total of 5 extractions. The aqueous phase was then extracted once with Phenol:chloroform (1:1), once with chloroform:isoamyl alcohol (24:1). To the aqueous phase I then added 1/10 volume of 2M Sodium acetate and 2 volumes of ethanol to precipitate the nucleic acid:

Fig 9.1

Sea urchin histone gene clones

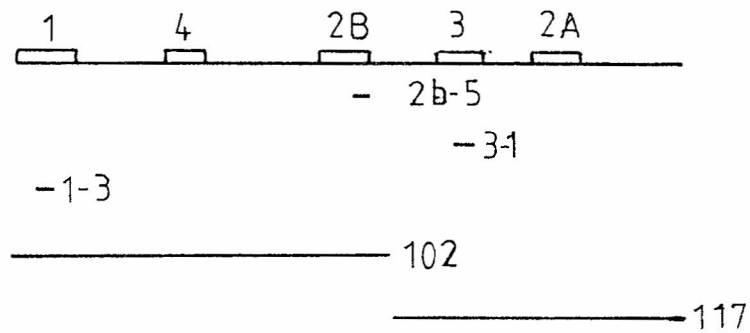


Fig 9.1 Sea urchin histone gene clones and subclones used in this study. From Sures et al (1978) and Kedes unpub.

9. Gene mapping

this was almost entirely RNA (but see section 11.3.3 below). The precipitate should be white and fluffy: appearance of grey or extremely fine precipitates indicated severe contamination, degraded RNA or both. The RNA pellet was redissolved in 20mls of 1xSSC, 0.1%SDS and made 10µg/ml in Proteinase K. This was incubated at 37°C for 45', then extracted vigorously with an equal volume of phenol, precipitated from ethanol, redissolved in water and aliquoted for storage as an ethanol pellet.

9.2.2 Filter hybridisation of RNA.

This was done under essentially the conditions described by Zernik et al (1981) after Ricciardi, Miller and Roberts (1979). DNA in TE, usually @50µg, was made 6xSSC by addition of 1/3 volume of 20xSSC, boiled briefly and dropped immediately onto 1cm circles of autoclaved nitrocellulose paper. These filters were dried, then baked for 2 hours at 80°C.

Each filter was broken up and placed in a 1.5ml eppendorf tube. To each was added 300µg of fresh total ovary RNA in 10mM PIPES (pH6.4), 0.4MNaCl, 0.2%SDS. The tubes were sealed and incubated at 50°C for 3 hours in a shaking waterbath. The RNA was then removed and the filters washed seven times in 1xSSC (0.5mls, 60°C, 1min), three times in 0.2xSSC(0.5mls, 60°C, 1min) and once in 20mMTris pH7.7, 5mMNaCl, 2mMEDTA (0.5ml, 50°C, 5'). The RNA was then eluted from them by boiling in two successive 200µl aliquots of water, pooling the eluant, adding 10µg of E.coli tRNA carrier, 40µl of 2M Sodium acetate and 1ml of ethanol, and leaving overnight at -20°C to precipitate the RNA.

9.2.3 In vitro translation.

Wheatgerm extract for this procedure was the generous gift of Drs R.W.Old and A.C.Coleman (results in Fig 9.1) and Dr. Andy Cummings (results in fig 9.3). The procedure is as described by Roberts and Patterson (1973). 300 μ Ci of Tritiated lysine (Amersham TRK-520) was freeze-dried in water to a volume of 30 μ l, and to this was added 5 μ l 'Energy Mix' (20mM Dithiothreitol, 10mM ATP, 0.2mM GTP, 80mM Creatine phosphate, 0.4mg/ml creatine phosphokinase), 5 μ l of 0.4M HEPES/KOH pH 7.6, 5 μ l of 1.1M KCl, and 5 μ l a mix of all 20 amino acids (minus the labelled one) at 20 μ M each. 5 μ l of this labelled reaction mix was mixed with 2.5 μ l of RNA dissolved in water and 5 μ l of wheatgerm extract for a standard reaction. The reaction was assembled on ice, then incubated for 1 hour at 25°C. 1 μ l of the result was taken to quantitate total label incorporation into acid-insoluble material, and to the rest was added 6 μ l of SDS-PAGE loading buffer. The samples were frozen at -20°C until required for electrophoresis. The aliquot for counting was incubated in 50 μ l of 0.1MKOH for 10' at 37°C, then, precipitated with 0.5ml of 10%TCA, filtered through nitrocellulose, washed with 5%TCA, dried and counted in POPOP scintillant. An incorporation of 15,000cpm in 1 μ l of a reaction to which 0.5 μ g of total ovary RNA had been added indicates a reaction which is sufficiently efficient.

9.2.4 Separation of the DNA strands of lambda.

This procedure separates the two strands of lambda genomes by complexing them to poly(rUrG) and then banding them on CsCl gradient. The procedure is exactly as described by Szybalski et al (1971). 375 μ g of DNA equivalent of freshly banded lambda phage as phage (in which form an OD(260) of 25 implies a DNA concentration of 1mg/ml) was

9. Gene mapping

dialysed overnight against 1mM EDTA pH8.0. To this was added 188 μ g of poly rUrG previously dissolved in 1mMEDTA, 7.5 μ l of 30% Sarkosyl and 1mMEDTA to a total volume of 1.2 mls. This was boiled for 2mins with swirling in a siliconised glass tube, then chilled in ice-water. These steps lyse the phage, releasing un-nicked lambda genomes into solution: normal preparations of lambda DNA are too severely nicked to be of use in this procedure. The volume was adjusted to 1.7mls with 1mMEDTA, and 6.8mls of saturated CsCl solution in water were added. The density was adjusted to 1.72g/ml (Refractive index=1.4015) using saturated CsCl solution and/or 1mMEDTA, and then loaded into a Beckman SW40Ti swing-out rotor. This was spun at 20,000 rpm, 12°C for 70 hours, a time calculated to be equivalent to that used in a different rotor by Szybalski et al (1971) (Birnie and Rickwood 1974), but which was clearly not exactly equivalent as the bands of nucleic acid were in the bottom 1/3 of the final gradient, not in the centre of the tube. Samples were collected by pumping from the bottom of the tube - 44 samples were collected. With a column length of 70mm and 44 fractions we would expect each of the fractions to represent 1.2mm of the gradient. As the peaks (density @1.760 and 1.740 (Szybalski et al 1971)) should be separated by 4-5mm (Beckman technical data) we can expect a separation of about 4 fractions between the peaks using these conditions.

9.3 Results.

9.3.1 What genes are present on Xbhl?

To answer this whole lambda recombinant DNA was loaded onto filters and used in a HRT assay. The final gel of the translation products is shown in Fig 9.2 . Note that all four core nucleosomes

Fig 9.2

HRT from Xbh1

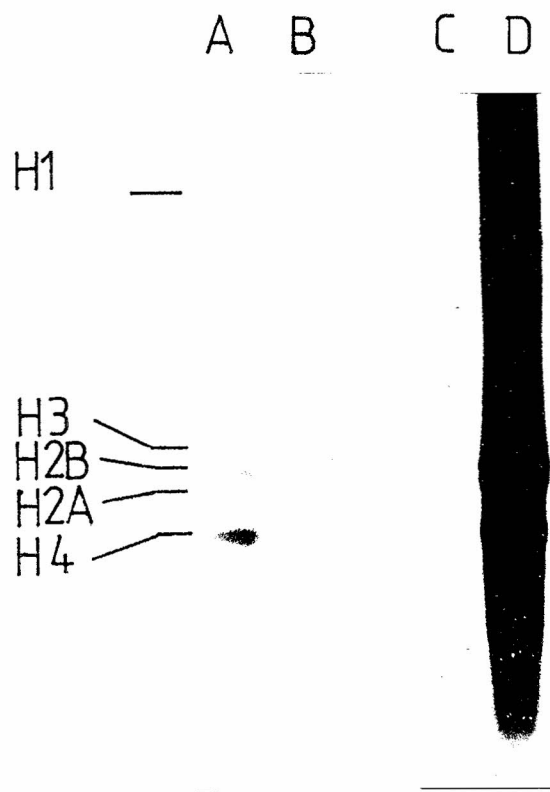


Fig 9.2 Hybrid release translation experiment - see text for details. Fluorograph of 18% SDS-PAGE of wheatgerm translation products of A) RNA hybrid released off Xbh1, B) RNA hybrid released off Xbh6, C) 1ng total ovary RNA, D) 1µg total ovary RNA.

can be seen in the Xbh1 track over their background in the 'Xbh6' track (recall that Xbh6 turned out not to be a histone gene cluster in section 7.3.2). There is also a suggestion that histone H1 can be seen here as well. Thus we can conclude that at least one copy of a gene (or part of a gene) for the four core nucleosomes is present in Xbh1, and possibly an H1 gene too.

9.3.2 What is their orientation?

To determine this, we need to determine from which strand the mRNA is synthesised, ie to which strand it hybridises. This is relatively easy, as standard methods exist for separating the DNA strands of lambda preparatively. These separated strands can then be used in HRT. I followed the procedure of Szybalski et al (1971) described above. The OD(260) profile of the gradient is shown in Fig9.3, together with the HRT results from samples taken from selected point on it. The sample in track 1 comes from the bottom of the tube, and represents small fragments of DNA bound to poly(rUrG). Tracks 3 and 5+6 represent the two separated strands on lambda-Xbh1. Clearly in the former are bands for H4, H2B, H1 and H3 while the latter contains only an H2A band. Thus all the genes on Xbh1 are coded by one strand except for H2A, which has the opposite polarity. If the DNA used in this experiment is banding the same way as does 'wild-type' lambda DNA (and the only reason it should not is if the insert in Xbh1 grossly distorted the pattern of poly(rUrG) binding), then we can relate this to the known poly(rUrG)-binding properties and transcriptional polarities of lambda, and deduce that H2A is transcribed right-to-left, H4, H2B, H1 and H3 left-to-right as the clone is written above in restriction mapping experiments.

Fig 9.3

HRT from separate Xbh1 strands

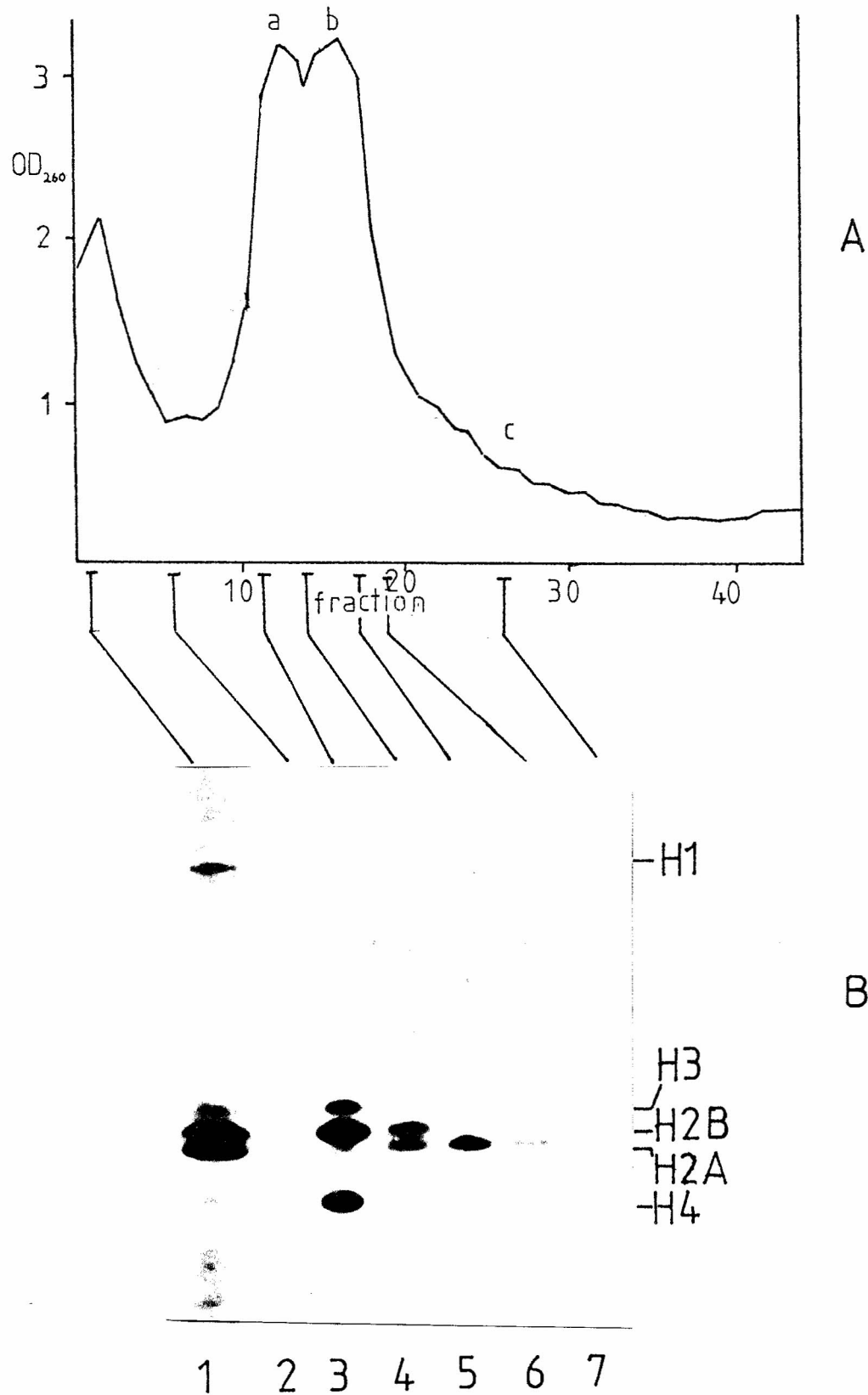


Fig 9.3 A) Strand separation of Xbhl. OD(260) profile of the poly(rUrG) gradient run as per text. Bottom of the gradient is to the left. a=heavy strand, b=light strand, c=banding position of double-stranded DNA. Seven of these fractions (marked below the graph) were taken and used in section B.

B) HRT from separate strands. Fluorograph of 18%SDS-PAGE of translation products of RNA hybrid released off DNA contained in fractions of the gradient. Track 1) fraction 3 (bottom of tube), 2) fraction 7, 3) Fraction 12 (heavy strand) 4) fraction 15, 5) Fraction 18 (light strand) 6) Fraction 21 (light strand) 7) Fraction 27 (double-stranded DNA).

9.3.3 What are the locations of the genes?

I used three methods to answer this question.

9.3.3.1 HRT.

I attempted to use HRT to identify which restriction fragments hybridised to which mRNAs. The result is shown in Fig 9.4 . The amounts of DNA available to do this experiment were apparently not sufficient to give a good signal in all cases. For example, the HindIII 1.54kb fragment gives a clear H4, H2B, and H3 signal, and maybe an H2A one as well. The XbaI 3.98kb fragment, in which the HindIII 1.54kb fragment is completely contained, shows none of these bands clearly. Thus the observation that the XbaI 3.98kb fragment does not hybrid release any histone message does not prove that there are no sequences complementary to histone genes within it. However we can be fairly confident that the HindIII 1.54kb fragment does contain sequences which hybridise to H4, H2B , H3 and maybe H2A, the XbaI 1.82kb fragment contains H2A and H2B, the XbaI 1.75kb probably contains H3, the XbaI 1.02kb contains H4 and the BamHI 1.62kb contains H3. There is also a faint H1 band in the BamHI 1.62kb fragment track.

This means either that this histone gene cluster has at least one H1, one H2B and two of each of H2A, H3 and H4 genes in it (or fragments of genes or pseudogenes: this technique detects sequence homologies, not functional genes), or that one or more of these results is spurious. Because of the data presented below, which is not consistent with the HindIII 1.54 fragment containing any genes, and considerations elaborated in section 10.3.5 and 12.3.2, I am inclined to regard the HRT result of the HindIII 1.54kb

Fig 9.4

HRT from regions of Xbh1

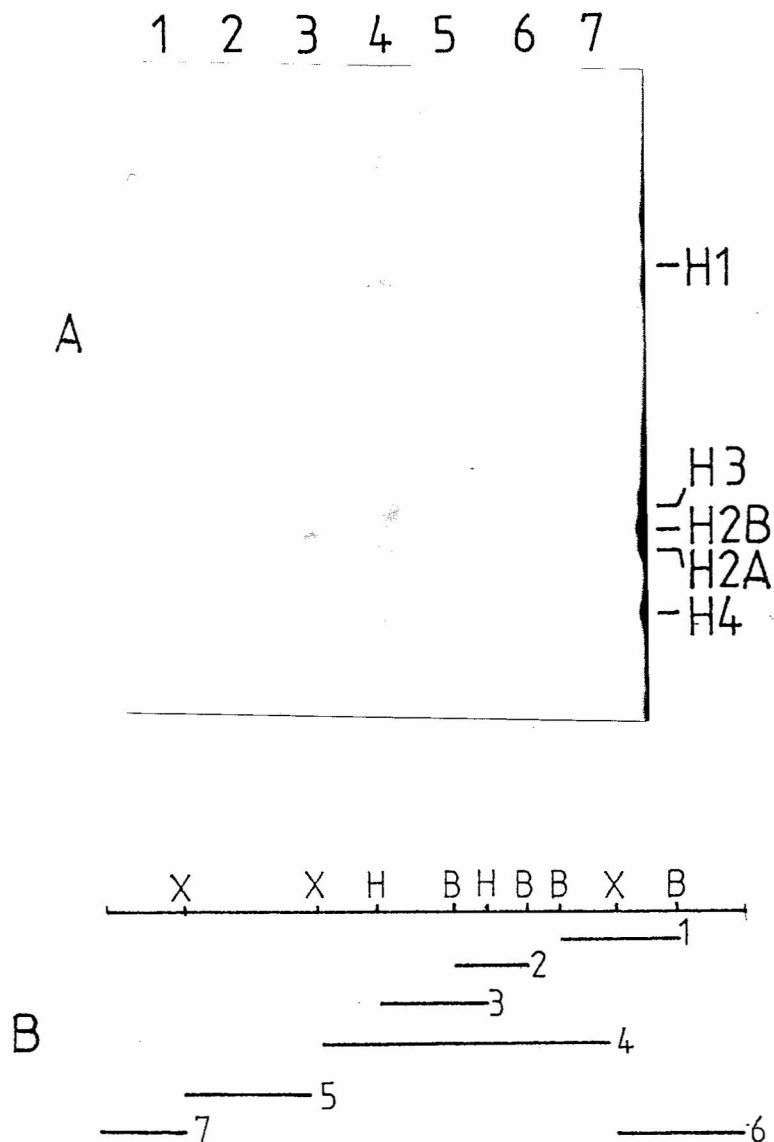


Fig 9.4 A) Fluorograph of 18% SDS-PAGE of translation products of RNAs hybrid released off restriction fragments of Xbh1. B) The restriction fragments used in section A)

9. Gene mapping

fragment as being spurious as far as mapping complete genes is concerned.

9.3.3.2 Hybridisation to heterologous probes.

This study again uses fragments of Xbhl as hybridisation probes: this time labelled fragments are used to probe onto Southern blots of restriction digests of the characterised Sea urchin histone gene clones pSp102 and pSp117 (Kedes et al 1975a,b , Sures et al 1978).

The rationale is as follows. pSp 102 and pSp117 may be digested with several enzymes to yield fragments containing only one gene, which do not co-migrate with any other fragments of the gene cluster. Specifically, digestion of pSp102 with HhaI+EoRI yields a 1.9kb fragment containing the H1 gene, a 1.6kb fragment containing the H2A and a 1kb fragment containing the H4. Digestion of pSp117 with HpaII gives a 1.25kb fragment containing H3 gene, and HinfI digestion leaves the H2A gene on a fragment 2kb long (Sures et al 1978). Thus by running these three digests on a number of duplicate gels, Southern Blotting and probing with fragments of Xbhl, I can determine which genes, if any, are on those fragments simply by the pattern of bands they light up.

The results are presented in Fig 9.5. Note that the HinfI digestion of pSp117 was not total, so the sizes of the bands cannot be accurately determined in this track. These are the positive results only. A number of the hybridisations gave no bands at all, either because there were no sequences homologous to histone genes on the probes used or because the hybridisation was too inefficient because of poor matching between the DNA sequences, inefficient Southern transfer or other causes: it is not possible to determine

Fig 9.5

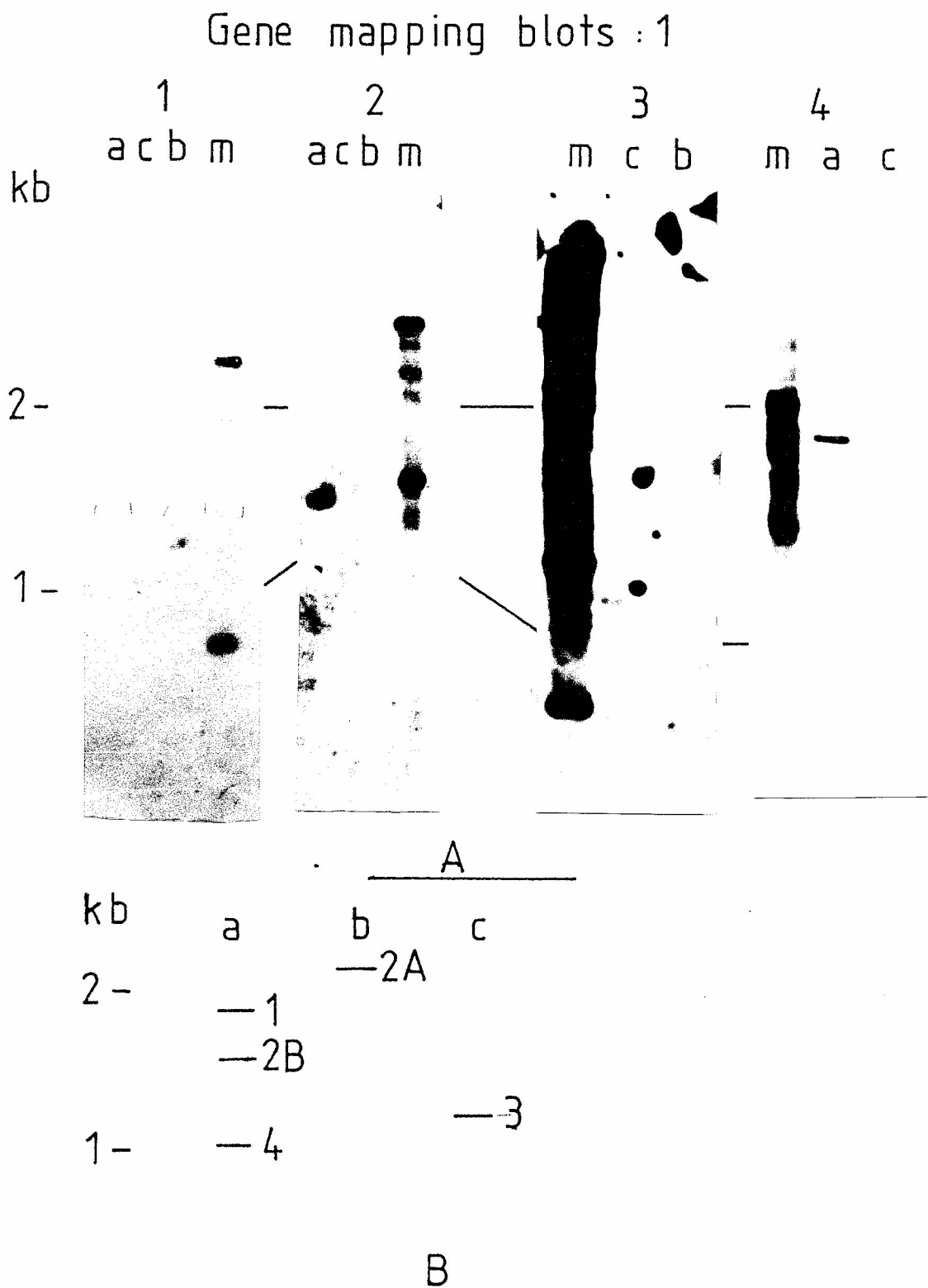


Fig 9.5 A) 0.8% agarose gels of a) EcoRI + HhaI digest of pSpl02, b) HinfI digest of pSpl17, c) HpaII digest of pSpl17, m) marker. These were Southern blotted and probed with 1) Xba 1.82kb fragment, 2) Xba 1.02kb fragment, 3) Sst0.97kb fragment, 4) BamHI 1.62kb fragment.

B) Schematic representation of the bands which would be lit up in digests a,b and c by probes containing 1-H1 gene, 2A- H2A gene, 2B-H2B gene, 3- H3 gene, 4- H4 gene. Bands smaller than 600bp have been ignored because they are poorly resolved in this system and would not have been detected.

9. Gene mapping

which of these possible factors was operating to produce any particular negative result.

9.3.3.3 Probing Xbhl with homologous and heterologous probes.

The converse experiment, probing restriction digests of Xbhl with cloned probes, yielded the clearest results for all genes except H1. The origins and characteristics of all the probes are which are relevant to this experiment. A variety of restriction digests was used: the most informative results are shown in Fig 9.6. Note that pSp102 contains H2B, H4 and H1 genes. Thus if a region is lit up by pSp102 which is not lit up by pcX1H4W1 (H4 only) or pSp2b-5 (H2B only) it should contain H1. However this is not a very clear method of detecting a gene, and attempts to probe Xbhl with the H1-gene subclones pSp1-4 and M13-H1 were not successful in supporting gene locations made on this basis. However sufficient consistent data was accumulated by the three approaches described here to assign the position of the H1 gene to the left-hand end of the BamHI 1.62kb fragment. Sequence data have subsequently confirmed this location (see section 15.1).

As with the restriction map data I will not go through the detailed logic by which the gene map is derived from the data presented in Figs 9.4 - 9.6. This data is summarised in Fig 9.7, which also presents the map derived from it.

We may note that the absolute gene polarity assignments are confirmed by the data above, on the basis of the H3 gene mapping data. LF60 contains only the extreme 3' end of an H3 gene (Turner and Woodland 1982), and hybridises to the terminal BamHI 0.97kb fragment (which in the above digest is part of a larger fragment containing

Fig 9.6

Gene mapping blots : 2

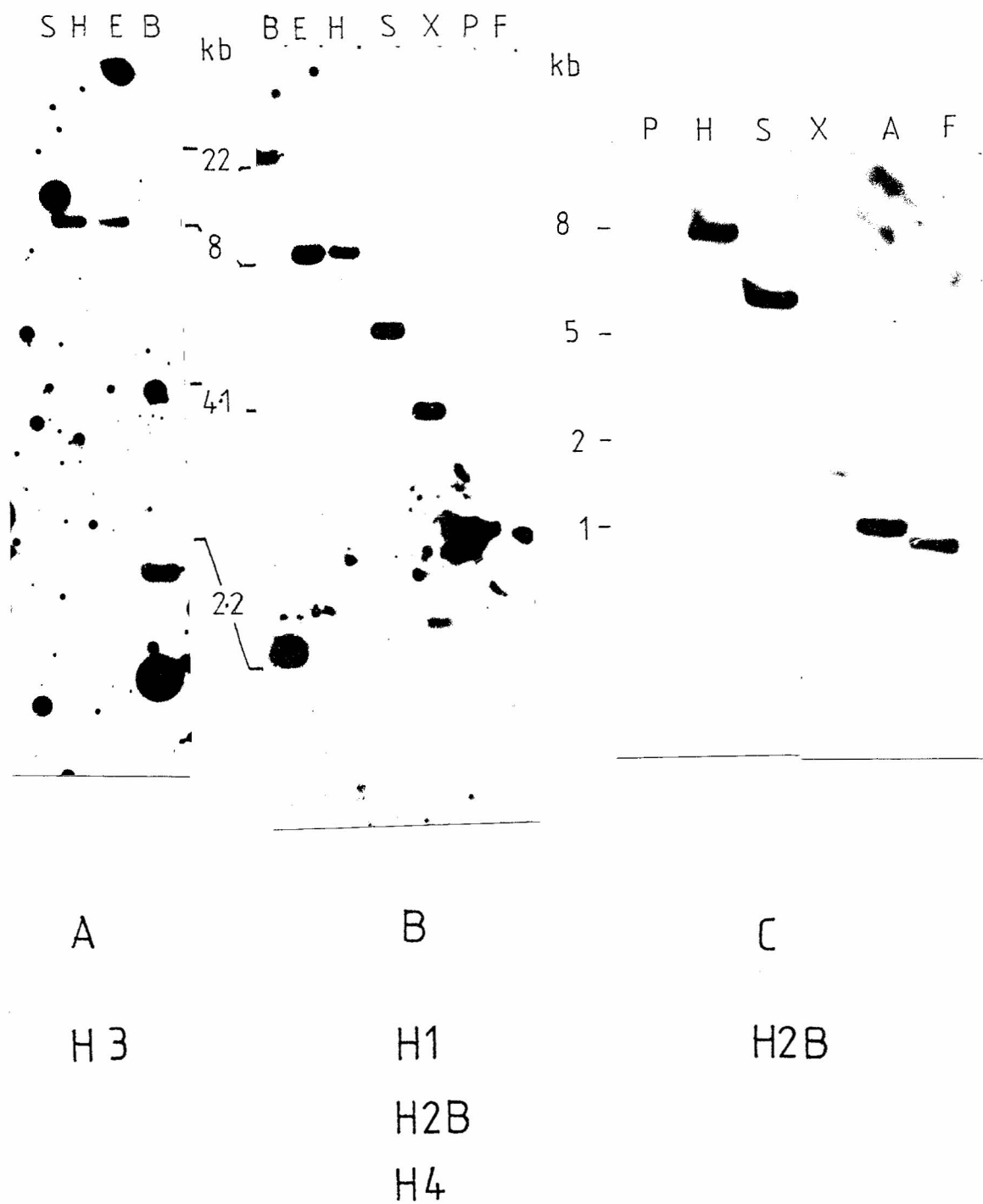


Fig 9.6 cont

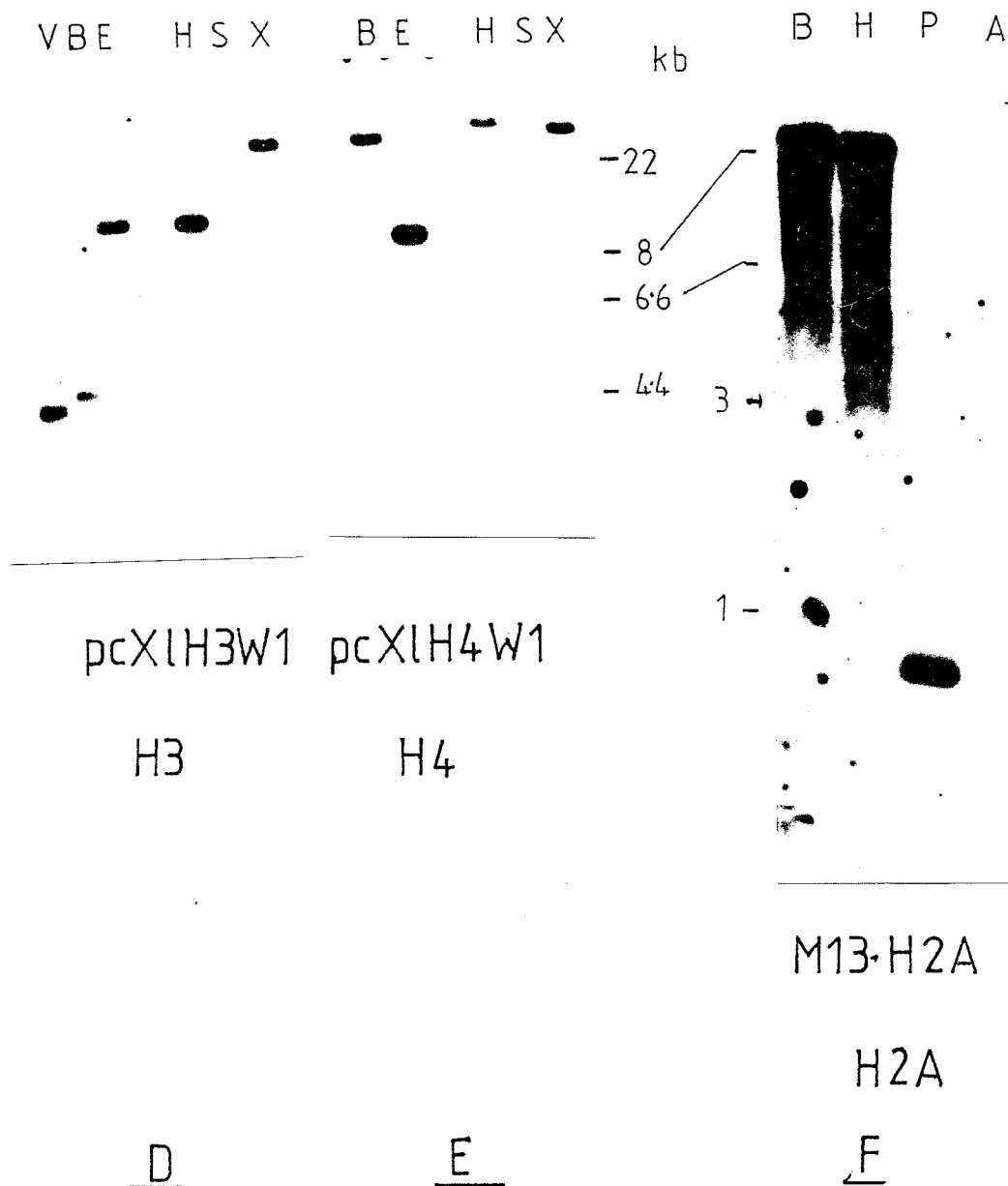


Fig 9.6 Restriction digests of Xbh1 (A,B,D,E) and pWB2 (C,F) probed with cloned gene probes. Probes: A) pSp3-1, B)pSp102, C) pSp2b-5, D) pcX1H3W1, E) pcX1H4W1, F) M13H2A. The genes contained in these probes are marked under the gels. Restriction digests are: A-HaeIII, B-BamHI, E-EcoRI, F-HinfI, H-HindIII, P-HpaII, S-SstI, V-AvaI, X-XbaI

9. Gene mapping

lambda sequences), but not to the BamHI 1.62kb fragment. pSp3-1, the Sea urchin H3 gene clone containing most of the central coding region of an H3 gene, does hybridise to the BamHI 1.62 kb fragment. Thus the BamHI 1.62kb fragment is nearer the 5' end of the H3 gene than the BamHI 0.97kb fragment, in agreement with the previous polarity assignments. This has also been confirmed subsequently by sequence studies (section 15.2).

9.4 Conclusions.

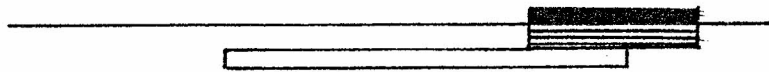
The map of the genes on Xbhl is presented in Fig 9.7. We can draw no conclusion from this on whether these 'genes' are complete, functional genes or whether they are pseudogenes, fragments of coding sequence analogous to the Orphone of Childs et al (1982) or even sequences with no relationship to histone genes other than chance sequence homology. Their homology with histone genes must be fairly precise in order to hybrid-release histone mRNAs under the conditions used (Zernik et al 1981), so this latter possibility is unlikely. Furthermore we know that the 3' end and the centre of the H3 gene is present. Proof that some of these genes contain complete coding regions free of spurious termination codons, the usual hallmark of pseudogenes (Proudfoot 1980) will be presented later.

I will reserve discussion of how this gene map may be related to other cloned histone gene clusters until the discussion section 17. Here, however, we can elaborate one specific similarity. Xbhl is not a clone of the major type of histone gene cluster described by Turner and Woodland (1982b). However the left 30% of this clone is remarkably homologous to the central region of the major cluster. If we allow a 600bp deletion between the H4 and H2A genes of the major cluster, the gene positions and 7 out of 14 restriction sites, 4 of

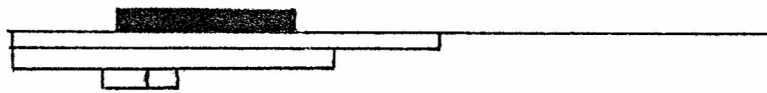
Fig 9.7

The gene map of Xbh1

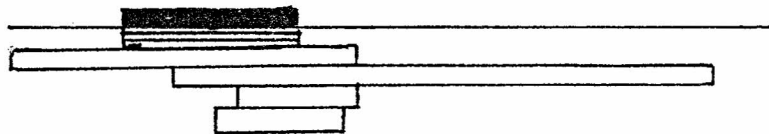
H1



H2A



H2B



H3

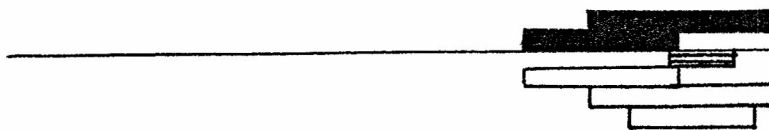


Fig 9-7 cont

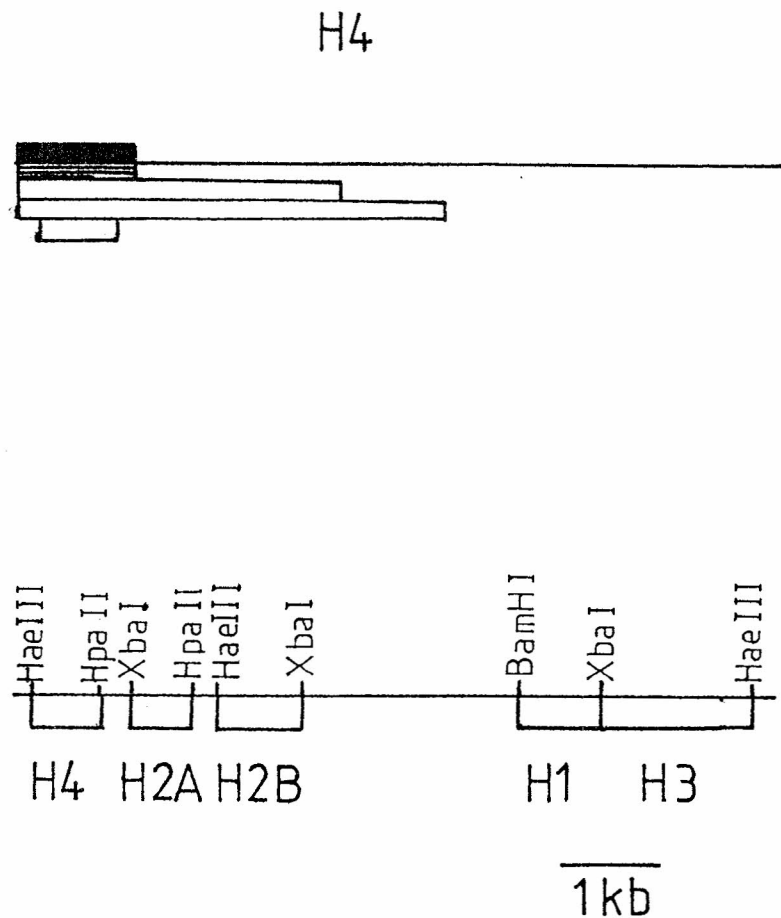


Fig 9.7 Derivation of the final gene map from hybridization data. The first five maps show the restriction fragments found to hybridize to the probes for the five genes: solid boxes, data from Fig 9.4 (HRT); Hatched boxes, data from Fig 9.5; Open boxes, data from Fig 9.6. The final map is shown below these data. The restriction sites marked on this map are the termini of the regions which hybridize to the various probes used above, and not necessarily the termini of the genes. This is particularly true of H3

them in non-coding DNA, are almost exactly aligned in Xbhl and the major cluster (Fig 9.8). The likelihood of this occurring at random within the resolutions of the mapping is less than one in 10 to the 12, something I therefore cannot assign to chance.

However the rest of Xbhl shows little homology to the major cluster. It is notable that the H3 gene of Xbhl contains a BamHI site at its 3' end, as do the H3 genes in many histone gene clusters from X.laevis, while the H3 gene in the major cluster contains no such site. The restriction pattern of the H2B gene coding regions are even more divergent between the two structures. I will leave further discussion of these points to later in the thesis, when other pertinent data have been presented.

Fig 9.8

Xbh1 and the 'major cluster'

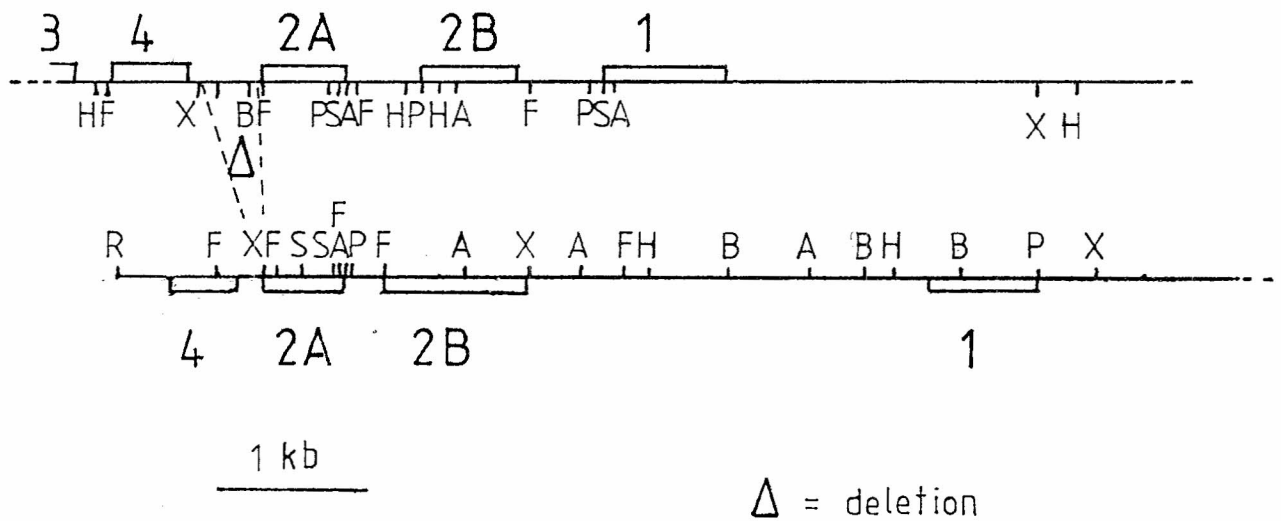


Fig 9.8 Similarity between the left 1/3 of Xbh1 and the centre of the X.borealis major cluster. Top line - central region of the X.borealis major cluster (Turner and Woodland 1982b). Lower line - Xbh1 . Restriction sites H-HindIII, R-EcoRI, F-HinfI, X-XbaI, S-SstI, A-AvaI, P-PstI, B-BamHI.

10. Repetitive DNA in Xbhl.

The Beaver had counted with scrupulous care,
Attending to every word,
But he fairly lost heart, and outgrabe in despair
When the third repetition occurred.

10.1 Introduction.

There are three reasons why we might wish to know the location and character of any elements in Xbhl which are homologous to sequences repeated in the genome several hundred to several million times, and hence falling in the category of 'repetitive DNA'. Firstly because such elements provide another locus on the clone by which it might be compared to other cloned DNAs, either repetitive genomic DNA clones or histone gene clusters. In particular we would wish to determine whether the histone gene cluster represented by Xbhl is embedded in satellite DNA as is true of the Notophthalmus histone gene repeat (Stephenson, Erba and Gall 1981a,b), or if similar repetitive elements are present in the major histone gene cluster of X.borealis (Turner and Woodland 1982b). Such elements could also be compared to other repetitive elements in Xenopus (Spohr, Rieth and Sures 1981) or other (Jelinek et al 1980, Schmid and Jelinek 1982) genomes.

Secondly, the location and nature of repetitive DNAs may give insights into the evolution of the surrounding DNA. Both as mediators of illegitimate recombination and as potential transposing or inserting elements, repetitive DNAs contribute to genome fluidity (vide supra section 2.7). Of course, they may not yield any such data - a considerable body of information has to be accumulated before

evolutionary pathways can be established for any gene.

Thirdly, and most importantly considering the use for which the clone was originally intended - the study of the control of gene expression - repetitive elements can both obscure the characterisation of a cloned DNA and diminish its usefulness in future experiments if not adequately characterised. Repetitive DNA close to the end of an H3 gene in a genomic gene clone has already been shown to bias gene number determinations in the studies of van Dongen et al (1981) (van Dongen 1982) on X.laevis. Analogous problems with extensive background hybridisation in studies using cloned DNA to probe genomic DNA (vide infra section 11.3.2) and using oocyte cDNA to probe onto cloned DNA (R.W.Old per comm and vide infra section 12.3.1), caused by the presence of an unsuspected repetitive element in the clones illustrate the problems which an unidentified repetitive element can cause.

The essence of the screening procedure used here for repetitive DNA is simple and quick. We are looking for sequences of the clone which are homologous to elements present in at least several hundred copies in the haploid genome. These can be isolated by self-annealing the genome to low Cot, and indeed such isolates have been used as probes in the analysis of cloned DNAs (Brison, Ardeshir and Stark 1982). However if the whole genome is hybridised to the clone without prior fractionation, the repeated DNAs in it will hybridise at lower Cot anyway, the reaction being pseudo-first order with respect to the genomic reiteration frequency. If the hybridisation is carried out with an excess of cloned DNA and for a short time, the extent of hybridisation becomes kinetically determined solely by the concentration of homologous sequence in the DNA probe. Thus our screen is simply to probe the clone with labelled genomic DNA: the extent to

10. Repetitive DNA

which this probe hybridises to a given region of the clone is a measure of the repetition of that region in the genome. In practice this means probing a Southern blot of restriction digests of the clone with nick-translated genomic DNA. Because hybridisation to DNA bound to nitrocellulose is considerably slower than DNA-DNA association rates in solution (Britten, Graham and Neufeld 1981), and because only small amounts of genomic DNA need be used to prepare a nick-translated probe, the standard hybridisation conditions given in section 5.7.2 will be suitable for this screen.

10.2 Methods.

10.2.1 Filter hybridisation.

This is after Turner and Woodland (1982b) and Kafatos, Jones and Efstradiatis (1979). Xenopus borealis genomic DNA and pWB2 plasmid DNA were sheared by sonication for three seconds and then boiled for 1 minute immediately prior to loading onto 1cm diameter nitrocellulose filters. 7.56µg of Xenopus and 3.7µg, 0.37µg, 0.037µg, 3.7ng, 0.37ng and 0ng of pWB2 DNA were loaded onto the filters in sets of six: in the first of these the ratio of X.borealis haploid genomes to pWB2 plasmids was therefore 1:34,000. The filters were dried and baked for 2 hours at 80°C. Each set of six was separately pre-hybridised for 24 hours in 3xSSC, 10xDenhardt, 50°C, then hybridised to a nick-translated probe DNA under the same conditions for 24 hours. The sets of six were incubated in separate 50ml beakers in a shaking water bath.

Filters were washed together in 1 litre of 2xSSC for 60-80 mins. Filters were washed thrice initially at 50°C, and then once at 55°C,

10. Repetitive DNA

60°C, 67.5°C, 71.5°C, 74°C, 78°C, and 87°C. After each wash they were counted by Cerenkov counting in water.

10.3 Results.

10.3.1 Locating repetitive DNA in Xbhl.

I ran several digests of Xbhl on agarose gels, Southern blotted them and probed with total genomic X.borealis DNA. The results of two gels are shown in Fig 10.1. This procedure fails to give significant hybridisation to pXlH4W1, showing that it will not detect the histone genes themselves with repetition frequencies of only 50-100. However it does light up pXlrl01, (Boseley et al 1979) a clone containing the X.laevis rDNA, about as well as it does Xbhl, which indicates that the genomic representation of some part of Xbhl is similar to that of the ribosomal repeat.

We can note that only a few bands are lit up in Xbhl by probing with genomic DNA. As there are large number of HpaII, HaeIII and HinfI fragments in Xbhl (vide supra Fig 7.5) this suggests that the hybridising region is located in only a small fraction of the clone. In fact, two intensities of hybridisation can be seen when 4-cutter digests of Xbhl are probed, illustrated by the HpaII bands of 1.75kb (intense) and @700bp (not so intense). This suggests two regions which hybridise with different efficiency: these are mapped onto the clone in Fig 10.1, and are designated IR1 and IR2 (for Intermediate Repetitive DNA).

10.3.2 Copy number of the repetitive elements.

Fig 10.1

Repetitive DNA in Xbh1

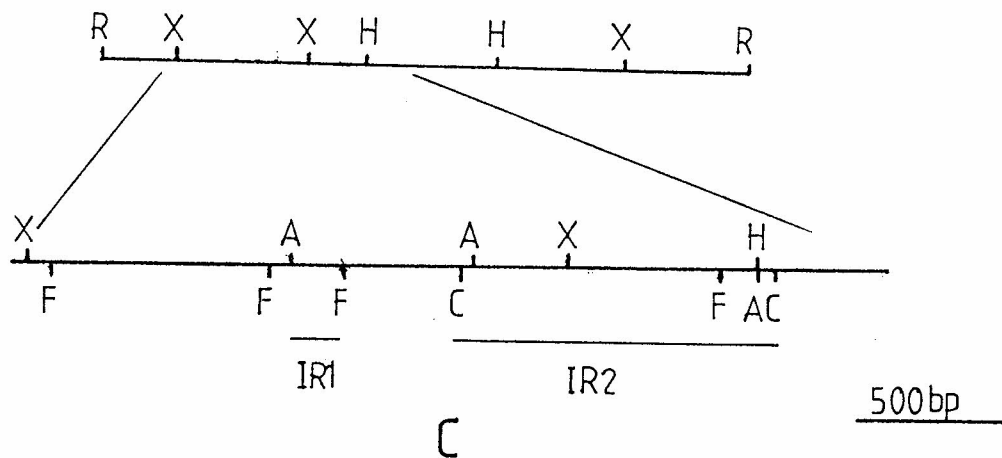
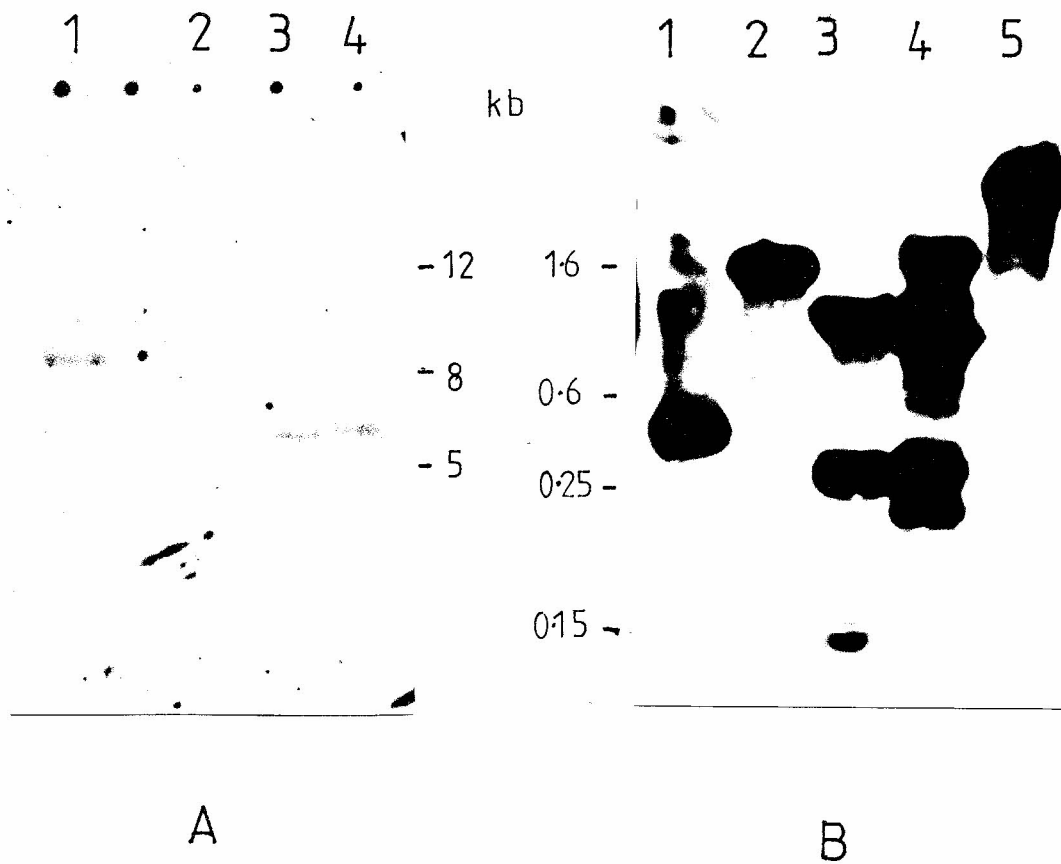


Fig 10.1 Probing restriction digests of Xbhl with nick translated X.borealis genomic DNA: hybridization and wash at 60°C, 2xSSC. A) 0.7% gel 1) EcoRI digest of Xbhl, 2) Pst digest of pcXlH4W1, 3) BamHI digest of pXLR101, 4) Pst digest of Xbhl. B) Digests of pWB2. 1) BamHI+HindIII, 2) HpaII, 3) HaeIII, 4) HinfI (note that this is not a total digest) 5) HhaI. C) Map of part of Xbhl showing the regions detected by this assay. R=EcoRI, X=XbaI, A=AvaI, F=HinfI, C=HincII, H=HindIII P=HpaII

10. Repetitive DNA

I used a variation of the Dot-blot hybridisation procedure of Kafatos, Jones and Efstradiadis (1979) to determine the copy-number of the repetitive elements of Xbhl. This method actually determines the total Representation of a sequence in the genome, ie its copy number multiplied by its length (vide infra Appendix 3). From the results above we already know this value is similar to the 18S+28S rDNA, ie 1.5Mb. The quantities of DNA used in this experiment were therefore adjusted to best enumerate representations of this order.

I used seven restriction fragments of Xbhl as probes, and LF53 and pAT153 as control: the former gave an H4 gene copy number of about 200 which is a fair result for a system designed to measure representations at least an order of magnitude higher.

The probes used are shown in Fig 10.2. Each of these probed 6 filters with varying ratios of pWB2:Xenopus genomic DNA loaded onto them. The experiment is described and its analysis discussed in Appendix 3. In brief, the labelled probe was hybridised to a set of six filters with a fixed amount of X.borealis genomic DNA and different amounts of pWB2 on each filter. The amount of label bound to each filter is related to the number of copies of homologous DNA on the filter, itself related to the copy-number of the genomic sequence in the DNA. The varying amounts of pWB2 on each filter provided an internal control which, by a simple arithmetic manipulation, can be used to convert the amount of label bound to the genomic DNA frequency of the hybridising species. The results and the preliminary data and its method of treatment are given in Appendix 3 : the representation figures are given in Fig 10.2 . Note that these are not copy-number figures. They are Representation, ie copy number multiplied by length, and are in kilobases per haploid genome. Notice also that they become rather erratic at 78°C. The

Fig 10.2

Genomic representation of IR2

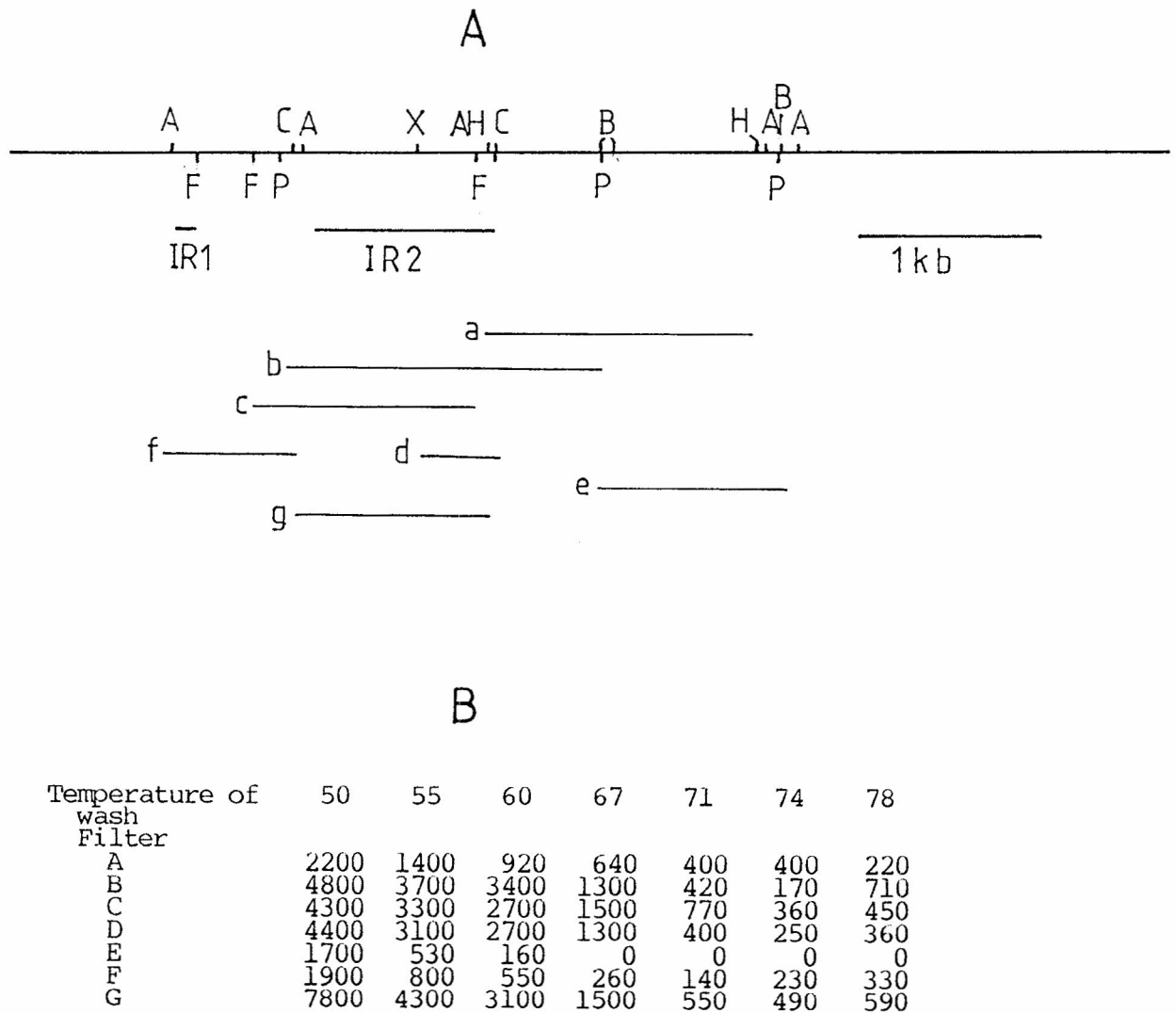


Fig 10.2 A) Location of probes used in the genomic representation experiment. Restriction enzymes as in 10.1. All probes were nick-translated. B) Genomic representation (in kb) of the probes shown in section A at seven wash temperatures. ND= not determined. The primary data and its manipulation to produce these figures are given in Appendix 3

10. Repetitive DNA

numbers were calculated relative to the pWB2 internal standard, and so are internally controlled for leaching DNA off the filters, variations in hybridisation efficiency and so on. However at 78°C the hybrids with the pWB2 standard themselves begin to melt (Data from Appendix 3) so this internal control breaks down.

We can see from this that the 0.9kb BamHI fragment is essentially present in zero copies (± 100) per genome. As this lies outside the repetitive region this is as expected. The short region of repetitive DNA between H2A and H2B (IR1) is present in about 500 kb representation at 60°C as detected by the 0.7kb AvaI fragment probe. As this probe contains only part of the 200bp IR2 region (as it is determined by restriction studies in Fig 10.1), the total representation of IR1 is $\approx 1\text{Mb}$, giving it a copy number of $\approx 3000/\text{haploid genome}$.

IR2 is not so easily analyseable. However we may argue as follows. Let us assume that IR2 is a single homogenous block of sequence with respect to its genomic repetition, so that different parts are represented to the same extent in the genome, ie it is one 'element'. This is the simplest model. Taking the figures for the 60°C wash, we can deduce what contribution this block of repetitive sequence makes to the genomic DNA hybridising with each of the probes which include a part of it. This is simplified by calculating representation rather than copy-number: if a fragment of representation 'x' is adjacent to a fragment of representation 'y', the representation of the combined stretch of DNA is $x+y$. This is not true of copy number, which depends on whether 'x' and 'y' are part of the same element or not.

Such deductions produce consistent results from the 60°C and 55°C

figure . However the figures are not consistant with a single, homogenous block in the cases of other temperatures: at least three blocks must be postulated (each homogenous within itself) to explain how the probes behaved at 50°C, 67°C and 71°C. The diagrams of total genomic representation at different tempemperatures are shown in Fig 10.3. These are the simplest possible explanations consistant with the data: other more complex ones, such as the proposition that the IR2 element contains three quite unrelated elements and relating this to the data for 55°C and 60°C, are also quite consistant.

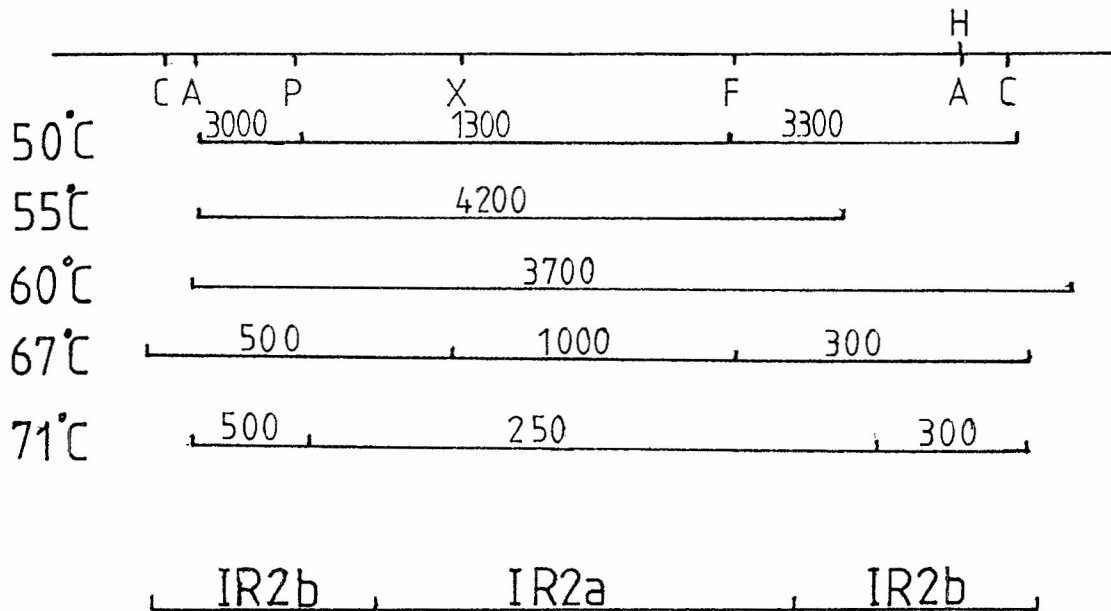
These data are all consistant with a single pattern, given at the bottom of Fig 10.3 . The IR2a element is present in ©1Mb representation in the genome, and its hybrids to pWB2 melt at a fairly sharp temperature between 67°C and 71°C, denoting ©5% mismatch (see Appendix 3 for hybrid melting data). The IR2b elements are present at a similar representation at 60°C, but their hybrids with pWB2 melt with much more complex temperature dependance, a significant proportion between 50 and 55°C denoting mismatch of >15%. Thus the different parts of IR2 have quite different homologies with families of repeated sequences in the genome.

It is curious that IR2a gives such a sharp melting point: this suggests that it is ©5% diverged from a very uniform family of sequences. By contrast the IR2b elements show a dispersed pattern of homology typical of the Alu-like sequences of mammals (Schmid and Jelinek 1982) where any one family is highly homogenous within itself, but is distantly related to a much wider circle of sequences. I will take up this IR2a/IR2b difference again later.

It is interesting to note that both of the IR2a 'end' regions have adjacent Aval and HincII sites at their ends. These are at

Fig 10.3

Structure of IR2



10.3 Potential structure of IR2. Assuming that IR2 is made up of uniform abutting blocks of repetitive DNA, this figure shows the minimum number of blocks needed to explain the data for each temperature in Fig 10.2, their genomic representation and their extent. At the bottom is a summary figure in which IR2 is divided into three areas. An interpretation along these lines is consistent with the data presented for all temperatures in Fig 10.2

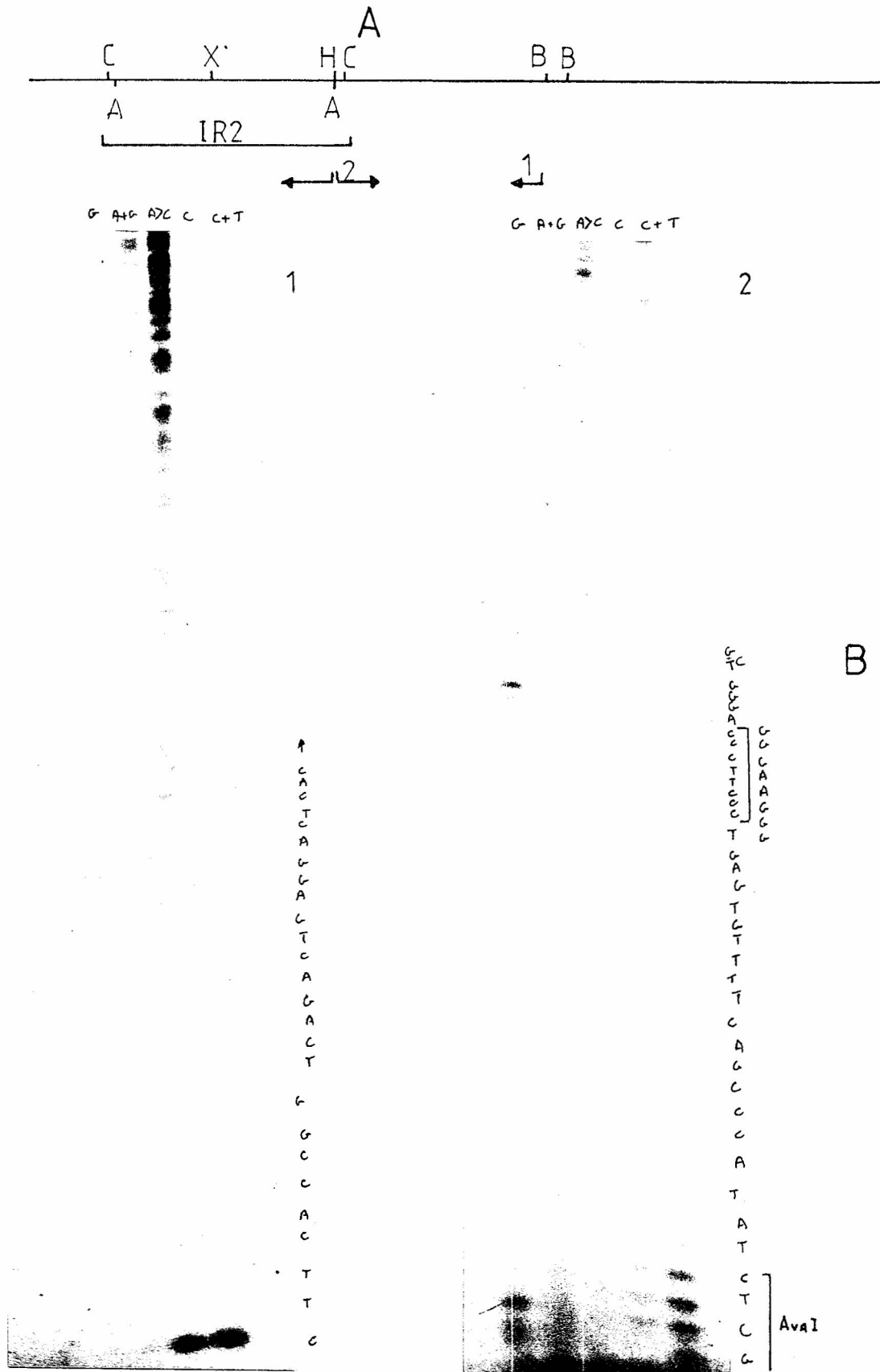
opposite ends in the two regions, and their order is inverted, suggesting that, if these two elements are related (and, although I have called them both IR2b there is no evidence relating them except similar melting curves with genomic DNA), they could be inverted with respect to each other.

10.3.3 Sequence studies on IR2b (right).

I used Maxam and Gilbert sequencing methodology to study the sequence around the HindIII restriction site in the right-hand end of IR2 (Maxam and Gilbert 1978). I end-labelled the HindIII ends of the left HindIII 3.6 fragment and the central 1.54 fragment of pWB2 with Klenow polymerase, cold dATP and p32-labelled dGTP, digested the products with MspI and isolated the fragments bearing those termini. I also labelled the large BamHI 4.2kb fragment, cleaved it with XbaI and isolated the labelled end containing the left-most HindIII site. This strategy and the sequence is given in Fig 10.4. Also given is the AT/GC richness of the sequence, calculated per 2bp step on a 14bp region. This shows the sequence to be @40% A+T with with some fluctuations. I have not sequenced across the HindIII site, and so I cannot say how many bp the gap between the two sequenced regions represents. However, if there is a small HindIII fragment between the central and leftmost ones, detailed restriction mapping of this area shows that it cannot be more than 20-30bp long. On the same gels as I used to sequence the DNA adjacent to the HindIII site I also ran the sequence of the other end of the fragment from which the left-hand fragment was obtained with its HindIII site labelled: this fragment was a vector fragment. I can tell from this approximately how many bases from the HindIII site the sequence presented in Fig 10.4 begins. Based on this and on the assumption of no small HindIII fragment between the two sequenced fragments, I have spaced 10-14

Fig 10.4

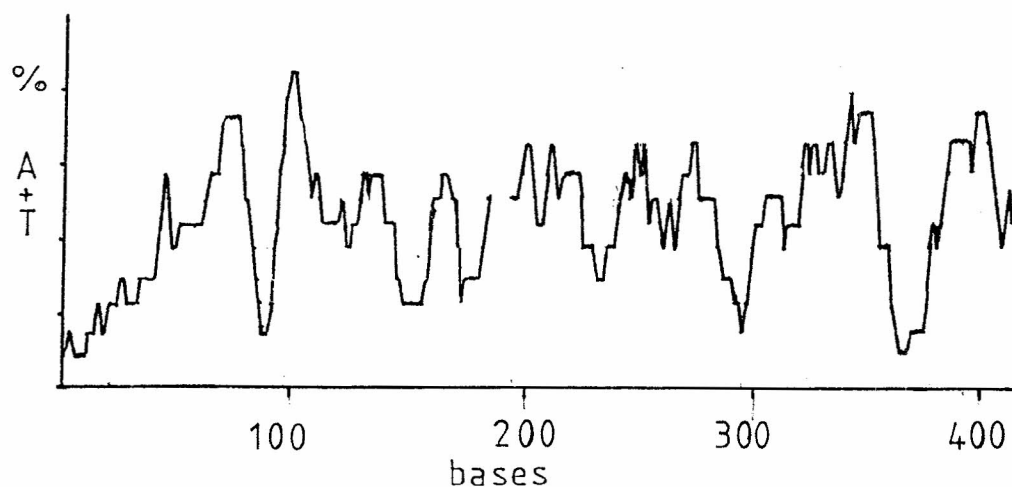
DNA Sequences from IR2



C

1
 CCACC GCCGG CCCAC CGGGG CCAGC CACNN GNGGA CGGCA 40
 TACGC GGACT CGAGA CACGA CNAAG AATAC AATAC AACCA 80
 CCCCC AGCCC CCCAA GTTTA AAGTT ATTGA CCCAG GTAGA 120
 AGGTG GGACT TGGCC AAAGG TTTAG GAGGT CGGCC GTGGC 160
 ACATT AANGG TTCCG CCNGN AAGCT T -4-6 bp -
 194 Ava II
 ATAGC TCGAG ATATG GGCTG AAAAC ACTCA GGGAA GGGTC 233
 CCAGT CNACT GTATT GGGTT TGCTT CCTCT GCCGT ATATC 273
 CCTAC TTGCT CGTTG AGCCG CACGG GGCAT GTATC CGTGG 313
 GACAG CTTAG GTTTA TCCTC CAGTA AGAGA AGAAA AAAGC 353
 TTTGC AGTGC CGGGC GGGTG TNGCG GATTC TCGTT ATTTG 393
 TTTGG TATTG TGTCG GCTAA GTTGT TGTT
 422

D



E

```

1
AAGNA AATTT GTAAA GGAAT ACAAN AAAAN 30
AAAGA AAATT CNACA ATTTT AGGAG ACAA 60
TACTA TTCTC TGTTT AAGCA CTCAG GATCA 90
GACTG GCCAC          TTC G  ↑
                        103 Bam

```

Fig 10.4 Sequence data from IR2 region.

A) Sequencing strategy. Symbols as in Fig 10.1 B) Sequence gels showing the determination of the sequence around the central HindIII site, including the GGGAAGGG motif. Reactions (G, A+G, A+C, C, C+T) are indicated on the top of the gel. The gels represent the stretches of sequence indicated in part A

C) Sequence around the left HindIII site of Xbhl, including the end of the IR2 region. HaeIII, AvaI and HincII restriction sites are marked. Sequence is written 5' to 3'.

D) A+T richness of the sequence. Stretches of 14bp are counted for their content of A+T. The 14bp frame is translated 2bp for every calculation.

E) Sequence from the left-most BamHI site of Xbhl.

unsequenced bases between the sequenced regions.

This sequence has no homology to human Alu (Schmid and Jelinek 1982), Drosophila 412 element terminal (Will, Bayev and Finnegan 1981), consensus TATA, CAAT or polyadenylation site (Breathnach and Chambon 1981) or Xenopus laevis rDNA spacer (Boseley et al 1978) sequences. There is a very weak homology between bp 306 to 322 inclusive and part of the histone 3' homology block (Hentschel and Birnsteil 1981), and a stronger one between bp224-252 and the internally repetitive, dispersed intermediate repetitive element found in X.laevis by Spohr, Reith and Sures (1981). These homologies are illustrated in Fig 10.5 . If the former were a true histone homology block it would point to a histone gene to the right of IR2, which gene is not in evidence. The latter homology contains a perfect match for the sequence GGGAAGGG, and contains the HincII site noted to terminate the IR2 element at both ends together with an Aval site. With the exception of the sequence GGGAAGGG neither homology is very impressive.

The sequence is not internally repetitive, although this statement cannot therefore be extended to all of IR2b (right). There is no obvious change in A+T-richness anywhere in this sequence to suggest an 'end' to the IR2 element, and the start of intergene spacer (which is often A+T-rich). However about 80bp right from the HincII site the nature of the sequence changes: short runs of 3 - 6 A or T residues are seen in contrast to the rest of the sequence where there is only one run of A or T residues longer than 3bp (AAAA at bases 214 - 218). The sequence from the BamHI site to the right of IR2 shows the same pattern of oligo(A) and oligo(T) runs: this region is outside the region of repetitive DNA, and so is part of the non-repeated intergene spacer unique to this cluster. It is likely,

Fig 10.5

Analysis of IR2 sequence

A



B



Fig 10.5 Short regions of IR2 are compared with A) Part of the repetitive element of Spohr, Reith and Sures (1981): 1) Repetitive element, 2) IR2 sequence bases 224-253 inclusive. B) The histone homology block (Hentschel and Birnstiel 1981) is compared to IR2 bases 322-306: note this latter is written backwards, IE as a 3' to 5' strand

10. Repetitive DNA

therefore, that the start of this oligo(A).oligo(T)-rich region does indeed mark the end of IR2 and the start of the non-repeated intergene spacer, and so the terminus of IR2 at the right-hand end can be more precisely positioned between 340 and 360bp from the start of the sequence presented in Fig 10(c).4 .

10.3.4 Phylogeny of IR2.

Alu-like sequences of man have conserved features found in other repetitive elements in primates and in similar elements in many other mammals (Schmid and Jelinek 1982). RATS also share some sequences between species (Calos and Miller 1980, Will, Bayev and Finnegan 1981). Is the repetitive element IR2 similarly conserved in amphibia or is it a more divergent element? To give some indication I repeated the screen performed in section 10.3.1 above, but instead of probing Xbhl with X.borealis DNA alone I probed a series of identical blots with a range of different genomic DNAs : X.borealis, Xenopus laevis (the gift of Dr.R.W.Old), Triturus cristatus and Ambystoma mexicana (the gift of Mr. John Adair), chick (The gift of Dr. Phil Turner) and calf and salmon sperm (the gifts of Drs John Norton and John Clewley) genomic DNAs. The results are tabulated in Fig 10.6: a positive score means that the test DNA hybridised to the same fragments of Xbhl as did X.borealis DNA. The filters were washed at successively higher temperatures in 2xSSC and autoradiographed after each wash. The nick-translated salmon sperm DNA seemed to bind to all of the restriction fragments, contributing a significant background over which any repetitive DNA hybridisation could not be detected after the 60°C wash. The reason for this high background is unknown.

These results demonstrate that the IR2 sequence is present in

Fig 10.6

Phylogeny of IR2

	55°	60°	65°	70°	75°
X.laevis	✓	✓	✓	✓	✓
X.borealis	✓	✓	✓	✓	✓
T.cristatus	✓	-	-	-	-
A.mexicanum	✓	?	-	nd	nd
Calf	?	-	-	-	-
Chick	✓	?	-	-	-
Salmon	✓	?	-	-	-

Fig 10.6 HpaII, HaeIII, HindIII, EcoRI and BamHI digests of Xbhl were separated on 0.7% agarose gels, Southern blotted, probed with genomic DNA from one of the seven species listed at 50°C, 2xSSC and then washed in 2xSSC at successively higher temperatures. A positive score indicates that, after that wash, autoradiography of the filter showed the same pattern of bands as was seen on the X.borealis-probed filter at 60°C. tick=positive score, -= negative score, ?= uncertain, nd= not determined.

both Xenopus species, in Ambystoma and Triturus. It is probably also present in Chick and maybe in Calf and Salmon, although if so it is extensively diverged in these species in the majority of the copies. Indeed, the only species tested which had DNA closely homologous to IR2 present in many copies besides X.borealis was X.laevis. Failure to wash off the genomic probe at 75°C implies a sequence divergence between IR2 and the hybridising species in X.laevis of 5% or less, similar to that deduced for the homologous X.borealis family. Thus the IR2-like sequences in X.laevis appear to have diverged no more from Xbhl than those in X.borealis, and hence probably have diverged from the original X.borealis borealis family itself by 5% or less in the 8My since the species diverged.

10.3.5 Isolating lambda-XIR clones.

To further characterise the IR2 family of sequences, I re-screened the genomic gene library whose construction I described in section 6.3 with a probe specific for the IR2 region of Xbhl. Screening was essentially identical to that performed to isolate Xbhl in the first place in section 7.3.1: phage were plated at @2700/plate on 8 square plates, nitrocellulose filters prepared after Benton and Davis (1977) and hybridised in 2xSSC at 60°C to the Hpal.75kb fragment from Xbhl (see Fig 10.2) labelled by nick translation. There were a number of positive plaques on each plate: on the plate used to select phage for subsequent screens there were 45 very strong positives, 47 intermediate strength positives and 91 weak positives; the division between intermediate and weak positives is a fairly arbitrary one as a range of spot intensities was seen between these two. One filters' autoradiographs is shown in Fig 10.7.

I picked twenty plaques and re-screened them with the same probe.

Fig 10.7

1st screen for XIR clones



Fig 10.7 Filters were prepared from phage plates as described (Benton and Davis 1977), and probed with Xbhl HpaII 1.75kb fragment, nick translated, at 60°C, 2xSSC and washed in the same conditions. Autoradiograph of one filter. s= strongly hybridising plaque, i=intermediate, w=weak.

10. Repetitive DNA

I grew up ten potential positives from this second screen. Their EcoRI restriction patterns are shown in Fig 10.8 . They were designated lambda XIR (for Xenopus Intermediate repetitive DNA).

These clones hybridised to the Xbhl HpaII 1.75kb fragment to very varying degrees. When strongly hybridising clones were run on the same gel as weakly hybridising ones, Southern blotted and probed, the former bound nearly all the labelled probe available, and the latter appeared not to hybridise at all. It is established from solution studies that sequence differences between two DNA species severely reduces their rate of hybridisation (Wetmur and Davidson 1968, Bonner et al 1973), which is probably a contributing factor to this effect. If used in a hybridisation on its own, a diverged sequence will hybridise to the labelled probe. If used together with a sequence homologous to the probe, the probe will hybridise to the homologous sequence faster than to the non-homologous one, thus depleting the pool of probe molecules before significant probe has bound to the non-homologous DNA.

The situation in lambda 21a was also confused by the apparent presence of two inserts of considerable size: if we study the photograph of the phage preparation of the XIR clones (Fig 6.5) we can see that XIR21a is actually a double isolate, and so the two inserts represent two different recombinants. The same is true of XIR 18, although the second band is much fainter in this case.

XIR14, 21a(top) and 23 all have 8.5kb inserts, hybridise strongly to the probe and, on other restrictions digests (XbaI, BamHI - not shown) give restriction patterns identical to those expected of Xbhl. In other words, these clones seem to be isolates of Xbhl. They

Fig 10.8

XIR clones

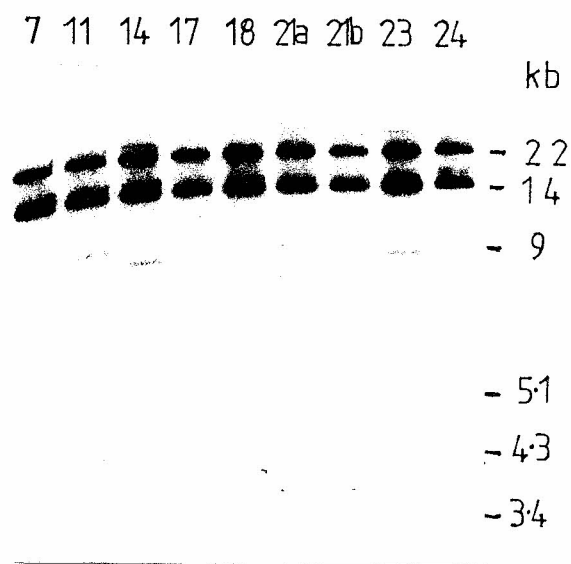


Fig 10.8 DNA from XIR clones (numbered as at the top of the track) was nick-translated, digested with EcoRI and electrophoresed on 0.7% agarose. Autoradiograph of gel.

10. Repetitive DNA

correspond to the strongly-hybridising species seen in the initial screen: their number, 47/2700, is about the same as I found in the earlier screen (section 7.3.4) confirming this supposition.

Discounting XIR 14,21a(top) and 23, we can see that XIR7,11,17,18 and maybe 21b all give different hybridisation patterns with the HpaII 1.75kb probe than does Xbhl (see Fig 10.9). I probed separate blots of XIR7, 11 and 18 with the HindIII 1.54kb fragment, showing that the homology was not due solely to homology with a sequence on the 3' end of the H2B gene

XIR7 and 18 also hybridise to pcX1H4W1, although not to pcX1H3W1, showing two clones to contain sequence homologous to an H4 gene. Thus not only are these clones homologous to a repetitive DNA present in the Xbhl histone gene cluster, at least 2 out of 5 are homologous to a histone gene as well. These data are presented in Fig 10.10 .

What can we deduce from this data? That, of the 5 positive XIR clones, two hybridise to an H4 cDNA is, to say the least, odd. Neither contain the right size EcoRI fragments to be representatives of the major cluster, and anyway they do not contain an H3-related sequence, while the H3 gene is adjacent to the H4 in the major cluster (Turner and Woodland 1982b). As there are only about 90 H4 genes in X.borealis including the major cluster (Turner and Woodland 1982b) it is very unlikely that these clones represent 'genuine' H4 genes. pcX1H4W1 does not hybridise to IR2 significantly when it is used to probe the whole of Xbhl (section 9.3.3.3). However Xbhl contains an H4 gene, which will probably bind all the probe in such a hybridisation, as the Xbhl clones did when they and XIR clones were present in the same hybridisation. Thus my failure to see a

Fig 10.9

Rescreen of XIR clones

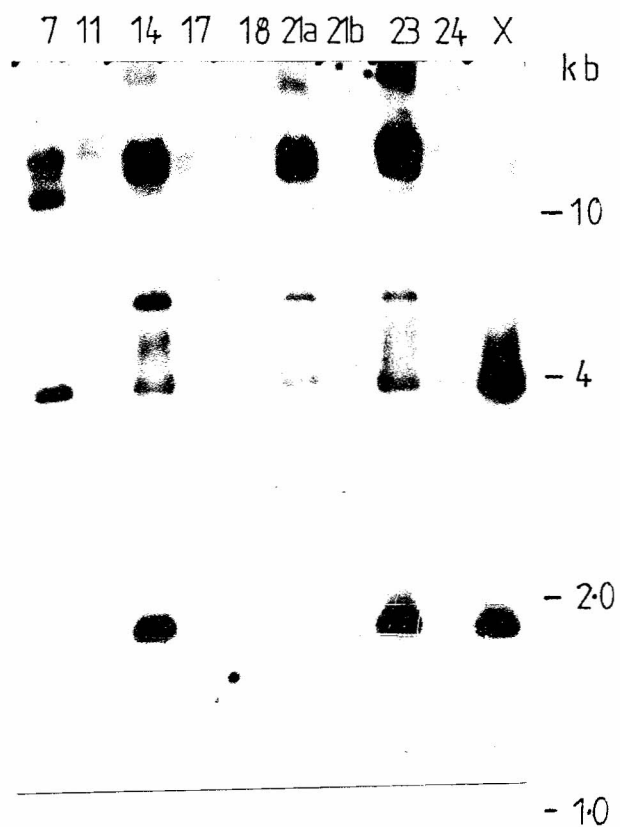


Fig 10.9 XIR recombinant DNAs were digested with XbaI, separated on 0.8% agarose gel, Southern blotted and probed with nick-translated Xbhl HpaII 1.75kb fragment. Hybridisation and wash were at 60°C in 2xSSC.

Fig 10.10

Homologies between XIR & histones

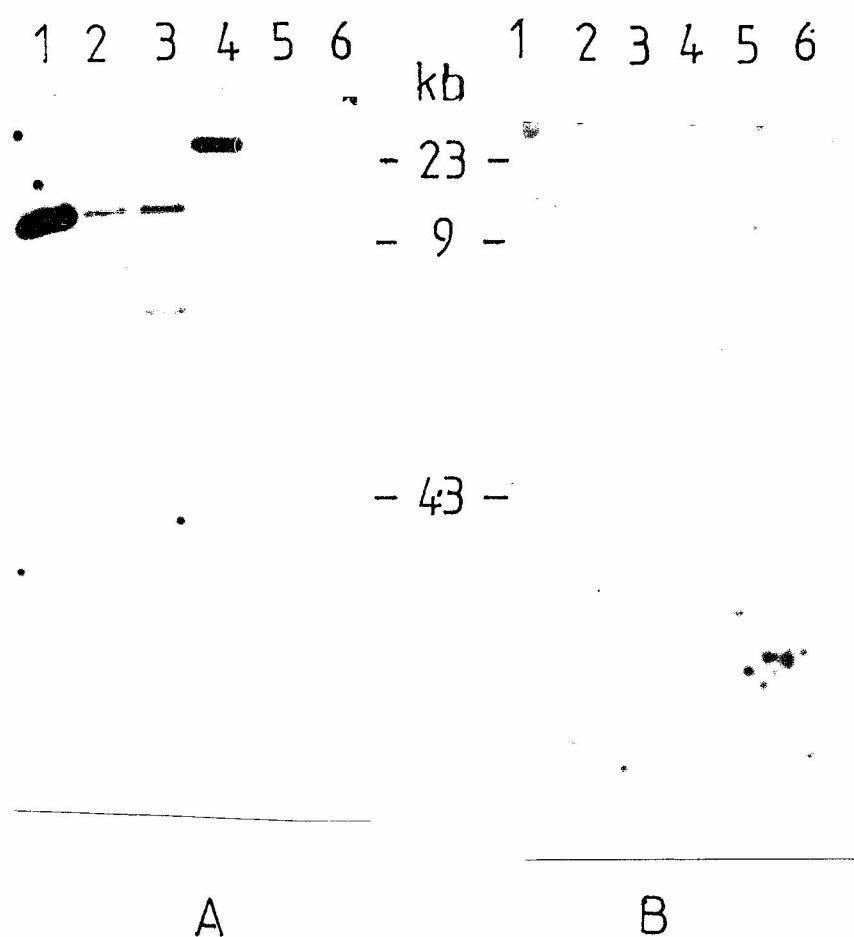


Fig 10.10 Southern blots of digests of XIR clones probed with A) Xbhl HindIII 1.54kb fragment, B) pcXlH4W1. Recombinants: 1 and 4=XIR7 2 and 5=XIR11, 3 and 6 =XIR18. Digests: 1,2 and 3 EcoRI, 4,5 and 6 BamHI.

trace hybridisation of pcXlH4W1 to IR2 in previous experiments need not eliminate the possibility that such hybridisation might occur under more favourable conditions. We might note here that these experiments do not show if the XIR clones are hybridising to the coding or the non-coding regions of pcXlH4W1 (they are not hybridising to the vector as pcXlH3W1, a different cDNA cloned in the same vector, fails to hybridise significantly to XIR clones).

Most of the XIR clones are unique isolates, (although XIR7 and XIR24 turned out to have identical EcoRI, BamHI and XbaI digestion patterns). Thus the library is not as excessively distorted with respect to XIR clones as it is with respect to histone gene clones. The distribution of hybridisation intensities seen in Fig 10.7 in the primary screen is also qualitatively like the distribution of homologies detected in the genome when it is probed with IR2 in the gene-number experiment above: a few closely-related (strongly hybridising) sequences being outnumbered 3:1 by more distantly related (weakly hybridising) species. If we assume the library is not distorted at all, an admittedly rash assumption, we can calculate that, if 138/2700 plaques contain IR2-related sequences and each phage contains on average 6.1kb of X.borealis DNA, then there are $\approx 12,000$ copies of those sequences scattered in the genome. If we postulate that the IR2 sequences are clustered, then this number must be multiplied by the mean number of IR2 elements within a 6kb cluster. The similarity of this former figure to the copy-number of IR2 we will derive below (section 17.1.5) suggests that IR2 is indeed scattered throughout the genome with little tendency to cluster.

10.4 Conclusion - the repetitive DNA families in Xbhl.

We can draw some conclusions about the repetitive DNA families

10. Repetitive DNA

cloned as part of Xbhl by drawing together what I have found in this section. I will not do more at this point in the thesis as I have other data which are relevant to IR2 which will be presented later.

There are two repetitive DNA regions in Xbhl, IR1 being present to about 500kb and IR2 to about 4Mb representation when hybridised at 60°C in 2xSSC. These numbers are quite sensitive to the conditions under which they are measured. IR1, lying between the H2A and H2B genes, is a short sequence related to a diverse family of genomic sequences quite highly diverged from it and, at least in some cases, from each other. Its proximity to the H2A gene is interesting: we know from functional studies that the H2A gene of Xbhl can produce a normal H2A protein (vide infra section 14.3) and so cannot be a pseudogene. Whether its promoter regions have been affected by IR1, whether they are some distance from IR1 or whether IR1 is actually intermingled with or part of the promoter region of the H2A gene we cannot say without precise mapping data to show where the H2A gene and IR1 are to within a few base pairs resolution. This makes the region upstream regions of the H2A gene a very interesting one to pursue with DNA sequence studies.

IR2 is a more complex structure, probably containing at least three elements as determined by gene number and divergence measurements. The presence of paired Aval and HincII restriction sites at its ends and the similarity of the melting profiles of those ends when hybridised to genomic DNA suggest that they may be related elements mutually inverted: further evidence below will bear on this point (section 16.3.2). Until we come to this, and thus decide how many elements IR2 really is, we cannot decide what its copy number is in the genome.

IR2 is probably dispersed throughout the genome, and appears to be associated with an H4 gene-like sequence in at least some cases.

The position of IR2 in Xbhl is interesting and, I believe, significant. As I noted at the end of section 9.4, the right-hand 1/3 of Xbhl is quite similar in structure to part of the major X.borealis gene cluster. The rest is dissimilar. The boundary between these two regions is marked by IR2. I would therefore like to introduce the model in which IR2, a dispersed repetitive DNA element, mediated a recombination event between the major histone gene cluster and some other histone-containing sequence (possibly one of the semi-minor clusters: the evidence on their restriction patterns is not plentiful enough to determine their restriction map), producing Xbhl, a minor type of histone gene cluster. Whether this was through illegitimate recombination, gene conversion or transposition is not indicated by this data, of course. It does tie in with the known genome-fluidizing action of intermediate repetitive DNA, however. This will be discussed further in section 13.

Further speculations may be educted from this data. However I will leave these to a later section when further relevant data has been presented.

11. Genomic distribution of Xbhl sequences.

They searched until darkness came on, but they found
Not a feather, nor button, nor mark
By which they might know that they stood on the ground
Where the Baker had met with the Snark.

11.1 introduction.

In this brief section I will examine where in the genome different sections of Xbhl have homologues by digesting genomic DNA with a variety of restriction enzymes, Southern blotting and probing them with labelled fragments of Xbhl. This technique has been extensively used on histone gene families (Crawford et al 1979, van Dongen et al 1981, Turner and Woodland 1982b) and other genes . In particular Turner and Woodland (1982b) have mapped the histone genes of the major cluster of X.borealis. The structure of this cluster is not the same as that of Xbhl. In this section I will not seek to repeat this approach, but will rather demonstrate that Xbhl is a very minor component of the X.borealis genome, and characterise the genomic distribution of the repetitive elements cloned in Xbhl.

11.2 Methods.

11.2.1 RPC-5 chromatography of DNA.

This is as described by Hardies and Wells (1976). 1.5mgs of X.borealis DNA was digested to completion with EcoRI. The digest was extracted once with phenol, and dialysed overnight at room

11.Genomic distribution

temperature against 1.25M Sodium acetate, 50mMTris pH7.5 . A 21ml column of RPC-5 was equilibrated with the same buffer, packed into a 15cm tube and the DNA loaded onto it. The DNA was then eluted by a 1.4 - 1.7M linear Sodium acetate gradient in 50mM Tris pH7.5: a total volume of 200 mls of sodium acetate was passed through the column under unaided gravitational flow at an average rate of 8.2mls/hour. The column was prepared, loaded and eluted at room temperature. 2ml fractions were collected, and the OD(260) of the eluant continuously monitored . At the end of the run the column was washed with 20mls of 2M Sodium acetate: this did not seem to elute any further DNA so it is unlikely that there was any DNA left bound to the column after elution with 1.7M Sodium acetate. The OD(260) profile of the eluant is shown in Fig 11.1

11.3 Results.

11.3.1 Genomic distribution of histone genes.

The genomic distribution of the majority of histone genes in X.borealis has already been described by Turner and Woodland (1982b). That this major cluster is not equivalent to Xbhl is shown in Fig 11.2 . Here several restriction digests of genomic DNA and an EcoRI digest of Xbhl have been Southern blotted and probed with size of the Xbhl. Not only is the Xbhl insert size different from that of the the EcoRI fragment which contains the major cluster, the other fragments are different from those generated from Xbhl, (remembering that we can only compare internal fragments, as the lambda vector sequences will not be present in the X.borealis genome.). Thus the genomic major histone-containing EcoRI fragment and the insert of Xbhl are not the same, as indeed we suspected they would not be from

Fig 11.1

RPC-5 column eluant

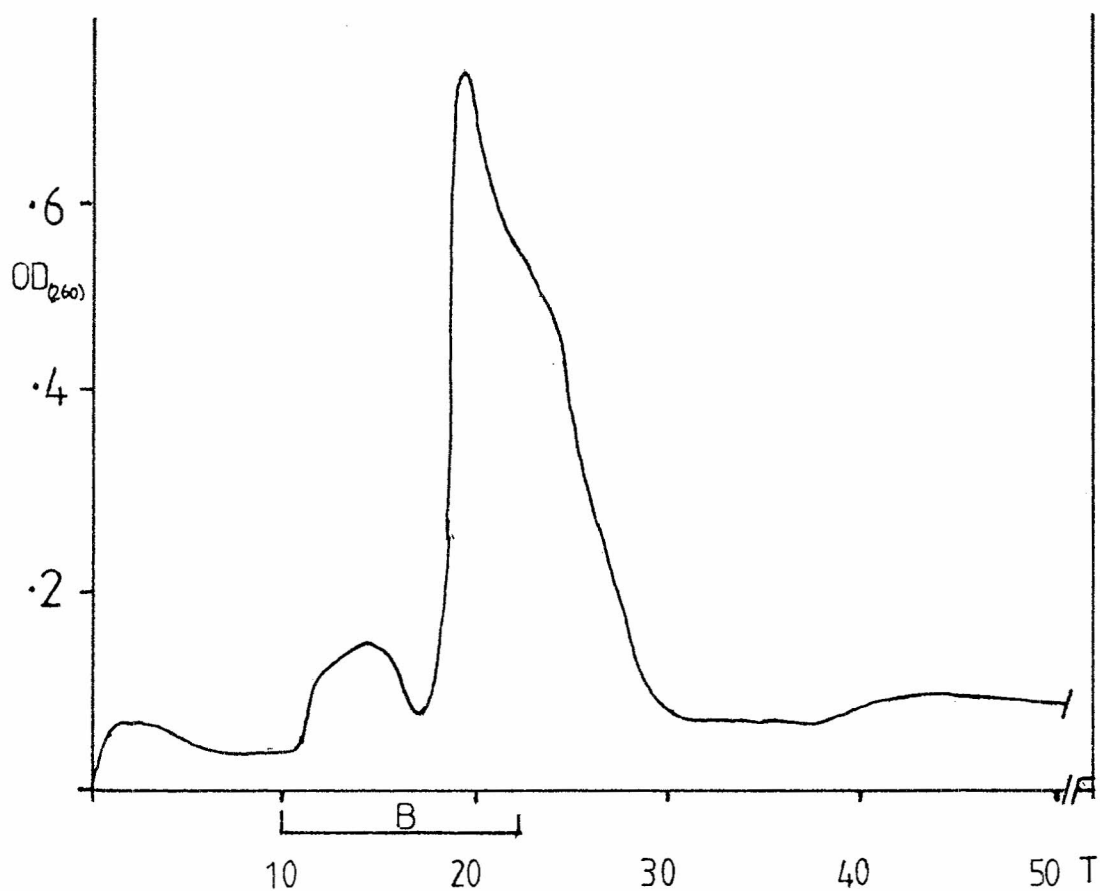


Fig 11.1 OD(260) profile of eluant from RPC-5 chromatographic column. Fractions 1-50 are shown, and the OD of the terminal fraction ('T', = fraction 92). 'B' is the region electrophoresed and probed in Fig 11.3

Fig 11.2

Genomic DNA probed with Xbh1

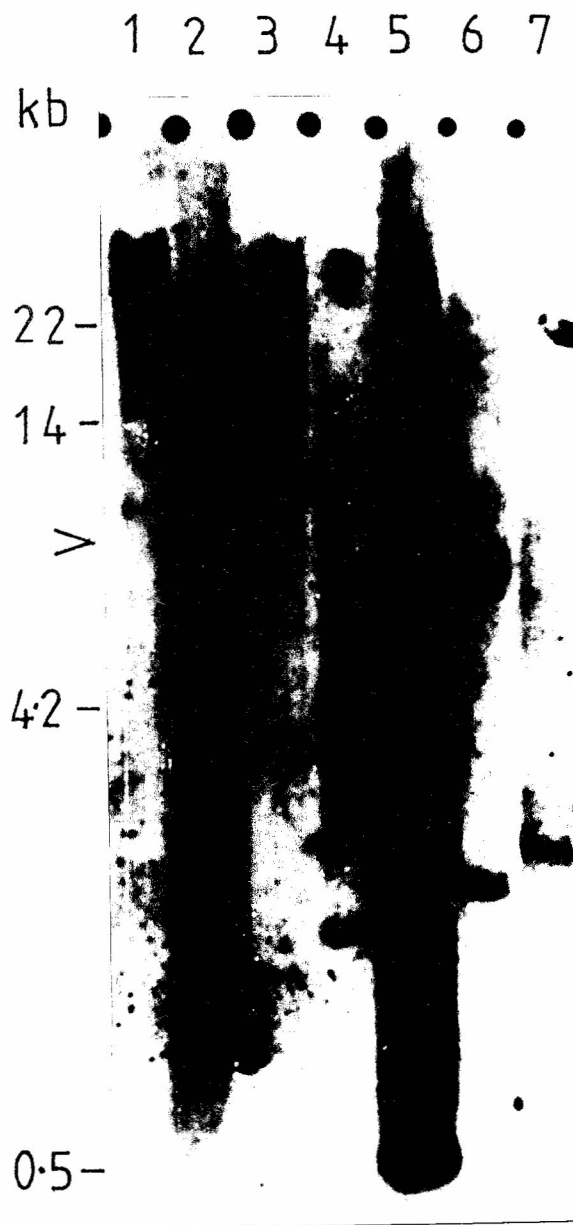


Fig 11.2 Genomic DNA from X.borealis was digested with 1) EcoRI (partial digest), 3) EcoRI (total digest), 4) HindIII, 6) BamHI, 7)PstI, and lug electrophoresed on 0.6% agarose. Track2: XbhI digested with EcoRI, track5 XbhI digested with HindIII (10ng each). The gel was Southern blotted and probed with whole nick-translated XbhI. 'V' marker= position of the XbhI 8.5kb insert.

earlier comparisons of restriction and gene maps.

I mentioned above (section 6.3.1) that I ran an RPC-5 column to separate the fragments of genomic DNA containing histone genes prior to cloning them. Although the cloning did not work the RPC-5 chromatography did. I ran some of the samples so produced on a gel, Southern blotted and probed with the H1, H2B and H4-containing sea urchin gene clone pSp102 (as pcX1H4W1 and pcX1H3W1 were not available at the time). The result is shown in Fig 11.3: this shows only some of the samples from the column - the others all gave only a faint background smear of hybridisation with no bands visible. pSp102, which contains the Sea urchin embryonic H4, H2A and H1 genes and spacer sequences between them, hybridises to the major cluster (©10kb), a semi-minor cluster (©4.5kb) and a minor cluster of ©8.4kb, approximately the same length as Xbhl (within measuring limits). Now, in the absence of other information it is impossible to say that the 8.4kb band is the sequence from which Xbhl was cloned, especially as pSp102 contains considerable A:T-rich spacer sequences and a homopolymer (T:C)₂₆ which could hybridise to homologous sequences in X.borealis which are quite unrelated to histones (Sures et al 1978, Callan and Old 1980). It is notable, however, that these three bands are the only ones seen in this experiment: the profusion of bands expected if pSp102 were cross-reacting extensively with the 'wrong' sequences such as satellite (Botchan and McKenna 1973) is not visible except as a background smear.

Note also that these bands all lie in the same fractions. The basis on which RPC-5 fractionates DNA fragments is not known. The size of small fragments is important (Hardies and Wells 1976), but becomes insignificant if the fragments are larger than ©1.5kb. The ends of the molecule - 5' or 3' protruding or blunt - can also

Fig 11.3

Histone genes in RPC 5 fractions

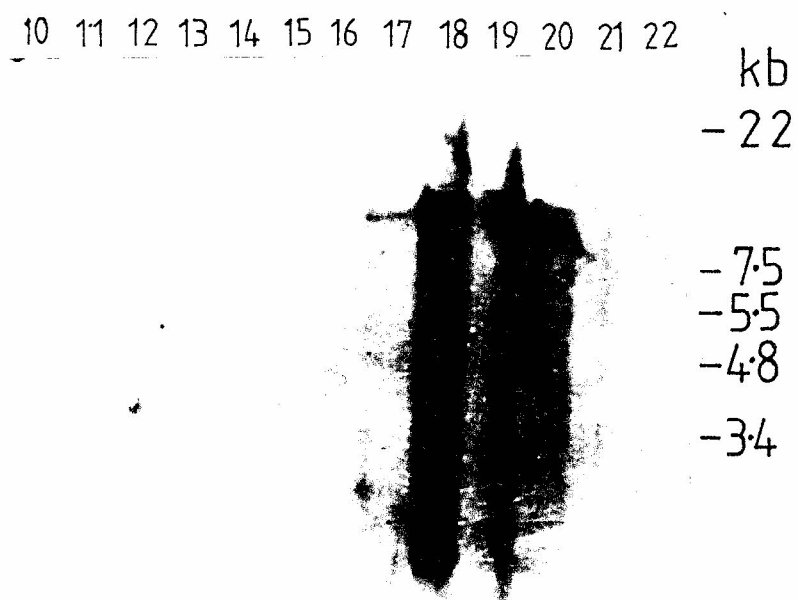


Fig 11.3 Samples of fractions 10 to 22 from the RPC-5 column eluant (see Fig 11.1) (containing 2ng in fractions 18-22) were separated on 0.7% agarose, Southern blotted, probed with pSp102 and washed at 55°C, 2xSSC

influence separation (Wells et al 1980). As the ends of the molecules in this experiment are all EcoRI ends and the molecules under discussion are over 2kb long, these effects will be negligible. However Wells et al (1980) also point to an effect of the A:T richness of a sequence - high A+T content or an A+T rich run in an otherwise G+C-rich sequence can cause the DNA to elute later than sequences without these features. Other factors are also involved, as these alone do not account for the observed elution pattern of DNAs whose sequence is known. Thus RPC-5 separates DNA fragments according to a complex, sequence-dependant criterion which is partially associated with A+T-richness. Although this does not tell us why any two fragments of DNA appear in the same RPC-5 fraction, it does tell us that they must share some features of sequence organisation. Thus it is likely that the pSpl02-hybridising sequences in RPC-5 fractions 18 and 19 are quite similar in overall sequence organisation. I would say from this that the 8.4kb band is therefore probably similar to the major histone gene cluster in X.borealis both in some of its DNA sequence (as it hybridises to pSpl02) and in its overall sequence organisation, and hence could well represent the original of the Xbhl sequence.

Why is an Xbhl-like sequence not detected in genomic DNA probed with Xbhl? Clearly the answer lies in the repetitive elements in Xbhl: these are hybridising to a background of heterogenous sequences contributing too great a background smear to the tracks to allow any but the most abundant histone genes to be detected. This is confirmed in Fig 11.4. Here I compare a genomic digest of X.borealis DNA with PstI probed with Xbhl to one probed with the 1.75kb HpaII fragment from Xbhl, which contains IR2 but no coding sequences. We can see some bands in the track probed with Xbhl (PstI cuts thrice in the major cluster), but none in the IR2-probed track, although the

Fig 11.4

IR2 in genomic DNA

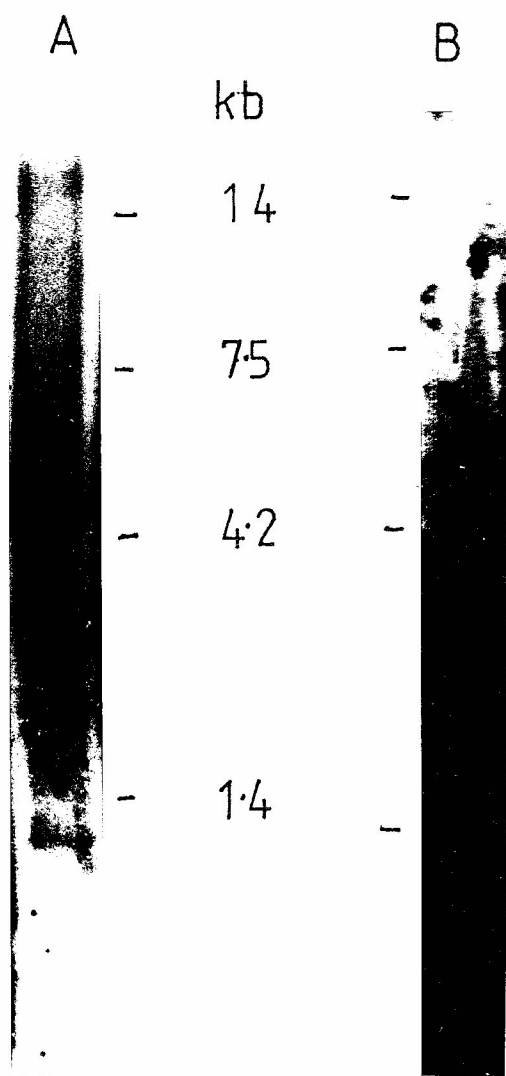


Fig 11.4 Pst I digests of genomic X.borealis DNA were electrophoresed on 0.8% agarose, Southern blotted and probed with A)Xbhl and B) Xbhl HpaII 1.75kb fragment, both hybridised and washed at 60°C, 2xSSC

background is the same in each.

11.3.2 Genomic distribution of IR2.

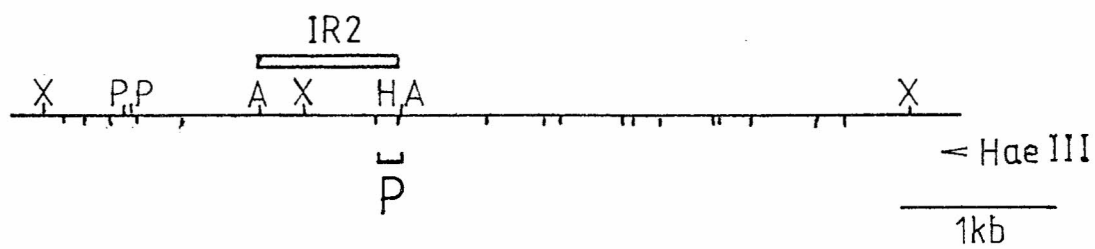
One implication of the results above is that IR2 is not present on a single fragment type in a PstI or EcoRI digest, but is dispersed among a number of fragments. This is confirmed in the Southern blot in Fig 11.5 . Digestion of the genomic DNA with a battery of enzymes gives no reproduceable bands except one at @200bp in the Sau3a track and an irregular series in the HincII digest. Both of these are very faint compared to the background smear. We may conclude, therefore, that the IR2 element(s) are dispersed throughout the genome and are not clustered either as a satellite-like repeat (in which case only one or two bands, or a regular ladder of bands, would be seen (Botchan, McKenna and Sharp 1973)), or in a few regions of the genome containing many copies of the sequence in a short total length of DNA, in which case a more complex pattern of discrete bands would be seen when 6-cutter digests were probed with IR2. This also supports the conclusion from the section cloning lambda XIR clones that the majority of 6kb regions of X.borealis DNA had only one IR2 element on them (section 10.3.5).

11.3.3 Extra-chromosomal distribution of IR2.

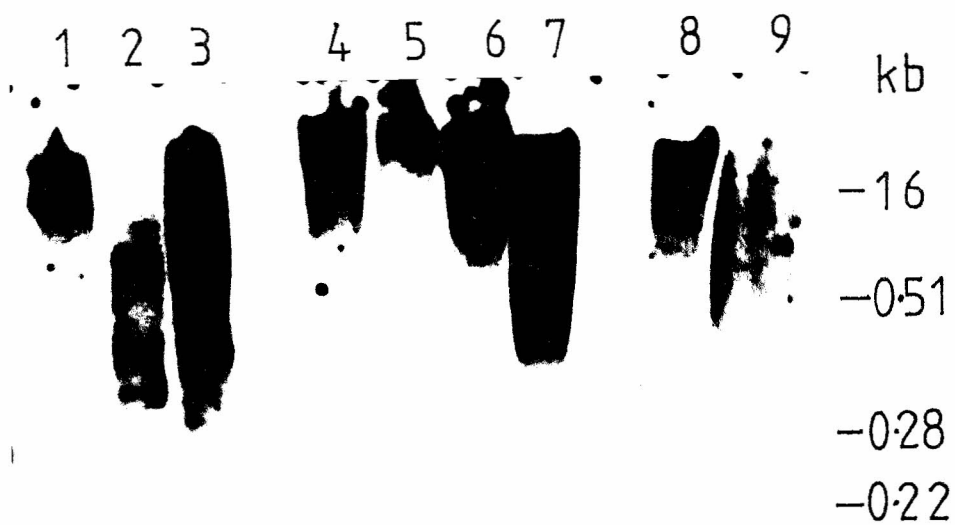
To see if I could detect IR2 on extra-chromosomal circles of DNA, I ran some total ovary RNA prep on an agarose DNA gel, Southern blotted it and probed with three probes covering IR2. The RNA prep had not been treated with DNase (vide supra section 9.2.1) and so if small circles of extrachromosomal DNA were present in the ovary they would also be present in this RNA preparation. The results were essentially the same for all three probes: HindIII 1.54kb, BamHI

Fig 11.5

Genomic distribution of IR2



A



B

Fig 11.5 A) HaeIII 250bp fragment used as a probe in part B.
X=XbaI, P=PstI, A=AvaI, H=HindIII

B) Digests of genomic DNA separated on 2% agarose, blotted and probed with the probe labelled 'P' in part A (HaeIII 250bp). Digests:
A)MspI, B) AluI, 3) HaeIII, 4) EcoRI , 5) HindIII, 6) RsaI, 7) HinfI, 8) HincII, 9) Sau3A.

4.2kb left terminal and the *Ava*I 1.01kb fragments. An illustrative result is shown in Fig 11.6: this blot has been probed with the *Ava*I 1.01kb fragment which includes nearly all the IR2 region but no genes outside it. No bands are visible, even though the smear of high molecular weight DNA is intensely labelled. I conclude that IR2 is not present on an abundant circular DNA.

11.4 Conclusions.

We now have a fairly clear picture of what the elements of *Xbh*I are and where they come from in the frog. The whole *Eco*RI fragment is present in *X.borealis* in only 1-3 copies per haploid genome (Dr.P.C.Turner estimates this from his studies). Its gene order and positions are related to part of the major cluster type of *X.borealis*, but the two clusters are different. Contained within *Xbh*I are two repetitive elements. One, IR2, is dispersed throughout the genome in a total representation of ≈ 4 Mb, If it is present in the major histone gene cluster its presence there would contribute only 1% to the total genomic representation of this element, so it is unlikely that it would be seen on Southern blots of genomic DNA probed with IR2. Conversely the fact that we do not see a 10kb band lit up by IR2 in an *Eco*RI digest of *X.borealis* genomic DNA does not mean that we can say whether the major cluster has a representative of IR2 in it or not. IR2 is not present on an abundant extrachromosomal element in ovary. I cannot say if it is present on one in other tissues, or if it is present on a non-abundant circle. As the levels of circular DNA discussed in section 2.7.4 above are in the range 0.1 - 10 molecules per cell, this preliminary search in *Xenopus* cannot be regarded as more than a pointer to further studies.

Fig 11.6

IR2 on circular DNA

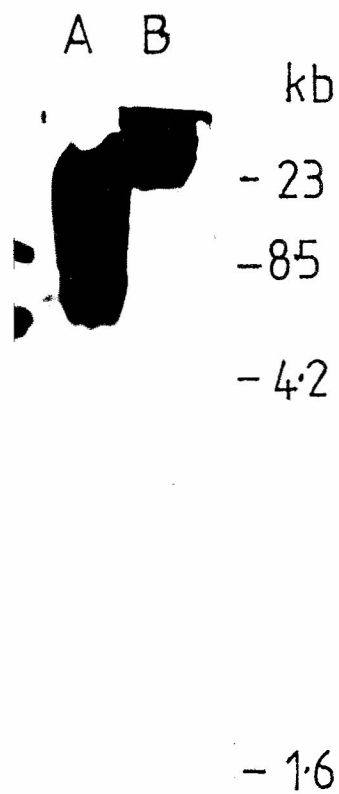


Fig 11.6 A) 50µg of 'total ovary RNA' and b) 5µg of genomic DNA were run on a 0.7% agarose gel, blotted and probed with Xbhl AvaI 1.02kb fragment containing IR2. Hybridisation and wash at 60°C, 2xSSC

12. In vivo transcription of IR2.

12.1 Introduction.

Two experiments relate to the expression of the repetitive element IR2 in vivo. These both look at RNA: the former examines the proteins made by RNA which hybridises to IR2 (Hybrid Release Translation, HRT), the second examines directly the amount of RNA complementary to IR2. The background and results for the HRT experiment have already been presented (Section 9 above). The second approach uses a variation of the gene-mapping techniques used several times before in this study. In this experiment several restriction digests of Xbhl are Southern blotted and probed with cDNA synthesised from radioactive precursors on a poly-A⁺ RNA template. Any sequences complementary to poly-A⁺ RNA will be lit up by this procedure.

12.2 Methods.

12.2.1 Synthesis of cDNA probe.

cDNA was synthesised according to Buell et al (1978). 10x Buell buffer is 500mM Tris pH8.3, 100mM MgCl₂, 300mM 2-mercaptoethanol, 1.4MKCl. In each labelling reaction 3ul of labeled dGTP (alpha-p(32), 10mCi/ml in water) were added to 1ul of poly-A⁺ RNA, 2.5ul of

12. IR2 transcription

10xBuell buffer, 1.2 μ l each of dATP, dCTP and dTTP (1mM each), 1 μ l of BSA (2.5mg/ml), 10 μ l of water, 2 μ l oligo-dT primer (1mg/ml) and 2 μ l of reverse transcriptase. The reaction was assembled on ice, then incubated for 1 hour at room temperature. The product was subsequently purified on a Sephadex column as for a nick-translation.

12.3 Results.

12.3.1 Hybridisation of Xbhl to cDNA probes.

The Poly-A⁺ RNA used in this experiment was from X.laevis, and was the gift of Dr. Elizabeth Sturgess and Ms. Elizabeth Balantine. PolyA⁺ RNA from Testis and Neurula was used: these two RNAs gave essentially the same results. The result for Testis RNA is shown in Fig 12.1. There is an horrendous background of blackening on this autoradiograph, but it is clear that there is little hybridisation to Xbhl sequences. By contrast pTCA19.7 does light up strongly with this probe. pTCA19.7 is a subclone of X1h19, a X.laevis genomic histone gene cluster clone known to contain a repetitive element which hybridises strongly to poly-A⁺ cDNA (Old per com), and so it acts as a positive control in this experiment. It is not homologous to IR2 (Tom Aldridge per com). Thus the failure of Xbhl to show any hybridisation in this experiment can be taken as good evidence that no sequences in Xbhl (including IR2) are present on an abundant poly-adenylated mRNA in testis or neurula.

12.3.2 Hybrid Release Translation from IR2.

The data for this have been presented in section 9.3.3.1 (vide supra). We noted there that the central region of the clone behaved

Fig 12.1

Expression of IR2

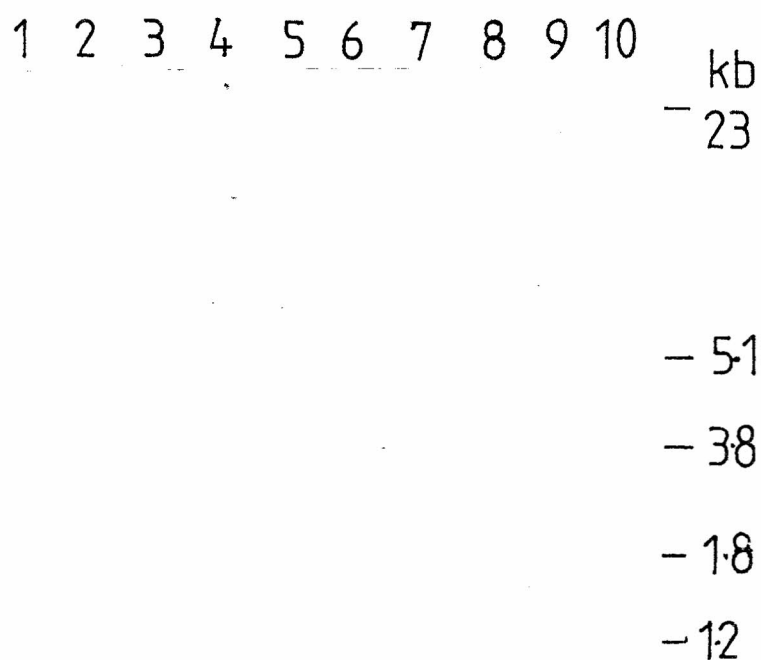


Fig 12.1 Digests of pWB2 (1-8) and pTCA19.7 (9 and 10) were separated on 0.8% agarose gel, Southern blotted and probed with labelled cDNA prepared from polyA+ testis RNA (see text). Tracks 1-8 digests of pWB2 1) EcoRI, 2-4) Isolated XbaI 3.98kb fragment, 2) HindIII, 3) HincII, 4) BamHI, 5)AvaI, 6) XbaI, 7)BamHI, 8) PstI. 9) HindIII digest of pTCA19.7, 10) undigested pTCA19.7

anomalously in the HRT assay, hybridising to the ovarian mRNAs for all four core histones. We can now confirm that the restriction fragment which gave these results, the HindIII 1.54 fragment, does not contain any of the core histone genes as determined by other criteria. Thus it is reasonable to propose that IR2 sequences at the end of this fragment are actually hybridising to the mRNAs. If this is so, then either the H2B, H3 and H4 mRNAs must be hybridising to the HindIII 1.54 fragment in the same DNA strand as they hybridise to in their respective genes in the strand-separation + HRT experiment, as only one DNA strand can be seen to hybrid release these three histones, or they must be hybridising much less efficiently to IR2 than to their respective genes. The HindIII 1.54kb fragment only hybrid releases an H2A message weakly.

That a repetitive element hybrid releases histone mRNA might be a peculiar feature of the oocyte mRNA used in this experiment. In Notophthalmus satellite DNA adjacent to the oocyte histone gene clusters is transcribed during the lampbrush stage of oogenesis due to readthrough from the histone genes themselves: this presumably does not occur in other tissues (Dia: et al 1981). An analogous occurrence in Xenopus oocytes may be the reason that IR2 is expressed on Xenopus histone message. However we must note that IR2 is quite unlike the transcribed satellite of Notophthalmus histone gene clusters (Dia: et al 1981) in its genomic distribution, both with respect to the histone gene cluster it is in and its repetition in the genome in general, and in its internal organisation. Even if the histone transcripts of oocyte do contain IR2-homologous sequences due to oocyte-specific failure of termination, this sequence must be specific to histone messages out of the abundant messages stored in the oocyte, as only histone mRNAs are hybrid released by the HindIII 1.54kb fragment.

12.4 Conclusion.

IR2 is not extensively transcribed, but has homologues on the histone mRNAs of oocyte. Repetitive DNA adjacent to genes coding for proteins is known to be transcribed in several other cases (reviewed by Davidson and Posakony 1982) where transcription of these sequences is not due to the unusual transcription pattern of the oocyte lampbrush chromosomes.

We have seen elsewhere (section 10.3.5) that 2 out of 5 XIR clones bearing homologues of IR2 hybridise to the full-length H4 cDNA clone pcX1H4W1, although they do not hybridise to the 3'-end H3 cDNA clone pcX1H3W1 or to an H1 gene clone containing coding sequence alone. These hybridisations are only apparent under favourable hybridisation conditions (low stringency and absence of other complementary sequences to compete for the labelled probe). If the IR2 repetitive element does hybridise inefficiently to non-coding regions of histone genes, it would explain how a significant number of the 10,000 IR2-related sequences (vide infra) can be related to H4 genes and can hybrid release H3, H4, H2B and possible H2A, while the total number of H4 genes in Xenopus is only ©90 (Turner and Woodland 1982b).

13. 'In his Image: ...' : Xbhl and Xlh8.

"Its a Snark" were the words that first came to their ears
And seemed almost too good to be true.
Then followed a torrent of lughter and cheers,
Then the ominous words "It's a Boo..."

13.1 Introduction.

In the introduction I noted briefly that the X.laevis histone gene cluster Xlh8 isolated by Dr. R.W.Old at Warwick was remarkably similar to the Xbhl described in this work. Throughout the course of this work doubts have been expressed that Xlh8 and Xbhl were actually different. It is clearly essential to the interpretation of this work that I establish that Xbhl is different from any X.laevis histone gene clone which might have contaminated my gene library at the start . The reasons for my conclusion that Xbhl is a clone derived from a X.borealis are summarised below. In this section I will show by means of restriction data that Xbhl and Xlh8 (and two other X.laevis histone gene clones very similar to it) are extremely similar but not identical.

13.2 Methods.

All methods used in this section have been described previously. The X.laevis histone gene cluster clones lambda- Xlh2, Xlh8 and Xlh202 are lambda-gtWES recombinants isolated from a X.laevis EcoRI genomic gene library by screening with the H4 cDNA probe pcXlH4W1 in the same fashion as Xbhl was isolated. They all contain 8.5kb EcoRI inserts with a similar restriction map to Xbhl, and all contain all

identical pair of clones will be referred to as Xlh202. They all contain 8.5kb EcoRI inserts with a similar restriction map to Xbh1, and all contain all five histone genes.

13.3 Results.

13.3.1 Xbh1 and Xlh8 are very similar.

Fig 13.1 shows restriction digests of the EcoRI inserts of Xbh1 and Xlh8. In all cases the majority of the fragments are of the same size, suggesting that the restriction map of these two clones would be very similar. The gene map of Xlh8 (deduced from Southern blot experiments similar to those presented above for Xbh1) is also very similar to that of Xbh1 (Tom Aldridge per com). Xbh1 and Xlh8 are therefore very similar clones.

They are not identical, however, as is shown by the differences between their restriction patterns. The smallest HindIII fragment is smaller in Xlh8 than in Xbh1, and the 2nd largest Sst fragment is significantly larger. The SstI difference is consistent with the absence in Xlh8 of the left-most SstI site in Xbh1. The HindIII difference is not so easily explained. The smallest HindIII fragments differ by 140bp, a difference also seen between the two largest HpaII fragments. The simplest explanation for this and several other differences is the presence of a small deletion in Xlh8 relative to Xbh1. The limits of this deletion may be mapped on Xbh1 by comparison of the 4-cutter restriction patterns of the two clones: this and other differences are mapped in Fig 13.2. These are only the restriction differences I can accurately assign, and do not account for all the observed band differences between Xbh1 and Xlh8 in HpaII

Fig 13.1

Restriction of Xbh1 & Xlh8

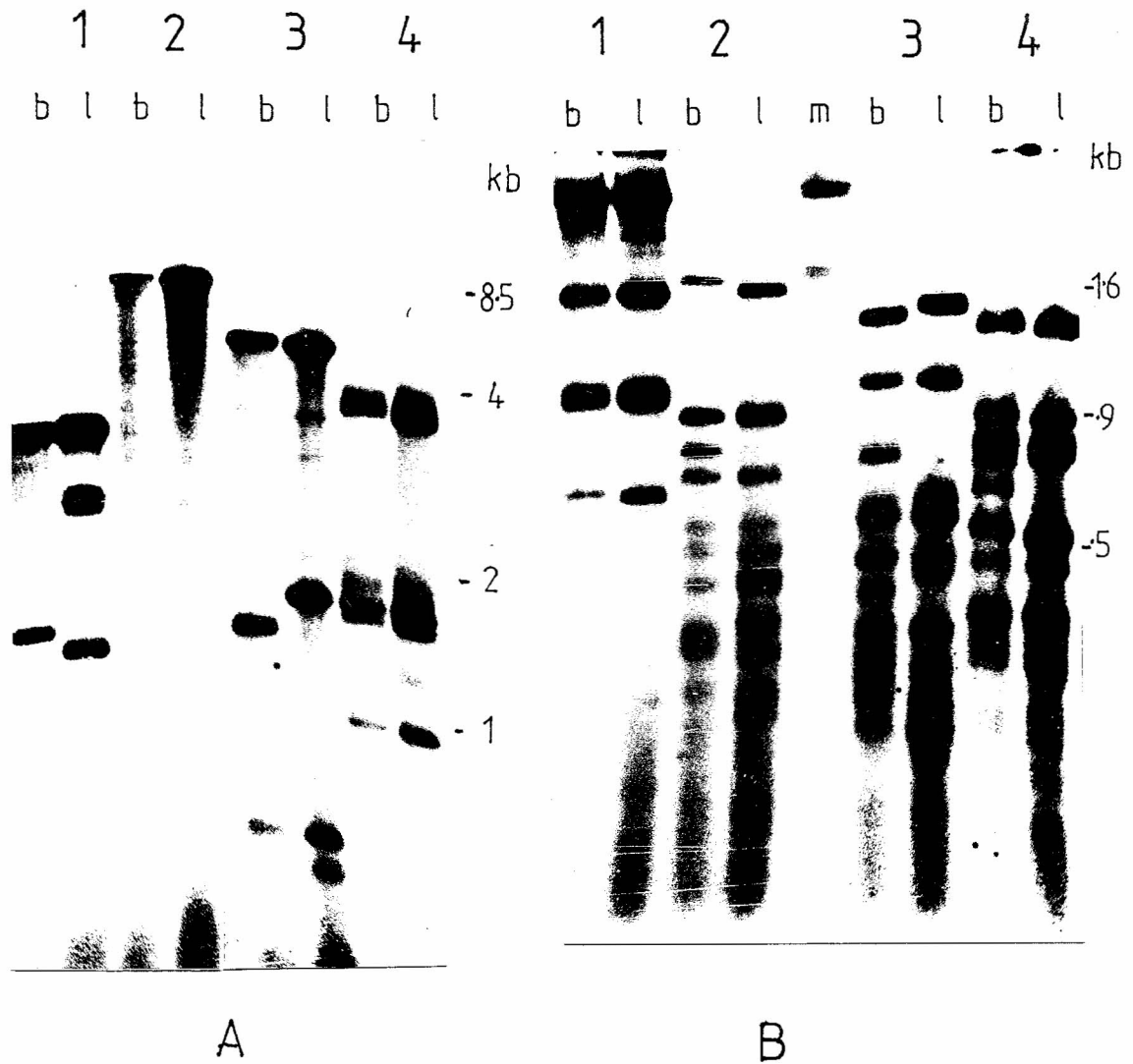
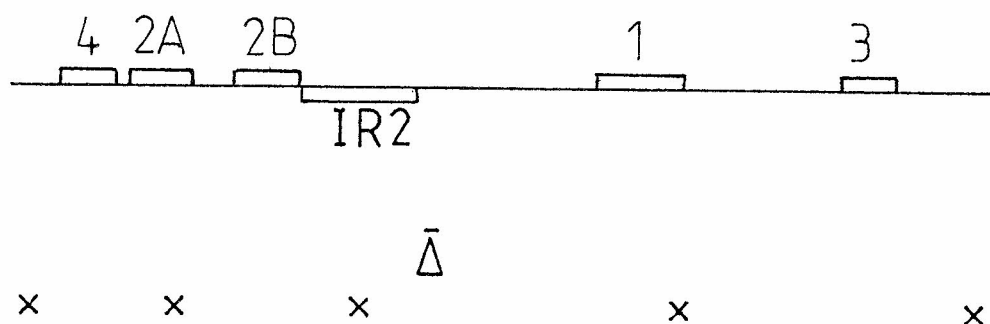


Fig 13.2 A) Autoradiograph of 0.7% agarose gel of digests with 1) HindIII , 2) EcoRI, 3) SstI, 4)XbaI. B) 2.5% agarose gel of digests with 1) BamHI, 2)HpaII, 3)HaeIII, 4) HinfI. DNA samples: isolated, nick translated inserts of 'l' Xlh8, 'b' Xbh1

Fig 13.2

Differences between Xbh1 & Xlh8



Δ = deletion

Fig 13.2 Map of the differences between Xlh8 and Xbh1 inserts which are locatable to specific loci. X=point differences, 'deletion' = extent and site of the region of Xbh1 proposed to be absent from Xlh8

or HaeIII digests.

13.3.2 Xbhl is similar to but different from Xlh2 and Xlh202.

Two other clones like Xlh8 have been isolated along with it: these were designated Xlh2 and Xlh202. They are extremely like Xlh8 and Xbhl as judged by restriction data: no gene mapping has been performed on these clones. They are, however, different from each other and from the latter two clones, as is illustrated in Fig 13.3 . The differences are small - only one or two site differences are seen with any given enzyme.

The BamHI digests in Fig 13.3, and other data (Tom Aldridge per com) suggest that the insert is oriented in the vector in the opposite orientation to Xbhl in Xlh8 and Xlh202. Thus these clones cannot be the source of any contamination. That the inserts of all three clones are different from each other is shown best by the HinfI digest. Note here that one band, which is shared between Xlh2 and Xlh202, is not shared by Xbhl. That this band is shared by the former two clones shows that it is not at the junction of the insert and the vector DNA, as these two inserts have the opposite orientation. Thus the fact that Xbhl does not share this band shows that at least one internal HinfI site is different between Xlh202 and Xbhl. The results shown in Figs 13.1 and 13.3 have all been repeated by the author or by other members of the laboratory.

Before deducing that the X.borealis gene library used in these experiments was not contaminated by Xlh2, Xlh8 or Xlh202 we can note the following. Firstly, the genomic DNA from which the library was cloned does not contain an excess of Xbhl-like sequences as judged by the pattern of EcoRI bands to which Xbhl hybridises in it (section

Fig 13.3

Restriction of Xbh1-like clones

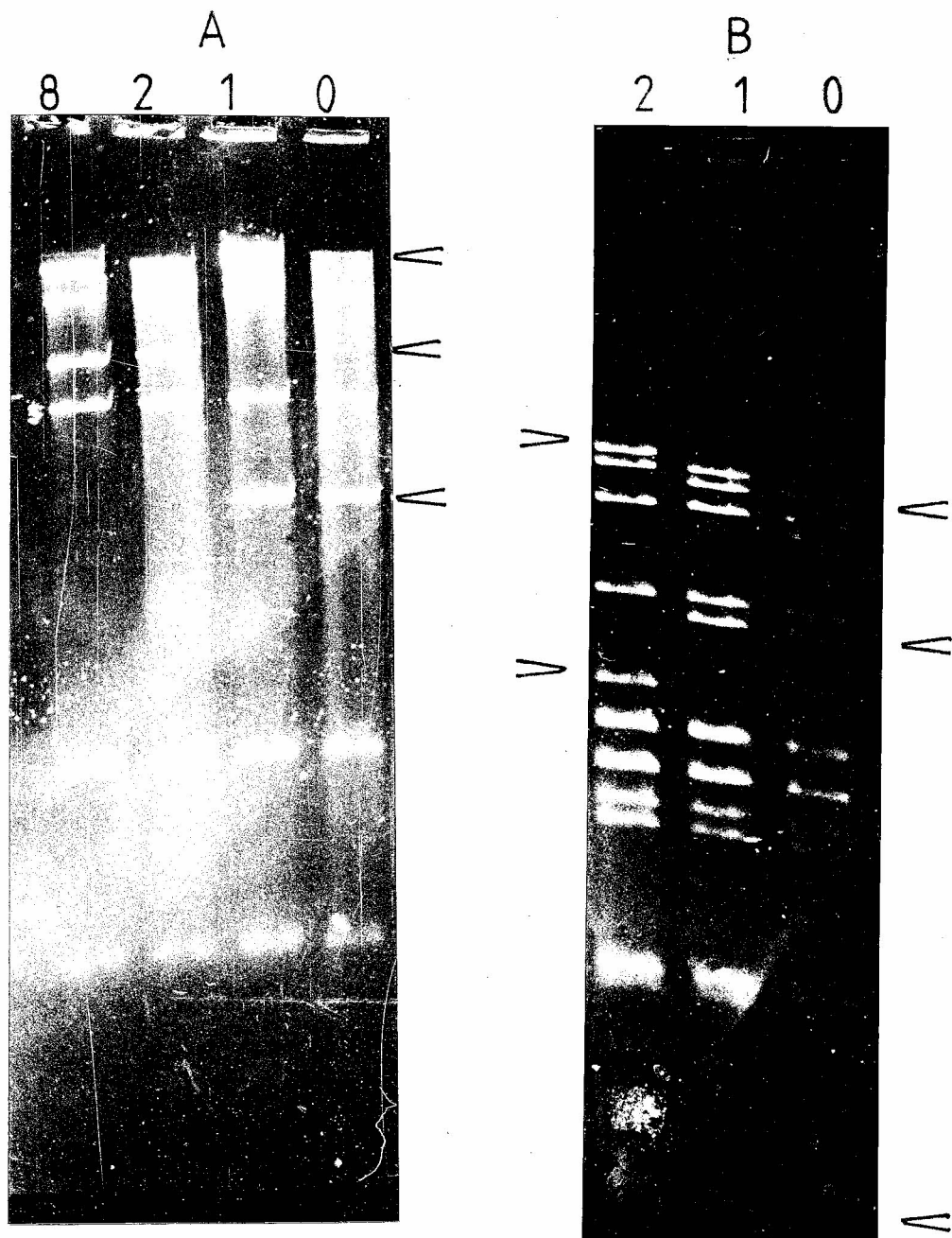


Fig 13.3 cont.

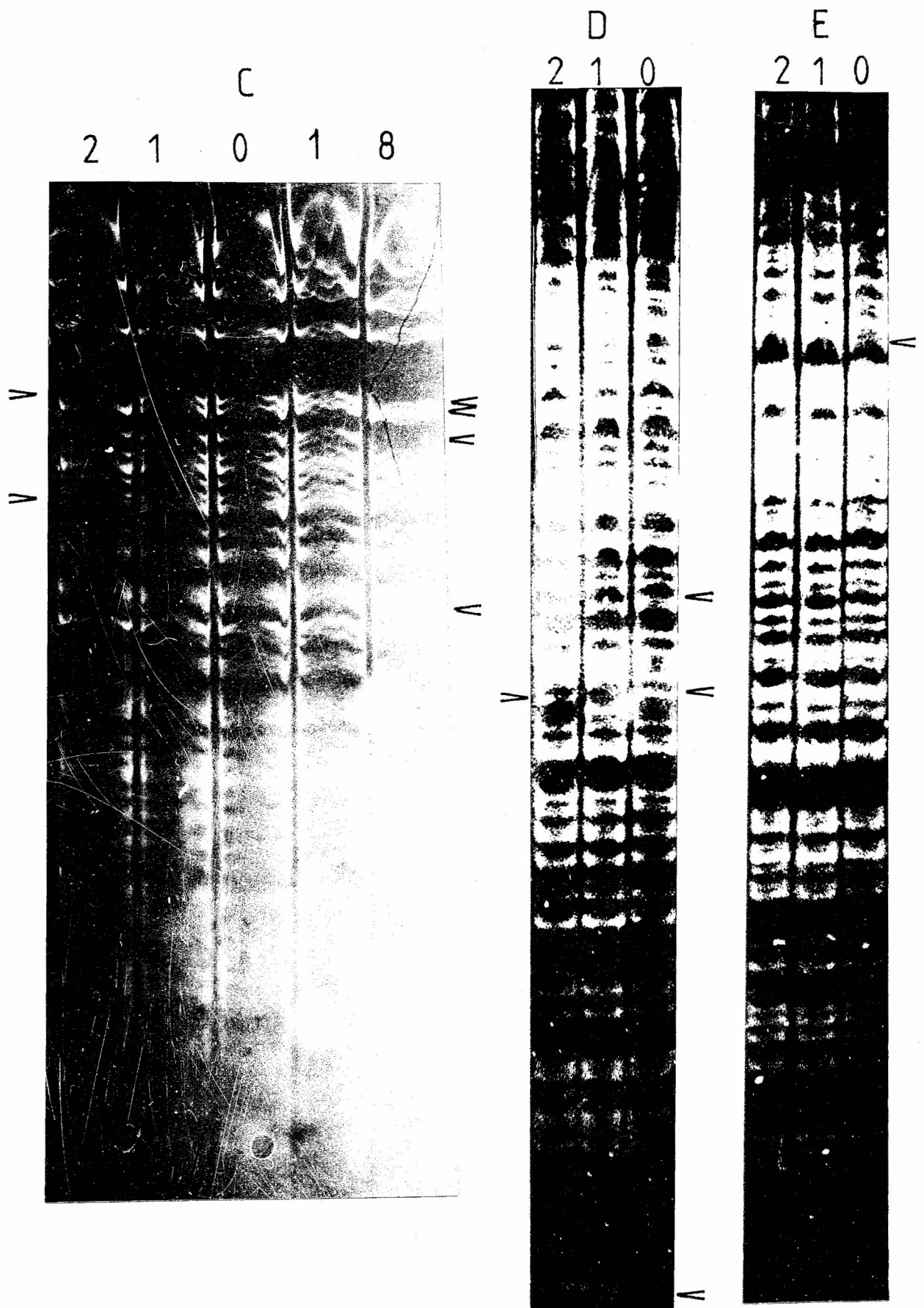


Fig 13.3. Restriction patterns of the Xbh1-like clones. The lambda recombinants Xbh1 ('1'), Xlh2 ('2'), Xlh202 ('0') and Xlh8 ('8') were digested with A) BamHI, B) HincII, C)HpaII, D) HinfI, E) HaeIII. The digests were separated on 1% agarose (A and B) or 7% polyacrylamide (C, D and E) gels and stained with ethidium bromide. Potential band differences between adjacent tracks are arrowed. Material for some of these digests was provided by Mr. Tom Aldridge.

11). Thus the contamination must have entered the library between the isolation of the DNA and the screening stage. During this phase of the work Xlh8 was present in the laboratory as DNA, but DNA had not been isolated from Xlh2 or Xlh202, which existed only as phage isolates (R.W.Old per com). Thus, if contamination occurred, either it was by Xlh8 DNA (in which case the orientation of the insert could have been altered by EcoRI restriction and religation) or by phage of any of the three phage strains (in which case the orientation of Xbh1 would have to be the same as that of the contaminating phage). To disprove contamination, then, we must prove that 1) all three possible contaminants are different from Xbh1 in some feature (orientation of insert, phage DNA restriction pattern or insert DNA restriction pattern), AND 2) that the insert of Xlh8 is different from that of Xbh1. Figs 13.1 and 13.3 demonstrate both of these points. Thus we may deduce that the X.borealis gene library described in section 6 is unlikely to have been contaminated with Xlh2, Xlh8 or Xlh202.

We can deduce the degree of sequence divergence between these clones from the number of bands we observe not shared between them in restriction digests by the same enzyme. (the logic of this derivation is presented in Appendix 4). Fig 13.4 shows these differences deduced from the results in Fig 13.3. The figures are not exact, as the number of band differences seen between Xlh2, Xlh202 and Xbh1 is only 0-3 per 4-cutter.

13.4 Conclusions.

I have shown above (Section 11.3.1) that an Xbh1-size EcoRI fragment exists in the genome of X.borealis which hybridises to a histone gene cluster. Here I show that Xbh1 is similar to a family of

Fig 13.4

Sequence differences between clones

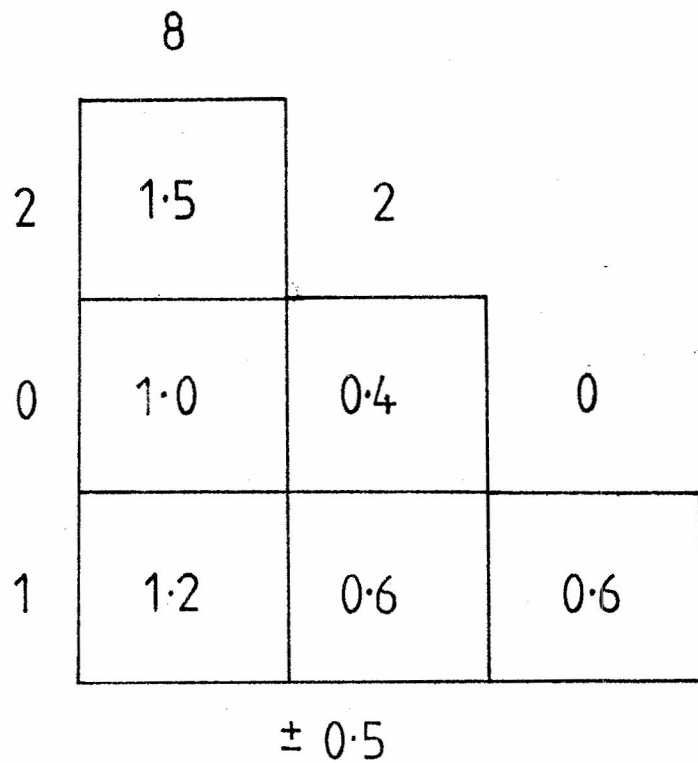


Fig 13.4. Percentage sequence difference between Xlh8 ('8'), Xlh2 ('2') Xlh202 ('0') and Xbh1 ('1') deduced from restriction pattern differences shown in Figs 13.1 and 13.3 and in other data (not shown) according to the method described in Appendix 4. Error = standard error of the mean.

histone gene clones from X.laevis. A sequence divergence of 1.2% between Xlh8 and Xbh1 is sufficient to account for all the 4-cutter restriction pattern differences seen. Dr. Philip Turner has found a similar level of divergence between the H1 gene and boardering sequences by direct sequencing of Xbh1 and Xlh8 (vide infra section 15.1) supporting this indirect measurement. I will leave discussion of the significance of this for the evolution of Xbh1 and related sequences to the Discussion section below (section 17.2.5), as these are not straightforward.

This concludes the results which my experiments have produced. The following results have been obtained in collaborative experiments with other workers at Warwick and elsewhere: I have had personal experience of all the techniques mentioned in the following sections.

14. Functional studies on Xbhl.

A collaborative study with Professor H.R.Woodland.

14.1 Introduction.

I mentioned in the Introduction that it was the original aim of this project to study the control of transcription of histone genes by micro-injecting cloned genes into Xenopus oocytes before and after alteration of their sequence in vitro. This aim is of course a long-term one, but this project cannot report anything except some preliminary data from these expression experiments. In these preliminary experiments we injected Xbhl, both in lambda-WES and pBR325 recombinants, into oocytes and labelled them with 35-S-Methionine or tritiated Lysine. The histones were extracted and analysed on PAGE. The object was to see which of the genes on Xbhl can be transcribed in oocytes, and to what degree. Some interesting results emerged from this, some of which have direct relevance to this thesis.

14.2 Methods.

14.2.1 Micro-injection of DNA into Xenopus oocytes.

DNA was made 150-250µg/ml in 80mM NaCl for injection. Micro-injection was carried out as previously described (Mert : and Gurdon 1977, de Robertis and Mert : 1977). Ovary was dissected out of Xenopus laevis or borealis females and suspended in Barth-X medium. Oocytes were stored and incubated at 22°C. Individual oocytes

14. Microinjection

were teased out of the ovary , and oocytes with uniformly pigmented animal poles and cream vegetal poles selected for micro-injection. The micro-injection needle was inserted into the animal pole (where it is most likely to encounter the Germinal Vesicle) and @20nl of DNA was injected. 20 - 40 oocytes were injected with each sample. (See Fig 14.1). The injected oocytes were incubated in Barth-X overnight and the dead ones removed, and then labelled in tritiated lysine or S(35)-labelled Methionine at @1mCi/ml in Barth-X overnight, in a typical volume of 20 oocytes in one sample of 70ul of labelled material.

14.2.2 Analysis of histones from Oocytes.

Germinal vesicles were dissected out of injected oocytes manually, suspended in 10ul of 5mg/ml Calf thymus histones in water , mixed with 10 - 20 unlabelled oocytes to provide carrier proteins and frozen at -70°C for future use. The germinal vesicle sample was homogenised in 1 ml of Guanidinium Chloride (10%) in 40% ethanol on ice by disrupting them in a 10ml glass homogeniser and then leaving the homogenate to stand for 1 hour on ice with occasional strokes of the homogeniser. The homogenate was centrifuged in an Eppendorf bench-top centrifuge, the supernatant taken and 12 mls of ethanol added. This was left overnight at -20°C to precipitate the histones before centrifugation at 2900 rpm, 10°C. The supernatant was removed very carefully, the pellet washed in 5mls of 90% ethanol, re-centrifuged and dried in vacuo. The almost invisible pellet was taken up in water for electrophoresis.

14.3 Results of microinjection experiments.

The results of one microinjection experiment is illustrated in Fig

Fig 14.1

Microinjecting oocytes

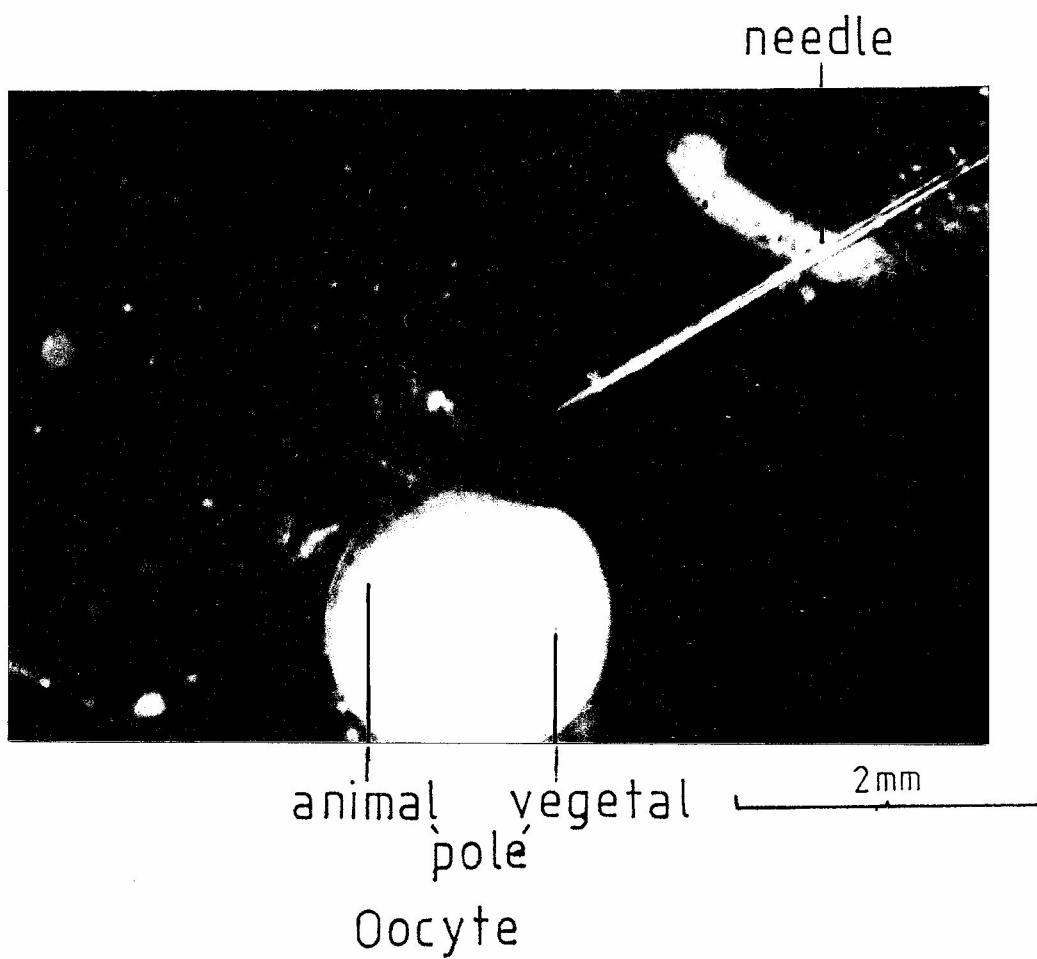


Fig 14.1 Needle for microinjecting oocytes, and an oocyte. Both are in Barth-X: note the needle is full of air.

14. Microinjection

14.2 . H1, H3 and H4 are expressed in both plasmid forms of Xbhl, although the latter is not strongly expressed, and is not evident at all when the original lambda recombinant is injected. The orientation of the EcoRI insert in the plasmid pBR325 makes no difference to these genes expression, which therefore suggests that this expression is relying on a promoter within the insert and not on plasmid sequences to synthesise the mRNA.

The mobility of the H1 protein is significantly faster on SDS-PAGE than that of the X.laevis erythrocyte H1. This is also true of the X.borealis oocyte and erythrocyte H1 proteins (Woodland per com): it is not possible at this stage to say which of the two X.borealis erythrocyte H1 proteins the Xbhl H1 is more like. It is not formally possible, of course, to say that it is an H1, although its H1-like mobility on both SDS-PAGE and Acid-urea PAGE makes it probable that this band does represent a genuine H1 protein product.

H2B is expressed weakly if at all. This is interesting, opening up the possibility that this H2B gene is a pseudogene. Remember that the restriction map of the H2B in Xbhl was quite different from that of the major cluster (Vide supra Section 9.4), testifying to considerable DNA-sequence divergence. However in the absence of DNA sequence studies this must remain speculative.

H2A presents a more interesting picture. pWB1 synthesises an H2A at low levels which migrates more slowly than the X.laevis erythrocyte H2A on SDS-PAGE and contains Methionine , which is unusual for H2As (Isenberg 1979). pWB2 synthesises much larger amounts of H2A as judged by Lysine-labelled material, and this H2A migrates differently from the background H2A on Acid-Urea-PAGE, I.E. it is not

Fig 14.2

Microinjection of Xbh1 clones

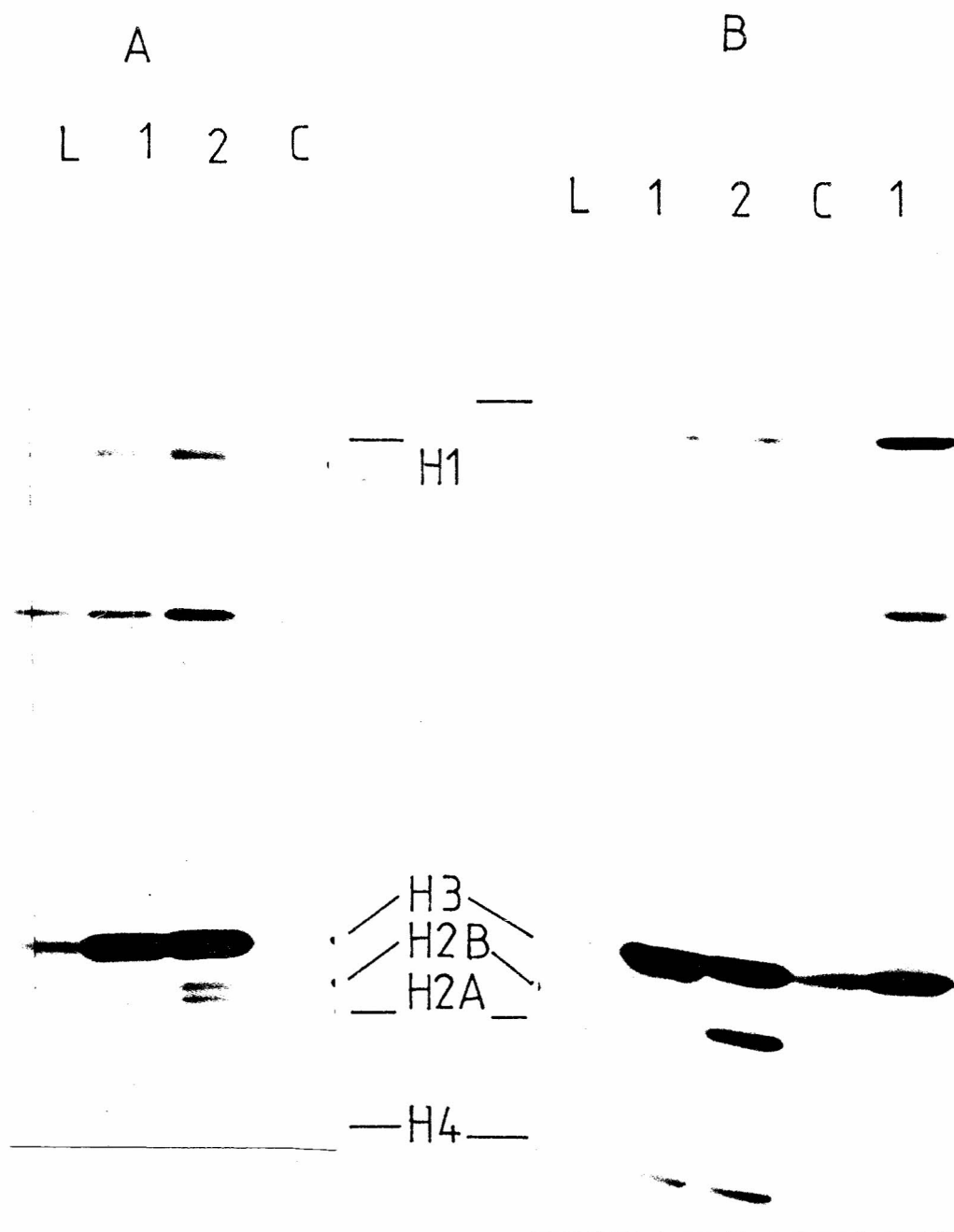


Fig 14.2 Microinjection of Xbhl clones into Xenopus oocytes. 20nl of DNA solution was injected into each of 20-40 oocytes, the oocytes were incubated with labelled amino acid and then the basic nuclear proteins extracted as described and separated on 18% SDS-PAGE. A) Autoradiograph of (35)S-Methionine- labelled material. B) Fluorograph of tritiated-lysine - labelled material. DNA: '1'=pWB1, '2'=pWB2, 'L'=lambda Xbhl, 'C'= control injection of 20nl of 80mM NaCl.

Fig 14.3

the oocyte-type H2A. But if oocytes are labelled with Methionine after injection of pWB1 and pWB2, the two plasmids give identical results. Thus the H2A synthesised by pWB2 in large amounts cannot contain Methionine. Notably pRW8a, a pBR325-recombinant containing the EcoRI insert of Xlh8 in the same relative orientation as the Xbhl insert in pWB2, does not show this methionine-less H2A, but does synthesise an H2A which migrates faster than the X.laevis erythrocyte standard on SDS-PAGE (Woodland per com). These various results are summarised in Fig 14.3



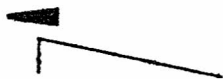
14.4 Conclusions.

The most obvious conclusion is that the insert of Xbhl strongly stimulates the oocyte to make H1 and H3, and in some cases H2A. Although this has been found before (Merz and Gurdon 1977, de Robertis and Merz 1977) it is still notable that a foreign DNA injected in relatively enormous amounts in a prokaryotic-eukaryotic recombinant molecule functions so as to produce recognisable proteins at all, let alone at high rates. This is a testimonial to the flexibility of the transcriptional apparatus of oocytes. It also suggests that these genes will be of use in future studies, especially the H3 gene which possesses unique BamHI, PstI and SstI sites in its coding region and BglII and AvaI sites in or near its 5' end, all potential sites for insertion of foreign DNA or for in vitro mutagenesis. Thus while this experiment has not performed any such studies itself, it has been a useful first step towards such studies.

The results shed more light on the structure of the clone. although raising as many questions as they answer in the case of the H2A gene. The transcription of the H1, H3 and H4 genes in Xbhl show these genes are intact in this clone. This point has been glossed

Fig 14.3

H2A expression in Xbh1 & Xlh8

Plasmid	pRW8A	pWB1	pWB2	
Clone	Xlh8	—Xbh1—		
H2A				
SDS	F	S	S	S
Met	+	+	+	-
Amount	+	+	+	+++

Summary of microinjection results as they apply to H2A. 'Plasmid' - plasmid recombinant microinjected. 'Clone' - genomic clone from which the insert in the plasmid was derived. 'H2A' - orientation of the H2A gene relative to the vector sequences. 'SDS' - mobility of H2A bands on SDS-PAGE: F=fast relative to the X.laevis erythrocyte standard, S=slow. 'Met' - + for proteins containing methionine, - for those without. 'Amount' - relative intensity of band seen on SDS-PAGE.

over slightly in the mapping experiments as nucleic acid hybridisation studies locate homologous sequences, not functional genes. The demonstration that Oocytes, when injected, make H1, H3 and H4 from Xbhl enables me to say that these genes are complete in Xbhl. With a caveat about its anomalous behaviour I noted above, this also applies to the H2A gene. This presumably also includes upstream regions essential for transcription such as the 'TATA' and cap boxes, as these have been found to be essential for transcription in oocytes (Grosschedl and Birnsteil 1980, Grosschedl et al 1981, Grosschedl and Birnsteil 1982). It is interesting that the H3 and H1 genes, which are strongly expressed in this assay, are on one end of the clone, and that the less well expressed H2A, H2B and H4 genes (H2A is being expressed in one plasmid in response to a recombinant-specific effect discussed below) are on the other. I will amplify on the possible significance of this spacial relationship later in section 17.1

The H2A results are the most difficult to analyse. They would suggest that Xbhl contains two H2A genes, one transcribed weakly in both pWB1 and pWB2 and coding a protein which contains methionine, and one transcribed only in pWB2 and which does not contain a methionine codon. There is sufficient room in the region mapped as an H2A gene in section 9.3.3 above to fit two H2A genes, although there could not be more than about 100bp between them, into which would have to be fitted one each of 3' homology block, 'enhancer' region if any, and TATA and cap boxes. They must be transcribed in the same direction as only one strand of the DNA hybridises to H2A mRNA (Section 9.3.2). Alternatively one or other of the "H2A" proteins might actually not be an H2A. The more likely candidate of the two is the low-level, methionine-containing H2A expressed in pWB1 and pWB2. H2A proteins do not usually contain methionine (although as saw in section 1.3 H2A is quite variable between and within species). The methionine-minus

protein can be detected on Acid/Urea-PAGE comigrating with authentic H2A markers, so this proteins is fairly likely to be an H2A, and so if we require to eliminate one of the two, the methionine-plus protein, which cannot be seen in the H2A region of Acid/urea-PAGE, is the better choice.

The most striking effect seen in H2A transcription is undoubtedly the enhancement of H2A synthesis in pWB2 over pWB1 and lambda. There are four possible explanations for this.

Firstly, the insert in pWB1 might have undergone a rearrangement during its re-cloning into pBR325 from the original lambda. This is unlikely, as these techniques are found to be remarkably faithful in replicating exact DNA sequence (Glover 1977, Dahl, Flavell and Grosveld 1981).

Second, this could be due to run-on of an RNA polymerase which binds to vector sequences and then either 'diffuses' along the DNA or transcribes it until it reaches an authentic eukaryotic termination site. Thus in one orientation the first terminator it would reach would be that downstream of the H2A gene. Two arguments suggest that this is not so. Firstly, this run-on would have to traverse the H1 and H3 genes, both of which are very actively transcribed in the opposite direction. Whether an RNA polymerase could 'run on' through such an oncoming barrage of other polymerases is uncertain. It is possible that the H2A genes are being transcribed on different molecules than are the H1 and H3 genes among the vast population of molecules being injected, in which case this argument would not apply.

Thus if the cause of the differential expression of H2A is not a mutation in the immediate promoter or a run-on effect, then it must be

14. Microinjection

a long-range effect of vector sequences on the H2A gene. There are two ways this may happen: an 'enhancer' may fortuitously be present in pBR325, or one may be created in one recombinant by a combination of vector and eukaryotic sequences. These hypotheses are testable, but as yet have not been tested.

Long-range enhancer effects have been found in other gene systems (Grosschedl and Birnstiel 1982, Moreau et al 1981, de Villiers 1981, Benoist and Chambon 1981, Jat et al 1982). However here the effect appears to be non-specific, enhancing transcription of any gene placed downstream of the enhancer sequence. By contrast this 'enhancer' appears to operate only on the H2A gene of Xbhl, despite the adjacency of other transcribable genes.

Fourthly, this could be an artefact. In this regard it is instructive that the expression of other genes on microinjection can vary significantly from experiment to experiment. Fig 14.2 (b) shows this. The results of two separate microinjection experiments on pWB1 are shown run on the same gel. In one (track 2) H4 is synthesised to a greater amount than H1. In the other (track 5) this is reversed, and the H4 protein band is little more intense than that in the 'control' track. The source of this variation is unknown, but it is commonly observed. Preliminary data from follow-up experiments suggest that a similar artefact might be responsible for the anomalous transcription of H2A from pWB1 and pWB2 (Woodland per com).

15. DNA sequence studies on the coding regions of Xbhl.

A collaborative study with Dr. P.C.Turner.

15.1 Sequence of the H1 gene of Xbhl.

The H1 and H3 genes of Xbhl have been partially sequenced. The sequence of some of the H1 gene region is presented in Fig 15.1. This sequence is under active investigation, and will shortly be published elsewhere. Dr.P.C.Turner has been sequencing the H1 genes of Xbhl and Xlh8 in parallel, and has detected 7 differences between them out of 450bp sequenced in the coding and adjacent regions, and one stretch of 6bp at the start of the H1 gene which is not shared between Xbhl and Xlh8. Curiously, several of these differences are in the first base position of a codon, and result in a change in the protein sequence being coded by this gene. The difference between Xbhl and Xlh8 is @1.6% (counting only the point mutations) as computed from this sequence, which is in good agreement with the figure calculated in section 13.3.2 above from restriction data. Perhaps this is curious, as this is a gene coding for a protein and so might be expected to be more rigidly conserved than the spacer sequences around it. The protein sequence is quite divergent from Sea urchin, trout testis and rabbit thymus H1 proteins, while sharing obvious homologies with them in the conserved region between amino acids 90-110, but bears strong homology in some regions with another H1 gene cloned from X.laevis by R.W.Old : this latter gene clone (lambda Xlh19) has no large-scale structural similarity to Xbhl or Xlh8, and the H1 gene in it appears to be only half a gene. The 5' half of the gene does not code for any recognisable protein, and there is no Intron-splice site coinsensus sequence boardering this half of the gene and the half which does code

Fig 15.1

H1 gene sequence

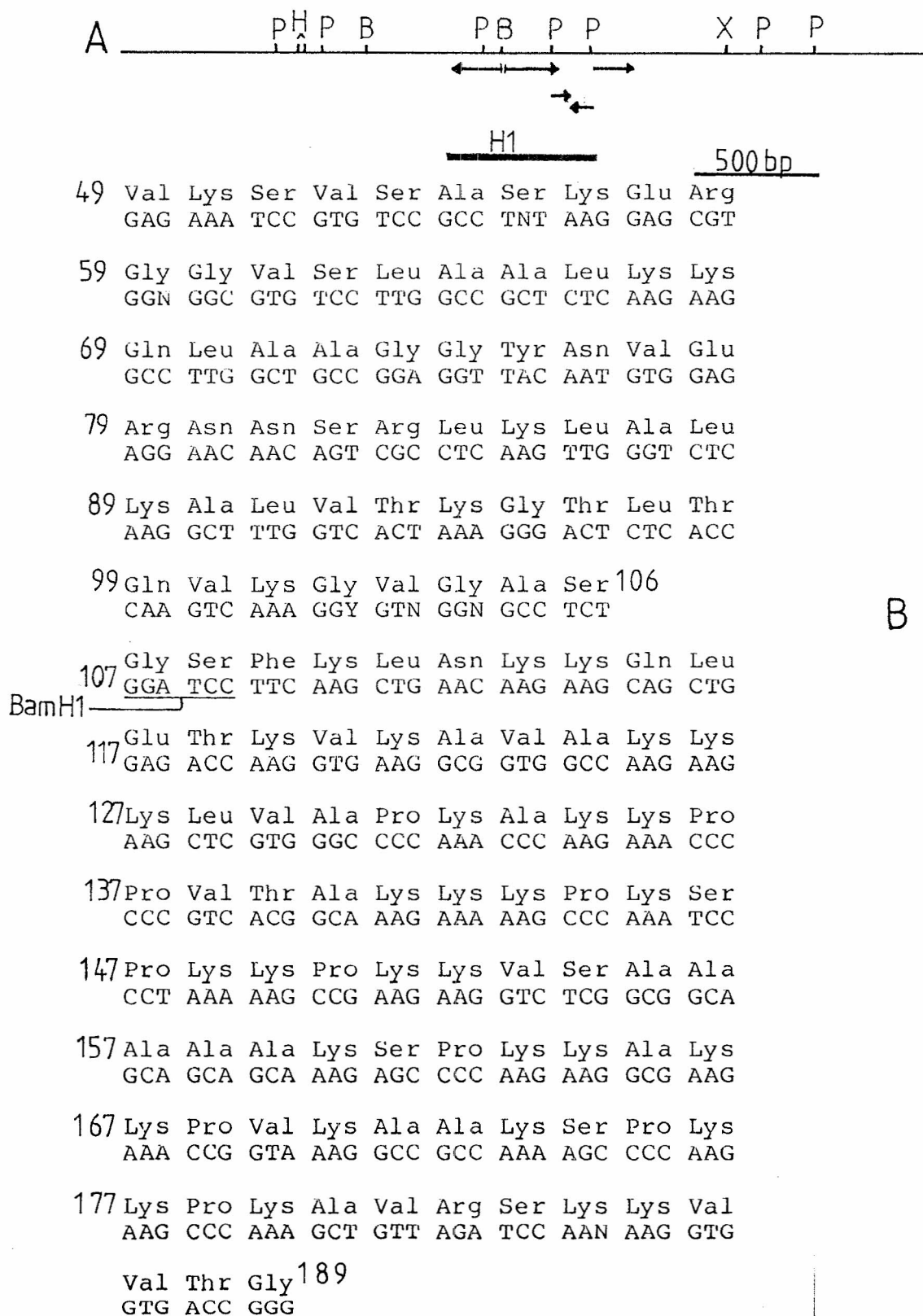


Fig 15.1 cont

C

```

CCGGC ACTGT CCCCC CCCGG CCTCA AAGGC TCTTT TCAGA
                ** * *** ***** *****
                ----- Homology -----
GCCGC CGCCG GCTCC GTCAG AAGAG CCGAT ACTGT CACTG
    ** * * *
    ----- block -----
CATTA TAGTC AGGGA GGCTT TCTGG TATCT GCTGA ATTGT
GCGAT GCCAG TCTGA CTTC

```

D

	80		89
Xbh1	Asn Asn Ser Arg Leu Lys Leu Ala Leu Lys		
	AAC AAC AGT CGC CTC AAG TTG GCT CTC AAG		
	* *** ** *** *** ** *** *** ***		
Xlh19	ACA AAC AGC CGC CTC AAG CTG GCT CTC AAG		
	Thr Asn Ser Arg Leu Lys Leu Ala Leu Lys		
1	Ala Leu Ala Thr Lys Gly Thr Leu Thr Gln		
	GCT TTG GCT ACT AAA GGG ACT CTC ACC CAA		
	*** ** * ** ** * * ** ** * ***		
19	GCT CTG GTC ACG AAG GAG ACC CTG CTC CAA		
	Ala Leu Val Thr Lys Glu Thr Val Leu Gln		
			99
1	Val Lys Gly		
	GTC AAA GGC		

19	GTC		
	Val		
	100		

Fig 15.1 A) Sequence strategy for sequencing the H1 gene.

B) Sequence of part of the H1 gene coding region. The amino acid sequence coded is written above the DNA sequence, and is numbered according to the best fit to the Rabbit and Trout H1 sequences (Isenberg 1979). There is no formal proof that the sequence is continuous over the BamHI site, but the protein coded here is strongly homologous to the rabbit and trout H1s, and indicates no gap at this point.

: B=BamHI, P=HpaII, X=XbaI, H=HindIII. C) H1 gene 3' homology block. Region downstream of the H1 coding region. The bases homologous with the 3' Homology block (Hentschel and Birnstiel 1981) are asterisked.

D) Comparison of protein (1st row) and DNA (2nd row) sequence of the H1 gene of Xbhl with the DNA sequence (3rd row) and protein (4th row) sequence of Xlhl9 H1 gene. Codons 80-100 of the Xbhl sequence are listed. Base homologies are asterisked.

Fig 15.2

H3 gene sequence

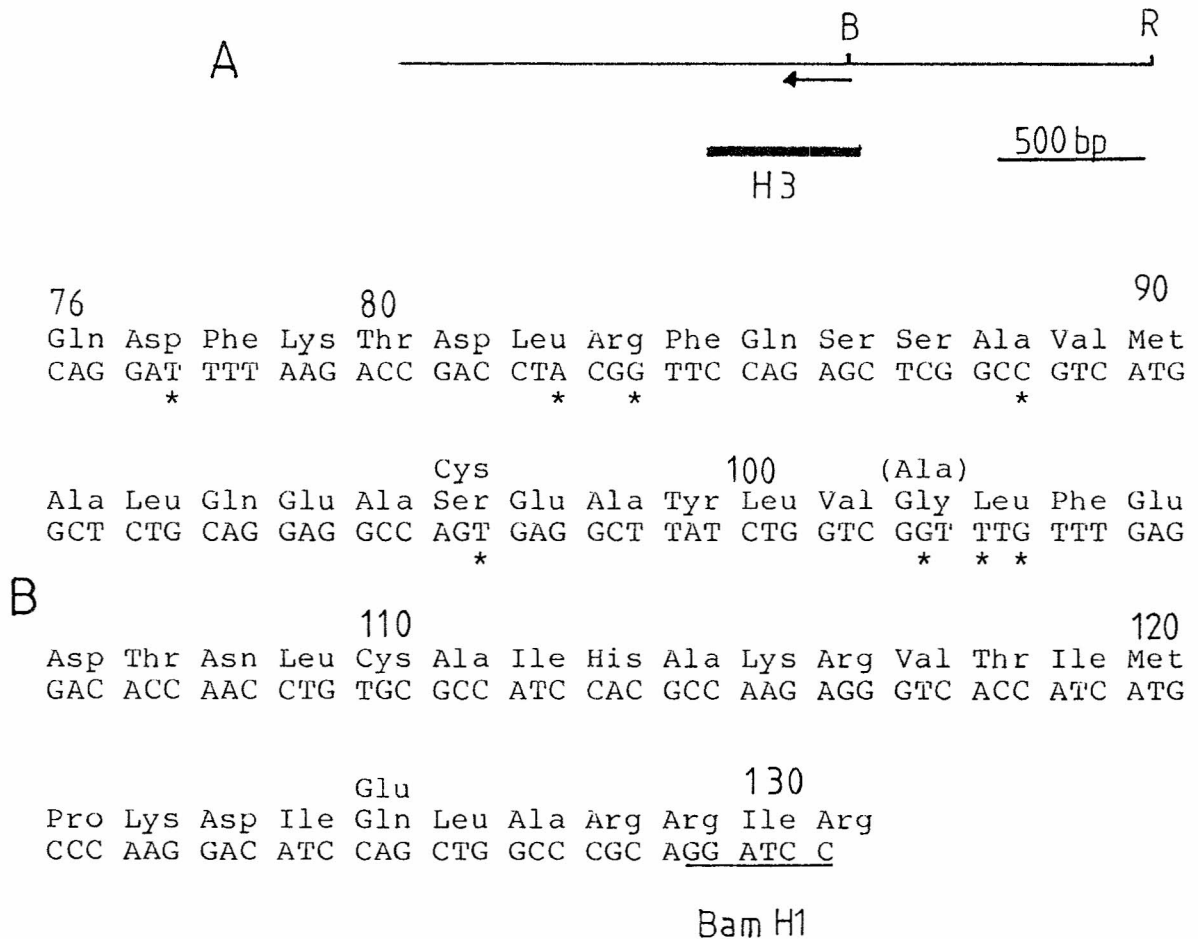


Fig 15.2 A) Sequence strategy for sequencing the H3 gene of Xbh1

B) Partial sequence of the H3 gene of Xbh1. The protein sequence coded by the gene is written above the DNA sequence. Amino acids written above the protein sequence are the corresponding amino in calf thymus H3 (Isenberg 1979) where these are different. The bracketed amino acid is different between Xbh1 and the X.laevis H3 gene's protein reported by Moorman et al (1981). Asterisked bases are different in Moorman et al (1981)'s gene.

for a recognisable H1 protein. This X.laevis H1 gene is probably therefore a pseudogene. These points are shown in Fig 15.1 .

The Xbhl H1 gene has a conventional 3'homology block downstream of its 3' end. The DNA sequence studies have allowed me to position the gene more precisely on the physical map of Xbhl: the summary map in the Discussion section 17.1 incorporates this information.

15.2 Partial sequence of the H3 gene of Xbhl.

A partial sequence of the Xbhl H3 gene is presented in Fig 15.2 . This was determined by sequencing from the BamH1 site in the 3' end of this gene. The region codes for amino acids 76-130 inclusive in which region the protein sequence differs in two amino acids from the Calf thymus H3 sequence (from Isenberg 1979): positions 96 (Ser instead of Cys) and 125 (Gln instead of Glu). It is extensively homologous with the X.laevis sequence determined by van Dongen (1982): differences are indicated by asterisks in Fig 15.2 . The proteins coded by the two genes differ only in position 102, where Xbhl codes a Gly instead of the X.laevis Ala.

15.3 Analysis of H3 genes from several species.

15.3.1 Coincident codons.

The majority of differences between the H3 sequences of van Dongen (1982) and that presented here are in 3rd base positions (2 are not). Turner and Woodland (1982) and Busslinger et al(1982) have pointed out that such positions evolve quite slowly when compared to spacer or even coding DNA of some genes, and Turner and Woodland

15. Gene sequences

(1982) have found some third-base positions in the codons of the H4 gene which are conserved in a wide range of organisms' H4 genes from Sea urchin to man. Is this true of H3 genes? The answer is given in Fig 15.3 . Here data from two chick (Engel, Sugarman and Dodgson 1982), one mouse (Sittman et al 1982), one human (Clark, Krieg and Wells 1981), two Psammechinus and one Paracentrotus (Busslinger et al 1982), one Stongylocentrotus (Sures et al 1978), one X.laevis (van Dongen 1982), one X.borealis (this work) and some very fragmentary sequence from Notophthalmus (Stephenson, Erba and Gall 1981a) H3 genes have been compared: the comparisons are not as informative as I might like because, as well as the Notophthalmus, the X.borealis and the human sequences are incomplete. I have scored the codons which are the same in all sequences, ie which have all three bases the same. Equivalent positions nearly always have the same first and second base, of course, as the protein sequences for which they code are so highly conserved. As we can see, there are 12 codons out of the 134 in the gene which can vary (the two other codons are for methionine) which are constant in all ten sequences. I have called these 'Coincident codons' to avoid the inherent bias of evolutionary viewpoint in called them 'conserved'.

Also on Fig 15.3 are equivalent comparisons performed with subgroups of the sequences. As would be expected, these subgroups of species, which are more closely related phylogenetically than the entire group, yield more coincident codons. The number of coincident codons seen within any group of species is roughly inversely proportional to the evolutionary distance between members of the group.

15.3.2 Codon usage and sequence conservation in H3 genes.

Fig 15.3

Coincident codons in H3 genes

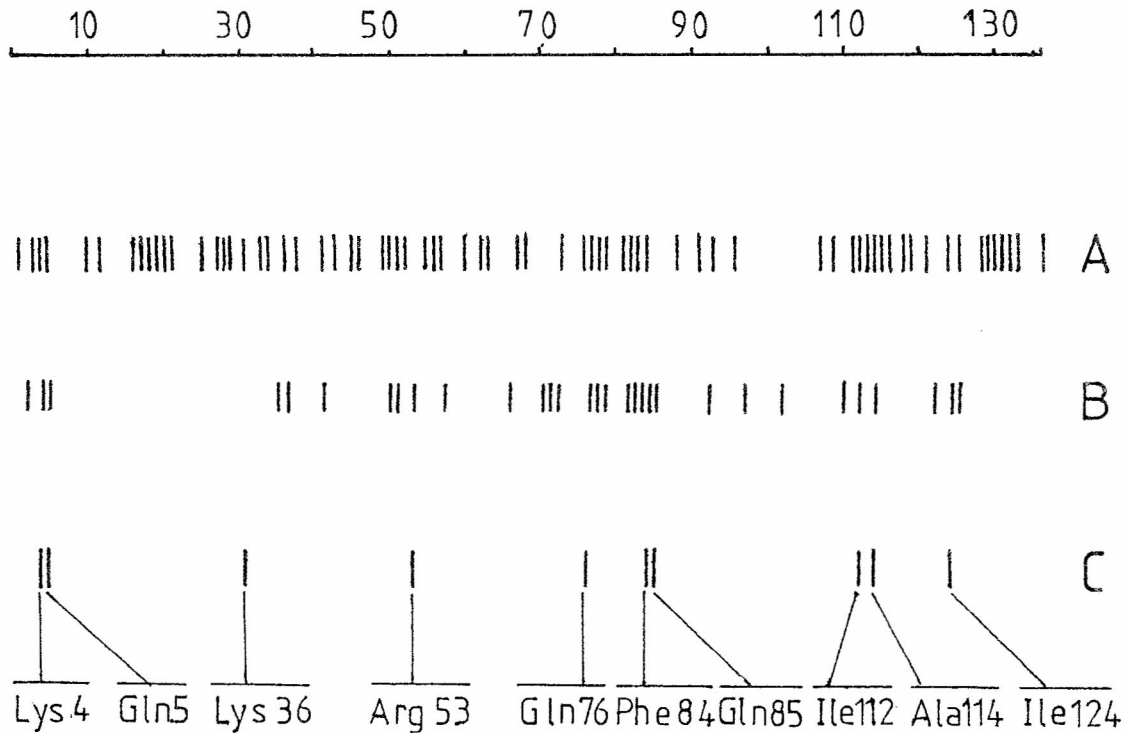


Fig 15.3 Coincident codons in 10 H3 genes. See text for details. A score on this figure means that the codons for that amino acid are the same in all the sequences tested. From Sures et al 1978, Busslinger et al 1982, Sittman et al 1981, Stephenson, Erba and Gall, 1981, Moorman et al 1981, van Dongen and human H3 gene sequences. 1982, Clarke, Krieg and Wells 1981, Engel, Sugarman and Dogson 1982 and this work. A) Comparison of 4 sea urchin gene sequences. B) Comparison of chick, mouse and human sequences. C) Comparison of sea urchin, Xenopus, Notophthalmus, chick, mouse, man

15. Gene sequences

Is the presence of these conserved codons due solely to 'fossilisation' of codons in the genes, implying that a very slow rate of random drift of third base positions has not fully randomised the sequence of H3 genes between Sea urchin and man but that the coincident codons are just the product of chance? Or is this the result of selection pressure acting to preserve certain sequences in the coding region of H3 genes? I favour the latter explanation for the following statistical reason.

If we look at the usage of codons in the H3 genes sequenced to date, we can see that the redundant bases are not a random distribution of A,T,G, and C, but rather that usage of synonymous codons is extremely non-random (See Fig 15.4). For example, 70% of lysine codons are AAG, only 30% AAA. This superficially explains the coincident codons noted above, which are invariably the most common synonym for the amino acid coded in that position. For example, there are 13 lysines in H3. The probability that all the codons at any one position in 9 H3 genes (not 10, as we do not have complete sequence for all the ten genes) will all be AAG is 4%. Thus the probability that two lysine codons out of 13 should be AAG in all 9 genes is @6%. Similar calculations for the other coincident codons shows the probabilities of their occurrence to be: 3 CAG out of 8 Gln - 15% , One CGC out of 17 Arg - 5% , one TTC out of 4 Phe - 17% , 2 ATC out of 7 Ile - 4% , one GCC out of 18 Ala - 0.5%. These figures suggest significance, but as any sufficiently prolonged search will turn up unlikely coincidences I would not consider this data compelling case for sequence selection, given the codon usage observed.

That the sequence is important is shown by considering not why the coincident codons occur but why the codon usage, which makes the coincident codons a likely event, is as it is. Such unequal usage of

Fig 15.4

Codon usage in 10 H3 genes

Amino acid	codon	number counted	frequency				
Phe	TTT	7	.28	Ala	GCT	48	.318
	TTC	18	.72		GCC	61	.404
Leu	TTA	3	.028	Tyr	GCA	26	.172
	TTG	5	.048		GCG	16	.404
	CTT	12	.114	His	TAT	6	.24
	CTC	21	.2		TAC	19	.76
	CTA	13	.124	Gln	CAT	6	.316
Ile	CTG	51	.2		CAC	13	.684
	ATT	11	.24	Asn	CAA	6	.17
	ATC	33	.73		CAG	29	.83
	ATA	1	.02	Lys	AAT	5	.5
Val	GTT	9	.177		AAC	5	.5
	GTC	20	.392	Asp	AAA	29	.30
	GTA	7	.137		AAC	68	.70
	GTG	13	.294	Glu	GAT	14	.42
Ser	TCT	8	.151		GAC	19	.575
	TCC	9	.17	Cys	GAA	22	.275
	TCA	2	.037		GAG	58	.725
	TCG	5	.094	Arg	TGT	7	.636
	AGT	11	.027		TGC	4	.364
Pro	AGC	18	.339	Gly	CGT	41	.287
	CCT	15	.278		CGC	53	.371
	CCC	27	.5		CGA	13	.109
	CCA	9	.167		CGG	9	.063
Thr	CCG	3	.055		AGA	10	.07
	ACT	10	.122		AGG	17	.119
	ACC	45	.549		GGT	11	.189
	ACA	17	.315		GGC	21	.362
	ACG	10	.122		GGA	20	.345
					GGG	6	.103

15.4 Codon usage in 10 H3 genes. Refs as in Fig 15.3

15. Gene sequences

synonymous codons is common. Turner and Woodland(1982) find a similar distribution of codon useage in H4 genes; the only notable difference between theirs and that given in Fig 15.4 is that in H4 genes serine AGPy codons are rarely used, whereas in H3 genes they comprise more than half the serine codons. In general differences between the useage of synonymous codons in H3 genes are not as great as those between H4. Amaldi et al (1982) also find the same assymetries of codon usage in X.laevis ribosomal as in histone genes for some but not all codons. Thus Ile, Gln, Lys, Ser and several others have the same pattern of codon usage in ribosomal and histone genes in Xenopus, but Phe and Arg have the opposite bias. Grantham, Gautier and Goug (1980, 1981) report that this bias in codon useage is widespread. It could be due to any one of three classes of reason : evolutionary accident, coding function or non-coding function. If the first reason is correct then the useage of codons in present day genes should reflect merely some original bias in an ancestral H3 gene's useage of synonymous codons, itself presumably due to chance. If the second is true then certain sequences will be favoured over others within the codon, but the choice of adjacent codons will be random within the requirements of the amino acid sequence. Thus under this hypothesis we might suppose that codons which may mutate to a 'stop' codon with a single base change are selected against in preference for codons which may only mutate to a 'stop' codon by two base changes. If the third explanation is correct then the sequence within and between codons will be conserved, so that adjacent codons are selected non-randomly.

I tested these various hypotheses in Fig 15.5 . Here I have taken data from three H3 genes only - X.laevis (van Dongen 1982), mouse (Engel, Sugarman and Dodgson 1982) and sea urchin (Sures et al 1978). To take, say, all the sea urchin sequences in this analysis could

Fig 15.5

Dinucleotide frequencies

dinucleotides : observed

2nd and third bases : XNN						3rd and 1st bases : XXN NXX					
2nd base	3rd base	A	T	C	G	3rd base	1st base	A	T	C	G
	A	23	13	20	53		A	12	9	16	33
	T	10	16	34	30		T	19	11	30	29
	C	22	33	51	8		C	60	12	40	28
	G	16	25	36	13		G	33	6	32	31

bases in codons

Frequencies of bases in codon positions,

	1st base	2nd base	3rd base
A	124	109	70
T	38	90	89
C	118	114	140
G	121	88	102

dinucleotides : expected

2nd and 3rd bases : XNN					3rd and first base : XXN NXX						
2nd base	3rd base	A	T	C	G	3rd base	1st base	A	T	C	G
	A	19	24	38	28		A	22	7	21	21
	T	16	20	31	23		T	28	8	26	27
	C	20	25	40	29		C	43	13	41	42
	G	15	19	31	22		G	32	10	30	31

$$\chi^2 = 69.6$$

$$\chi^2 = 38.1$$

15.5 Statistical analysis of dinucleotides in H3 genes. See text for description.

15. Gene sequences

lead to a false impression of statistical significance by boosting the sample size with four closely related, and hence not random, sequences. I asked:

1) with what frequency does each base occur in the 1st, 2nd and 3rd positions in the codons and

2) which base follows each of the bases in the 2nd and 3rd positions (ie how often do each of the 16 possible dinucleotides occur in the 2nd and 3rd), and 3rd and 1st base-pair positions in two adjacent codons. Now, given the observed base frequencies at each position, we can calculate how often we would expect each base to be followed by each other base at any one position. Thus, as 124/401 1st bases are A, and 70/401 3rd bases are A, we would expect the two-codon combination NNA-ANN to occur 22 times in 401 codons. It actually occurs 12 times. I did the same calculation on each of the 16 possible dinucleotides: the results are also presented in Fig 15.5. Chi-squared testing this contingency table against the observed frequencies shows that in the actual genes the distribution of third bases is extremely non-random with respect to both the codon the base is in and the adjacent codon. Thus what is being observed here is not evolutionary accident (which would not lead to significant deviation from chance and could not explain how ribosomal protein genes and histone genes could come to have similar patterns of codon usage for some amino acids) nor an effect of the gene's coding role (which would lead to non-random base distribution within the codons only) but a conservation of sequence which does not relate to the gene's protein-coding capacity.

15.3.3 CpG and sequence conservation in H3 genes.

How this putative sequence conservation effect I state is acting in H3 genes actually is acting, this analysis does not show. Nor does it show what the sequence might be. It is notable that in both 2nd-3rd and 3rd-1st base pair comparisons the dinucleotide CpG is strongly disfavoured. This is a general feature of eukaryotic DNAs (Subak-Sharpe et al 1966, Grantham 1978, Nussinov 1981). This sequence is sometimes methylated on the C-residue in eukaryotes, and the methyl group, once introduced, is perpetuated by modification of the new C-residue by a CpG-specific methylase when the CpG is replicated (Vanyuchin et al 1973, Bird and Southern 1978, Bird 1978, Bird and Taggart 1980, Wigler 1981, Doerfler 1981, Stein et al 1982). The absence of methyl groups on CpG sequences in otherwise methylated genomes has been correlated with gene transcription (van de Ploeg 1980, Bird, Taggart and Macleod 1981, Mandel and Chambon 1979) and removal of the methyl groups by 5-azacytidine during replication with gene activation (Niva and Sugahara 1981, Taylor and Jones 1979). This has lead to the suggestion that methylation at CpG groups is concerned with the control of transcription. That this is not necessarily so has been shown by Macleod and Bird (1982) who found no inevitable correlation between rDNA transcription and de-methylation.

Bird (1980) has suggested that the known mutational lability of 5-Me-C residues is the reason why CpG is underrepresented in the genome. C and meC residues undergo spontaneous deamination to U and T residues respectively. The U reidues are replaced by C by a specific correction mechanism, but the T residues are not recognised as erroneous, and so are left in the DNA. Thus a meC to T transition is effected. Bird (1980) demonstrates a direct correlation between low CpG, high TpG and CpA and high DNA methylation, and suggests that this is because C(me)pG mutates to TpG (and its complement CpA) more

15. Gene sequences

rapidly than these sequences mutate back again, hence denuding the highly methylated genome of CpG. This is not supported by the data presented in Fig 15.5 . Although TpG and CpA occur slightly more frequently than expected by chance, this difference varies according to whether the 2nd-3rd or 3rd-1st base pairs are being considered, is not correlated well with the paucity in CpG and does not nearly account for the CpG depletion. Furthermore several other dinucleotides which have no connexion with the lability of CpG are severely under-represented in H3 genes. ApA is one we mentioned above. The results of several other surveys of gene sequences also fail to support the idea that CpG-paucity alone explains unequal codon usage (Graham 1978, Graham, Gautier and Goug 1981,1982, Nussinov 1981, Turner and Woodland 1982, Amaldi et al 1982). Thus we can be sure that, while CpG is under-represented in these H3 genes, this is due to selection for some sequence more complex than 'not-CpG'. This is not to say, of course, that CpG paucity elsewhere in the genome is due to the same mechanism as in the H3 genes. As I mentioned, transcribed genes have lower methylation than non-transcribed regions, and so might undergo different pressures with respect to their CpG dinucleotides.

Turner and Woodland (1982) note that in H4 genes Serine codons at any given position are either TCN or AGPy, and never are both types found in different genes in homologous amino acid positions. They suggest that this is due to the improbability of the simultaneous occurrence of a T-A and a C-G point mutation in the same codon to convert one type of codon into the other. In the H3 genes compared here, all serines bar serine 57 follow this rule. In the latter position, all vertebrates have TCN, all Sea urchins AGPy. It is possible that this represents a double mutation event, but more likely that the genes have evolved through a gene which does not code for

15. Gene sequences

Serine at position 57. In this regard it is notable that Serine 28 is not invariably a serine all the genes sequenced to date. In mouse it has changed to the GCG alanine codon (Engel, Sugarman and Dodgson 1982) and in Xenopus laevis to a TCG cysteine codon (van Dongen 1982). The latter is, of course, a potential intermediate between TCN and AGPy. This work and the H3 gene sequence of van Dongen (1982) also show a Cys to Ser change in an H3 protein, this time at position 96.

16. Heteroduplex analysis of Xbh1 and Xlh8.

A collaborative study with Dr. Lesley Coggins (Beatson Inst.).

16.1 Introduction.

Electron microscopic hybridisation techniques have been widely used as a complement to other hybridisation procedures and to DNA sequencing for determining the precise limits of homology between nucleic acid molecules (Davis, Simon and Davidson 1981). In all the techniques nucleic acid molecules are hybridised together, then spread out on an electron microscope grid and visualised directly. The spreading technique is such that the length of the molecules as measured in the microscope is proportional to their length in bases (or base pairs), so that a physical map of their homologies may be obtained by measuring the length of the structures seen in the E.M..

I embarked on a heteroduplex analysis of Xbh1 and Xlh8 to find the differences, rather than the similarities, between them. As I noted in section 13.3.1 several of the differences between the restriction patterns of Xbh1 and Xlh8 are explicable if we postulate a small deletion in the latter relative to the former. To demonstrate that this is the correct explanation I intended to hybridise Xbh1 to Xlh8 and see what regions of the former, if any, were not represented in the latter. Mapping these regions would give a precise location to any potential deletion or insertion.

While I was performing some preliminary experiments on this, however, it became clear that there was a lot of intra-molecular hybridisation going on as well. This can occur if the molecules are

16. Heteroduplex analysis

internally repetitious (Wu and Davidson 1973, Ferguson and Davis 1975). This is known as 'snapback' hybridisation, and the structures so illuminated as 'snapback structures', because, being an intra-molecular reaction, it is independent of the concentration of DNA being used and occurs extremely fast under the dilute conditions employed for heteroduplex analysis.

The usefulness of electron microscopic analysis of DNA homologies for this study is that homologies between an entire molecule and another entire molecule can be mapped without fragmenting the molecule into defined pieces. To map regions of homology using radiolabelled tracer hybridisation to DNA bound to filters, as I have done in several experiments presented above, the region to be used as a probe must be accurately excised from the rest of the DNA. Hybridisation in solution or on filters tells us only that two molecules share sequence, not how much or where it is. Both these questions are answered by electron microscopic techniques. However they are prone to artefacts of DNA-DNA aggregation, and so do require careful DNA preparation and analysis of a number of results to obtain reliable data.

16.2 Methods.

16.2.1 Preparation of reagents.

DNA was prepared by equilibrium CsCl gradient centrifugation as described. DNA which had been prepared some weeks previously was found to be too extensively nicked to be of use for these procedures, as was DNA which had been kept at room temperature.

16. Heteroduplex analysis

Formamide was purified by recrystallisation: formamide was chilled, then stirred at 0°C until ~30% had crystallised. These crystals were filtered out and used as the formamide stock. All stock reagents were kept at -20°C except cytochrome C solution and DNA, which were kept at 4°C.

16.2.2 DNA:DNA hybridisation.

These conditions were after Davis, Simon and Davidson (1971) modified by Dr. L.Coggins (per com). Hybridisation was carried out in 50% formamide at 37°C: the formamide reduces the melting temperature of the DNA helix, so these conditions are roughly analogous to 60 to 65°C aqueous hybridisations used in most of this study. 0.1 - 0.2µg of DNA was used per hybridisation. 10µl of formamide and 2µl of Tris/EDTA (1MTris pH8.5, 0.1MEDTA) were mixed with 8µl of water containing the DNA and taken up into a surface-siliconised 50µl capillary. The ends of the capillary were sealed with a bunsen, the capillary was heated for 2-5 mins at 75°C to denature the DNA and then was incubated in a 37°C water bath. It was incubated there for ~30 mins for snapback structure examination, 2-3 hours for heteroduplex analysis.

16.2.3 'Kleinschmidt' electron microscopy of DNA.

In this technique of Kleinschmidt (1968, Davis, Simon and Davidson 1971) the DNA is spread out in a film of basic protein which both separates the molecules and binds to them, making them more visible in the e.m. : this film is then taken up on a plastic-coated e.m. grid for visualisation. The grids are coated with Collodian thus. The grids were placed on a wire mesh which was immersed in distilled water in a Buchner funnel. The neck of the funnel was

16. Heteroduplex analysis

suitably stoppered with a rubber tube and screw clip. One drop of 2% Collodion in amyl acetate was put onto the water surface and allowed to spread and dry. When interference colours were no longer visible in the film, the water was slowly released from the funnel, dropping the film onto the grids. These may then be dried under vacuum and, if kept clean, kept indefinitely.

DNA from the hybridisation is emptied into a repelcoated tube, and 40 μ l of water and 30 μ l of formamide were added. Just before spreading 10 μ l of 1mg/ml cytochrome-C in water were added. This is the Hyperphase. The hypophase, on which the hyperphase solution is spread is made thus: 30mls (per spread) of 10% formamide, 0.01MTris, 1mM EDTA is made up and tipped into a 9cm diameter glass petri dish. A light dusting of clean talcum powder was blown onto the surface to visualise the protein film when it forms. A clean glass slide was then slid onto this solution so that it rested on the base and the side of the petri dish as a ramp leading through the surface. It is essential that this glassware be scrupulously clean and detergent free: chromic acid washing followed by overnight rinsing in tap water and then two rinses in distilled water was found adequate in these experiments.

When the talc on the hypophase surface was stationary, the hyperphase solution was dropped gently onto the glass slide 'ramp' at the point where it entered the hypophase. The hyperphase spread spontaneously over most of the hypophase surface, pushing the talc ahead of it. With a pair of tree-pruners a coated grid was touched to this film coated-side down, then drained of hypophase on filter paper, stained in 90% ethanol containing 1/100 vol of 5mM Uranyl acetate in 50mM HCl for 20 seconds, washed in 95% ethanol and dried in air. I took 4 - 5 grids from each film at @2 minute intervals, as

16. Heteroduplex analysis

the properties of the film alter slightly with time. The grids were shadowed with Pt/Pd: although several authorities say that this is unnecessary (Kleinschmidt 1968, Davis, Smon and Davidson 1971) , I found it essential.

The grids were examined under a Joel electron microscope at 6000x magnification. Specific molecules were photographed for detailed examination.

16.3 Results.

16.3.1 Heteroduplex analysis of Xlh8 and Xbhl.

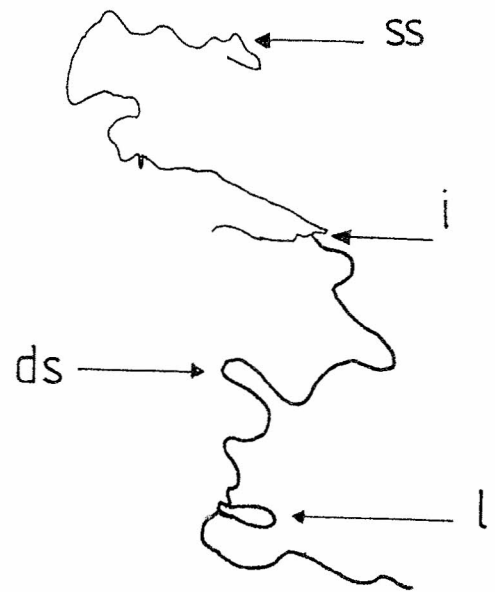
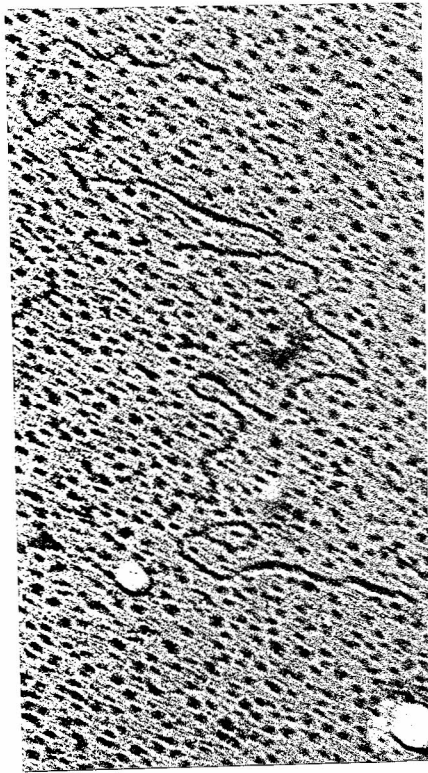
Fig 16.1 shows a representative of a heteroduplex between pRW8a and pWB1, and a schematic plot of eight such duplexes. pRW8a and pWB1 have the Xlh8 and Xbhl inserts respectively cloned in the opposite relative orientation in pBR325. So, if the inserts hybridise to each other, the plasmid vector regions of both molecules will be in opposite physical orientations and so will not hybridise. Thus the end of the insert can be accurately mapped by locating the point where the double-stranded insert:insert hybrid splits into two single strands. Of course, several other hybridisations are possible, so the structures summarised in Fig 16.1 are by no means the only ones seen on the grids. Complete double-stranded circles of pWB1 or pRW8a provide an internal control for the length of the mapped segments.

The data in Fig 16.1 is consistent with a deletion in one of the plasmids relative to the other 3.1 ± 0.22 (=S.D.) kb from one end. Which end is not revealed, of course. However this is consistent with a deletion occurring where we expected it to from restriction data.

Fig 16.1

Heteroduplex of pWB1 & pRW8A

A



0.5 μ m
1 kb

B

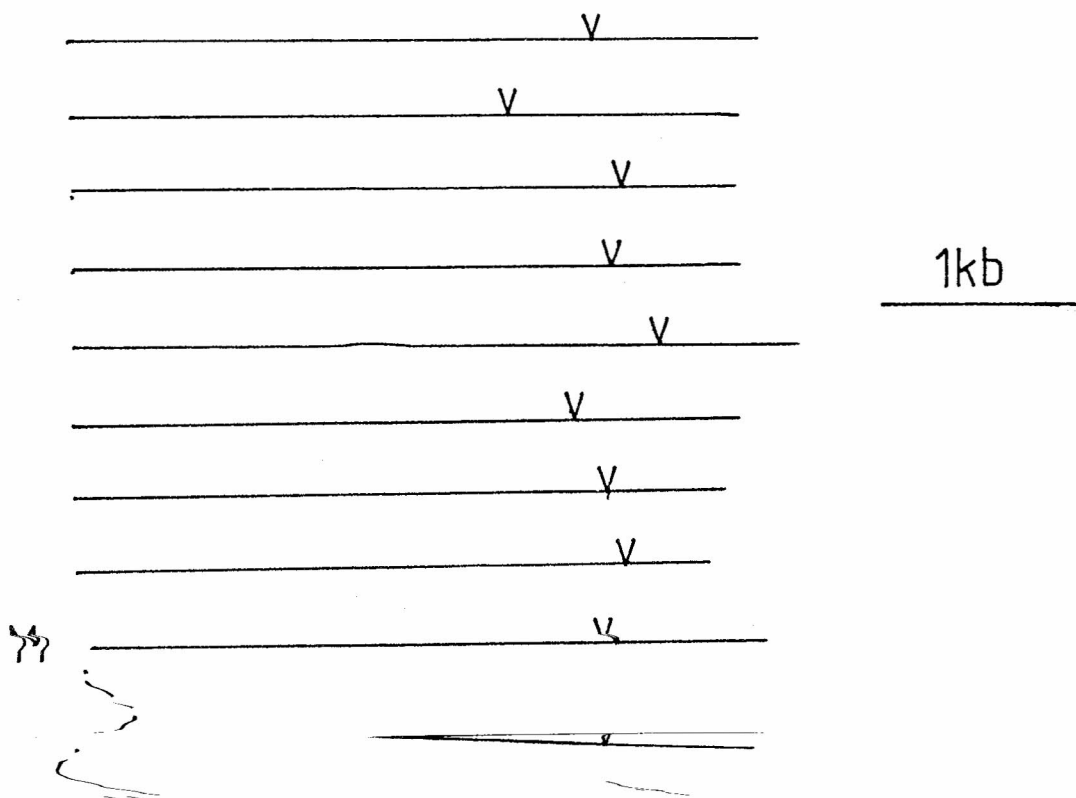


Fig 16.1 Heteroduplex analysis of Xlh8 and Xbh1 (see text for DNA species used and conditions for this experiment.) A) Photograph of one molecule and tracing of the photograph. ds= double-stranded DNA, ss=single stranded, i=end of the insert of pWB2/pRW8A , l=loop.

B) Analysis of eight molecules. 'V' = location of the loop. M= plot of the mean position of the loop structure.

16. Heteroduplex analysis

Although not conclusive, this is good supportive evidence for the deletion postulated in section 13.3.1.

It is notable that in two of the eight molecules examined in detail, a region on one or either side of the loop (it is not possible to say which) was also single stranded. It is probable that this is a spreading artefact. However it is possible that this actually points to a limited region in which the homology between Xlh8 and Xbhl is not as good as elsewhere, causing some melting of the hybrid beside the deletion in Xlh8. A shorter region of 6bp of complete mismatch has been found in the Hl gene region (vide supra section 15.1), which lends credence to this idea.

16.3.2 'Snapback structures' in Xbhl.

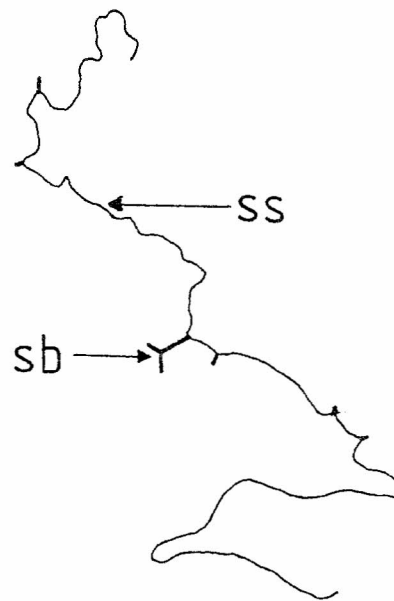
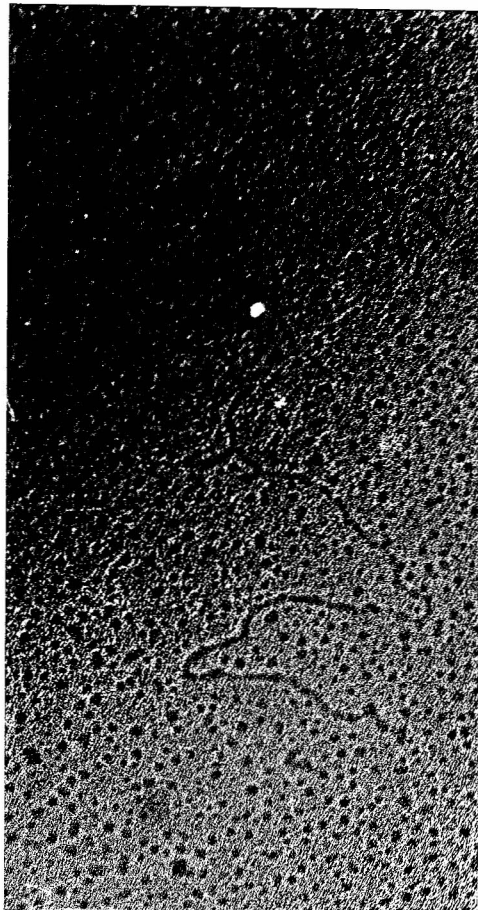
Fig 16.2 shows a Sal-digested pWB1 molecule which has been denatured and allowed to 'snap-back', together with a schematic plot of the structures in 9 such molecules. Some of these are clearly not full-length: I have assumed on these that a single nick has removed one or other end of the molecule. The molecules are lined up by their ends. Closed, denatured circles of pWB1 provided an internal control from which the sizes of these molecules could be calculated. All snap-back structures or potential structures have been plotted. Only three regions show snapback structures in more than two molecules, so I suggest that the apparent structures elsewhere are due to non-specific association of DNA, random folding of the DNA molecules across themselves or aggregates of stain or cytochrome-C imitating a short stretch of DNA.

Two regions show short duplex stretches in 4 out of 9 molecules. It is not implausible that this is chance association of artefacts.

Fig 16.2

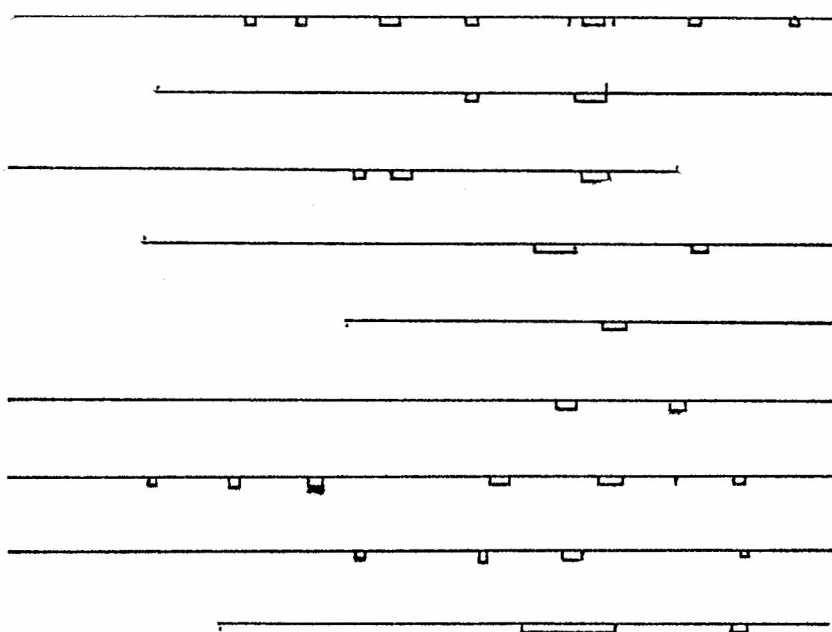
Snapbacks in pWB1 : 1

A



0.5 μ m
2 kb

B



1kb

M



Fig 16.2 Snapback analysis of SalI digested pWB1. A) One molecule and a tracing of it. sp= snapback structure (double stranded) ss=single stranded DNA.

B) Analysis of nine such molecules. Boxed regions are lengths of DNA involved in snapbacks or potential snapbacks. M= proposed 'average' structure, indicating the three regions referred to in the text as having more than a random number of snapback structures.

It is notable that one aligns approximately with the repetitive element IR1. Whether this is truly significant must await further study.

The region ①1/3 from one end of the molecule contains duplex structures in all 9 molecules. These structures are not simple, however, in that at least three of the molecules show two distinct snapbacks in this region - one of these, shown in Fig 16.2, shows apparent complex internal structure in one of them. Again, how much of this is due to chance association of DNA and how much to genuine hybridisation I cannot say. However it is probable that there are at least two positions in this region which can form duplex, snapback structures, as not only are two structures seen in some molecules, but also the distribution of all the snapback structures seen is incompatible with a single locus within the limits of accuracy of this experiment. Thus there are probably two inverted repeat sequences in this region of the molecule.

Snapback analysis of a Sall digest shows where the snapback structures are relative to the Sall site in pWB1, but not their absolute orientation. So I also studied snapback structures in KpnI, BamHI and XbaI digests of pWB1. An example of the first is shown in Fig 16.4. The same structures as we observed in Fig 16.2 can be seen and in a similar relative position: thus they must lie approximately equidistant from the Sall and KpnI sites. This places this snapback structure in the repetitive element IR2.

The BamHI digest yielded too few DNA molecules for analysis. The XbaI digest, done on a previous preparation of pWB1, was extremely fragmented due to nicking of the original DNA. However in about half the molecules where snapback structures were seen they were very near

Fig 16.3

Snapbacks : 2

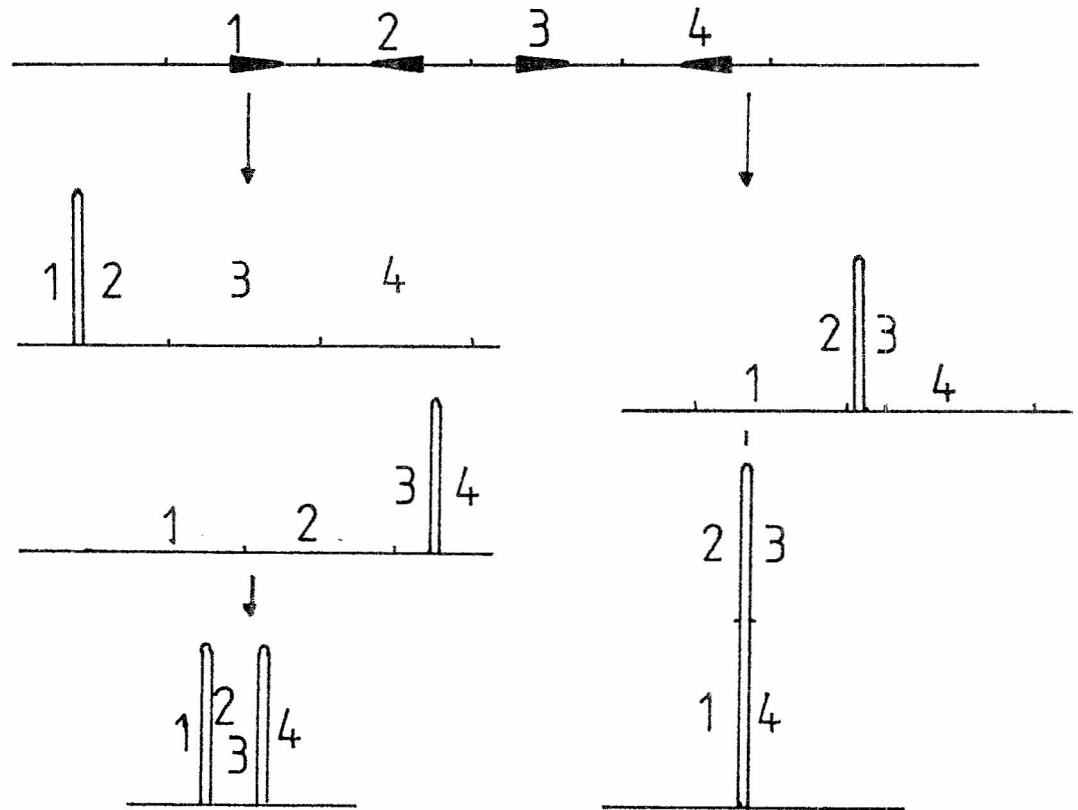


Fig 16.3 Illustration of how 4 copies of an element in an alternating array can form a variety of snapback structures by self-hybridising in different ways.

Fig 16.4

Snapbacks : 3

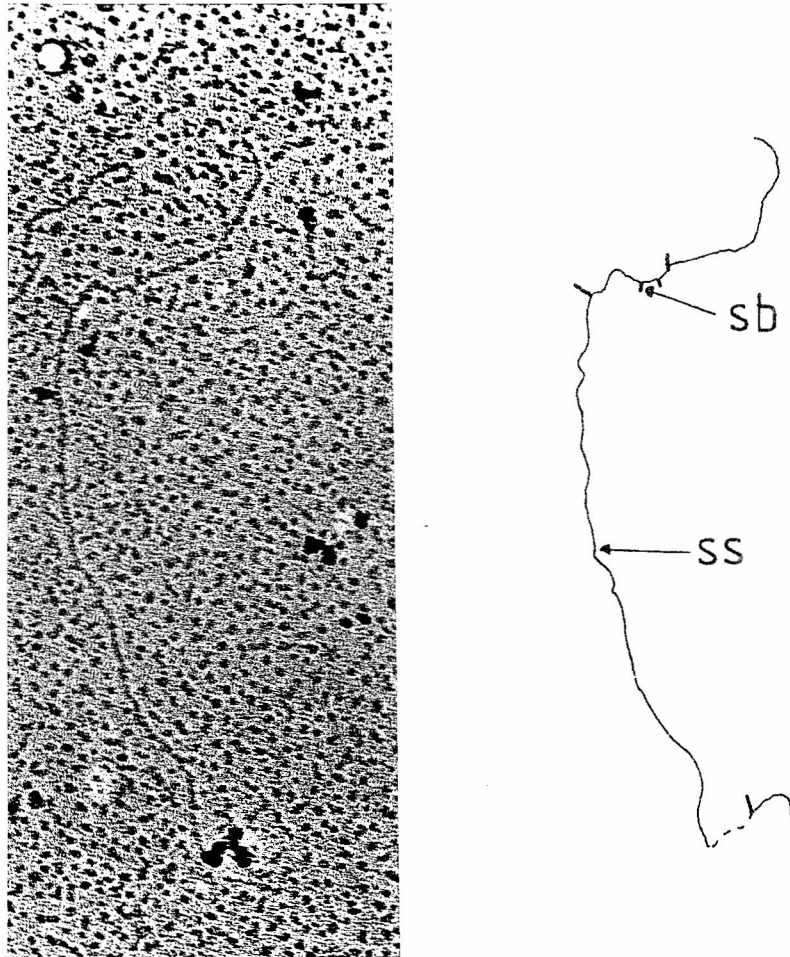


Fig 16.4 Analysis of a molecule of KpnI-digested pWB1. Symbols as in Fig 16.2

16. Heteroduplex analysis

one end of a DNA molecule. This strongly suggests that they derive from DNA near an XbaI site, again fitting with our location in IR2. Notably, fewer of the XbaI-digested molecules showed two snapback structures than was true of the Sall-digested ones. This might indicate that XbaI cuts in the middle of a sequence which gives rise to one of the snapback structures seen in a Sall digest. However it might also reflect slight differences in the conditions of hybridisation and spreading between experiments. The logic of location of these elements is shown in Fig 16.5 .

16.4 Physical structures in Xbhl.

This section presents the results of a preliminary analysis only. However the following points may be reasonably confidently made.

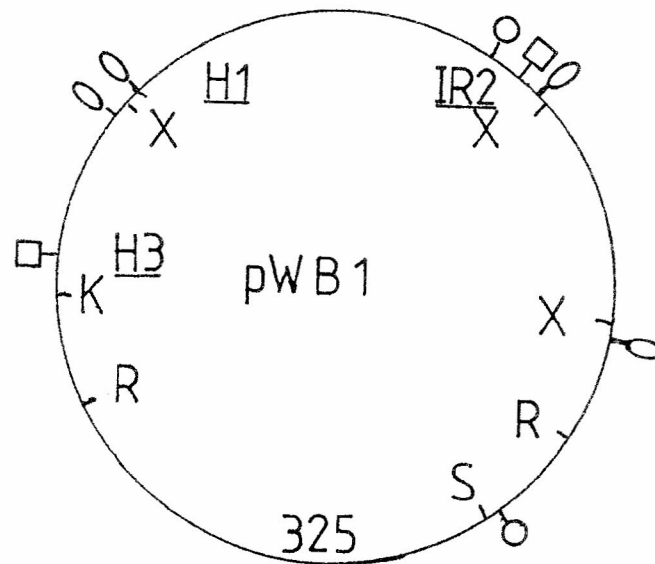
Xlh8 has a small deletion relative to Xbhl, @3.4 kb from the left-hand end as defined by restriction mapping data.

Xbhl has at least one inverted repeat sequence in the repetitive DNA region IR2, and probably two. We can therefore say that IR2 is not as simple, homogenous block of sequence, but rather probably contains more than one element. We reached the same conclusion from another line of reasoning previously (section 10.3.2): the discussion of the implications of this for IR2's structure and copy-number in the genome will be taken up shortly.

It is interesting that the deletion in Xlh8 occurs next to IR2. It is not in IR2 itself. However it has been noticed in other contexts that intermediate repetitive elements are associated with increased polymorphism in other organisms (Mirault et al 1979, Young 1979, Calos and Miller 1980, Potter 1982, Thompson and Woodruff 1982, Calbretta

Fig 16.5

Analysis of snapbacks



□	sb	(SalI)
○	sb	(KpnI)
⊖	sb	(XbaI)

Fig 16.5 Plot of the points on pWB1 where the snapbacks seen under the e.m. could be. Plotted are all points that are a) about 5kb from a SalI site, b) about 5kb from a Kpn site, c) near the end of a large XbaI fragment. Only one site fulfils all three conditions. Also plotted are H3 and H1 genes and IR2. K=KpnI, S=SalI, X=XbaI, R=EcoRI.

16. Heteroduplex analysis

1982, Bell, Selby and Rutter 1982), which itself has been suggested to be a product of these elements' transposition (see section 2.7 above). This would appear to be another example of the general rule that intermediate repetitive DNA elements can in some way 'induce' variation in the DNA adjacent to them.

It is possible that the IRI region contains a short inverted repeat . This possibility is based on the close coincidence of snapback structures in this region in 4/9 of the molecules examined. The probability of 4 out of 33 actual or potential snapback structures coinciding in 4/9 molecules is @6%. This is low enough to suggest that sequence studies of IRI might reveal interesting internal structure. Another region @500bp upstream of the H1 gene shows snapbacks in 4/9 molecules too - the same 6% probability applies here. However this region does not coincide with any repetitive DNA element, nor with any obvious duplications in the restriction map.

Discussion

17. Discussion.

"In a moment I've seen what has hithertoo been
Enveloped in absolute Mystery,
So I'll give you at large, at no extra charge,
A lesson in natural history."

In this section I will summarise and integrate the results described in sections 6 to 16 above, and then see to what extent these results may be integrated into the current knowledge on Xenopus histone genes in particular, and the eukaryotic genome in general.

17.1 Structure of Xbhl.

The structure of Xbhl is summarised in Fig 17.1 . I will discuss the evidence for this structure briefly.

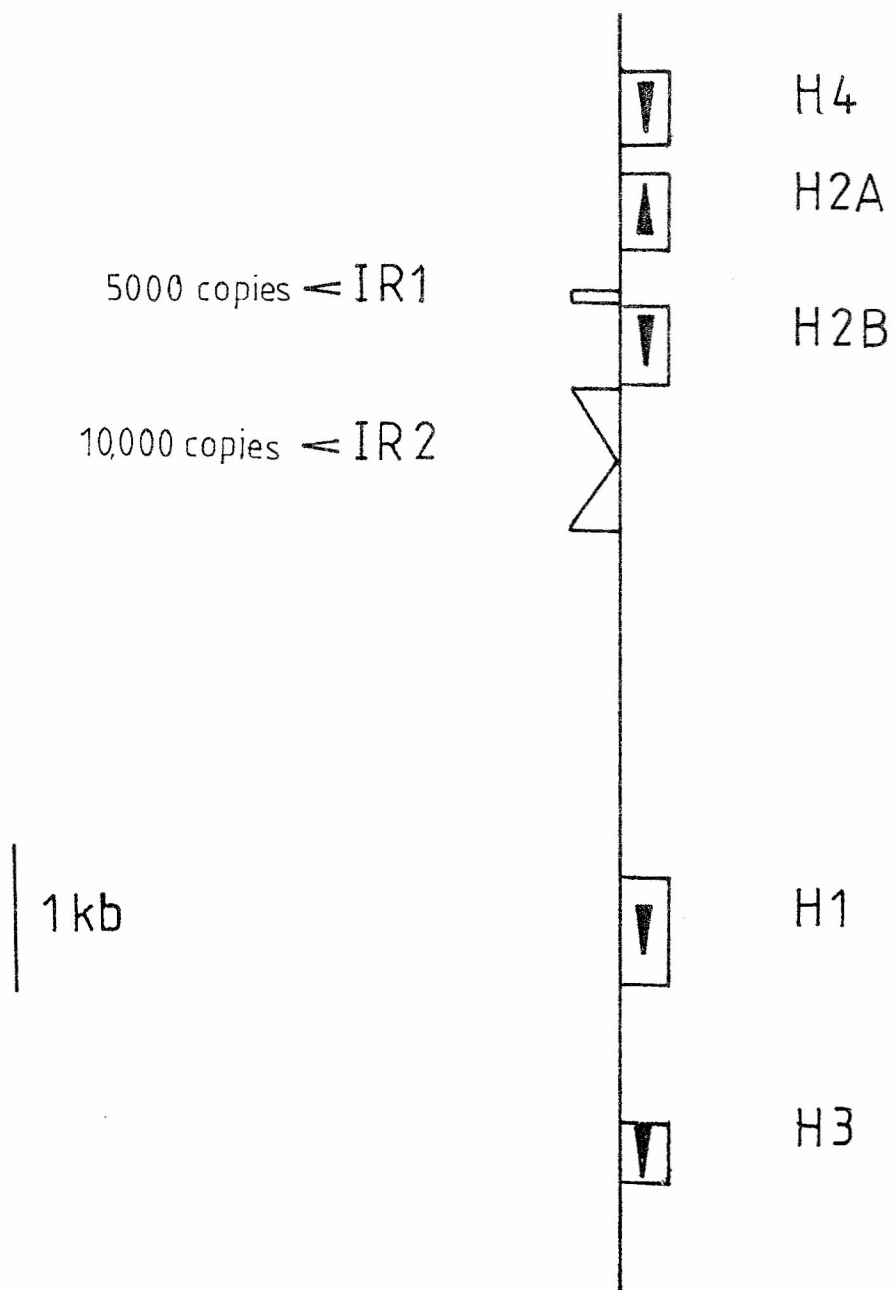
17.1.1 H3 and H4 genes.

I have relatively little information on the H4 gene of Xbhl apart from its location. This may be established fairly precisely by analogy with Xlh8, where the H4 gene region has been partially sequenced, and by Southern blot mapping with the homologous H4 cDNA clone pcXlh4W1. The restriction map of this gene region is identical to the analagous region in Xlh8, although at least one neighbouring restriction site is different between these clones (section 13).

The H4 gene in Xbhl is complete, is not a pseudogene and contains at least some of the 5' and 3' non-coding regions which control the correct initiation and termination of transcription. This is shown by the gene's expression on microinjection into

Fig 17.1

Xbh1



oocytes, as described in section 14. Interestingly, despite being the gene nearest to the vector DNA in these recombinants, the expression of the H4 gene does not seem to be affected by the presence of lambda or pBR325 DNA sequences. Thus it is probable that the H4 gene is being transcribed off its own promoter.

The H3 gene is also transcribed when the clone is microinjected into oocytes, again apparently off its own promoter. It should be noticed that the H3 gene extremely active in this assay, and directs the synthesis of as much protein as the other four cloned genes put together. Whether this is due to extremely efficient transcription or to more efficient translation of the mRNA produced has not been investigated. The microinjection studies show that the H3 gene of Xbhl could be used as a source of a strong promoter for functional studies on this and other eukaryotic genes, using the microinjection assay for function, and so will be of considerable use in future studies. This is being followed up (A.Wilson and R.W.Old per com.)

The H3 gene is a 'variant' in that it does not code for a protein identical to the Calf thymus protein (Section 15). The protein coded by the H3 gene cloned from X.laevis by Moorman et al (1980) (Moorman et al 1981) is also different from both that coded by Xbhl and the calf thymus protein sequence. We noted in the introduction that some sequence variant H3 proteins are known (Isenberg 1979, Brandt, Patterson and van Holt 1980). Whether the 'variant' H3 from Xbhl is functionally different from that of Calf thymus, or that coded by the H3 gene described by Moorman et al (1981), is difficult to say. However we might note that the X.laevis and Xbhl sequences have a cysteine replaced by a serine relative to calf thymus H3 at position 96, which might point to different interactions within the nucleosome or to an altered

17. Discussion

pattern of modification of these two proteins. Langan, Rall and Cole (1971) have documented an H1 protein where a serine-to-alanine change is known to occur at a phosphorylation site, and hence to cause an altered pattern of phosphorylation on this protein. Raman spectroscopy shows that it is this cysteine, rather than Cys 110, which forms the disulphide bridge between H3 molecules in calf thymus protein (McGhee and Felsenfeld 1980), again suggesting that alteration at this position could be functionally significant.

17.1.2 The H1 gene.

The sequence data on the H1 gene is incomplete at the time of writing. As well as confirming that Xbhl contains an H1 gene (it will be recalled that the Southern blot data presented in Section 9 was not totally convincing on this point) this data accurately locates the gene. Sequence information on H1 genes and H1 proteins is harder to analyse than that for H3 genes, because, as we saw in Section 1, the H1s exist as many sequence variants both within and between species, some of which have functional significance, so it is seldom certain that differences or similarities between the sequences of two H1 proteins are due to functional constraints, evolutionary relationships or neutral polymorphism.

However we have two gene sequences from the Xenopinae with which to compare the Xbhl gene. The H1 in Xlh8 contains a number of point differences to that in Xbhl, several of which actually alter the protein coded by the gene. This is curious, in view of the observed similarity of these cloned gene clusters, which might suggest that they encode H1 proteins evolutionarily and functionally related. It is possible that the sequence of some H1 proteins might be free to undergo changes within certain limits

without affecting the fitness of the protein for its role. A similar flexibility has been suggested for the actin proteins (Fornwald et al 1982) on the basis of gene cloning and sequencing data.

Another H1 gene has been cloned from X.laevis in an unrelated histone gene cluster (Xlh19). This is only a partial H1 gene, the sequence coding for the initial 80 amino acids being absent from the clone. Some of the remaining gene has been sequenced, and found to code for an H1 protein which is homologous to that of Xbhl in 16/21 amino acids in the conserved region. This is a better homology than the calf or sea urchin H1 genes show (see Section 15). Thus we find that the H1 genes from different gene clusters in the Xenopinae are more closely related than the H1 proteins from Xenopus and calf or sea urchin, as we would expect. It is interesting that H1 proteins from related gene clusters in these two species (Xbhl and Xlh8) are more closely related than two proteins from different clusters within the same species (Xlh8 and Xlh19). This might be related to the Xlh19 'gene's' probable pseudogene nature. However the lack of stop codons in the reading frame (Proudfoot 1980) and the fact that the third bases in codons are more diverged between these genes than other bases, suggest that there is a continuing coding restraint on their sequence divergence, and hence that these genes have diverged while still under the selective pressure of being required to code an H1 protein. Such variation between H1 genes of a species would be expected if the sequences of variant H1s are functionally significant, and hence evolutionarily constrained, as I suggested in section 1.3 .

Of the 16 amino acids conserved between Xbhl and Xlh19, 10 are coded by the same codon. This is a higher frequency than would be expected by chance, but not nearly as striking statistically as the

conservation of the third bases in the codons of the H3 genes noted in section 15.3. It does suggest, however, that an analogous conservation of non-coding DNA sequence is occurring in H1 as well as in H3 and H4 genes to limit the sequence divergence of the H1 genes between X.laevis and X.borealis.

The protein coded by the H1 gene in Xbhl may be examined by microinjecting the gene into oocytes, and examining the protein product (Section 14). It migrates in the same region on SDS-PAGE and Acid/Urea-PAGE as authentic oocyte and erythrocyte H1s from Xenopus, and in particular very similarly to the oocyte H1 of X.borealis.

The H1 gene is expressed strongly on microinjection. It may be significant that the other strongly expressed gene, H3, is downstream of the H1 gene, and transcribed in the same direction as it. We could envisage, for example, that much of the H3 gene transcription we observe in microinjection studies is actually due to run-on of the RNA polymerase from the H1 gene, the transcription of the latter failing to terminate correctly. Alternatively, some 'enhancer' sequence near both genes might enhance the transcription of H3 and H1, but be unable to similarly affect that of the more distant H2B, H2A and H4 genes. A third, simpler alternative is that the H1 and H3 genes have promoters which are very effective in the oocyte. I favour this last alternative as being the most straightforward, and also being in line with evidence mentioned below for a bipartite origin for Xbhl.

17.1.3 The H2A and H2B genes.

The genes which revealed the most interesting structure so far

are the H2A and H2B genes. Microinjection evidence may be summarised thus (see also Fig 14.3): H2B - no protein produced. H2A - a methionine-containing H2A is synthesised at low levels from both plasmid recombinants, and another, methionine-free H2A, at high levels when the Xbhl insert is placed in one orientation in pBR325 only.

This raises a number of questions. The gene mapping data presented in Section 9 gives no indication about whether a given region contains one or more than one gene, merely that it contains at least one. It also gives no indication as to whether that gene/those genes are functional or are pseudogenes, merely that they are not so diverged from the probe that they no longer hybridise to it. The gene mapping data in Section 9 is reasonably conclusive on what it does say, however: we can be sure that the H2A- and H2B-homologous sequences of Xbhl are confined to the regions shown in Fig 17.1. We can also be sure that, if there are two H2A genes in Xbhl, then they are both transcribed off the same strand. In view of the apparent multiplicity of H2A proteins synthesised from Xbhl, we should therefore ask whether other data give us clues about the internal organisation of the regions marked as H2A and H2B genes in Fig 17.1.

Two further points. H2A usually does not contain methionine, although H2B usually does - the low level H2A-like protein detected on microinjection of either plasmid into oocytes does contain methionine. Secondly, the restriction map of the H2B gene of Xbhl differs significantly from those of Moorman et al (1980) and Zernik et al (1980) from X.laevis, Stephenson, Erba and Gall (1981) from Notophthalmus and Turner and Woodland (1982b) from X.borealis.

These genes all contain either AvaI and HindIII sites, or BamHI sites in the coding region (when sites for these enzymes have been mapped): sites for these enzymes are permitted in several positions by the calf H2B protein sequence (Isenberg 1979). This is not true of all clones isolated from X.laevis (T.Aldridge and R.W.Old per com), and is not true of the H2B gene of Xbhl, which has sites for none of these enzymes. The restriction map of the H2B genes of the major gene cluster of X.borealis (Turner and Woodland 1982b) and Xbhl differ in other restriction sites as well. Thus it is likely that the sequence of the H2B gene of Xbhl differs significantly from that of many other amphibian H2B genes.

It may also be noted that the H2A protein made by Xlh8 migrates differently on SDS PAGE to either H2A-like protein made by Xbhl. The H2A gene regions of these two clones differ by at least one restriction site (section 13).

Thus we have several options. Either there are two H2A genes in Xbhl, which must therefore be closely linked and transcribed in the same direction, or one of the 'H2A' proteins produced by microinjection of the clone into oocytes is not really an H2A. Either the H2B gene is a variant, or it is a pseudogene. Or the microinjection result is artefactual. Until at least some of these ambiguities have been eliminated we cannot resolve the functional arrangement of the H2A/H2B gene region.

H2A and H2B are adjacent and divergently transcribed in Xbhl, as in a number of other histone gene clusters (discussed in section 2.3.6). In section 2.3.6 I hypothesised that this might be concerned with a requirement for coordinate control of H2A and H2B genes. Unfortunately the transcription of the H2A and H2B genes in

microinjected oocytes is rather excentrically controlled, and so we cannot confirm or deny this hypothesis from these studies.

One fascinating aspect of the microinjection study is that an H2A- like protein is strongly expressed in oocytes injected with pWB2, but is not expressed in those injected with pWB1. The inserts in these recombinants are meant to be the same, although experiments to rule out a small alteration in one of the two during the re-cloning of the original EcoRI fragment into pBR325 have not been performed. The possibility that this effect is artefactual is too strong to allow us to base much speculation on its validity. However if it is reproduceable, pointing to the 'enhancement' of transcription of H2A by distant, vector sequences, then ,despite this effect having little in common with any controls which might operate on the Xbhl sequence in the genome, it is still of great interest as being a selective, long-range transcription enhancing effect, an almost unique transcriptional control in eukaryotic molecular genetics (Dylan and Tjian 1982).

17.1.4 Snapback elements of Xbhl.

In sections 8 and 16 I indicated several regions of Xbhl which had potential inverted repeats, or 'snapback' structures, in them. The only region which has consistantly shown a snapback structure in electron microscopic studies is the region 3.5kb from the lefthand end of the clone, the location of the intermediate repetitive element IR2. The two other potential snapbacks detected by electron microscopic studies (section 16) are too small and statistically insignificant to be more than pointers to be used to direct further study, and the potential inverted repeat around the H3 gene suggested by restriction mapping studies (section 8.4) was not

detected in the electron microscopic studies at all. This is not to say that these regions do not contain inverted repeat structures, just that this study has not convincingly demonstrated them.

The inverted repeat found in IR2 is seen in all electron microscope studies and is also suggested by the repetition of the pair of HincII-AvaI site pairs bordering IR2 (Section 8 and 16), and by the symmetric pattern of genomic repetition found in this region in section 10. However electron microscopic studies indicate that this is not a single inverted repeat, but instead probably consists of a more complex region of mutually inverted sequences. I will return to this in a moment.

17.1.5 Intermediate repetitive sequences of Xbhl.

Two regions of Xbhl are homologous to families of repeated DNAs in the genome of X.borealis. A small region of @200bp between H2A and H2B genes (IR1) is present in about 3000 copies per X.borealis genome (assayed at 60°C, 2xSSC). This sequence might contain an inverted repeat.

A much larger region designated IR2 lies to the 3' end of the H2B gene. The genomic representation experiment presented in section 10 suggests that this is composed of at least three elements - the experiment sets no upper limit on its complexity, of course. The end two blocks are related to genomic sequences of similar number and divergence from the cloned probes used in section 10. Electron microscopic evidence suggests this region contains two or four blocks of sequence, all related to each other. This is summarised in Fig 17.2. Fig 17.2 also presents the simplest structure for IR2 which is consistent with this data. In this model there are two

Fig 17.2

IR2

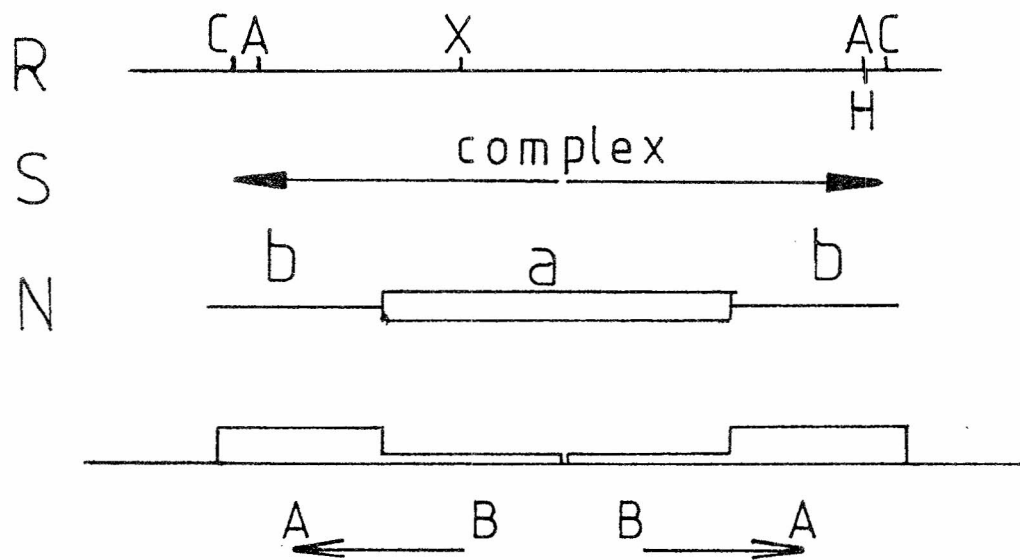


Fig 17.2 Analysis of IR2. Structure as revealed by restriction mapping ('R'), snapback analysis ('S'), and genomic representation analysis ('N') are plotted together with the minimum structure needed to explain these results. See text.

elements, each composed of two sequence blocks (A and B) which are partly homologous to each other and mutually inverted. The two A-B elements are mutually inverted as well. The 'outside' end of the A blocks contain a HincII site and an Aval site in both cases, and a sequence GGGAAGGG which is found in other Xenopus repetitive DNAs (section 10 and Spohr, Reith and Sures 1981). In this respect it is like the Alu-like elements, which also share a short region of homology with other repeated DNA families. Each block is about 500bp long, and we can deduce from the genomic representation figures that the A-B monomer is present in about 10000 copies per haploid genome. It is possible to estimate the divergence between the two A-B units, and between the A and B halves of each, on the basis of the electron microscopic data and restriction map of this region. : these suggest that the two A elements are more closely related than the two central B elements.

I cannot say exactly where IR2 ends from this data, specifically at the left-hand end where it abuts the H2B gene region. It is possible that it overlaps the H2B gene itself, so that the latter is a pseudogene (Proudfoot 1980).

It is also notable that the right-hand end of IR2 is where there is a deletion in Xlh8 relative to Xbhl. Intermediate repetitive DNA has been strongly implicated in the generation of polymorphisms in eukaryotes (reviewed in Section 2.7), and so it is tempting to see a causal link between the deletion in Xlh8 and IR2.

IR2 is distributed throughout the genome (sections 10 and 11), and is not obviously clustered or in a satellite-like array. It is probable that some regions of IR2 are weakly homologous to histone mRNAs (sections 9,10 and 12). The non-translated regions of some

17. Discussion

mRNAs have been found to be conserved between similar genes (vide infra section 17.3); if IR2 is similar to non-coding regions of histone mRNA then this familial similarity would explain why it hybrid releases histone mRNA but does not hybridise to the bulk of cellular RNA, and why some IR2-like elements hybridise weakly to pcX1H4W1. Why they apparently do not hybridise to pcX1H3W1 is not clear. The relevance of this homology to the evolution of Xbhl will be taken up below.

In view of its pattern of dispersal around the genome and failure to hybridise effectively to most cellular mRNAs, it is unlikely that IR2 is a control or 'sensor' element of the Britten and Davidson type (Britten and Davidson 1969, Davidson and Britten 1979). The association with histone-like sequences would be more impressive evidence for a 'sensor' role for IR2 if IR2 did not outnumber all the histone genes 20:1 in X.borealis (Turner and Woodland 1982b and Section 10). It would seem more reasonable to suppose that IR2 is 'selfish' or 'useless' DNA in the style discussed in section 2.7. I did not detect IR2-homologues on circular extrachromosomal DNA (section 11), but the experiment described in that section is not sufficiently sensitive to detect circular DNAs present in less than 1-5 copies per haploid chromosome set (ie 2-10 copies per diploid cell), and so does not rule out IR2's presence on a low frequency population of circular molecules.

This completes the description of Xbhl and its components. How does this gene cluster fit in with other gene clusters in the Xenopinae?

17.2 Xbhl and other histone gene clusters.17.2.1 Xbhl and the X.borealis major histone gene cluster.

The similarities between the X.borealis major histone gene cluster as described by Turner and Woodland (1982b) and Xbhl are summarised in Fig 17.3 from data in sections 8,9,10 and 14. We can note several observations that suggest that Xbhl is a chimera of parts of two other gene clusters.

Firstly, as we noted before, the restriction map of Xbhl's left third matches well with that of the centre of the major cluster. The likelihood of obtaining such a good match between any two DNA sequences is extremely low.

Secondly, the genes are positioned in the same order and, allowing one insertion in Xbhl relative to the major cluster, in the same relative positions in the overlapping regions of these two clusters. This is not quite so conclusive: the gene map data is not so precise as that of the restriction map, as the former has been derived from Southern blotting experiments and hence can only define genes in terms of outside limits of regions of homology, rather than as exact coding or transcribed regions. However, the alignment is suggestive.

Thirdly expression studies show the left-hand three genes of Xbhl to be poorly expressed in oocytes, while the right-hand two are strongly expressed (excluding the recombinant-specific expression of H2A, which must be due to the influence of vector sequences). The

Fig 17.3

Xbh1 and the 'major cluster'

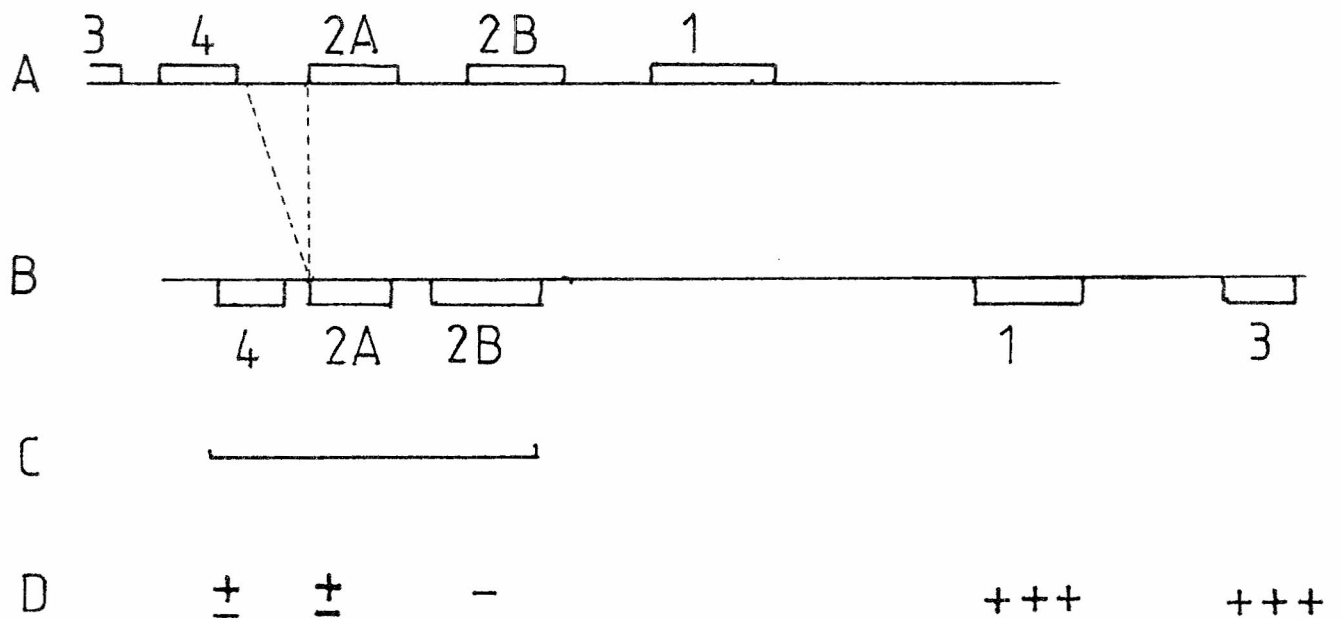


Fig 17.3 Similarity between Xbh1 and the X. borealis major cluster. A) Major cluster after Turner and Woodland (1982b). B) Xbh1 C) region of partial restriction site homology, D) relative rate of protein production from the genes in Xbh1 on microinjection (discounting the vector effects seen in pWB2).

most obvious explanation for this is that the promoters on the H1 and H3 genes function well in oocytes while those on the H4, and H2B genes function poorly (section 14), and the H2A gene appears to function poorly also in three out of four microinjections: as I mentioned above the interpretation of this result is still uncertain. This is another division of the genes into two classes, and accords with the physical division into left-hand and right-hand sections of the clone. What this means for the in vivo expression of these genes we cannot say, as microinjection of relatively enormous amounts of cloned DNA into oocytes is likely to provoke a different response from the systems controlling transcription than they would give to a single copy of this sequence integrated into a chromosome.

Lastly, the end of the homology between Xbhl and the major cluster is marked by IR2, an intermediate repetitive element known to be dispersed throughout the genome and probably to have weak homology with core histone mRNAs and which is suggestively associated with a 'polymorphism' between Xlh8 and Xbhl.

The obvious evolutionary interpretation of this is that Xbhl is the product of recombination between the major cluster of X.borealis and some other histone gene cluster. This would be mediated by IR2, through 'illegitimate recombination' (either because both clusters contain sequences homologous to IR2 or because IR2 is partly homologous to histone genes), or by transposition. These various options are illustrated in Fig 17.4 .

There is no evidence as to what the structure of the second gene cluster might be, other than that part of it presumably must have been similar to the right-hand half of Xbhl at the time of Xbhl's

Fig 17.4

Formation of Xbh1

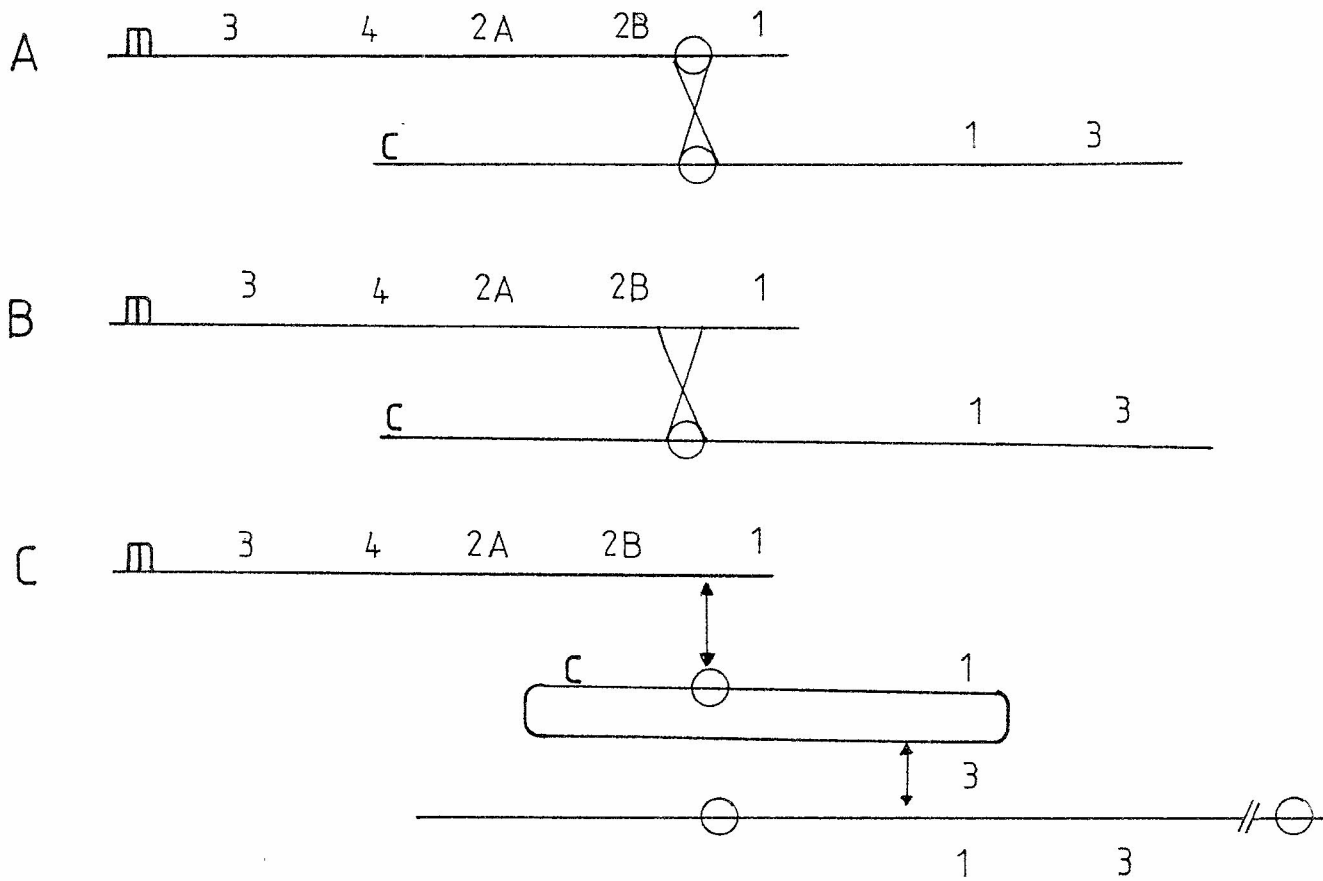


Fig 17.4 Possible mechanisms involving IR2 in the formation of Xbh1 from the *X.borealis* major cluster ('m') and another histone gene cluster ('c'). A) Recombination between IR2 elements B) Recombination between IR2 and the 3' end of the H2B gene, C) Transposition.

formation. However as the only gene clusters from X.borealis to have been characterised are Xbhl and the major cluster, this is unsurprising.

Note that here we are assuming that the similarity of structure between parts of Xbhl and the major cluster is not due to convergent evolution. If selection were to act not only on the transcribed sequences and such sequences as the TATA and AATAAA 'boxes' near them, but also on all the intergenic spacer in a gene cluster, then we could argue that the similarity between these two clusters is due to similar functional constraints upon them and might not reflect anything concerning their evolutionary relationship. That such sequence selection might occur in these gene clusters is suggested below. However the observation that only some of Xbhl is related to the major cluster suggests that this similarity reflects evolutionary changes as much as functional constraints, which would be expected to cause the major cluster and Xbhl to converge to the same degree along their whole length.

Xbhl is a minor species of histone gene cluster in the genome of X.borealis, present in 1-3 copies per haploid chromosome set.

17.2.2 The X.laevis 'major cluster'.

As I noted in section 2, a substantial number of histone gene clones from X.laevis have similar structures which are also consistent with the 'major' histone gene cluster structure proposed by van Dongen (1982) on the basis of Southern blotting experiments on genomic DNA (Fig 2.7 and 2.8). The clones of van Dongen (1982) and Zernik et al (1981), which contain H3-H4 gene pairs at both ends, are consistent with this 'major cluster' being tandemly repeated at least

once in the genome, with a repeat length of about 13kb. Zernik et al (1981) note that the H1 genes on their clone are homologous to oocyte H1 mRNAs, and suggest that this tandemly repeated structure codes the stored histones and histone messages in the oocyte. Its copy number of about 50% of the histone genes in X.laevis (van Dongen et al 1981, van Dongen 1982, Zernik et al 1981, Turner and Woodland 1982b), ie 40-50 copies per haploid set (Hilder et al 1981, Turner and Woodland 1982b) is consistent with the minimum number of genes calculated to be required to provide a template for histone mRNA synthesis in the oocyte (Woodland and Wilt 1980a,b, Woodland 1980). Thus it is attractive to think that these cloned genes do represent tandemly repeated clusters of histone genes which have been reiterated to meet the demands of oogenic transcription, in the same way as the embryonic gene clusters have in sea urchins (Kedes and Birnstiel 1971, Hieter et al 1979, Childs, Maxson and Kedes 1979, Spinelli et al 1980) and Drosophila (Lifton et al 1977).

The X.borealis major cluster might be thought of as an analogous tandemly repeated array of embryonic genes - certainly the possibility that they are tandemly repeated remains (Turner and Woodland 1982b). However the failure of an ovarian H3 cDNA clone to hybridise effectively with this cluster (Turner and Woodland 1982b) must cast some doubt on this. Among other possible interpretations is that this shows that the major cluster in X.borealis does not code ovary H3 mRNA.

17.2.3 The circular histone gene map of Xenopus laevis.

The major clusters in X.laevis and X.borealis are not the same. However their gene orders are cyclic permutations of each

other. If the gene order from the X.laevis major cluster is written in a circle a structure similar to many of the cloned DNAs shown in Fig 2.7 may be obtained by breaking the circle at the appropriate point (Fig 17.5). Exceptions are Xlh19 (H4 - H1) , and Xlh25 with the gene order H4 - H2B - H2A - H3 - H4 and the closely related Xlh24 (H4 - H2B - H2A - H3) of R.W.Old (unpub), and Xlh4 of Zernik et al (1981) which contains genes for H4, H2B and H1 only. This last clone contains a gene homologous to H1a, whereas the 'major cluster' clone contains an H1 more similar in sequence to the H1b as judged by its hybridisation behaviour towards the mRNAs of H1a and H1b. Both H1a and H1b are found in the oocyte (Flynn and Woodland 1980).

Thus histone gene clusters in the Xenopinae seem to fall into two classes: those clusters which have gene order which is a cyclic permutation of the sequence 3-4-2A-2B-1, with variable spacing between the genes and those which have a different order. Xbhl falls into the first class. If the X.laevis or X.borealis major cluster types are tandemly repeated, then the variants of this gene order such as Xbhl, with inter-gene spacings varying widely between each other, could be generated by combinations of spacer expansion and contraction (together with some 'point' mutation causing the restriction site polymorphism seen today). Indeed this conservation of a cyclically permutable gene order is evidence that at some time in the past a common ancestor of all these gene clusters was tandemly repeated.

Also of note is that the order of the core histones in the major Notophthalmus gene cluster is the same as that in the circular map in Fig 17.6 (Stephenson, Erba and Gall 1981a). The polarities of the genes differ, however, as does the position of the H1. If we postulated that these two gene clusters are evolutionary

Fig 17.5

The *X.laevis* 'circular map'

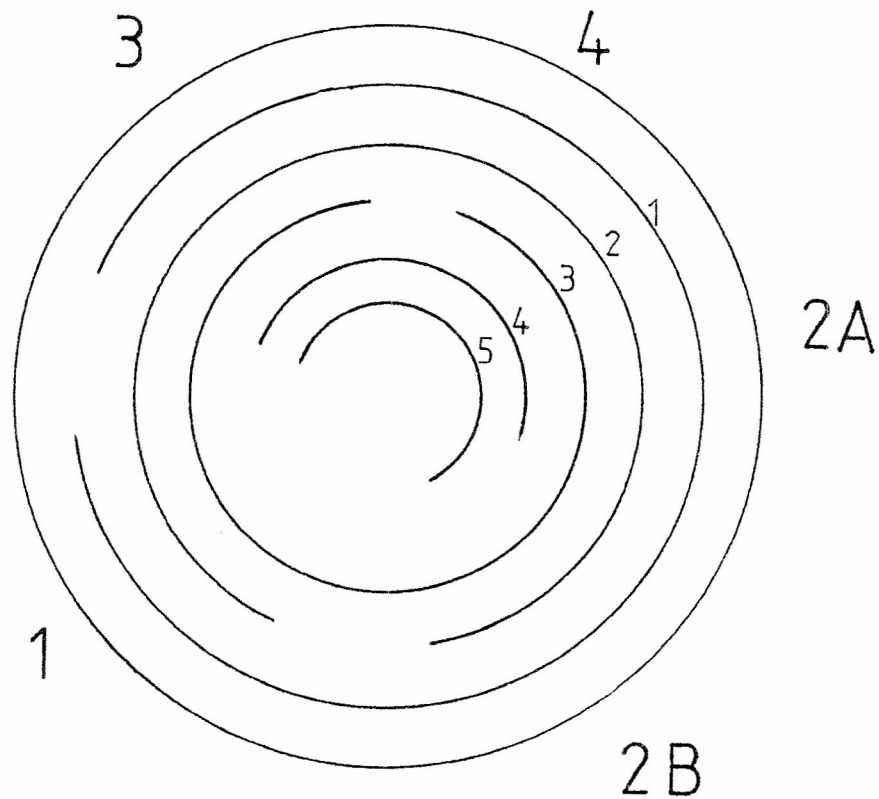


Fig 17.5 The 'circular map' of *Xenopus* histone gene clusters. The following gene clusters are mapped onto this circle as arcs extending around that segment of the circumference on which the genes are organised in the same order as in that clone. 1) The *X.borealis* major cluster. (Turner and Woodland 1982b) 2) Structure of several very similar cloned histone genes (Zernik et al 1981, van Dongen 1982,), 3) Xbh1 and Xlh8, 4) Xlh11 (R.W.Old per com), 5) Xlh7 (R.W.Old per com), X-hi-1 (Moorman et al 1980).

derivatives of a common ancestor, we would have to imagine a mechanism which allowed a gene's polarity to be reversed while not allowing it to move position relative to its neighbours. While recombination between flanking snapback elements could perform this, there is no evidence for such elements flanking the relevant histone genes in Notophthalmus or Xenopus.

I hypothesised above that Xbhl was the product of illegitimate recombination or transposition between the X.borealis major gene cluster and another, unidentified histone gene cluster. We could now postulate that this could be another repeat unit of the X.borealis major cluster, so that the recombination was between tandem repeats of a DNA sequence rather than between separate clusters. This cannot easily explain how the left 30% of Xbhl remains similar to the major repeat while the right 70% has diverged from it, however.

This picture of relationships between histone gene clusters suggests that recombination and spacer deletion and insertion occur quite rapidly in evolutionary terms. No-one has isolated a clone from X.laevis like the X.borealis major cluster, and what mapping and gene cloning data there is available on the X.borealis histone gene clusters shows that only one X.borealis gene cluster (Xbhl) is similar to an X.laevis one (Xlh8). This suggests that the 'half-life' for a given histone gene cluster's disruption by spacer alteration is less than the 8My since X.laevis and X.borealis diverged (Bisbee et al 1977). This would be in keeping with the polymorphism of histone gene cluster sequences seen between X.laevis individuals (Turner and Woodland 1982, van Dongen 1982), which could be seen as this cluster-disrupting tendency at work in the present population. Two observations are in rather direct

conflict with this. Firstly, Notophthalmus has the same core histone gene order as the circular X.laevis map, despite having diverged at least 100My ago. Secondly, Xbh1 (isolated from X.borealis) is extremely similar to Xlh2, Xlh8 and Xlh202 (isolated from X.laevis). The first observation could be assigned to chance similarity of gene orders, but this second could not. This is a rather significant observation, and one which has implications outside the field of Xenopus histone genes.

To appreciate this point it is necessary to return to the gene clusters from X.laevis which do not fit a circular permutation of the major cluster gene order. Why do they not fit? Two theories may be called 'Accident' and 'Design'. 'Accident' suggests that they are simply the product of more extreme cases of the action of whatever mechanism heterogenises the other gene clusters. Thus extreme amplification of all the spacers in the 'circular map' clusters would lead to genes which were essentially isolated in the genome. Recombination between or deletion of these could then produce clusters of any desired shape or size. In this case their variance has no significance. The 'Design' hypothesis states that these clusters hold genes whose requirements for the regulation of transcription differ from those in the 'major' circular map clusters, and that their different cluster structure reflects this, and is therefore concerned with the control of expression of their component genes. Thus they might code the proteins produced during S-phase, unlike the 'major cluster' genes which, if active during oogenesis, would not be co-regulated with DNA synthesis (Adamson and Woodland 1974). In this case, the variation in clusters is significant for their function.

Of course, we do not know which hypothesis is correct. However

17. Discussion

some strong arguments push us towards the second. One of these is the observation of the Xlh8-like family of gene clones.

17.2.4 Xlh8 and Xbhl - chance or conservation?

Xlh8 has two sister clones derived from X.laevis, and so must represent at least two loci in the frog. The gene library from which these clones were isolated was prepared from the DNA of one individual. All three Xlh8-like clones are different. They therefore cannot be alleles of a single locus. There is a family divergence of about 0.5% - 1% between them (Section 13). They are all diverged from Xbhl by a similar amount. A maximum value for the sequence divergence of Xbhl from its ancestor since the time X.borealis and X.laevis diverged from a common ancestor is therefore 1.5% (this in 8My (Bisbee et al 1977)) making a divergence rate of less than 0.15%/My. This is unusually low (Wilson, Carbon and White 1977, Jeffreys 1981): rates of 0.5% /My or greater are usually deduced for DNA sequence divergence rates, and greater values than this are usual for non-transcribed DNA. That this is the maximum divergence rate is shown by Templeton, Salle and Walbot (1981), who point out that the observed sequence difference between homologous genes in two species is the sum of their mutual divergence since the ancestors of the species ceased to interbreed added to the sequence difference between their ancestor genes at that time due to polymorphisms in the original, ancestral population. Thus, if the Xlh8-like sequences were already a multigene family 10My ago in the Xenopus laevis/borealis ancestor, as the presence of such a family in X.laevis today would suggest, then since that time Xbhl has essentially not diverged from the gene family at all, and so has experienced a net mutation rate of zero.

A similar case between extreme conservation of sequence of two histone gene clusters from different species has been found by

Busslinger et al (1982), who found that an H3 gene and 300bp of adjacent non-transcribed spacer were almost totally conserved between Strongylocentrotus drobachiensis and a minor histone gene clone from Psammechinus miliaris, two species which diverged about 65My ago according to paleontological data. The overall structure of the entire gene cluster (designated h19) was also extremely conserved between these two. Other histone gene clusters did not show such extreme conservation between these species.

We can readily think of four reasons for the observed similarity between Xlh8-like clones and Xbhl, and for the analagous sea urchin case. 1) Chance. Out of all the genes cloned, some pairs of clones will be more similar to each other than others. The extremes of similarity seen here render this explanation improbable. 2)Convergent evolution. 3)Conservation of an ancestral structure. 4) Horizontal gene transfer.

17.2.5 Eclectic evolution and horizontal gene transfer.

The fourth option deserves some further attention. Busslinger et al (1982) suggest that there might have been a transfer of the h19 gene cluster between P.miliaris and S.drobachiensis in the recent evolutionary past, much more recently than the divergence of the two species. This is known as 'horizontal gene transfer', IE transfer of genetic information between rather than along lines of descent.

Five other cases of potential horizontal gene transfer are known. We mentioned above the two cases of transmission of retroviruses between species (section 2.7.4): this is a rather special case as the virus genomes are specialised to be able to invade cells and overcome their normal regulatory machinery for controlling gene expression,

whereas we are looking for examples of gene transfer where a non-viral gene is integrated into the normal, functioning gene set in the recipient. Three cases bear on this.

Progenitor cryptocides, an endo-cellular bacterial pathogen which has been associated with several neoplastic diseases in man, has been shown to secrete a protein immunologically indistinguishable from and functionally similar to human chorionic gonadotrophin (Livingston and Alexander-Jackson 1970, Livingston and Livingston 1974), and is presumed to have obtained the genetic information to do this from its human victims. Martin and Fridovich (1981) found that the bioluminescent symbiont bacterium of the Ponyfish has a Cu/Zn super-oxide dismutase immunologically and functionally similar to that found in fish, but quite different from the SODs found in lower eukaryotes and in prokaryotes, which use a Mn reactive centre. Again, a gene transfer from Ponyfish to symbiont is inferred. Thirdly, and most convincingly, the leghaemoglobin gene of Soybean has been cloned (Hyldig-Nielsen et al 1982, Brisson and Verma 1982) and has been found to contain three introns, two in exactly the same position as in mammalian globin genes (Efstradiatis et al 1980). A gene transfer from a parasitic insect, either from one with a globin gene with a different structure to the mammalian archetype, or one followed by an insertion event creating a new intron, is postulated to account for the presence of this globin gene in a higher plant (Hyldig-Nielsen et al 1982). The possibility that RATS may transfer themselves and possibly other, host genes between species (section 2.7.4) lends credence to these suggestions by postulating a mechanism by which inter-species gene transfer might occur.

While such Eclectic Evolution may be satisfying here, these three cases are qualitatively different from the potential histone gene

cluster transfers. In all three cases the donor and acceptor species are separated by vast distances in evolutionary time. Even if the proto-plants 3000My ago possessed a gene which could give rise to a globin gene, it is improbable that this gene would have survived the intervening time essentially unaltered and still functional in the absence of selection pressure for a globin protein until the Leguminosae evolved a requirement for it. Although conservation of molecular structure (Matthews et al 1981, but see Keim, Heinrikson and Fitch 1981) and even protein sequence (Goldstein et al 1975, Schlessinger, Goldstein and Niall 1975) over such lengths of time have been suggested, it is implausible that the selection pressure required could have acted on higher plant and mammalian globin genes to conserve similar gene structures over 3000My. In these three cases an organism has been proposed to acquire a specialised gene for which its ancestors had no need, but which some unusual feature of its present lifestyle requires. To satisfy this requirement it would either have to 'invent' the gene for itself from whatever suitable genetic material it possessed (recall that the globins are believed to have evolved from the myoglobins (Jeffreys 1981) which are also not common plant proteins), or to acquire the gene from another organism. Thus there would be strong selection pressure on the recipient organism to accept a horizontally transferred, foreign gene. It is notable that, in the first two cases, the donor and recipient organisms are known to exist in very intimate physical relationships, making the proposed gene transfer physically more probable.

Such is not the case in the sea urchins or the Xenopinae. In both cases (Bisbee et al 1977, Barnes 1974) the species have diverged very recently compared to the prokaryotic/eukaryotic split and have not undergone morphological or ecological divergence in that

time even to the extent that the mammals have in the same time, let alone that *Rhizobacter* and Soya bean have since the archeic. So despite the three examples above, the following lines of argument lead us away from the idea that horizontal gene transfer has occurred in the former two pairs of species.

Let us assume that the overall structure of gene clusters is important to their function. If an *Xlh8*-like gene cluster was advantageous to the *Xenopus laevis/borealis* ancestor, and is advantageous to both *X.laevis* and *X.borealis* today, then there is no reason to suspect that in the 8My since these creatures diverged one of them has lost the requirement for an *Xlh8*-like cluster, lost their *Xlh8*-like cluster and all clusters sufficiently similar to it so that they may mutate to form an *Xlh8*-like cluster, and then regained the requirement for such a cluster so that it had to 'import' it from another species. If the ancestor and both descendants needed an *Xlh8*-like cluster, then selection pressure can quite adequately account for its conservation.

It is possible that the non-transcribed structures in gene clusters are not selected, and hence that the presence of *Xbhl* in the *X.borealis* genome is due to interspecies gene transfer at random, and was not driven by selection. This is counter-indicated by a large body of information outlined below, but might be true anyway. In which case we can estimate an expected frequency of gene transfer between any two species on the basis that, of the 20-odd histone gene clusters studied in *X.laevis* and *borealis* one (*Xbhl* or *Xlh8*) has been transferred between these two species in the last 8My. If applied to all other pairs of species (and recall that *X.laevis* and *X.borealis* are not in any particularly intimate relationship in nature) this suggests a rate of about 1 in 160 histone gene clusters

per pair of organisms per My, or of the order of 1 histone gene cluster per 10,000 years taking all species in a reasonably profligate habitat into account. Hence essentially no histone genes in a species should be specific to that species: nearly all would be shared with other species with no regard to their phylogenetic relatedness. This is clearly contra-indicated by experiment in histone and other gene families.

In short, any explanation invoking horizontal gene transfer to explain the Xlh8-Xbhl similarity, or the occurrence of hl9 in S.drobachiensis, requires there to be strong selective pressure for that particular gene cluster in the recipient, and in that case selection alone can explain the sequence's presence. Thus Occam's razor (to the extent that it may be applied to evolution (Crisci 1982)) requires us to reject the horizontal transfer hypothesis.

But this opens up another problem, and one we have encountered before. It is clear that coding sequences of genes could be very strongly conserved by selection, especially in the histone genes, which code for such highly conserved proteins. Also it could be said that co-regulated genes could be required to be clustered. However selection appears to work on more than just coding sequence and overall proximity of co-expressed genes. The embryonic histone genes of the sea urchin have the same gene order, polarity and overall organisation in several species, despite these species' divergence up to 150My ago (Kedes et al 1975a,b, Portmann, Schaffner and Birnstiel 1976, Sures et al 1978, Kedes 1979, Cohn and Kedes 1979, Overton and Weinberg 1978, Busslinger et al 1982). The overall structure of the 'circular map' of the histone gene clusters of X.laevis discussed above has been conserved over the 8My since this species divergence

from X.borealis, despite ample evidence that mechanisms exist to produce a daedal diversity of histone genetic structures in this species. Sequence constraints on H3 genes (Section 14) and H4 genes (Busslinger et al 1982, Turner and Woodland 1982) and ribosomal protein genes (Amaldi et al 1982), unrelated to their coding capacity, appear to have been at least partly conserved in all the vertebrates. The unequal use of codons is universal in eukaryotes, but which codon is preferred for any amino acid depends on the organism and the gene type concerned (Graham 1978, Grantham, Gautier and Goug 1980,1981), so the sequence conservation implied must be concerned with a function whose detailed nature varies between species and between genes. Busslinger et al (1982) show that the sequence similarity of spacer sequences in sea urchin histone gene clusters bears little relationship to their evolutionary relatedness. Finally we noted that the majority of clones H2A and H2B genes have been found to be adjacent and divergently transcribed, this correlation even extending to the fungi. Thus it seems as if the whole of a gene cluster's structure and sequence is selected, and not just the coding regions'. This is a general point, although the histone genes are one of the most thoroughly characterised systems in which it has been shown. To show that selection does indeed act upon whole gene clusters, and indeed that the unit of function, and hence of selection, in the eukaryotes may be 50-150kb long, I will review briefly the evidence leading us to suspect that the average eukaryotic 'gene' is a stretch of more than 50kb of DNA, before returning to discuss how this relates to the problem of the conservation of Xlh8-like gene clusters.

17.3 The 50kb gene.17.3.1 Multigene families and multigene clusters.

It has become apparent in the last few years that many protein-coding genes in eukaryotes are members of multigene families. Histones are an obvious example of this. The families largely fall into two, well-defined categories, typified by actin and globin gene families. For what follows it will be necessary to describe these family types in a little more detail.

The globin genes are all expressed in a single cell lineage, the erythropoietic cell line of bone marrow (Harrison 1977, Perutz and Lehmann 1968, Efstradiatis et al 1980, Patient et al 1980, Cline and Golde 1979). In all organisms studied to date the genes are clustered (Patient et al 1980, Efstradiatis et al 1980, van Dommelen et al 1980, Tiemeier et al 1978, Jeffreys et al 1982) and they have two introns interrupting the coding region of each beta-like gene at the same position in the reading frame and of the same length. Between closely related species the introns in equivalent positions are accurately conserved in size and sequence (Jeffreys et al 1982), although within a species introns, 5' and 3' non-coding regions can vary a great deal between 'duplicate' genes (Jeffreys 1979, Michelson and Orkin 1980, Jeffreys et al 1982) even although those genes have diverged recently in evolution. That this conservation between species is not due to mutational 'cold spots' around expressed sequences is shown by the fact that small alterations in sequence at the beta globin locus are quite common in man (Jeffreys 1979, Bernards and Flavell 1980, Orkin et al 1982) and larger ones

are found, but are invariably detrimental to the locus' function . I alluded to this comparison of inter-species conservation coupled with intra-species variation before in section 2.5.2.

This pattern is a common one in multigene families coding for proteins expressed in a single cell lineage (note that this need not imply coordinate expression), although all of its characteristic features have been demonstrated in few gene families. The pattern of introns has been preserved in ovomucoid genes in chick since the bird's divergence from the mammals (Stein et al 1980), and in vitellogenin (Arnberg et al 1981) and alpha-collagen (Shaffer et al 1980, Weiss et al 1982) for about the same length of time, in larval serum proteins of Drosophila (McClelland, Smith and Glover 1981) and mouse alpha-amylase (Schibler et al 1980). 3' and 5' non-coding regions of mammalian pro-opiomelanocortin (Chang, Cochet and Cohen 1980) and alpha-lactalbumin (Hall, Davies and Craig 1981) and rodent Whey protein (Henninghausen and Sippel 1982) genes have also been strongly conserved between species. The sequence between members of a cluster of genes expressed in a common lineage is not so severely conserved in chick ovalbumin, X and Y genes (Royal et al 1979, Heilig, Murakowsky and Mandel 1982, Heilig et al 1982) and rat alpha-2u-globulin genes (Kurt : 1981) . Polymorphism between such genes has also been found in chick ovalbumin (Hughes et al 1979, Lai et al 1979) and human glycohemone alpha-subunit (Fiddes and Goodman 1981) to be solely restriction site polymorphism rather than larger deletions or insertions of sequence.

Several other families of genes with restricted patterns of expression are known to be closely linked: Bithorax and Decapentaplegic in Drosophila (Marx 1981, Spencer, Hoffman and Gelbart 1982) and H2 and HLA, and T/t gene complexes in mice and man

(Steinmetz et al 1982, Klein 1978, Barnstable, Jones and Bodmer 1979, Artz, McCormick and Bennett 1982), (although these are enormous gene clusters compared to the globin or ovalbumin loci, and are not strictly limited in their expression to single cell lineages) and alpha-fetoprotein and albumin genes in mouse (Ingram, Scott and Tilghman 1981). As well as genes which have been shown by gene cloning to exist in clusters, the mouse casein (Gupta et al 1982) and rat globulin (Kurtz 1981) co-expressed gene families have been shown by cytological studies to be linked on at least a chromosomal level.

These are not invariant rules. Prostatic steroid binding hormone genes show extreme conservation between duplicate genes within an organism (Parker, Needham and White 1982), and two protein hormones expressed in pituitary (Fiddes and Goodman 1981) show extensive homology between their 5' non-coding regions in the human genome. These homologies might be related to sequence requirements for binding of regulatory hormone receptors to these genes for their activation: analogous DNA-sequence specific binding of a steroid receptor protein to DNA has been found in several genes elsewhere (Westley and Knowland 1978, Mulvihill, LePennec and Chambon 1982, Compton, Scraber and O'Malley 1982), where the receptor-DNA interaction has been implicated as a major requirement for transcription of the adjacent gene (Lee et al 1981, Robins et al 1982).

That this pattern of conservation of cluster structure between species is directly concerned with the control of gene expression is suggested by the globin locus deletions in man. Here several deletions in the globin locus are known to degrade cluster function, even that of genes distant from the termini of the deletion (Fritsch, Lawn and Maniatis 1979, Bernards and Flavell 1980, Weatherall and

Clegg 1982). This implicates the conserved intron and spacer sequences in the control of the transcription of the whole cluster.

This pattern of inter-species conservation and intra-cluster divergence in tightly clustered genes is not the rule for all multigene families. Another pattern is typified by the actin gene family. Here a number of different genes code for a variety of proteins which are expressed in a tissue-specific pattern in nearly every type of animal cell (Sorti, Coen and Rich 1976, Garrels and Gibson 1976, Mannherz and Goody 1976, Clarke and Spudis 1977, Schwartz and Rothblum 1981). Actin genes have been cloned from chick (Fornwald et al 1982), Xenopus (G. Cross per com), Drosophila (Fryberg et al 1980), sea urchin (Durica, Schloss and Crain 1981, Cooper and Crain 1982), Dictyostelium (McKeown et al 1978), rat (Nudel et al 1982) and man (Engel, Gunning and Kedes 1981, 1982a,b). Only in this last species has more than one actin gene been found on a clone: HRL51 has two closely adjacent actin genes whose flanking sequences suggest that they are the product of a recent duplication (Engel, Gunning and Kedes 1982b). Rat and chick alpha actin genes have introns at equivalent positions, as do two beta genes from these species: apart from these two instances there is no similarity between the position of introns in any of the actin genes cloned (Fornwald et al 1982). There is a certain degree of sequence homology between the transcribed, non-coding regions of the cytoplasmic actins of mammals (Dodemont et al 1982).

This different pattern of distribution in the genome and reduced conservation of non-coding sequences between species is presumably connected with the non-cell lineage - specific pattern of expression of these genes. This is borne out by other cloned members of multigene families which follow the actin pattern and also are

expressed in a variety of cell lineages.

Human and rat prolactin and growth hormone genes, although evolutionarily related, are on different chromosomes (in man) and have very different intron sizes and sequences, although exons are of similar size (Cooke and Baxter 1982). Rat myosin heavy chain genes are probably similarly divergent (Nudel et al 1980). Drosophila alpha-tubulin genes expressed at different times during development are located distant from each other on the same chromosome (Mischke and Pardue 1982). Recently Leicht et al (1982) have shown that chick ovalbumin and human antitrypsin genes, although coding for evolutionarily related proteins and showing considerable protein homology, have introns which do not correspond in size, sequence or location between the two genes. Needless to say, they are not expressed in the same tissue: ovalbumin is an oviduct product (Royal et al 1979), antitrypsin is a more generally expressed proteinase inhibitor of plasma (Carrel et al 1982).

This suggests that histone genes might fall into both classes, in that clusters of genes might be coordinately expressed in a restricted number of cells while different clusters, like actin-type genes, will be expressed in different cell lineages. Although this is true of embryo and oocyte gene clusters (vide supra) too little is known about the expression of other types of histone gene clusters to say from experimental results if this is generally true. It is also unclear how 'housekeeping' genes, of which the histones may be an example, fit into this picture, which is derived largely from the study of the genes for differentiated products. However the histone gene data does lend a little support to the strong correlation between the conservation of extensive inter-gene spacer organisation, gene clustering and co-regulation of the clustered genes.

These data suggest that a unit of control of transcription is a whole gene cluster and not just a coding region. Can this be supported by other data?

17.3.2 How many genes in a eukaryote?

This depends on the eukaryote, of course, but two lines of evidence in a number of species suggest that only about 5-10% of most eukaryote's genomes is actually transcribed onto mature mRNA, which has implications for the sizes of the genes concerned.

The first is the study of how many RNA species are found in tissues and cultured cells. A measure of RNA complexity, the primary measure in these experiments based on several nucleic acid hybridisation techniques, is more use to us at the moment. Thus oocyte and early embryo mRNA from sea urchins and amphibia have a complexity of around 35Mb (Lewin 1975, Galau, Britten and Davidson 1974, Galau et al 1976), and some mammalian cultured cell lines have similar RNA complexities of 20 - 40Mb (Getz et al 1976, Williams and Penman 1975, Lewin 1975, Getz et al 1975, Bishop et al 1974). In sea urchin this complexity drops through development as later embryonic stages express fewer and fewer of the oocyte sequences, but do not start the synthesis of extensive new mRNA species to replace them (Rodger and Gross 1978, Wold et al 1978, Galau, Britten and Davidson 1974, Galau et al 1976, Galau et al 1977). In adult tissues in rodents, chick and sea urchin mRNA complexities are typically 4 - 20Mb (Axel, Feigelson and Schutz 1976, Higgins et al 1978, Galau et al 1976), and each tissue has only a 400-2000kb of mRNA sequence complexity not found in any other tissue. Thus almost all sequences found in liver may also be found elsewhere in the animal, especially ovary (Axel, Feigelson and Schutz 1976, Higgins et al 1978, Galau et

al 1976, 1977, Hough-Evans et al 1977, Xin et al 1982) : in ovary, in sea urchins and amphibia, the majority of the genes active in any one other tissue are also active. Tissue culture cells support this: on changing from resting to proliferating cultures mouse EC cells (Getz et al 1976) and fibroblasts (Williams and Penman 1975) retain nearly all their mRNA sequences, and only a few hundreds - 1000kb of new mRNA is specific for the new 'developmental' state.

An exception to this is brain (Chikaraishi, Deeb and Suecka 1978, Bantle and Hahn 1976) which in rodents expresses in excess of 100Mb of RNA sequence. This is a very much higher figure than for other tissues tested, but is this really surprising? I think not. Liver is a relatively homogenous tissue (Elias and Sherrick 1969), and Axel, Fiegelson and Schut (1977) state that their preparation of it was largely hepatocytes, a single cell type. By contrast brain contains numerous types of nerve cells morphologically and biochemically distinct from each other, glial cells to support them and epithelial tissue (Jordan 1954) and none of these contribute a predominant cell type to the tissue. Thus 'Liver' is essentially one type of cell, 'Brain' is at least 10. If cells have a common set of expressed sequences plus about 1 - 2Mb of cell-specific sequence for each cell type, it is unsurprising that total Brain RNA shows a higher complexity than total Liver RNA. Indeed, the high message complexity in brain is a good support for such a model.

So we arrive at the consensus that there is a 'base' transcription of 20 - 40Mb of sequence in growing cells, and a cell-type specific transcription of 1 - 2Mb of sequence not transcribed anywhere else in the organism.

Whether this complexity figure can be related directly to the

complexity of transcribed DNA sequence is unclear, because of the possibility that different genes could direct the synthesis of mRNAs indistinguishable under the hybridisation conditions used in these studies. The different histone H3 and H4 genes of mammals are known to direct the synthesis of similar RNAs coding almost identical proteins (Sittman et al 1981, Sierra et al 1982, Stephens et al 1977), and similar findings with somatostatin (Warren and Shields 1982), feather keratin (Kemp 1975), and mouse eye lens crystallin (Shinohara 1982) gene families suggest that the presence of several genes which direct the synthesis of nearly identical RNAs could be common. The mRNAs from these closely related genes would be detected as a single species under non-stringent hybridisation conditions. Thus if mRNA complexity is equated with the complexity of DNA sequence coding for the mRNA, the number of active genes could be underestimated. The degree to which this effect might bias gene number values is uncertain.

Converting the general statements above into an actual gene number is difficult, as we require to know the number of 'different' cell types in an organism. However, if we assume the fairly low estimate of 50 cell types (in the sense that hepatocytes are one 'cell type'), each expressing 1 - 1.5Mb of cell-specific RNA, and a 'background' of 30Mb of common/oocyte RNA, this gives a total complexity of 100Mb for the RNA-coding genomic sequences, or about 5-10% of a mammalian genome (Lewin 1974,1980).

The other approach to counting genes measures the number of genetically defined gene loci rather than the complexity of their product by attempting to locate all the mutable loci in a section of a genome. This has only been performed in Drosophila. Here various mutagens have been used to saturate the diminutive chromosome 4

(Hochman 1971, 1972, 1973) or small sections of the X-chromosome (Judd, Shan and Kaufman 1972, Judd and Young 1973, Kaufman et al 1975) or an autosome (Lifschy: and Falk 1968, 1969) with lethal mutations. In all three cases the number of loci which could give rise to lethal alleles was about the same as the number of chromomeres (Hochman 1973, Judd and Young 1973, Lewin 1974). It is likely that only 50 - 75% of all genes are detected by this approach, the rest being incapable of producing lethal alleles (O'Brien 1973, Young and Judd 1978). It is also known that some of these complementation groups are very tightly clustered (Marx 1981, Lefevre 1981, Homyk, Pye and Park 1982) and probably represent duplicated regions analagous to the globin or ovalbumin loci of vertebrates in structure. Thus the general rule of 1 'gene' = 1 band stands if we count closely linked, multigene clusters as functional units, suggesting a total of 5000 - 10,000 genes in Drosophila (and a not very significantly larger number without this assumption).

On a naive calculation both these measurements suggest that there are 50 to 150kb of genomic DNA to each gene.

17.3.3 The size of a 'gene'.

One interesting facet of the work counting lethal alleles in Drosophila is the rate at which X-rays induce mutations in this species. This has been measured for the Drosophila Dumpy locus at $10 \exp(-7)/\text{locus/rad}$ (Carlson and Southin 1962), and at similar values are reported during other investigations involving mutagenesis. Similar figures have been found in many somatic cell lines in culture from rodents and man (Bridges, Huckle and Ashwood-Smith 1970, Bridges and Huckle 1970, Chu 1971, 1974, Wolff et al 1974, Albertini and DeMass 1973), and in intact mice (Selby 1973).

Equivalent rates in prokaryotes are 20-100 times less (Bridges, Huckle and Ashwood-Smith 1970, Hopwood and Sermonti 1962, Setlow and Setlow 1972, Wolff et al 1974, Kaplan et al 1960). The mechanism of induction of mutations by X-rays is believed to be similar in all organisms (Newcombe 1971, Setlow and Setlow 1972, Parker and Wilkinson 1974, Searle 1974), so this discrepancy in rate has been explained as an effect of the environment, metabolic or physical, of the DNA in eukaryotes (Bridges and Huckle 1970), although no similar effect is seen between bacterial and viral genomes (Brown 1966, Kaplan et al 1960). However the simplest explanation for the different rates of X-ray mutagenesis is that the X-ray target size is 20-100 times as large in eukaryotes as in prokaryotes. This explanation has been accepted for the nematode Caenorhabditis (Brenner 1974), where 'genes' are consequently deduced to be 34kb long on average. A similar figure for mammals would be 60 - 150kb, and 80-180kb for Drosophila, depending on the mutation rate chosen for the calculation.

Clearly the coding regions of eukaryotic genes are not 150kb long, so there are two possible deductions we can draw from this figure: a) the figure is incorrect or b) most of a eukaryotic gene is non-coding. This latter is supported by the observation that most natural (McDonald and Kelley 1971, Upchurch et al 1975, Ghangas et al 1975) and induced (Beaudet, Roufa and Cashey 1973) mutations of the HPRT locus in man and mouse respectively result in a lowering of the level of enzymes in the cells (as measured by enzyme activity or immunological cross-reacting material) and not in the alteration of the immune or enzymological properties of the enzyme, ie they are 'control' mutants. Similar analyses for hamster DHFR (Graf and Chasin 1982) and rat albumin (Capetanaki, Flytzanis and Alanco 1982) gene mutants in cell cultures may most readily be explained this way.

Genetical mapping of a control mutant 2.8kb from the Rosy locus XDH-coding region in Drosophila, showing that the control of this transcription unit extends at least half way to the next transcription unit, supports this class of interpretation (Chovnik et al 1977).

Several other lines of evidence hint that a 'gene' in eukaryotes is considerably larger than a coding region.

Firstly, electron microscopic examination of non-nucleosomal, transcribed regions in chromatin find that chromatin in an open, relaxed conformation occasionally studded with nascent RNA strands is present in 'domains' up to 20kb long in insects (Foe, Wilkinson and Laird 1976, McKnight and Miller 1976, Lamb and Daneholt 1979) and amphibia (Scheer 1978).

Second is the pattern of methylation, which as we noted in section 15.3.3 is correlated with gene activity (Mandel and Chambon 1979, Bird, Taggart and Gehring 1981, Kuo, Mandel and Chambon 1979, van der Ploeg and Flavell 1980, McCleod and Bird 1982). In the sea urchin and mouse DNA hypermethylated regions are not distributed at random, or just in non-transcribed coding regions, but in long tracts of DNA 20 to over 50kb long, separated by less methylated sequences (Bird, Taggart and Smith 1979, Naveh-Manly and Ceder 1982). Some smaller units would be expected if expressed, hypomethylated regions are arranged at random in the genome rather than being segregated to different 'domains' of over 50kb.

Thirdly, transcription (Jackson, McCready and Cook 1981, Robinson, Nelkin and Vogelstein 1982) and DNA replication (McCready et al 1980, Pardoll, Vogelstein and Coffey 1980, Hunt and Vogelstein

1981, Buongiorno et al 1982) are associated with the domains of DNA bound to the non-histone nucleoskeleton. These domains are estimated from electron microscopic, sedimentation and nuclease digestion studies to be 45-120kb long, supercoiled, and appear to be universal to eukaryotes (Cook and Brazell 1975, 1980, Ide et al 1975, Benyajati and Worcel 1976, McCreedy, Cox and McLaughlin 1977, Paulson and Laemmli 1977, Marsden and Laemmli 1979, Okada and Comings 1979).

Fourthly as we noted above, we may calculate from mRNA complexity data and from gene-counting experiments in Drosophila that a eukaryotic 'gene' has 20-200kb of DNA to it. Previously this was thought to show that 90% of eukaryotes DNA was 'junk', but now we can suggest other reasons.

Lastly, a ratio of 1kb coding:10 - 50 kb non-coding (the magnitude of ratio we would expect of a 'domain' 50 - 100kb long containing 1 to 5 coding regions) is just what we observe in human globin (Efstradiatis et al 1980), chick ovalbumin (Royal et al 1979) and mouse H2 (Steinmetz et al 1982) loci from the cloning studies. These loci, we noted before, are co-regulated in a restricted number of cell types and are conserved between species. In the beta globin locus a direct correlation between the integrity of the intergene spacer and the control of transcription has been seen. We may now suggest, therefore, that the co-regulation of these clusters and the conservation of their intergenic spacer are results of the fact that the whole gene cluster is, in some sense, a functional unit, and that its sequence is required to be conserved for that function. This also ties in with the observed conservation of sequence within genes, which is not related to their coding capacity, seen in a variety of genes mentioned above (Grantham 1978, Grantham, Gautier and Goug 1980, 1981, Amaldi et al 1982, Nossinov 1982, Turner and Woodland

1982 and this work section 15). Note that this also explains why the actin-type gene families are different from the globin-type ones. They are not expressed in the same coordinated way as the latter and consequently are segregated to different 'domain' regions where they experience selection for different spacer sequence: consequently gene cloning experiments find them to be unlinked and not to show the same extensive conservation of flanking sequences.

17.3.4 The 50kb gene and repetitive DNA.

If the basic functional unit of the eukaryotic genome is a 50 - 150 kb length of DNA whose sequence must be at least partially conserved if the transcribed regions within it are to be functional, then this sheds a little light on some problems I raised in section 2. At that time I concluded that intermediate repetitive DNA showed many of the characteristics of 'selfish' DNA rather than those of regulatory elements or a structural genome component, but noted that the observed stability of the genome did not fit in well with this hypothesis. On the contrary, we would expect the 'selfish' elements to disrupt the genome, so rather than conserved genomes exhibiting limited polymorphism we should see conserved coding regions floating in a shifting sea of selfish sequences. However if the genes are actually 50 - 150 kb long, and therefore most if not all of the genome comprises overlapping or abutting genes, then there are no 'inter-gene' sites into which these 'selfish' elements could insert without killing the cell or organism in which they were. Only during speciation or in the course of evolution, when gene functions are being modified or discarded will a 'selfish' element be able to insert into a functional domain, then to be frozen there as the domain takes up new functional constraints. That this is the correct analysis is suggested by the observed stability of primate (Yunis and

Prakash 1982) and Xenopinan (Tymowska and Fischberg 1977) chromosomes, which have preserved most of their cytological structure over 15My (Harrison et al 1977) and 30My (Bisbee et al 1977) respectively, and by the constancy of the relative positions of centromere, heterochromatic regions and nucleolar organiser seen in a wide variety of animals and plants (Lima-de-Faria 1977).

17.4 The 50kb histone gene.

What has this to do with the structure of cloned histone genes of Xenopus borealis?

Two things. Firstly, it suggests that the structure of cloned histone gene clusters is related to their function and not just to their history. Thus the major and minor clusters of X.borealis could well have quite different roles because of their different structure. It is interesting in this respect to recall that several of the restriction site differences between the homologous regions of Xbhl and the major cluster are in the H2B gene, suggesting that this gene is either a pseudogene or that it codes for a variant protein. If variant histone proteins do have functional significance in the control of gene activity as I suggested they might in section 1 above, then they would need to be differentially synthesised in different tissues (as, indeed, they are observed to be). This in turn would mean that their genes would not be intermingled in gene clusters with other variants but would be segregated into different clusters. Thus it is unsurprising that Xbhl and the Xenopus laevis gene cluster have different H2B, H1 and H3 genes despite being apparently related.

We can say the same of the heterogenous 'non-circular-map' gene

clusters of X.laevis, which under this interpretation would be functional at different times from the oocyte genes and from each other. The 'major cluster' genes, by contrast, would be expressed at the same time, as they are linked into a large unit and so will be present in only a few 'domains'. This is consistent with the idea that they are expressed together in the oocyte. The same point is illustrated better in a system where the tandemly repeated genes' function is known and does not have to be guessed : the sea urchin and Drosophila embryonic genes, which are grouped into large clusters and are co-regulated, but are separate from the non-embryonic histone genes, and not co-regulated with them.

This leaves some uncertainty about the other 'circular map' clusters with the same gene order but different spacer organisation from the X.laevis 'major cluster'. The cluster polymorphism in X.laevis would suggest that many of the variant clusters are not essential, as individuals lacking them can be found. If their structure was uniquely related to their function we would expect every histone gene cluster variant to be essential, as they would all have unique functions. It is possible that the polymorphisms generated by doubling up the entire chromosome set by tetraploidisation of X.laevis (Wahli et al 1979, Patient et al 1980, Bozzoni et al 1981, Widmer et al 1981, see section 3 above) offers an explanation for this: the polymorphism of X.laevis histone genes could be due to a low level of point differences causing the observed restriction site heterogeneity in otherwise very similar clusters. Thus many of the 'variant' clusters detected by Southern blots of genomic DNA might be only trivially different from each other. The finding of pairs of minor gene clusters of clones differing by only a few restriction sites (Xlh8, Xlh2 & Xlh202, Xlh24 & Xlh25, Xlh11 and Xlh23) supports this interpretation, and suggests that the diversity of histone genes in

X.laevis might be less if measured in terms of gene spacing and order than if measured by restriction mapping.

Secondly, this makes explicable the similarity between Xbh1 and the Xlh8 family as being due to selection acting on inter-gene spacers in this cluster (although it does not really 'explain' this as the nature of the selection, apart from its being concerned with gene function, is unknown). Indeed, I cited the similar observation of extreme conservation of a histone gene cluster in sea urchin (Busslinger et al 1982) as potential evidence to support the 50kb-gene idea above, and the case of Xlh8 and Xbh1 is little different from this sea urchin one. The idea that long stretches of DNA around coding regions are under almost as severe sequence-conserving constraints as the coding region itself offers a mechanism by which one gene cluster could mutate at less than 0.15% /My in the last 8My, and suggests that other examples of the same thing could well be found in Xenopus and elsewhere when sufficient numbers of histone gene clones have been examined. In this interpretation the Xlh8-Xbh1 similarity could be a normal product of the evolutionary selection of DNA sequence imposed by the mechanisms of the control of gene activity in eukaryotes, as opposed to a statistical freak, retroviral transductant or product of a contaminated gene library.

Appendices

18. Appendix 1. Cosmid cloning: an attempt and its failure.

"But, oh, beamish nephew, beware of the day
 If your Snark be a Boojum, for then
 You will softly and silently vanish away,
 And never be met with again."

18.1 Introduction.

Cosmids are plasmid gene cloning vectors which have been designed for cloning large stretches of genomic DNA with high efficiency. They are small plasmids in which the cohesive ends of lambda phage DNA have been inserted (Collins and Bruning 1978, Collins and Hohn 1978, Hohn and Collins 1980, Chia, Scott and Rigby 1982). These plasmids replicate like other high copy-number plasmids, and have been manipulated to contain single restriction sites for several 6-cutters commonly used in gene cloning. In vitro lambda DNA-packaging systems can package any DNA molecule containing the lambda cohesive ends separated by 45-50kb of DNA into 'infectious' lambda particles (vide infra, Appendix 2 for discussion and refs.), but the non-recombinant cosmids are well below this size limit. However if they are ligated with long fragments of exogenous DNA, their cohesive end - to - cohesive end distance becomes great enough for them to be packaged. The exogenous DNA is usually a collection of restriction fragments of genomic DNA from an organism for which a gene library is required, so a spectrum of ligation products is produced. Only those containing the plasmid sequences with inserts of ≤ 30 kb are packagable. So to produce a gene library all that needs to be done is to ligate high molecular weight DNA at high DNA concentration (as the packaging systems' in vivo substrate is linear concatomeric DNA, and it is unclear whether

it can package covalently closed circles (Hohn 1975, Hohn and Murray 1977)), package this product into phage particles and then infect susceptible E.coli with them. Only recombinant plasmids containing large inserts of genomic DNA will appear as colonies, as non-recombinants will fail to be packaged. Thus the library is guaranteed to contain only plasmids with large inserts.

This is a very favourable method of generating a gene library, especially when large genes (Weiss et al 1981) or a number of clustered genes (Royal et al 1979, Steinmetz et al 1982) which are likely to occupy a considerable length of DNA are sought. Histone genes clearly fall into the second category, and so I attempted to make a cosmid gene library from X.borealis. This did not work, for reasons which appear to be specific to Amphibian DNA.

18.2 Methods

18.2.1 Packaging DNA into lambda phage particles.

This is the method of Hohn and Hohn (1974): the method of Enquist and Sternberg (1979) was also used in this section, but has been already described (section 5.5.2).

Two 10ml cultures of BHB2688 and BHB2690 were grown overnight at 30°C. These were used to inoculate two 1litre cultures of L.broth, which were shaken at 30°C until their OD(600) reached 0.3 . They were then incubated at 45°C for 15mins, and then shaken vigorously at 37°C for 3 hours. The two cultures were then mixed.

The cultures were centrifuged at 6000 rpm, 10' 0°C in an MSE

6x250ml rotor. The pellet was resuspended in a minimum volume of Complementation beffer (40mM Tris pH8.0, 10mM Spermidine, 10mM Putrescine (both as HCl salts), 0.1% 2-mercaptoethanol, 7%DMSO) : 20-30mls was typical. The bacteria were re-centrifuged and re-suspended in the smallest volume of complementation buffer consistant with the suspension being pipettable. The bacterial suspension was then pipetted into 50µl aliquots and frozen radidly in liquid nitrogen for storage at -70°C.

The standard packaging reaction was as follows. A 50µl aliquot of the mix was thawed out on ice, re-frozen in liquid nitrogen and re-thawed. To this was added 1ul of 100mM ATP and DNA in less than 5µl of water or TE to a final DNA concentration of 2µg/reaction or less. The reagents were added quickly and the mix mixed thoroughly on ice, so as to effect the mixing before the bacterial cells lysed and the viscosity of the mixture increased. The mix was incubated for 30mins at 37°C. 2µl of 0.25µg/ml DNase I was then added with 0.5ml of phage buffer, and the mix incubated for a further 30 mins. The mix was then ready for plating as a bacteriophage stock.

18.3 Results.

18.3.1 Initial cloning attempt.

70µl of X.borealis DNA at 400µg/ml, digested to 25% completion with EcoRI and 1.00µl of 150µg/ml pHC79 (Hohn and Collins 1980) cosmid vector linearised with EcoRI were ligated in a total volume of 180ul for 36hrs at 18°C. 15µl of this was packaged into lambda phage heads in the system of Hohn and Hohn (1974), and plated on

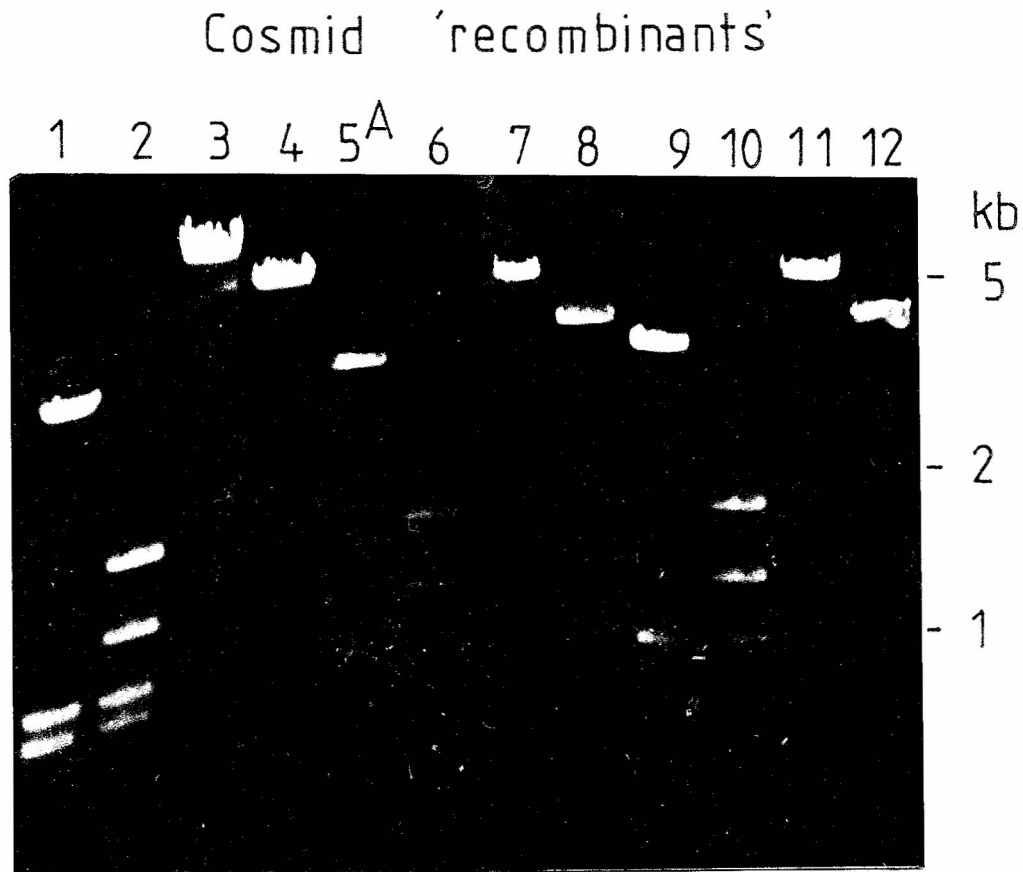
ampicillin plates. 15 colonies were selected and the plasmid DNAs from them prepared. All but one plasmid were indistinguishable from pHC79 in size and restriction pattern, and the one exception ('E122') proved to have no EcoRI insert, carrying instead a duplication in the region of the lambda cohesive ends. An illustrative digest is shown in Fig 18.1 .

This was rather worrying. In theory, no non-recombinant molecules should be present in the gene 'library': in practice the library contained nothing but non-recombinants. The following sections describe what I have been able to discover of the reason for this discrepancy.

18.3.2 A successful E.coli gene library.

It is possible that the DNA I was using to construct the gene library was in some way inadequate due to its sequence, its chemistry or to contamination. Thus I tried cloning some E.coli DNAs. E.coli C DNA (From Dr. Ian Jones) was digested to 10% completion with EcoRI and ligated in a 5:1 mass ratio to Eco-RI digested Homer1 vector (Chia, Scott and Rigby 1982). 2µg of this was packaged into lambda heads, the results plated on ampicillin, plates and 10 colonies were selected. The plasmid DNA from these were prepared, EcoRI digested and electrophoresed on agarose. The result is shown in Fig 18.2 . Clearly at least 8/9 of these contain one or more EcoRI fragments not derived from Homer1 vector. Hence the cloning procedure is working here. As, apart from the exogenous DNA and the vector (pHC79 and Homer1 are both pBR322 derivatives) all factors in this experiment are the same as in the failed attempt to construct a X.borealis gene library, then the DNAs used must be responsible for the latter failure.

Fig 18.1



B

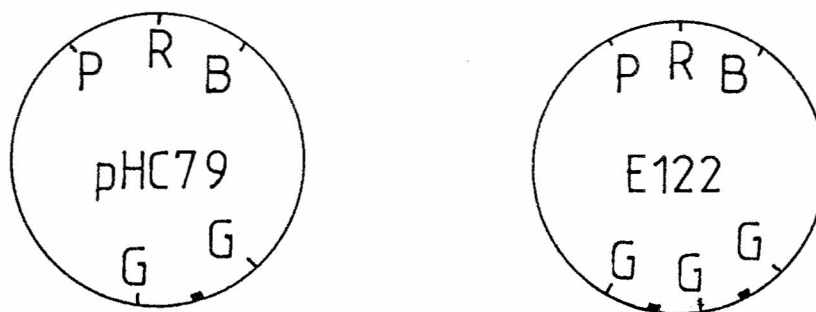


Fig 18.1 A) Analysis of E122 (1-4), pHC79 (5-8) and a 'recombinant' from the Xenopus borealis cosmid gene bookshelf (9-12). Digests: 1,5,9) BglII, 2,6,10) BglII+EcoRI, 3,7,11) PstI, 4,8,12) PstI+BamHI. 0.7% agarose gel stained with ethidium. B) Probable restriction structure of E122 compared to pHC79

Fig 18.2

Cosmid recombinants

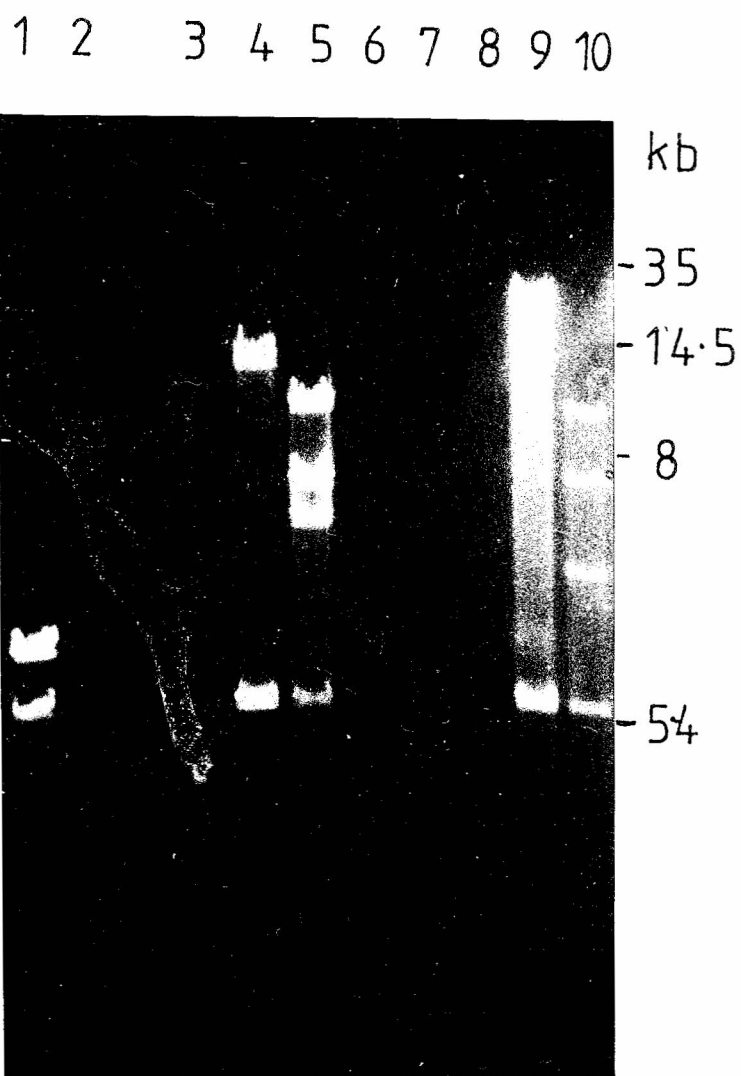


Fig 18.2 EcoRI digests of 10 potential recombinant cosmids isolated from the E.coli cosmid library. 0.7% agarose-ethidium gel.

18.3.3 Non-recombinants are concatomer derivatives.

One way in which cosmid vector-only plasmids may become incorporated into phage heads is for them to concatomerise in the ligation reaction: a 5kb vector joined head-to-tail 9 times could form a molecule packagable by the in vitro system. Such an occurrence is improbable, but not impossible. To see if this is responsible for the non-recombinant plasmids I observed to comprise >90% of the X.borealis 'gene library' I treated the vectors with bacterial alkaline phosphatase (BAP) to remove their terminal phosphates, thus rendering them incapable of self-ligation, and repeated the cloning experiments. The results are shown in Fig 18.3 . Clearly, the BAP treatment has little effect when the E.coli DNA is ligated into the vector, but ligating X.borealis DNA to BAP-treated cosmids gives no colonies at all. Thus we postulate that all the colonies I saw in the initial gene cloning attempt were due to packaging of concatomerised vector molecules.

These clones contained plasmids which were not 45-50kb long, instead consisting of mixed monomer and oligomer vector molecules with up to 5 or more monomer units in the undigested molecules. This is probably due to recombination of repeated vector units within these multi-vector circles to produce smaller molecules. That this can occur even in the RecA- host we use is demonstrated in Fig 18.4 . Here EcoRI-linearised pCR1, a plasmid containing the kanamycin resistance gene, was ligated to pHC79 in a 2:1 mass ratio, and the result packaged and plated on kanamycin-containing plates. All colonies therefore derive from bacteria containing a recombinant plasmid containing at least one pCR1 unit: as pCR1 is 13kb long 3 pCR1 molecules and one pHC79, or some equivalent length of DNA with some other ratio of monomers, are needed to fill the phage heads.

Fig 18.3

Analysis of cosmid methodology

Vector	B.A.P.	DNA	colonies
Homer1	-	none	312
"	+	"	0
"	-	X.borealis	270
"	+	"	1
"	-	E.coli C	> 2000
"	+	"	> 2000
pHC79	-	none	23
"	+	"	0
"	-	X.borealis	100
"	+	"	0
"	-	E.coli C	> 2000

Fig 18.3 Analysis of cosmid cloning techniques. The number of colonies produced by plating 500ng of ligation mix packaged into lambda heads in vitro on ampicillin plates is given for several treatments of the component DNAs. DNA was EcoRI restricted as described, and then ligated in a mass ratio of 1 vector: 10 genomic DNA. The table shows the effect of varying the vector used ('vector'), bacterial alkaline phosphatase treatment of the linearised vector ('BAP' - +=vector treated, -= vector not treated) and the exogenous DNA ligated to the vector ('DNA').

Initially the pCR1 and pH79 must have been covalently linked to allow a Kn-resistance gene to be packaged in vitro, and in accordance with this all Kn-resistant colonies were found to be Ap-resistant. If the packaged DNA was plated on Ap plates, only 60% of the colonies were Kan-resistant, supporting the idea that pH79-only concatomers could form and be packaged in such an experiment.

One Kn-resistant colony was picked, grown in 10mls of L.broth containing ampicillin for 8 hours until it reached stationary phase and aliquots removed at three times and plated on Ap and Kn plates. The titres obtained are shown in Fig 18.5. Clearly, the Kan-resistance is being lost from the culture, even though the Kn-r and Ap-r genes must have been covalently linked originally (pCR1 is not packageable on its own (data not shown)). Thus some recombinational event between multiple copies of plasmid sequences in the concatomeric primary clones (pCR1 is a ColE1 derived plasmid, as is pBR322 and its derivatives) is causing loss of sequences from the original 45-50kb plasmid.

18.3.4 Failure to clone is due to X.borealis DNA.

Clearly the X.borealis cosmid recombinants are failing to clone when E.coli recombinants and plasmid concatomers are succeeding. Why is this? Fig 18.5 and 18.6 show a number of experiments which locate the problem. The first set show that cosmid cloning does not work efficiently on X.borealis or Ambystoma mexicana DNA (the latter was the gift of Dr. John Adair). This effect is not due to gross contaminants in the DNA. Other results showed that re-banding the plasmid vectors on CsCl, re-extracting the X.borealis DNA with phenol and/or 2-methoxyethanol (Kirby and Cook 1967) and

Fig 18.4

Plasmid instability

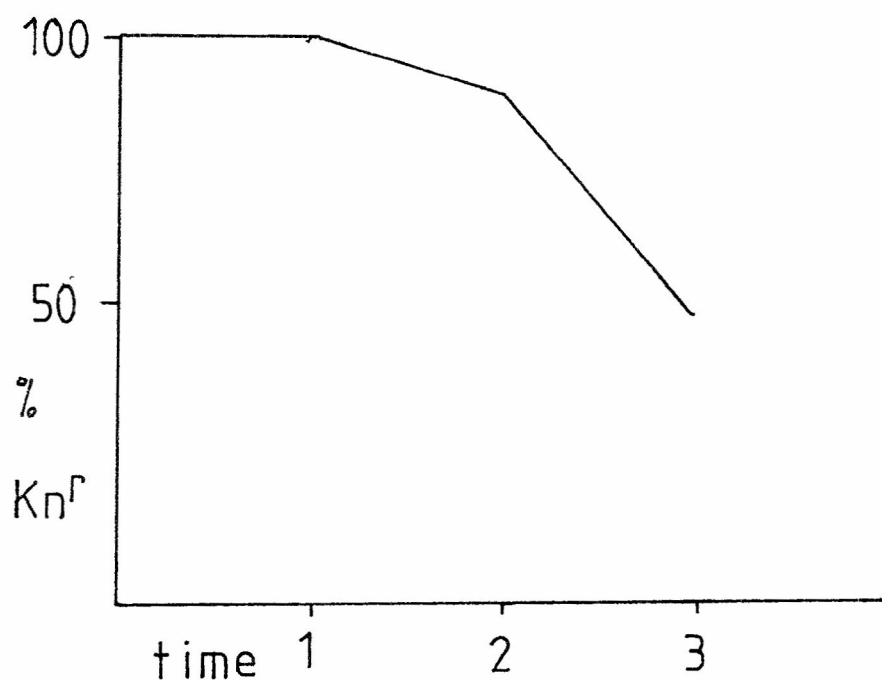


Fig 18.4 A Kn^r , Ap^r cosmid/pCRI recombinant was grown in a 10ml L.Broth culture containing ampicillin, and samples were taken at 1) Early log phase (4 hours after inoculation), 2) Mid log phase (7 hours after inoculation), 3) stationary phase (14 hours after inoculation). These were plated on ampicillin plates, colonies were picked and replated on Kanamycin plates. Plot of fraction of Am^r isolates which were also Kn^r .

Fig 18.5

Cloning several DNAs in cosmids

DNA	colonies
Ambystoma: liver	418
: blood	198
X. borealis: prep 1	588
: prep 2	198
: prep 1 } precipitated	49
: prep 2 } from ethanol	45
E.coli K	> 2000
E.coli C	> 2000

Fig 18.5 400ng of exogenous DNA were ligated to 40ng of pHC79, packaged and plated as described. The colonies formed when ligations of the same amount of several exogenous DNAs were plated are listed ('colonies')

Fig 18.6

Cloning several DNAs in λ gtWES

λ DNA	DNA	plaques
λ cI857S7	-	113000
incubated in ligation buffer with X.borealis		119,200
λ gtWES	E.coli C	37800
X borealis : prep 1		750
prep 2		234
X.laevis		1425
X.b. + E.coli C	mixed before ligation	24,000
X.b. + E.coli C	mixed after ligation	126

Fig 18.6 400ng of exogenous DNA were ligated with 80ng of lambda-gtWES 'arms', packaged and plated as described. Some were then mixed with another exogenous DNA species not ligated to any vector. The plaques formed when ligations of the same amount of several exogenous DNAs were plated are listed ('plaques'). The packaging reactions involving lambda-cI857S7 used 90ng of lambda DNA which had not been

changing the packaging system used from that of Hohn (1975) to that of Enquist and Sternberg (1979) had no significant effect on the cloning efficiency.

The second set of experiments are results from the work which lead to the construction of the lambda-gtWES gene library described in section 6.3. Here a variety of DNAs are cloned in lambda-gtWES, and again *E.coli* DNA is much more efficient at producing plaques than Xenopus DNA. The ligation of X.borealis DNA 'prepl' and lambda-gtWES was subsequently processed as a gene library, and shown to be largely non-recombinant (vide supra section 6.3.1).

What is notable, however, is that if E.coli and X.borealis DNA are mixed before ligation to lambda-gtWES, the efficiency of packaging drops dramatically. If they are mixed after ligation, there is little drop in efficiency. Furthermore, if packagable lambda DNA is incubated for 48hours with X.borealis DNA in ligation conditions, no significant inhibition of packaging is seen.

Other results suggested that the X.borealis DNA was not degraded by the packaging procedure. Samples of DNA were mixed with packaging extracts, phenol-extracted and electrophoresed on agarose. The results are not clear, however, and are not presented here.

18.4 Conclusions, and a speculation.

I conclude that X.borealis DNA as prepared for these experiments contains some factor which inhibits the packaging reaction. All other steps in the cloning procedure work normally, and all other reagents work satisfactorily in model reactions. Any cloning using X.borealis DNA and lambda-derived packaging systems is apparently doomed to

partial failure. In the case of lambda-gtWES enough recombinants can be produced to make a useable gene library. In the case of cosmid cloning, which selects inserts 5 times the size of those optimal for lambda-gtWES, no gene library can be produced. The same effect is seen with Axolotl DNA and, to a lesser extent, X.laevis DNA.

The 'factor' is not diffusible, is not removed from the DNA by several standard purification procedures and only acts in cis. This points to an extremely tightly bound ligand or to a chemical modification of the DNA. In the latter respect it is notable that X.borealis DNA is very highly methylated in erythrocytes, X.laevis DNA slightly less so (Bird and Taggart 1980). No effect of extensive DNA methylation has been noted in in vitro lambda packaging systems to date, but then as no-one has had reason to look for one this is unsurprising. We might note that cosmid gene libraries have been successfully constructed from chick (Royal et al 1979), Human (Weiss et al 1981), mouse (Steinmetz et al 1982) and Drosophila (Hogness unpub) DNAs, none nearly as highly methylated as those of the Xenopinae, and of course from E.coli (this work), which, on the basis of fraction of total DNA bases methylated, is hardly methylated at all (van der Ploeg and Flavell 1980, Bird and Southern 1978, Bird and Taggart 1980, Doerfler 1981).

19. Appendix 2. Critical variables in lambda-gtWES cloning.

But the Barrister, weary of proving in vain
That the Beaver's lace-making was wrong
Fell asleep, and in dreams saw the creature quite plain
that his fancy had dwelt on so long.

19.1 Introduction.

In this brief section I will describe a theoretical analysis of the DNA restriction, ligation and packaging reactions which were performed during the generation of the X.borealis gene library described in section 6.3. This analysis was performed to see if an 8.5kb EcoRI fragment from the genome would be expected to be over-abundantly represented in the gene library just from the dynamics of cloning.

19.2 Methods.

Computer analysis was performed on the author's Nascom-2 microcomputer.

19.3 Results

19.3.1 The approach.

I assumed that EcoRI sites are randomly distributed in X.borealis DNA: with the exception of satellite sequences (Botchan, McKenna and Sharp 1973) this is a good approximation. The program

calculated the number of molecules of DNA in each of a series of size classes after partial EcoRI digestion, subtracted from this the number which will self-ligate to form closed circles (Duguic:yk, Boyer and Goodman 1975) and then computer the product of ligating this with a known mass of purified lambda-gtWES arms. (Sgarebella and Khorana 1972, Sugino et al 1977). I assumed that only molecules consisting of monomers of lambda left arms, Xenopus DNA and lambda right arms ligated in that order are packagable. The program scored these according to insert EcoRI size (Hohn 1975, Hohn and Murray 1977, Hohn and Hohn 1974).

Computing how efficiently these molecules actually are packaged into viable phage particles is difficult, as no consistent experimental data are available. Enquist and Sternberg (1979) and Maniatis et al (1978) using similar systems find that only a narrow band of DNA sizes are packageable. However these workers note that by adding spermidine and putrescine to the packaging system these narrow limits may be broadened considerably (Enquist and Sternberg 1979). In this case they match those of the in vitro system developed by Hohn (Hohn and Hohn 1974, Hohn 1975, Hohn and Murray 1977, Patient et al 1980), which uses putrescine and spermidine, and those of Feiss et al (Feiss et al 1977, Feiss and Siegele 1979) using an in vivo system, packaging DNA between 80% and 105% of the lambda-cI857S7 DNA size with fairly uniform efficiency. The system of Enquist and Sternberg (1979) as modified by Ish-Horowitz (per com) used in this study also uses spermidine and putrescine, and so this latter, broad size response was assumed for this study.

19.3.2 The model results.

The results are presented in Fig 19.1 . The distribution of sizes

Fig 19.1

Output from cloning model

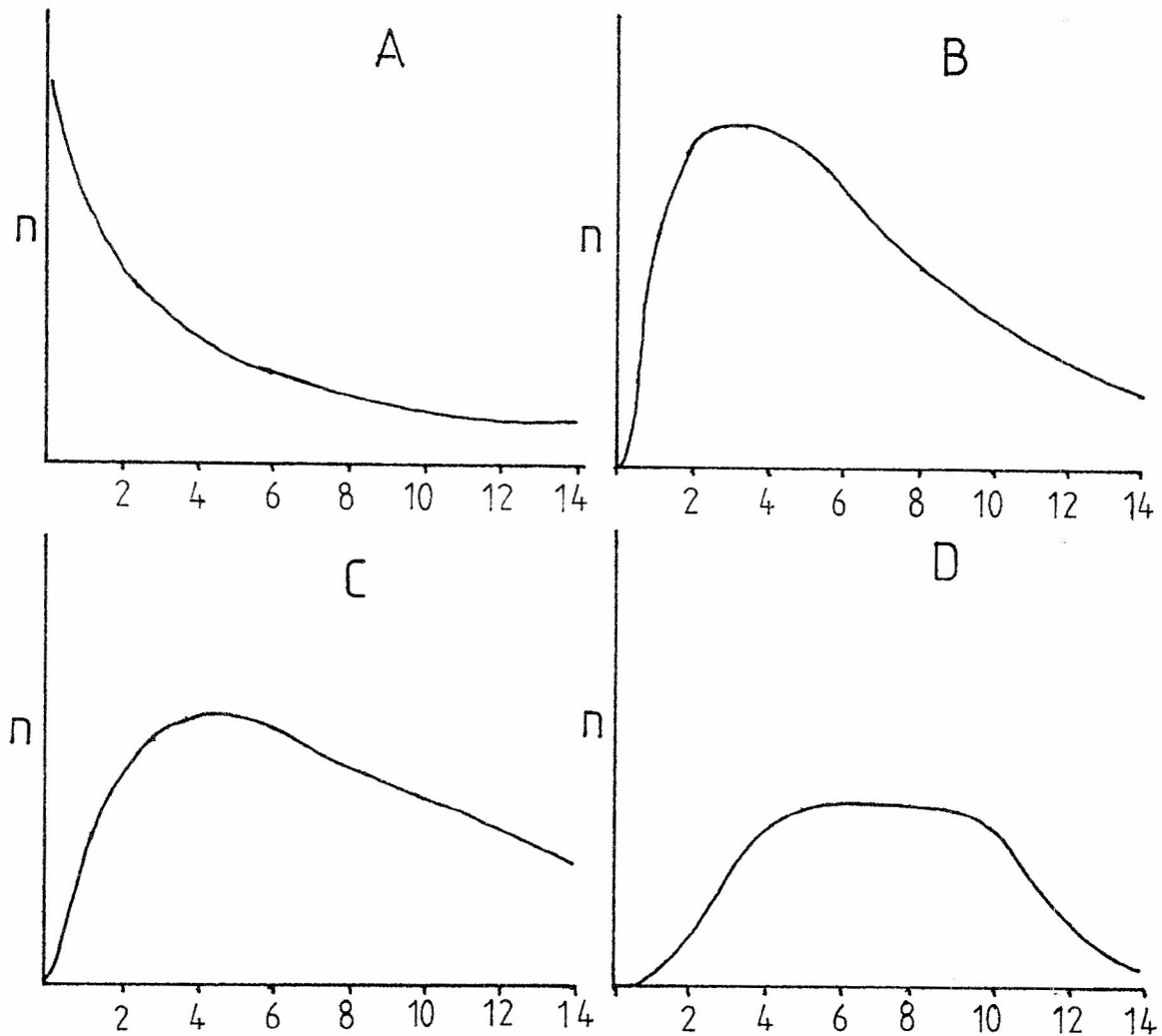


Fig 19.1 Presents output from the computer model. Vertical axes - relative number of molecules, horizontal axes - length of EcoRI fragments/ inserts in kilobases. A) Restriction digest with EcoRI. B) linear remnant of A) after self-ligation of small molecules. C) Potentially packageable molecules after B) had been ligated to lambda-gtWES arms. D) relative efficiency with which C) is packaged into phage, measured as the likelihood of a given size of EcoRI fragment occurring in a phage divided by the likelihood of it occurring in the initial digest A.

predicted fits with the limited data in section 6.3.5 concerning the actual distribution obtained, suggesting that the model is not severely inaccurate. The final distribution is a measure of the frequency with which a given EcoRI fragment size occurs in the gene library divided by the frequency with which it occurs in the initial digest, ie the efficiency with which it is cloned. Although this clearly shows the system to be more efficient at cloning fragments of 6-12kb, the effect is hardly startling, considering that ~40% of the X.borealis EcoRI-digested DNA lies in this range. Thus, although the cloning system used could have caused Xbhl to be over-represented in the gene library 1-3 times, the size-specificity of the ligation and packaging reactions can not explain the observed gross over-representation of this sequence.

20. Appendix 3. Genomic representation data for IRI and IR2.

Taking three as the number to reason about,
 A convenient number to state,
 We add seven and ten, and then multiply out
 By one thousand diminished by eight.....

20.1 Introduction.

Hybridisation of a labelled probe to DNA bound to nitrocellulose filters is an established method for establishing the number of a given type of sequence in a mixture of sequences, usually genomic DNA or an RNA population (Kafatos et al 1979, Britten, Graham and Neufeld 1981). It is not as accurate as solution hybridisation methods (Britten, Graham and Neufeld 1981), but it is quicker and produces useable results if performed carefully.

In this study probes containing cloned DNA were hybridised to a mixture of cloned and genomic DNAs bound to nitrocellulose (vide supra section 10.3.2). I shall assume that DNA hybridises equally fast to all homologous sequences on the filter in any given hybridisation: this may not necessarily be true as fragment length (Wetmur and Davidson 1968), probe secondary structure (Studier 1969) and sequence divergence (Bonner et al 1973) can all affect the rate of reassociation, but it is a reasonable approximation here where at least the first two variables have been largely eliminated.

Consider then a filter with m copies of pWB1 on it and n copies of the X.borealis genome, of size 5.45kb (Bolivar 1978) and 3250 Mb (Bisbee et al 1977) respectively. Note that if there are

two copies of the probe sequence in pWB2, the effective concentration of pWB2 (m) is doubled. The samples have been sonicated to approximately the same size to eliminate the effect of chain length of hybridisation rate (Wetmur and Davidson 1968). F bases of a probe of total length L hybridise to an homologous sequence in the genome, while its whole length L hybridises to the pWB1 mixed in with the genome. Thus the total number of counts bound to a filter will be given by:

$$\begin{aligned} \text{Total Counts} = & \text{specific activity} \times \text{efficiency of binding} \\ & \times \text{number(homologous sequences)} \end{aligned}$$

Assume specific activity and efficiency of hybridisation are the same for all sequences in one hybridisation.

Then $\text{Counts} = K(L.m + F.n)$ Where $K = \text{unquantified constant}$

If $L.m \gg F.n$, ie there is a large excess of pWB1 (filter A)

$$\text{Counts} = K.L.m = A$$

If $L.m = 0$, ie there is no pWB2

$$\text{Counts} = K.F.n = B$$

Hence the total length of complementary DNA in the genome

$$= F.n = A/B \times L.m$$

Thus by plotting a graph of [pWB1] vs counts bound to the filter, the highest [pWB1] gives A and the intercept gives B. In practice I plotted these values as Counts vs log [pWB1], as this made the points more easily visible over the range of concentrations used.

20.2 Results.

The data are given in Fig 20.1. The value for A is off the graphs, whose scale has been adjusted to show the intercept only. These intercept values, A and corrected B values with the background of non-specific binding subtracted have been tabulated in

Fig 20 1 : prologue

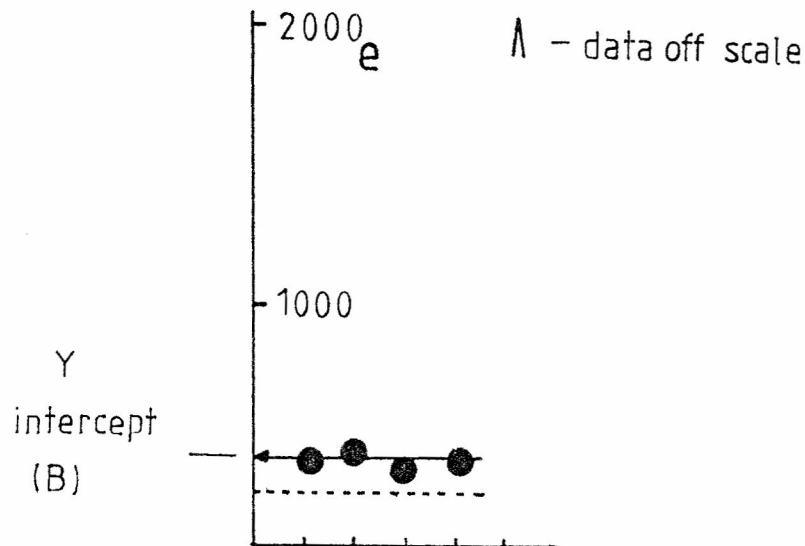


Fig 20.1 Presents the data used to calculate the genomic representation of IR2. Above is an example of one of the graphs. The figures are a plot of counts/10 mins (Y axis) vs filter number (X axis). The filters were loaded with 7.56µg of X.borealis genomic DNA and 0, 0.37ng, 3.7ng, 37ng, 370ng of pWB2 respectively. A sixth filter, loaded with 3700ng of pWB2 is not shown in this figure: results for this filter ('A' in the calculations) are presented in Fig 20.2 . The graphs show these five data points, the Y-axis intercept ('B' in the calculations, the solid line) and the background counts bound to the filters. (dotted line). Eight probes are used: restriction fragments a-g as described in section 10 and pAT153 (h) as a control. There are seven sets of 8 graphs, one for each wash temperature. The vertical scale is the same for each particular temperature, and is marked on the 'a' graph only.

These axes are chosen to display the relative values of the Y-axis intercept ('B'). Thus they are not suited to display all the data, which ranges over several orders of magnitude. Data points which lie off the plotted regions are indicated by an arrow.

Fig 20.1

Data to measure IR2's Representation

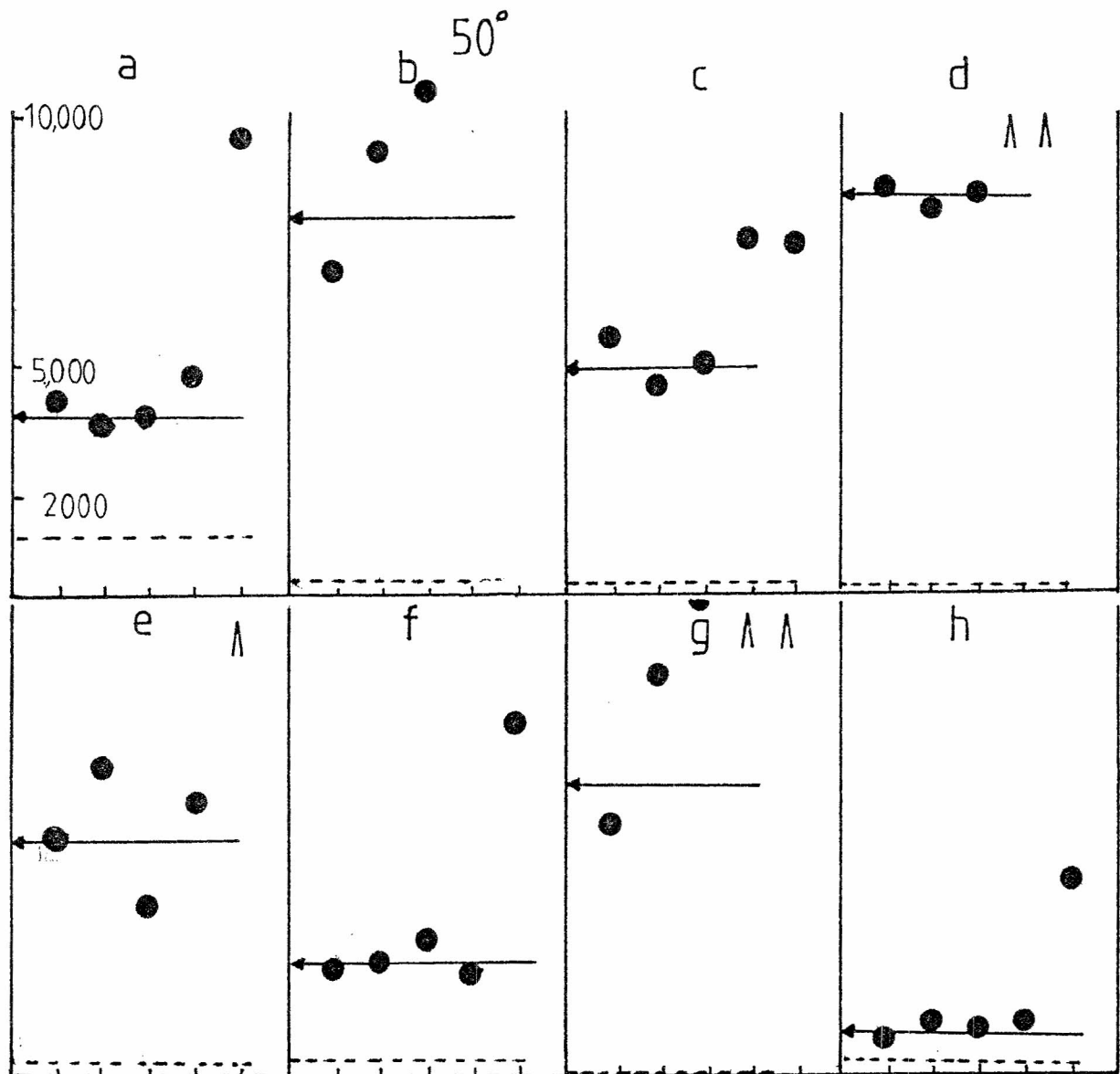


Fig 20.1 cont
55°

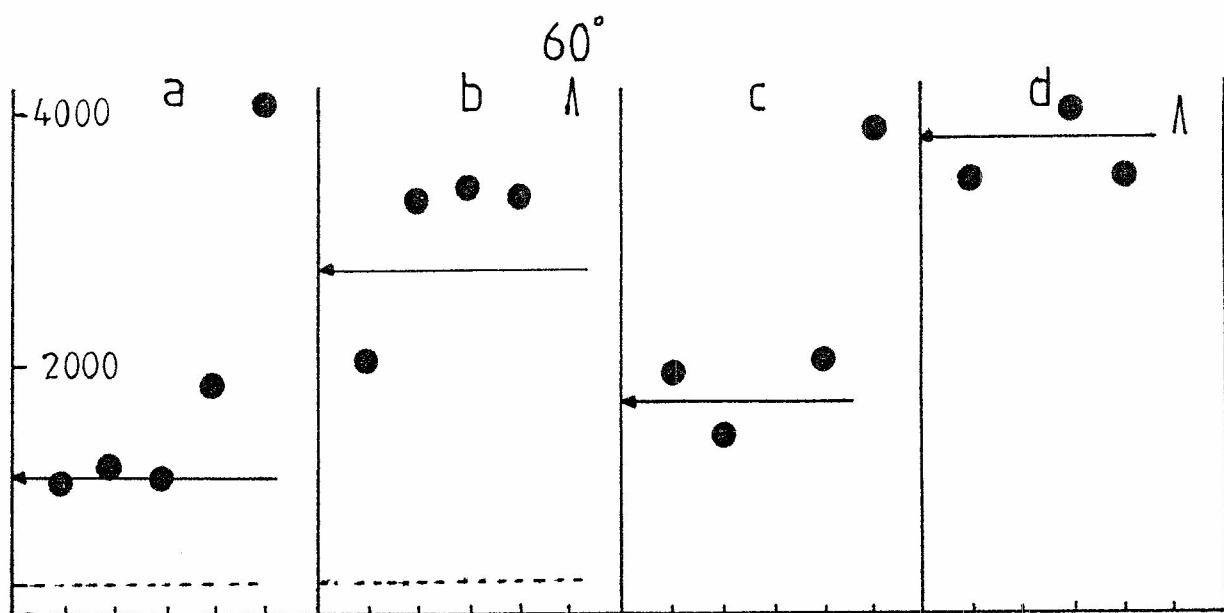
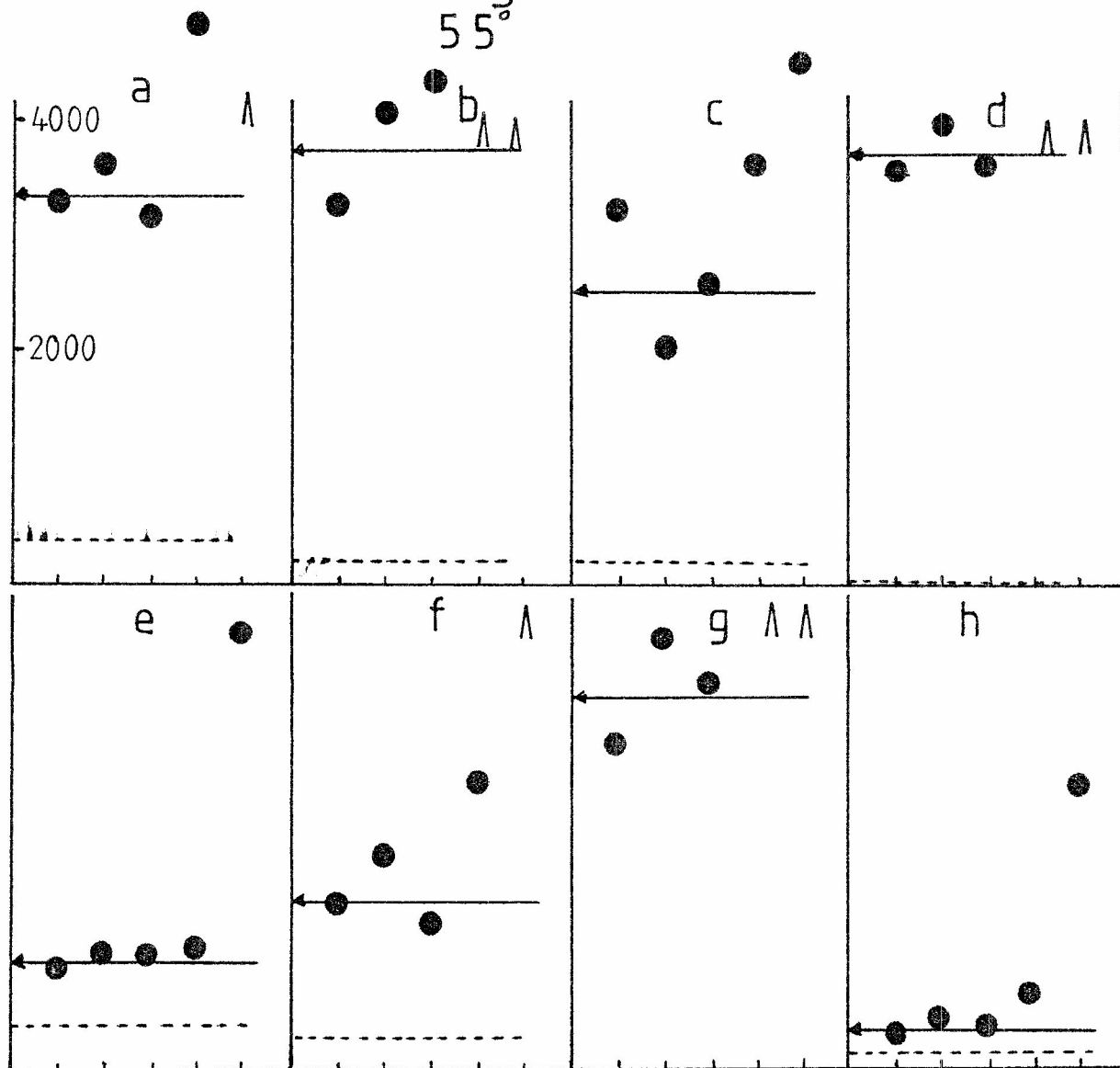
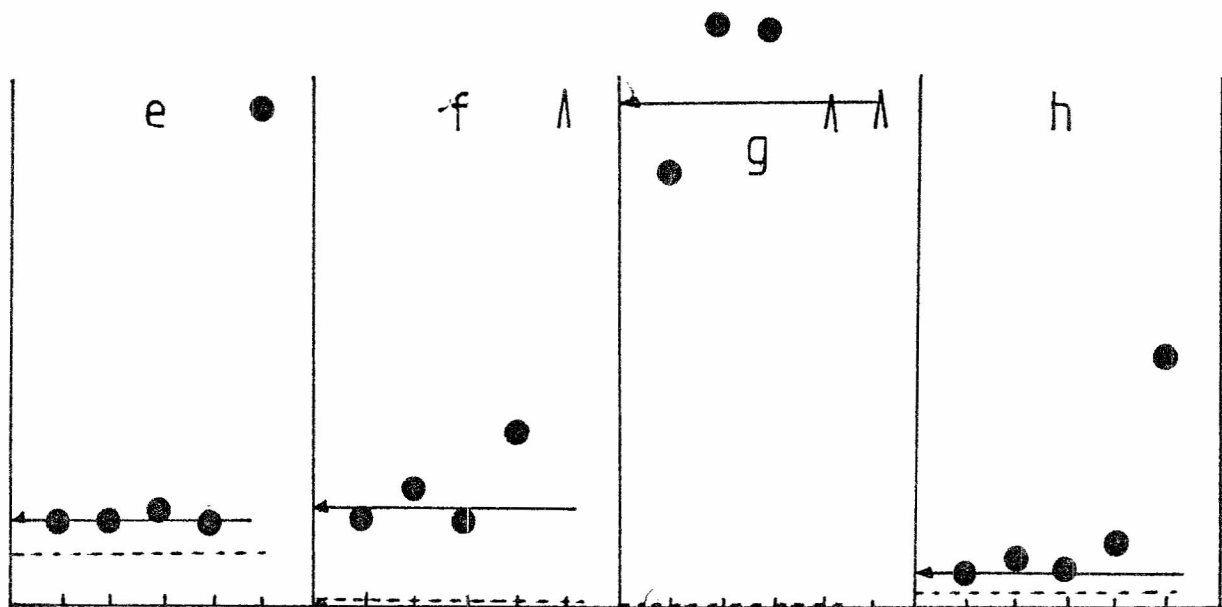


Fig 20.1 cont



67°

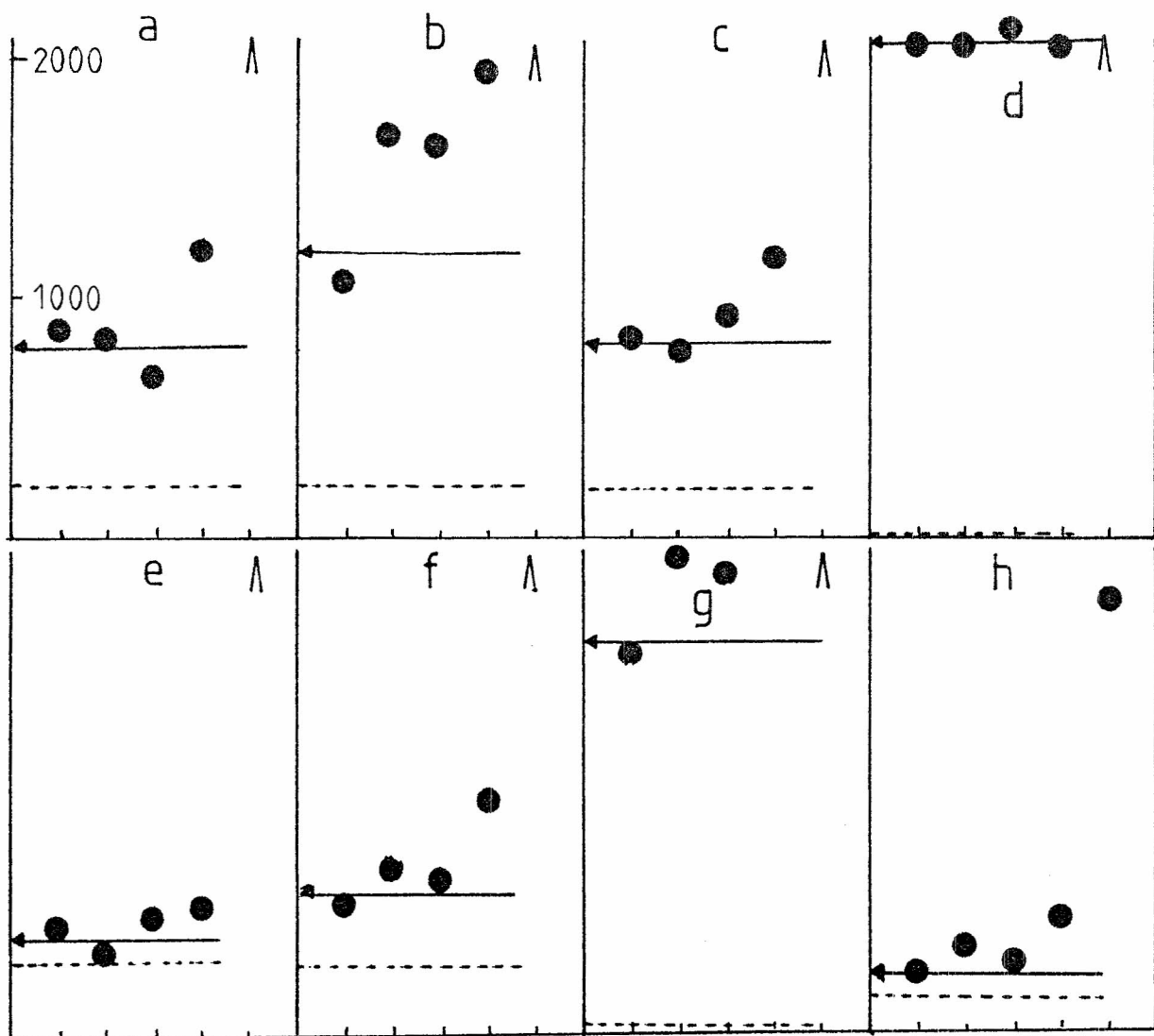


Fig 20.1 cont

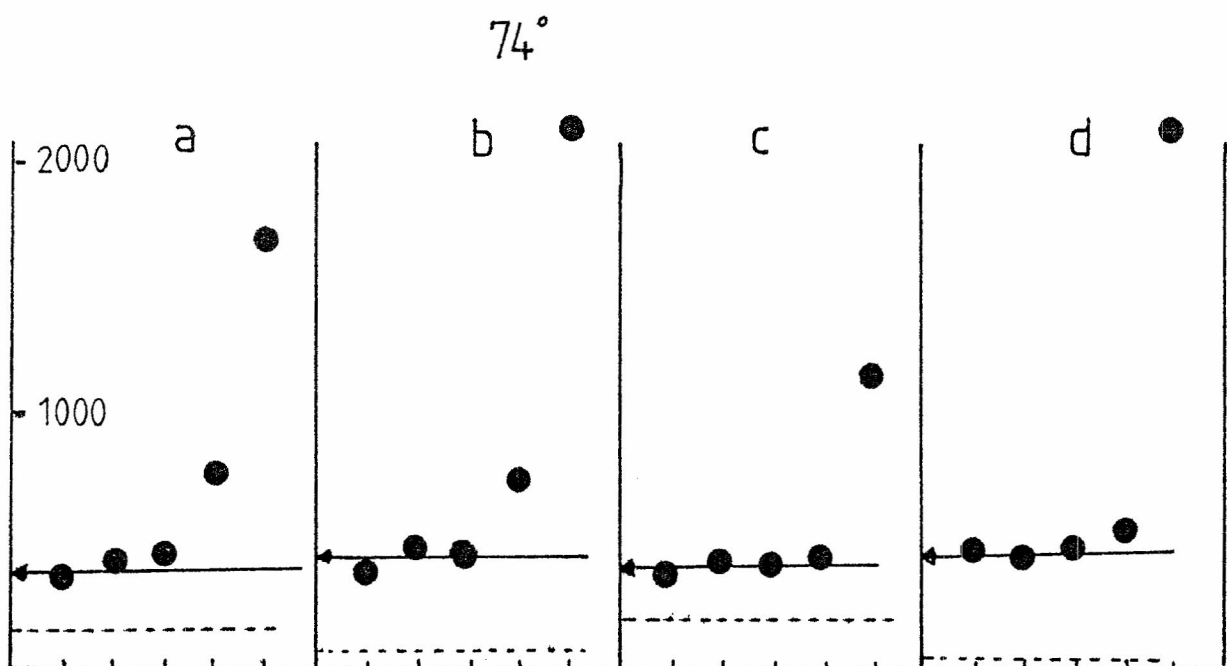
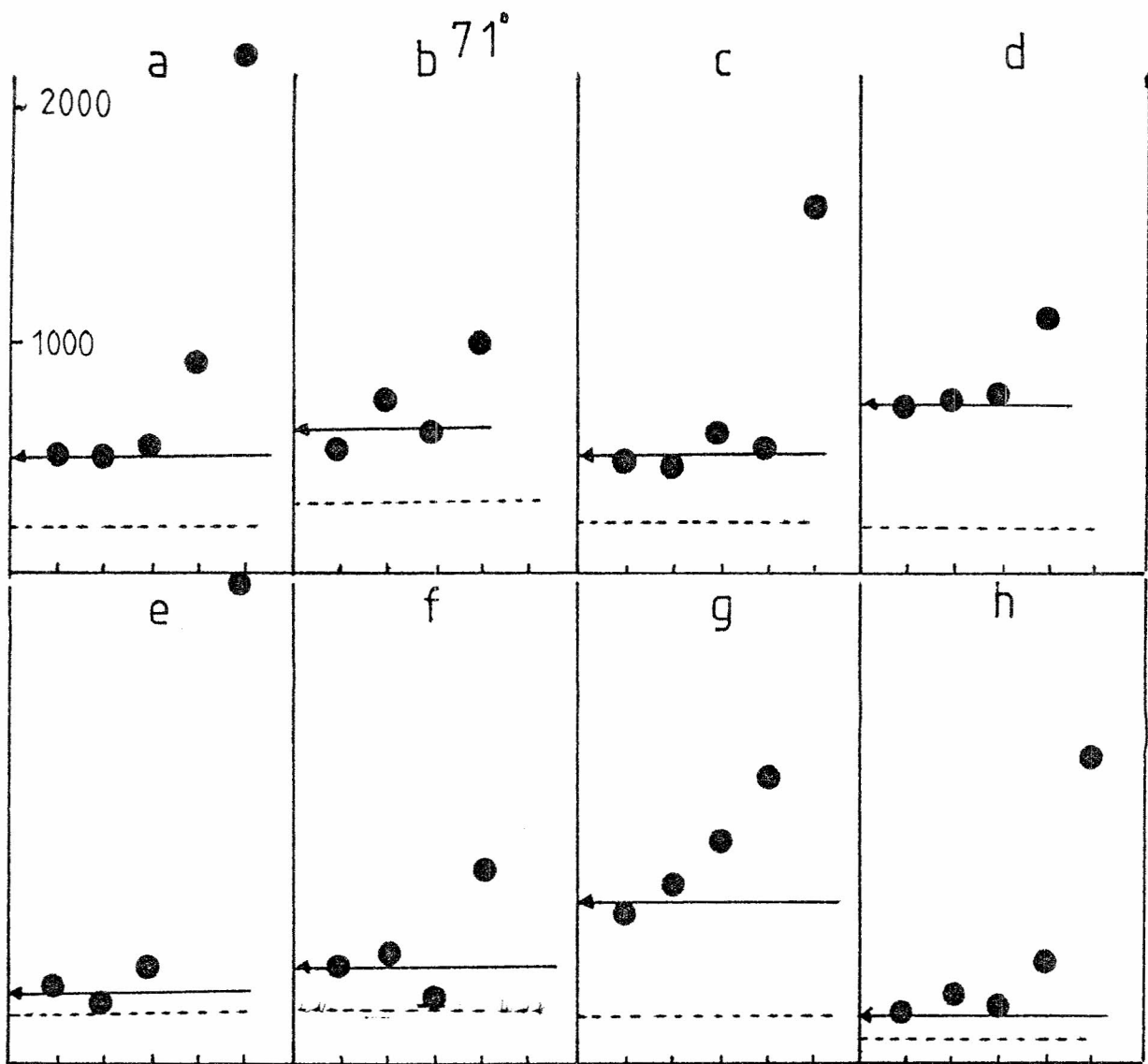
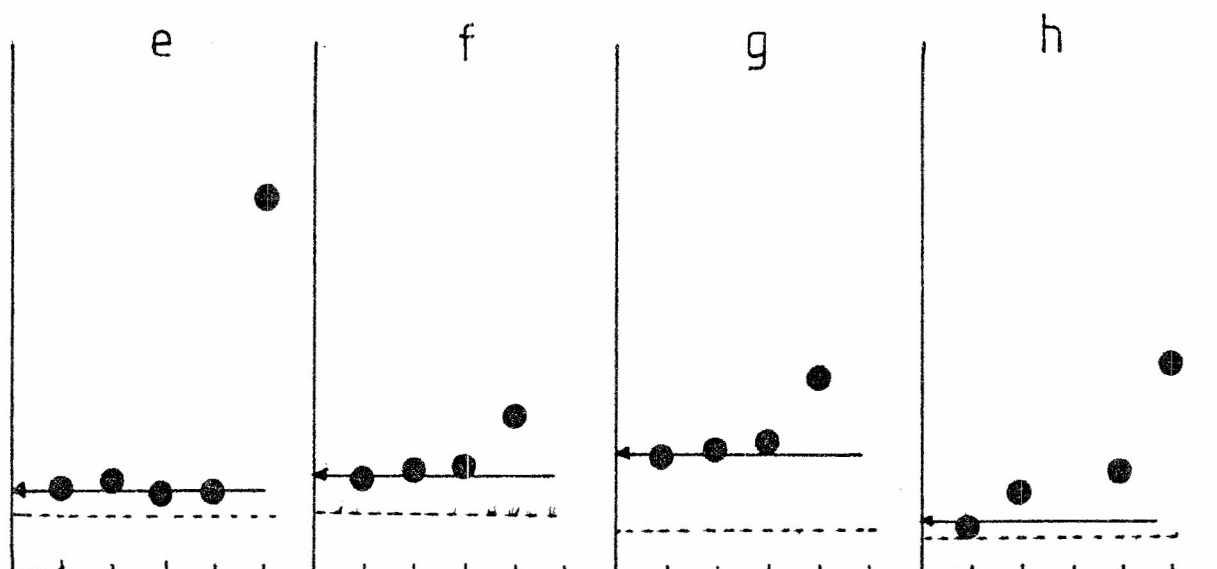


Fig 20-1 end



78°

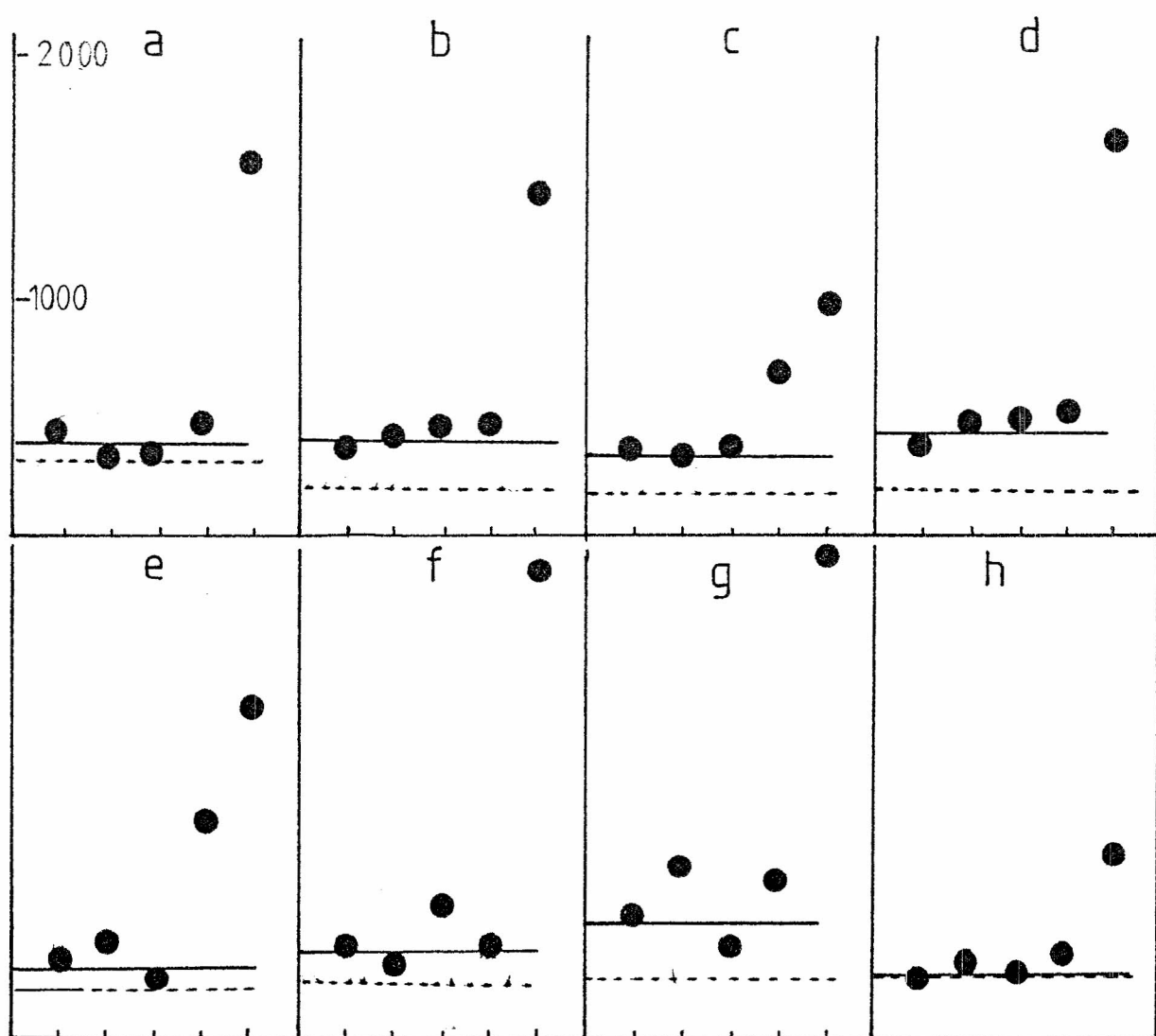


Fig 20.2 . From these figures, by application of the formulae above, the gene numbers listed in Fig 10.2 may be readily derived.

As well as these filters, I hybridised LF53 (Turner and Woodland 1982) to itself, lambda-Xlh19 (a lambda genomic H4-containing gene clone (Old unpub)) and pSp102 (the Sea urchin histone gene clone (Sures et al 1978)). These clones have sequence divergences from LF53 of 0%, 3% and 17% respectively in the H4-coding regions. The relative number of counts on these hybridisations, corrected for non-specific DNA-binding, at various temperatures are given in Fig 20.3 . This confirms the observations of Turner and Woodland (1982) that 1% mismatch causes $\approx 2^{\circ}\text{C}$ drop in T_m for well-matched sequences: the relationship may not be effective for highly divergent sequences. This gives an indication of the degree of mismatch being detected when the apparent number of copies of sections of Xbh1 decreases with increasing hybridisation temperature. It also shows why the gene numbers measured above 78°C are not reliable in this system: nearly all the internal control has been lost at this temperature, as the probe no longer hybridises well to pWB1.

Fig 20.2

Data for genomic representation

	Filter number						
	a	b	c	d	e	f	g
Temp.							
50 °C							
'A'	75000	84000	39000	86000	27800	41200	62500
'B'	3400	7450	4200	17400	1700	3580	15800
55 °C							
'A'	45000	48200	26000	56000	19000	25800	44000
'B'	1280	3300	2150	8080	360	950	6100
60 °C							
'A'	34000	37000	18500	39200	15500	19500	38500
'B'	660	2300	1250	4920	90	500	3870
67 °C							
'A'	34000	32600	16500	33800	12300	16800	30000
'B'	460	770	500	2040	0	200	1490
71 °C							
'A'	25500	24100	9900	21700	10500	12100	22000
'B'	210	190	190	400	0	80	390
74 °C							
'A'	18400	15800	7550	18800	7480	8400	15800
'B'	160	50	140	220	40	90	250
78 °C							
'A'	12800	12950	5790	12719	5050	5800	11000
'B'	60	170	140	210	90	90	210

Fig 20.2 Values of the Y-axis intercept ('B') from Fig 20.1, and 'A' for each of the seven filters probed with restriction fragments of Xbhl. These values have been corrected for the background, non-specific label-to-DNA binding revealed by the pAT153 control filters. From this data the genomic representation figures shown in Fig 10.2 may be directly calculated.

Fig 20.3

Hybrid melting curves

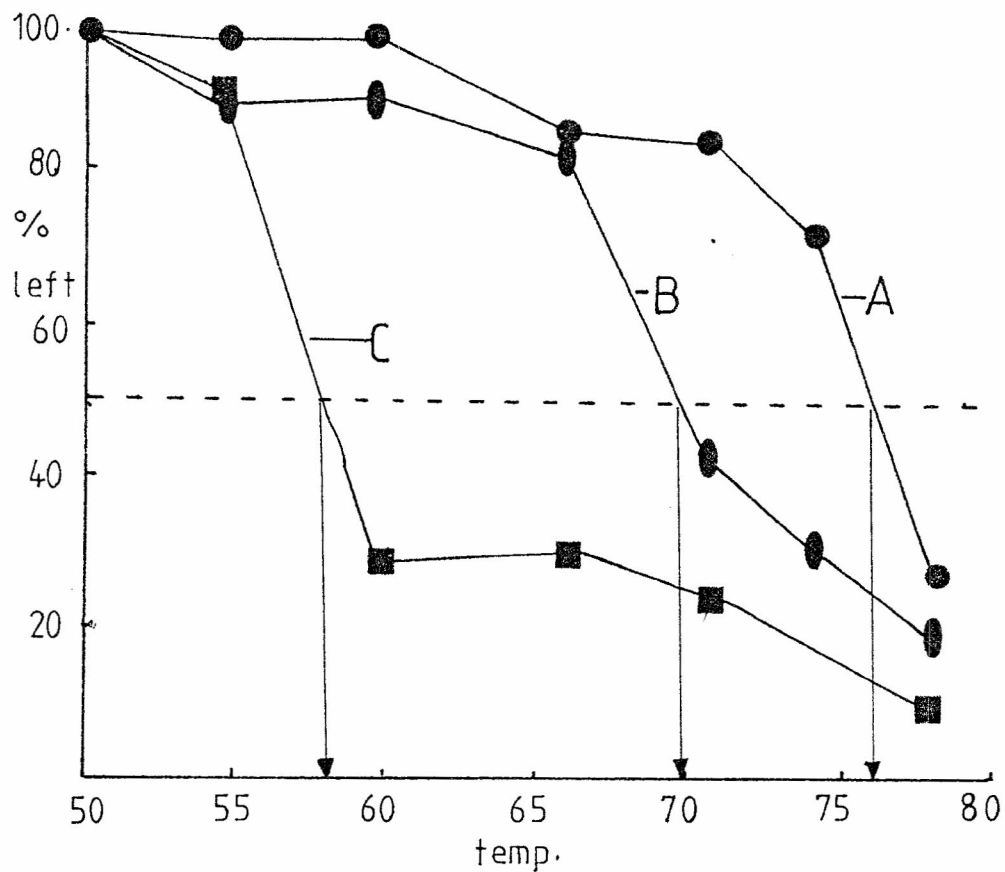


Fig 20.3 pcX1H4W1 was used to probe filters to which A) pcX1H4W1, B)X1h19, c)pSpl02 had been bound. The filters were washed at successively higher temperatures, and the bound radioactivity counted by cherenkov counting. These are plotted above as % of the counts bound at 50°C, corrected for non-specific binding. The sequence divergence between the H4-coding regions of these clones is A)0%, B)3%, C)17%. The temperature at which 50% of the label has been washed off the filters is indicated.

21. Appendix 4. Sequence differences determined by restriction.

...The result we proceed to divide, as you see,

By nine hundred and ninety and two:

Then subtract seventeen, and the answer must be

Completely and perfectly true.

21.1 Introduction.

The problem is how to relate differences observed in a restriction digest of two similar DNAs to the degree of sequence difference between them, without having to construct a restriction map of the DNAs. The problem falls into two parts: how many band differences are generated in a restriction digest by a given number of restriction site differences, and what average number of base changes will generate a known number of site differences.

21.2 Band differences caused by a new site.

I will not consider large-scale deletions or insertions for a moment. So:-

consider a restriction fragment of length L , into which we introduce n new sites. Clearly, the band L will disappear, and $n+1$ new bands will appear. The number of bands not shared between these two species will therefore be $n+2$. Thus the number of band differences caused by a site difference depends on how many site differences are likely to appear between two old sites. This analysis is, of course, just as true of sites 'disappearing' as of ones 'appearing'.

In general,

$$\begin{aligned}
n(\text{band changes}) &= n(\text{changes caused by one new site/frag}) \\
&\quad + n(\text{changes caused by two new sites/frag}) \\
&\quad + n(\text{changes caused by three new sites/frag}) \\
&\quad + \dots \\
&= n(\text{changes caused by one site}) \times (\text{prob}(\text{one site occurring})) \\
&\quad + n(\text{changes caused by two sites}) \times (\text{prob}(\text{two sites occurring})) \\
&\quad + \dots = \sum_{n=1}^L \frac{n+2}{n} \cdot {}^L C_n \cdot (p)^n \cdot (1-p)^{(L-n)}
\end{aligned}$$

where p is the probability that a given point in the DNA will be mutated to form a new site. By numerical solution of the above for low p , we can calculate that the mean number of band changes caused by one site change will be @2.7 .

Thus if the DNAs are very similar, the differences between them are number of band differences/ 2.7 site differences.

21.3 Site differences caused by base differences.

Restriction enzymes do not recognise the whole sequence of a stretch of DNA, only a sub-set of it. Consider a 4-cutter. This cuts at a specific 4-base sequence, which will occur once every 256 bases on average. Hence it recognises $4/256 = 0.0156$ of the total length of a DNA molecule. However, it will also just fail to recognise a whole lot of other sequences which have 3 out of the 4 bases correct, but not the fourth. 1 out of 3 mutations in this fourth base will give rise to a new site, thus also generating a restriction difference. 1/64 bases will be part of 'nearly-4-cutter sites', and 1/3 of mutations in these will also be detected. Thus in all the 4-cutter will detect changes in 3.5% of a given length of DNA. Thus one change in a 4-cutter restriction map represents an average of 50 changes in the whole clone. Similarly, 6-cutters detect changes in 0.34% of the

sequence.

Thus if two clones share all except 8 bands generated by one particular 4-cutter, then these represent $8/2.7 \approx 3$ site changes, themselves representing $3 \times 27 = 80$ base differences between the clones. If these clones are 8kb long, then this is about 1% mismatch.

This analysis takes no account of deletions or insertions, which always affect the sizes of the restriction fragments in which they occur even if no new sites are created or destroyed. These may be detected by comparing 4-cutter and 6-cutter restriction patterns. Clearly from the above there should be ≈ 10 times as many 4-cutter band differences as 6-cutter ones: a surplus of 6-cutter differences points to a deletion or insertion. The bands generated or removed by this should be discounted from the calculation of the point-mutation divergence. Of course, this is a rather vague caveat, as it only applies to insertions or deletions which give detectable changes in the migration of bands in the gel system being used. If the change is too small to be picked up (say 10bp in a 2kb fragment) it is operationally equivalent to a point mutation. This is a limitation imposed by the electrophoresis technology, and limits the sort of DNA fragments which may be analysed by the above reasoning. However this limit need not concern us here. Discounting differences in the 6-cutter pattern of Xbhl vs Xlh8 due to the known deletion in the latter, there is only one 6-cutter difference between the two clones. This compares with 4-8 with any 4-cutter on a comparable gel system, which suggests that there are no further major deletions or insertions in either clone relative to the other, and that the differences being detected are effectively point differences. This is borne out by studies of the H1 gene region (section 15.1).

References

References

- Adamson, E.D. and Woodland, H.R. (1974) "Histone synthesis in early amphibian development: histone and DNA synthesis are not co-ordinated." J. Mol. Biol 88, 263-285
- Adamson, E.D. and Woodland, H.R. (1977) "Changes in the rate of histone synthesis during oocyte maturation and very early development of X.laevis." Dev Biol. 57, 136-149
- Adolph, K.W., Cheng, S.M. and Laemmli, U.K. (1977) "Role of non-histone proteins in metaphase chromosome structure." Cell 12, 805-816
- Albertini, R.J. and De Mass, R. (1973) "Detection and quantification of X-ray-induced mutation in cultured, diploid human fibroblasts." Mutat. Res. 1, 199-224
- Allan, J., Harborne, H., Rau, D.C. and Gould, H. (1982) "Participation of core histone 'tails' in the stabilisation of the chromatin solenoid." J. Cell. Biol. 93, 285-297
- Amaldi, F., Beccari, E., Bozoni, I., Luo, Z.X. and Pierandrei-Amaldi, P. (1982) "Nucleotide sequence of cloned cDNA fragments specific for Xenopus laevis ribosomal proteins." Gene 17, 311-316
- Angerer, R.C., Davidson, E.H. and Britten, R.J. (1975) "DNA sequence organisation in the mollusc Aplysia californica." Cell 6, 29-39
- Arnberg, A.C., Meijlink, F.C.P.W., Mudler, J., van Bruggen E.J.F., Gruber, M. and Geert, A.B. (1981) "Isolation and characterisation of genomic clones carrying the chicken vitellogenin gene." Nucl. Acid Res.

9, 3271-3286

Artzt, K., McCormick, P. and Bennett, D. (1982) "Gene mapping within the T/t complex of the mouse . 1 t-lethal genes are non-allelic." Cell 28, 463-470

Asao, T. (1969) "Behaviour of histones and cytoplasmic basic proteins during embryogenesis of the Japanese newt Triturus pyrrhogaster." Exp. Cell Res. 58, 243-254

Asao, T. (1970) "Regional histone changes in embryos of the newt Triturus pyrrhogaster during early development." Exp. Cell Res. 61, 255-265

Axel, R., Feigelson, P. and Schutz, G. (1977) "Analysis of the complexity and diversity of mRNA from chicken liver and oviduct." Cell 7, 247-254

Baldacci, P., Royal, A., Bregegere, P., Abastader, J.P., Canci, B., Daniel, F. and Kourilksy, P. (1981) "DNA organisation in the chicken lysozyme gene region." Nucl. Acid Res. 9, 3575-3581

Bantle, J.A. and Hahn, W.E. (1976) "Complexity and characterisation of polyadenylated RNA in mouse brain." Cell 8, 139-150

Barnes, R.D. (1974) "Invertebrate Zoology." (W.B.Saunders)

Barnstable, C.J., Jones, E.A. and Bodmer, W.F. (1979) "Genetic structure of the major histocompatibility regions." in MTP Int. Rev. Bioch II vol 22, 151-226

References

Beadet, A.L., Roufa, D.J. and Cashey, C.T. (1973) "Mutations affecting the structure of HGPRT in cultured chinese hamster cells." P.N.A.S. 70, 320-324

Bell, G.I., Selby, M.J. and Rutter, W.J. (1982) "The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences." Nature, 295, 31

Benezra, R., Blankstein, L.A., Stollar, B.D. and Levy, S.B. (1981) "Immunological and organisational heterogeneity of histone H2A variants within chromatin of cells at different stages of Friend leukemia." J. Biol. Chem. 256, 6837-6841

Benoist, C. and Chanbon, P. (1981) "In vivo requirements of the SV40 early promoter region." Nature 290, 304-310

Benton, W.D. and Davis, R.W. (1977) "Screening lambda-gt recombinant clones by hybridisation to single plaques in situ." Science 196, 180-182

Benyajati, C. and Worcel (1976) "Isolation, characterisation and structure of the folded interphase genome of Drosophila melanogaster." Cell 9, 393-407

Berg, B.L. (1982) "High mutability and hybrid dysgenesis." Genetics 100, s4

Bernard, O., Honami, N. and Tonegawa, S. (1978) "Sequences of mouse immunoglobulin light chain genes before and after somatic changes." Cell 15, 1133-1144

References

- Bernards, R. and Flavell, R.A. (1980) "Physical mapping of the globin locus deletion in H.P.F.H." Nucl. Acid Res. 8, 1521-1535
- Bertlesen, A.H., Humayin, M.Z., Karfopoulos, S.G. and Rush, M.G. (1982) "Molecular characterisation of small polydisperse circular DNA from an African green monkey cell line." Biochem. 21, 2076-2085
- Birchmeier, C., Grosschedl, R. and Birnstiel, M.L. (1982) "Generation of authentic 3' termini of an H2A mRNA in vivo is dependant on a short inverted repeat and on spacer sequence." Cell 28, 239-245
- Bird, A.P. (1978) "Use of restriction enzymes to study eukaryotic DNA methylation." J. Mol. Biol. 118, 49-60
- Bird, A.P. (1980) "DNA methylation and the frequency of CpG in animal DNA." Nucl. Acid Res. 8, 1499-1504
- Bird, A.P. and Taggart, M.H. (1980) "Variable pattern of total DNA and rDNA methylation in animals." Nucl. Acid Res. 8, 1485-1497
- Bird, A.P., Taggart, M.H. and Smith, B.A. (1979) "Methylated and unmethylated DNA compartments in the sea urchin genome." Cell 17, 889-901
- Bird, A.P., Taggart, M. and McCleod, D. (1981) "Loss of ribosomal DNA methylation accompanies the onset of ribosomal gene activity in early development of X.laevis." Cell 26, 381-390

References

- Biro, P.A., Carr-Brown, A., Southern, E.M. and Walker, M.B. (1975)
"Partial sequence analysis of mouse satellite DNA: evidence for
short-range periodicities." J.Mol. Biol. 94, 71-86
- Bisbee, C.A., Baker, M.A., Wilson, A.C., Hadji-Azimi, I. and
Fischberg, M. (1977) "Albumin phylogeny of the clawed toad X.laevis."
Science 195, 785-787
- Bishop, J.O., Morton, J.G., Rosbash, M. and Richardson, M. (1974)
"The abundance classes in Hela cell mRNA." Nature 250, 199-204
- Bishop, J.M. (1981) "Enemies within: the genesis of retrovirus
oncogenes." Cell 23, 5-6
- Birnie, G.D. and Rickwood, D. (1974) "Centrifugal separations in
molecular and cell biology." p178
- Bogart, J.P. (1980) "Evolutionary implications of polyploidy in
amphibians and reptiles." in 'Polyploidy', ed W.H.Lewis, pp341-378
- Bohm, L., Briand, G., Sautiere, P. and Crane-Robinson, C. (1982)
"Proteolytic digestion studies of chromatin core-histone structure."
Eur. J. Bioch. 123, 299-303
- Bolivar, F. (1978) "Construction and characterisation of new
cloning vehicles III." Gene 4, 121-136
- Bolivar, F. and Backman, K. (1979) "Plasmids of E.coli as cloning
vehicles." Met. in Enz. 68, 245-267

References

- Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlack, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) "Construction and characterisation of new cloning vehicles II." *Gene* 2, 95-113
- Bonner, T.I., Brenner, D.J., Neufeld, B.R. and Britten, R.J. (1973) "Reduction in the rate of DNA reassociation by sequence divergence." *J.Mol.Biol* 81, 123-135
- Bonner, W.M. and Laskey, R.A. (1974) "A film detection method for tritium-labelled proteins and nucleic acids in polyacrylamide gels." *Eur. J. Bioch.* 46, 83-88
- Borun, T.W., Ajiro, K., Zweidler, A., Dolby, T.W. and Stephens, R.E. (1977) "Studies on human histone mRNA." *J. Biol. Chem.* 252, 173-180
- Boseley, P., Moss, T., Machler, M., Portmann, R. and Birnstiel, M. (1979) "Sequence organisation of the spacer DNA in a ribosomal gene unit of X.laevis." *Cell* 17, 19-31
- Botchan, M., McKenna, G. and Sharp, P.A. (1973) "Cleavage of mouse DNA by a restriction enzyme as a clue to the arrangement of the genes." *C.S.H.S.Q.B.* 38, 383-395
- Bonni, I., Beccari, E., Luo, Z.X. and Amaldi, F. (1981) "Xenopus laevis ribosomal protein genes: isolation of recombinant cDNA clones and study of the genomic organisation." *Nucl. Acid Res.* 9, 1069-1086
- Bradbury, E.M. (1976) "Current ideas on the structure of chromatin." *Trends in Bioch. Sci.* Jan 1976

References

Brack, C., Hiram, M., Lenhard-Schuller, R. and Tonegawa, S. (1978) "A complete immunoglobulin gene is created by somatic recombination." *Cell* 15, 1-14

Brandt, W.F., Patterson, K. and van Holt, C. (1980) "The histone genes of yeast." *Eur. J. Bioch.* 110, 67-76

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

Brandt, W.F. and von Holt, C. (1982) "The primary structure of Yeast histone H3." *Eur. J. Bioch* 121, 501-510

Brasch, K. (1976) "Studies on the role of histones H1 and H5 in chromatin structure." *Exp. Cell Res.* 101, 396-410

Breathnach, R. and Chanbon, P. (1981) "Organisation and expression of eukaryotic split genes coding for proteins." *An. Rev. Bioch.* 50, 349-383

Bredehorst, R., Wielckens, K., Adamietz, P., Steinhagen-Thiessen, E. and Hill, H. (1981) "Mono(ADP-ribosyl)ation and Poly(ADP-ribosyl)ation of proteins in developing liver and hepatomas." *Eur. J. Bioch.* 120, 267-274

Brenner, (1974) "The genetics of Caenorhabditis elegans." *Genetics* 77, 71-94

References

Bridges, B.A. and Huckle, J. (1970) "Mutagenesis of cultured mammalian cells by X-radiation and UV light." *Mutat. Res.* 10, 141-151.

Bridges, B.A., Huckle, J. and Ashwood-Smith, M.J. (1970) "X-ray mutagenesis of cultured chinese hamster cells." *Nature* 226, 184-185

Briggs, R.C., Rothstein, H. and Wainwright, N. (1976) "Cell cycle variation in chromosomal proteins of the lens epithelium." *Exp. Cell Res.* 99, 95-105

Brison, O., Ardeshir, F. and Stark, G.R. (1982) "General method for cloning amplified DNA by differential screening with genomic probes." *Molec. and Cell. Biol.* 2, 578-587

Brisson, N. and Verma, D.P.S. (1982) "Soyabean leghemoglobin gene family: normal, pseudo and truncated genes." *P.N.A.S.* 79, 4055-4059

Britten, R.J. and Davidson, E.H. (1969) "Gene regulation for higher cells: a theory." *Science* 165, 349-357

Britten, R.J., Graham, D.E. and Neufeld, B.R. (1981) "Analysis of repeating DNA sequences by reassociation." *Meth. in Enz.* 21, 363-418

Brown, D.F. (1966) "X-ray induced mutations in extracellular phage." *Nature* 212, 1595-1596

Brown, D. and Weber, C.S. (1968) "Gene linkage by DNA-RNA hybridisation.1) Unique DNA sequences homologous to 4S, 5S and ribosomal RNA." *J. Mol. Biol.* 34, 661-680

References

- Buell, G.N., Wickens, M.P., Payvar, F. and Schimke, R.T. (1978) "Synthesis of full-length cDNAs from four partially purified mRNAs." J. Biol. Chem 253, 2471-2482
- Buongiorno, N.M., Micheli, G., Carri, M.T. and Marilley, M. (1982) "A relationship between replicon size and supercoiled loop domains in eukaryotic genomes." Nature 298, 100-102
- Busslinger, M., Rusconi, S. and Birnstiel, M.L. (1982) "An unusual evolutionary behaviour of a sea urchin histone gene cluster." EMBO J. 1, 27-33
- Bustin, M. and Cole, R.D. (1968) "Species and organ specificity in very lysine-rich histones." J. Biol Chem. 243, 4500-4505
- Bustin, M., Goldblatt, D. and Sperling, R. (1976) "Chromatin structure visualisation by immunoelectron microscopy." Cell 7, 297-304
- Calbretta, B., Robberson, D.L., Barrera-Saldana, H.A., Lambrou, T.P. and Saunders, G.F. (1982) "Genome instability in a region of human DNA enriched in Alu repeat sequences." Nature 296, 219
- Callan, H.G. and Old, R.W. (1980) "In situ hybridisation to lampbrush chromosomes: a potential source of error exposed." J. Cell Sci. 41, 115-123
- Calos, M.P. and Miller, J.H. (1980) "Transposable elements." Cell 20, 579-595

References

- Capetanaki, Y.G., Flytzanis, C.N. and Alomo, A. (1982) "Repression of the Albumin gene in Novikoff hepatoma cells." *Mol. Cell. Biol.* 2, 258-266
- Carlson, M. and Brutlag, D. (1979) "Different regions of a complex satellite DNA vary in size and sequence of the repeating unit." *J. Mol. Biol.* 135, 483-500
- Carlson, E.A. and Southin, J.L. (1962) "Comparative mutagenesis of the Dumpy locus of Drosophila melanogaster." *Genetics* 47, 321-336
- Carrell, R.W., Jeppsson, J.-O., Laurell, C.-B., Brennan, S.O., Owen, M.C., Vaughan, L. and Boswell, D.R. (1982) "Structure and variation of human alpha(1) antitrypsin." *Nature* 298, 329-334
- Cary, P.D., Hines, M.L., Bradbury, E.M., Smith, B.J. and Johns, E.W. (1981) "Conformational studies on histone H1 in comparison with histone H1 and H5." *Eur. J. Bioch.* 120, 371-377
- Cavalier-Smith, T. (1978) "Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the C-value paradox." *J. Cell Sci.* 34, 247-278
- Chakravarti, A. and Beutow, K. (1982) "Estimation of the rate of nucleotide substitution under concerted evolution." *Genetics* 100, s11
- Chandler, M.E., Kedes, L.H., Cohn, R.H. and Yunis, J.J. (1979) "Genes coding for histone proteins in man are located on the distal end of the long arm of chromosome 7." *Science* 205, 908-910

References

Chang, A.C., Cochet, M. and Cohen, S.N. (1980) "Structural organisation of human genomic DNA encoding the pro-opiomelanocortin peptide." P.N.A.S. 77, 4890-4894

Chia, W., Scott, M.R.D. and Rigby, P.W.J. (1982) "The construction of cosmid libraries of eukaryotic DNA using the Homer series of vectors." Nucl. Acid. Res. 10, 2503- 2520

Childs, G., Maxson, R. and Kedes, L.H. (1979) "Histone gene expression during sea urchin embryogenesis: isolation and characterisation of early and late mRNAs of Strongylocentrotus purpuratus by gene-specific hybridisation and template activity." Dev. biol. 73, 153-173

Childs, G., Maxson, R., Cohn, R.H. and Kedes, L. (1981) "Orphans: dispersed genetic elements derived from tandem repetitive genes of eukaryotes." Cell 23, 651-663

Chikaraishi, D.M., Deeb, S.S. and Suecka, N. (1978) "Sequence complexity in nuclear RNAs in adult rat tissues." Cell 13, 111-120

Chovnik, A., McCarron, M., Hilliker, A., O'Donnel, J., Gelbart, W. and Clarke, S. (1977) "Gene organisation in Drosophila." C.S.H.S.Q.B. 77, 1011-1021

Chu, E.H.Y. (1971) "Characterisation of X-ray-induced forward mutations in chinese hamster cell cultures." Mutat. Res. 11, 23-34

Chu, E.M.Y. (1974) "Induction and analysis of gene mutations in cultured mammalian somatic cells." Genetics 78, 115-132

References

- Chiechanover, A., Elias, S., Heller, H., Ferber, S. and Hershko, A. (1980) "Characterisation of the heat-stable polypeptide of the ATP-dependant proteolytic system from reticulocytes." J. Biol Chem. 255, 7525-7528
- Clark, S.J., Krieg, P.A. and Wells, J.R.E. (1981) "Isolation of a clone containing human histone genes." Nucl. Acid. Res. 9, 1583-1590
- Clarke, M. and Spuddich, J.A. (1977) "Non-muscle contractile proteins." An. Rev. Bioch. 46, 797-822
- Clewell, D.B. (1974) "Nature of ColE1 plasmid replication in E.coli in the presence of chloramphenical." J. Bact. 110, 667-676
- Clewell, D.B. and Helinski, D.R. (1970) "Properties of supercoiled DNA-protein relaxation complex and strand specificity of the relaxation event." Biochem. 9, 4428-4440
- Cline, M.J. and Golde, D.W. (1979) "Cellular interactions in haematopoiesis." Nature 277, 177-181
- Coen, E.S., Thoday, J.M. and Dover, G. (1982) "Rate and turnover of structural variants in the rDNA family of Drosophila ." Nature 295, 564
- Coggins, L.W., Grindlay, G.J., Vass, J.K., Slater, A.S., Montague, P., Stinson, M.A. and Paul, J. (1980) "Repetitive DNA sequences near 3 human beta-type globin genes." Nucl. Acid. Res. 8, 3319-3334

References

Cohen, L.H., Newrock, K.M. and Zweidler, A. (1975) "Stage-specific switches in histone synthesis during embryogenesis of the sea urchin." Science 190, 994-997

Cohn, R.H. and Kedes, L.H. (1979) "Nonallelic histone gene clusters of individual sea urchins (Lytechinus pictus): polarity and gene organisation." Cell 18, 843-853

Coleman, A., Byers, M.J., Primrose, S.B. and Lyons, A. (1978) "Rapid purification of plasmid DNAs by hydroxyapatite chromatography." Eur. J. Bioch 91, 303-310

Collins, J. and Bruning, H.J. (1978) "Plasmids usable as gene-cloning vectors in an in vitro packaging by coliphage lambda : 'Cosmids'." Gene 4, 85-107

Collins, J. and Hohn, B. (1978) "Cosmids: a type of plasmid gene-cloning vector that is packagable in vitro in bacteriophage lambda heads." P.N.A.S. 75, 4242-4246

Compton, J.L. and McCarthy, B.J. (1978) "Induction of the Drosophila heat shock response in isolated polytene nuclei." Cell, 14, 191-201

Compton, J.G., Schrader, W.J. and O'Malley, B.W. (1982) "Selective binding of chicken progesterone receptor A subunit to a DNA fragment containing ovalbumin gene sequences." B.B.R.C. 105, 96-104

Cooke, N.E. and Baxter, J.D. (1982) "Structural analysis of the prolactin gene suggests a separate origin for its 5' end." Nature 297, 603-606

References

- Cook, P.R. and Brazell, I.A. (1975) "Supercoils in human DNA." J. Cell. Sci. 19, 261-279
- Cook, P.R. and Brazell, I.A. (1980) "Mapping sequences in loops of nuclear DNA by their progressive detachment from the nuclear cage." Nucl. Acid. Res. 8, 2895-2906
- Cooke, H.J. and McKay, R.D.G. (1978) "Evolution of a human Y-chromosome specific repeated sequence." Cell 13, 453-460
- Cooper, A.D. and Crain, W.R.Jr. (1982) "Complete nucleotide sequence of a sea urchin actin gene." Nucl. Acid. Res. 10, 3910-3919
- Couturier, M. (1976) "The integration and excision of the bacteriophage Mu-1." Cell 7, 155-163
- Crabtree, B. (1976) "Theoretical considerations of the sensitivity conferred by substrate cycles in vivo." Bioch. Soc. Trans. 4, 999-1002
- Crawford, R.J., Krieg, P., Harvey, R.P., Hewish, D.A. and Wells, J.R.E. (1979) "Histone genes are clustered with a 15kb repeat in the chicken genome." Nature 279, 132-136
- Crisci, J.V. (1982) "Parsimony in evolutionary theory: law or methodological prescription?" J. Theor. Biol. 97, 35-41
- Dabon, J.-R. and Cantor, C.R. (1982) "Role of histone pairs H2A, H2B and H3, H4 in the self-assembly of nucleosome core particles." J. Mol. Biol. 156, 771-789

References

Dahl, H.H., Flavell, R.A. and Grosveld, F.G. (1981) "The use of genomic libraries for the isolation and study of eukaryotic genes." in 'Genetic Engineering II' ed R.Williamson (Academic Press) 49-127

Darlington, C.D. (1942) "Chromosome chemistry and gene action." Nature 149, 66-69

Davidson, E.H. (1976) "Gene activity in early development."

Davidson, E.H. and Britten, R.J. (1979) "Regulation of gene expression: possible role of repetitive sequences." Science 204, 1052-1059

Davidson, E.H., Graham, D.E., Neufeld, B.R., Chamberlin, M.E., Amerson, C.S., Hough, B.R., and Britten, R.J. (1973) "Arrangement and characterisation of repetitive sequence elements in animal DNAs." C.S.H.S.Q.B. 38, 295-301

Davidson, E.H., Hough-Evans, B.R. and Britten, R.J. (1982) "Molecular biology of the sea urchin embryo." Science 217, 17-26

Davidson, E.H. and Possakony, J.W. (1982) "Repetitive sequence transcripts in development." Nature 297, 633-635

Davidson, N. and Szybalski, W. (1971) "Physical and chemical characteristics of lambda DNA." in "The Bacteriophage Lambda" ed Hershey, 45-78

Davis, R.W., Simon, M. and Davidson, N. (1981) "Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids." Meth. in Enz. 21, 413-427

References

- Delegeane, A.M. and Lee, A.S. (1982) "Coupling of histone and DNA synthesis in the somatic cell cycle." *Science* 215, 79-81
- Denhart, D.T. (1966) "A membrane-filter technique for the detection of complementary DNA." *B.B.R.C.* 23, 641-646
- de Robertis, E.M. and Meri, J.E. (1977) "Coupled transcription-translation of DNA injected into Xenopus oocytes." *Cell* 12, 175-182
- de Villiers, J. and Schaffner, W. (1981) "A small segment of polyoma DNA enhances the expression of a cloned beta-globin gene over a distance of 1400 bp." *Nucl. Acid. Res.* 9, 6251-6264
- Diaz, M.O., Barsacchi-Pilone, G., Mahon, K.A., and Gall, J.G. (1981) "Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt Notophthalmus." *Cell* 24, 649-659
- Dodemont, H.J., Soriano, P., Quax, W.J., Ramaekers, F., Lenstra, J.A., Graenen, M.A.M., Bernardi, G. and Bloemendal, H. (1982) "The genes coding for the cytoskeletal proteins actin and vimetin in warm-blooded vertebrates." *EMBO J.* 1, 167-171
- Doerfler, W. (1981) "DNA methylation - a regulatory signal in eukaryotic gene expression." *J. Gen. Virol.* 57, 1-20
- Donoghue, D.J., Goldfarb, M.P., Sharp, P.A. and Weinberg, R.A. (1979) "Organisation of murine sarcoma virus genomes." *C.S.H.S.Q.B.* 44, 721-726

References

Doolittle, W.F. and Sapein, C. (1980) "Selfish genes, the phenotype paradigm and genome evolution." *Nature* 284, 601-603

Dunsberg, D.H., Donoghue, D.J., Goldfarb, M.P., Sharp, P.A., and Weinberg, R.A. (1979) "Organisation of murine sarcoma virus genomes." *C.S.H.S.Q.B.* 44, 13-29

Durica, D.S., Schloss, J.A. and Crain, W.R.Jr. (1981) "Organisation of actin gene sequences in the sea urchin: molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin." *P.N.A.S.* 5683-5687

Dynan, W. and Tjian, R. (1982) "Transcription enhancer sequences: a novel regulatory element." *Trends in Bioch. Sci.* 7, 124-125

Easton, D. and Chalkley, R. (1972) "High resolution electrophoretic analysis of the histones from embryos and sperm of Arbacia ." *Exp. Cell Res.* 72, 502-510

Efstradiatis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connel, C.O., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Barelle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) "The structure and evolution of the human beta-globin gene family." *Cell* 21, 653-668

Eickbush, T.H. and Kafatos, F.C. (1982) "A walk in the Chorion locus of Bombyx mori." *Cell* 29, 633-643

References

- Eichbush, T.H., Watson, D.K. and Moudriankis, E.N. (1976) "A chromatin-bound proteolytic activity with unique specificity for H2A." Cell 9, 785-792
- Elias, H. and Sherrick, J.C. (1969) "Morphology of the liver." pp 13-16
- Elgin, S.C.R. and Weintraub, H. (1975) "Chromosomal proteins and chromatin structure." An Rev. Bioch. 44, 725-774
- Endow, S.A., Polan, M.L and Gall, J.G. (1975) "Satellite DNA sequences of Drosophila melanogaster." J. Mol. Biol. 96, 665-692
- Engel, J.D. and Dodgson, J.B. (1981) "Histone genes are clustered but not tandemly repeated in the chicken genome." P.N.A.S. 78, 2856-2860
- Engel, J., Gunning, P. and Kedes, L. (1982a) "Human cytoplasmic actin proteins are encoded by a multigene family." J. Cell. Molec. Biol. 2, 674-684
- Engel, J., Gunning, P. and Kedes, L. (1982b) " Human actin proteins are encoded by a multigene family." Cold Spring Harb. (submitted)
- Engel, J.D., Sugarman, B.J. and Dodgson, J.B. (1982) "A chicken histone H3 gene contains intervening sequences." Nature 297, 434-436
- Engelke, D.R., Ng, S-Y., Shastry, B.S. and Roeder, R.G. (1980) "Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes." Cell 19, 717-728

References

Enquist, L. and Sternberg, N. (1979). "In vitro packaging of lambdaDam vectors and their use in cloning DNA fragments." Meth. in En.: 68, 281-298

Enquist, N., Tiemeier, D., Leder, P., Weisberg, R. and Sternberg, N. (1974) "Safer derivatives of bacteriophage lambda-gt.lambda-c for use in cloning recombinant DNA molecules." Nature 259, 596-598

Feiss, M., Fisher, R.R., Crayton, M.A. and Egner, G. (1977). "Packaging of the bacteriophage lambda chromosome: effect of chromosome length." Virol 77, 281-293

Feiss, M. and Seigele, M.A. (1979) "Packaging of the bacteriophage lambda chromosome: dependance of cos cleavage on chromosome length." Virol 92, 190-200

Felsenfeld, G. (1978) "Chromatin." Nature 271, 115-122

Ferguson, J. and Davis, R.W. (1975) "An electron microscopic method for studying and mapping the region of weak homology between SV40 and polyoma DNAs." J. Mol. Biol. 94, 135-149

Fiddes, J.C. and Goodman, H.M. (1981) "The gene encoding the common alpha subunit of four human glycoprotein hormones." J. Mol. Appl. Gen. 1, 3-18

Finch, J.T. and Klug, A. (1976) "Solenoidal model for superstructure in chromatin." P.N.A.S. 73, 1897-1901

References

Finnegan, D.J. (1981) "Transposable elements and retroviruses." *Nature* 292, 800-801

Firtel, R.A. and Kindle, K. (1975) "Structural organisation of the genome of the cellular slime mould Dictyostelium discoideum: interspersed of repetitive and single-copy DNA sequences." *Cell* 5, 401-411

Flynn, J.M. and Woodland, H.R. (1980) "The synthesis of histone H1 during early amphibian development." *Dev. Biol.* 75, 222-230

Foe, V.E., Wilkinson, L.E. and Laird, C.D. (1976) "Comparative organisation of active transcription units in Oncopeltus fasciatus." *Cell* 9, 131-146

Ford, P. (1978) "Pseudogene structure in 5S RNA genes." *Nature* 271, 205-206

Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) "The complete nucleotide sequence of the chick alpha-actin gene and its evolutionary relationship to the actin gene family." *Nucl. Acid. Res.* 10, 3861-3876

Franklin, S.G. and Zweidler, A. (1977) "Non-allelic variants of histones H2A, H2B and H3 in mammals." *Nature* 266, 273-275

Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1979)
"Characterisation of deletions which affect the expression of fetal globin genes in man." *Nature* 279, 598-603

References

Fryberg, E.A., Kindle, K.L., Davidson, N. and Sodja, A. (1980) "The actin genes of Drosophila: a dispersed multigene family." Cell 19, 365-379

Gabrielli, F., Hancock, R. and Faber, A.J. (1981)
"Characterisation of a chromatin fraction bearing pulse-labelled RNA."
Eur. J. Bioch. 120, 363-369

Galau, G.A., Britten, R.J. and 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., Klein, W.H., Davis, M.M., Wold, B.J., Britten, R.J. and Davidson, E.H. (1976) "Structural gene sets active in embryos and adult tissues of the sea urchin." Cell 7, 487-505

Galau, G.A., Lipson, E.D., Britten, R.J. and Davidson, E.H. (1977)
"Synthesis and turnover of polysomal mRNAs in sea urchin embryos." Cell 10, 415-432

Gall, J.G., Stephenson, E.C., Erba, H.P., Diaz, M.O. and Barsacchi-Pilone, G. (1981) "Histone genes are located at the sphere loci of newt lampbrush chromosomes." Chromosoma 84, 159-171

Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) "Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements." Nature 294, 626-631

References

- Garrels, J.I. and Gibson, W. (1976) "Identification and characterisation of multiple forms of actin." *Cell* 9, 793-803
- Getz, M.J., Birnie, G.D., Young, B.D., MacPhail, E. and Paul, J. (1975) "A kinetic estimation of base sequence complexity of poly(A)-containing RNA in mouse Friend cells." *Cell* 4, 121-129
- Getz, M.J., Elder, P.K., Benz, E.W.Jr., Stephens, R.E. and Moses, H.L. (1976) "Effect of cell proliferation on levels and diversity of poly(A)-containing mRNA." *Cell* 7, 255-265
- Ghangas, G.S. and Milman, G. (1975) "Radioimmune determination of HPRT crossreacting material in erythrocytes of Lesch-Nyhan patients." *P.N.A.S.* 72, 4147-4150
- Gillespie, D., Pequinot, E. and Strayer, D. (1980) "An ancestral amplification of DNA in primates." *Gene* 12, 103-111
- Glover, D. (1976) "Maintenance and evolution of repeated genes in eukaryotes." *Nature* 263, 9
- Glover, D.M. (1977) "Gene cloning: a new approach to understanding relationships between DNA sequences." *Int. Rev. Biochem.* vol 15, 1-44
- Gluman, Y. (1982) "Eukaryotic viral vectors." Cold Spring Harbour (in press).
- Goldstein, G., Scheid, M., Hammerling, U. Boyse, E.A., Schlesinger, D.H. and Niall, H.D. (1975) "Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells." *P.N.A.S.* 72, 11-15

References

- Graff, L.H.Jr. and Chasin, L.A. (1982) "Direct demonstration of genetic alterations at the DHFR locus after gamma irradiation." *Mol. Cell. Biol.* 2, 93-96
- Graham, D.E., Neufeld, B.R., Davidson, E.H. and Britten, R.J. (1974) "Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome." *Cell* 1, 127-137
- Grantham, R. (1978) "Viral, prokaryotic and eukaryotic genes contrasted by mRNA sequence indexes." *F.E.B.S. Letts* 95, 1-11
- Grantham, R., Gautier, C. and Goug, M. (1980) "Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type." *Nucl. Acid. Res.* 8, 1893-1912
- Grantham, R., Gautier, C. and Goug, M. (1981) "Codon catalog usage is a genome strategy modulated for gene expressivity." *Nucl. Acid.* 9, r43-r74
- Green, M.M. (1980) "Transposable elements in Drosophila and other diptera." *An. Rev. Genet.* 14, 109-120
- Grosschedl, R. and Birnstiel, M.L. (1980) "Identification of regulatory sequences in the prelude sequence of an H2A histone gene by the study of specific deletion mutants in vivo." *P.N.A.S.* 77, 1432-1436
- Grosschedl, R. and Birnstiel, M.L. (1982) "Delimitation of far upstream sequences required for maximal in vivo transcription of H2A histone gene." *P.N.A.S.* 79, 297-301

References

Grosschedl, R., Wasylyk, B., Chambon, P. and Birnstiel, M.L. (1981) "Point mutation in the TATA box curtails expression of sea urchin H2A histone gene in vivo." *Nature* 294, 178-180

Grunstein, M. (1978) "Hatching in the sea urchin Lytechinus pictus is accompanied by a shift in histone H4 gene activity." *P.N.A.S.* 75, 4135-4139

Grunstein, M., Diamond, K., Knoppel, E. and Grunstein, J.E. (1981) "Comparison of the early H4 gene sequence of Stongylocentrotus purpuratus with maternal, early and late H4mRNA sequences." *Bioch.* 20, 1216-1223

Gupta, P., Rosen, J.M., D'Eustachio, P. and Ruddle, F.H. (1982) "Location of the casein gene family to a single mouse chromosome." *J. Cell. Biol.* 93, 199-204

Gurley, L.R., Walters, R.A. and Tobey, R.A. (1974) "Cell cycle specific changes in histone phosphorylation associated with cell proliferation and chromosome condensation." *J. Cell. Biol.* 60, 356-364

Gurley, L.R., Walters, R.A. and Tobey, R.A. (1979) "Sequential phosphorylation of histone subfractions in the chinese hamster cell cycle." *J.B.C.* 250, 3936-3944

Hall, L., Davies, M.S. and Craig, R.K. (1981) "The construction, identification and characterisation of plasmids containing human alpha-lactalbumin." *Nucl. Acid. Res.* 9, 65-84

References

- Halleck, M.S. and Gurley, L. (1982) "Histone variants and histone modifications in chromatin fractions from heterochromatin-rich Peromyscus cells." *Exp. Cell. Res.* 138, 271-285
- Hardies, S.C. and Wells, R.D. (1976) "Preparative fractionation of DNA restriction fragments by reverse phase chromatography." *P.N.A.S.* 73, 3117-3121
- Harrison, P.R. (1977) "The biology of the Friend cell." *MTP Int. Rev. Sci.* II vol 15, pp227-268
- Harrison, G.A. Weiner, J.S., Tanner, J.M. and Barnicot, N.A. (1977) "Human biology."
- Harvey, R.P., Krieg, P.A., Robins, A.J., Coles, L.S. and Wells, J.R.E. (1981) "Non-tandem arrangement and divergent transcription of chicken histone genes." *Nature* 294, 19-53
- Hayaishi, O. and Ueda, K. (1977) "Poly(ADP-ribose) and ADP-ribosylation of proteins." *An. Rev. Bioch.* 46, 95-116
- Hedgpeth, J., Goodman, H.M. and Boyer, H.W. (1972) "DNA nucleotide sequence restricted by the R1 endonuclease." *P.N.A.S.* 69, 3448-3452
- Heilig, R., Muraskowsky, R., Koepfer, C. and Mandel, J.L. (1982) "The ovalbumin gene family: complete sequence and structure of the Y gene." *Nucl. Acid Res.* 10, 4363-4382
- Heilig, R., Muraskowsky, R. and Mandel, J-L. (1982) "The ovalbumin gene family. The 5' end region of the X and Y genes." *J. Mol. Biol.* 156, 1-19

References

- Heintz, N., Zernik, M. and Roeder, R.G. (1981) "The structure of the human histone genes: clustered but not tandemly repeated." *Cell* 24, 661-668
- Hennighausen, L.G. and Sippel, A.E. (1982) "Comparative sequence analysis of the mRNAs coding for mouse and rat Whey proteins." *Nucl. Acid. Res.* 10, 3733-3744
- Hentschel, C.C. and Birnstiel, M.L. (1981) "The organisation and expression of histone genes." *Cell* 25, 301-313
- Hentschel, C., Probst, E. and Birnstiel, M.L. (1980) "Transcriptional fidelity of histone genes injected into Xenopus oocyte nuclei." *Nature* 288, 100-102
- Herman, R., Weymouth, L. and Penman, S. (1978) "Heterogenous nuclear RNA-protein fibres in chromatin-depleted nuclei." *J. Cell. Biol.* 78, 663-674
- Hers, H.G. (1976) "The usefulness of futile cycles." *Bioch. Soc. Trans.* 4, 984-998
- Hickey, D.A. (1982) "Introns as selfish DNA." *Genetics* 100, s29
- Heiter, P.A., Hendricks, M.B., Hemminki, K. and Weinberg, E.S. (1979) "Histone gene switch in the sea urchin embryo." *Biochem.* 18, 2707-2716
- Higgins, S.J., Burchell, J.M., Parker, M.G. and Herries, D.G. (1978) "Effects of testosterone on sequence complexity of polyadenylated RNA from rat seminal vesicle." *Eur. J. Bioch.* 91,

References

Hilder, V.A., Livesey, R.N., Turner, P.C. and Vlad, M.T. (1981)
"Histone gene number in relation to C-value in amphibians." Nucl. Acid.
Res. 9, 5737-5746

Hill, and Stone, (1976) "Poly(ADP ribose)." Rev. Physiol. and
Chem. Pharm.

Hochman, B. (1971) "Analysis of chromosome 4 in Drosophila
melanogaster. II EMS induced lethals." Genetics 67, 235-252

Hochman, B. (1972) "The detection of 4 more vital loci on
chromosome 4 of Drosophila melanogaster." Genetics 71, s71

Hochman, B. (1973) "Analysis of a whole chromosome in
Drosophila." C.S.H.S.Q.B. 38, 581-589

Hoeijmakers-van Dommelen, H.A.M., Grosveld, G.C., de Boer, E.,
Flavell, R.A., Varley, J.M. and Jeffrey, A.J. (1980) "Localisation of
repetitive and unique sequences neighbouring the rabbit beta globin
gene." J. Mol. Biol. 140, 531-547

Hohn, B. and Hohn, T. (1974) "Activity of empty, headlike
particles for packaging DNA of bacteriophage lambda in vitro."
P.N.A.S. 71, 2372-2376

Hohn, B. (1975) "DNA as a substrate for packaging into
bacteriophage lambda." J. Mol. Biol. 98, 93-106

Hohn, B. (1979) "In vitro packaging of lambda and cosmid DNA."
Meth. in Enz. 68, 299-309

References

- Hohn, B. and Collins, J. (1980) "A small cosmid for efficient cloning of large DNA fragments." *Gene* 11, 291-298
- Hohn, T., Flick, H. and Hohn, B. (1975) "Petit lambda, a family of particles from coliphage lambda infected cells." *J. Mol. Biol.* 98, 107-120
- Hohn, B. and Murray, K. (1977) "Packaging recombinant DNA molecules into bacteriophage particles in vitro." *P.N.A.S.* 74, 3259-3263
- Hollis, G.F., Heiter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) "Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing." *Nature* 296, 321-329
- Homyk, J., Pye, T.Q. and Pak, W.L. (1982) "Examination of a short region of the Drosophila genome for visual and behavioural mutants." *Genetics* 100, s30
- Hopwood, D.A. and Sermoniti, G. (1962) "The genetics of *Streptomyces coelicolor*." *Adv. Genet.* 11, 273-
- Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) "Ubiquitous family of repeated DNA sequences in the human genome." *J. Mol. Biol.* 132, 289-306
- Hough-Evans, B.R., Wold, B.J., Ernst, S.G., Britten, R.J. and Davidson, E.H. (1977) "Appearance and persistence of maternal RNA sequences in sea urchin development." *Dev. Biol.* 60, 258-277

Hsieh, T-s. and Brutlag, D. (1979) "Sequence and sequence variation within the 1.608g/cc satellite DNA of Drosophila melanogaster." J. Mol. Biol. 135, 465-481

Hughes, S.M., Payvar, F., Spector, D., Schimke, R.J., Robinson, H.L., Payne, G.S., Bishop, J.M. and Varmus, H.E. (1979) "Heterogeneity of genetic loci in chickens: analysis of endogenous viral and non-viral genes by cleavage of DNA with restriction endonucleases." Cell 18, 347-359

Hughes, S.M., Voght, P.K., Stubblefield, E., Robinson, H., Bishop, J.M. and Varmus, H.E. (1979b) "Organisation of endogenous and exogenous viral and linked non-viral sequences." C.S.H.S.Q.B. 44, 1077-1089

Hunt, B.F. and Vogelstein, B. (1981) "Association of newly replicated DNA with the nuclear matrix of Physarum polycephalum ." Nucl. Acid Res. 9, 349-363

Hyldig-Nielsen, J.J., Jensen, O., Paludon, K., Wiborg, O., Garrett, R., Jorgensen, P. and Marcker, K.A. (1982) "Primary structure of two leghemoglobin genes from soybean." Nucl. Acid. Res. 10, 689-702

References

Ide, T., Nakane, M., Anzai, K. and Andoh, T. (1975). "Supercoiled DNA folded by non-histone proteins in cultured mammalian cells." *Nature* 258, 445-447

Imoh, H. (1978) "Changes in H1 histone during development of newt embryos." *Exp. Cell Res.* 108, 57-62

Ingolia, T.D., Craig, E.A. and McCarthy, (1980) "Sequence of three copies of the gene for the major Drosophila heat shock-induced protein and their flanking regions." *Cell* 21, 669-679

Ingram, R.S., Scott, R.W. and Tilghman, S.M. (1981) "Alpha-fetoprotein and albumin genes are in tandem in the mouse genome." *P.N.A.S.* 78, 4694-4698

Isenberg, I. (1979) "Histones." *Ann. Rev. Bioch.* 48, 159-191

Jackson, D.A., McCready, S.J. and Cook, P.R. (1981) "RNA is synthesised at the nuclear cage." *Nature* 292, 552-555

Jackson, V., Shires, A., Tanphaichitr, N. and Chalkley, R. (1976) "Modification to histones immediately after synthesis." *J. Mol. Biol.* 104, 471-483

Jacob, E. (1976) "Histone-gene reiteration in the genome of the mouse." *Eur. J. Bioch.* 69, 275-284

Jacob, E., Malacinski, G. and Birnstiel, M.L. (1976) "Re-iteration frequency of the histone genes in the genome of the amphibian Xenopus laevis." *Eur. J. Bioch.* 69, 45-54

Jacobsen, H., Klenow, H. and Overgaard-Hansen, K. (1974) "The N-terminal amino acid sequence of DNA polymerase I from E.coli and of the large and small fragments obtained by limited proteolysis." *Eur. J. Bioch.* 45, 623-627

Jagadeeswaran, P., Tuan, D., Forget, B.G. and Weissman, S.M. (1982) "A gene deletion ending at the midpoint of a repetitive DNA sequence in one form of H.P.F.H." *Nature* 296, 469-470

Jat, P., Novak, U., Cowie, A., Tyndall, C. and Kamen, R. (1982) "DNA sequences required for specific and efficient initiation of transcription at the polyoma virus early promoter." *Mol. Cell. Biol.* 2, 737-751

Jeffreys, A.J. (1979) "DNA sequence variants in the gamma, delta and beta globin genes of man." *Cell* 18, 1-10

Jeffreys, A.J. (1981) "Recent studies of gene evolution using recombinant DNA." in 'Genetic Engineering II' ed R. Williamson.

Jeffreys, A.J., Barrie, P.A., Harris, S., Fawcett, D.H. and Nugent, Z.J. (1982) "Isolation and sequence analysis of a hybrid delta globin pseudogene from the brown lemur." *J. Mol. Biol.* 156, 487-503

Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) "Ubiquitous, interspersed repeated sequences in mammalian genomes." *P.N.A.S.* 77, 1398-1402

Jordan, H.E. (1954) "A textbook of histology." ppl17-150

Judd, B.H., Shen, M.W. and Kaufman, T.C. (1972) "The anatomy and function of a segment of the X.chromosome of Drosophila melanogaster," Genetics 71, 139-156

Judd, B.H. and Young, M.W. (1973) "An examination of the one cistron:one chromomere concept." C.S.H.S.Q.B. 73, 573-579

Kafatos, F.C., Jones, W. and Efstradiatis, A. (1979). "Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridisation method." Nucl. Acid. Res. 7, 1541-1552

Kaplan, R.W., Winkler, U. and Wolf-Ellmauer, (1960) "Induction and reversion of c-mutations by irradiation of the extracellular kappa-phage of Serratia." Nature 186, 330-331

Kaufman, T.C., Shannon, M.P., Shen, M.W. and Judd, B.H. (1975) "A revision of the cytology and ontogeny of several deficiencies in the 3A1-3C6 region of the X chromosome of Drosophila." Genetics 79, 265-282

Kawaichi, M., Ueda, K. and Hayaishi, O. (1980) "Initiation of poly(ADP-ribosyl) histone synthesis by poly(ADP-ribose)synthetase." J. Biol Chem. 255, 816-819

Kay, B.K., Schmidt, O. and Gall, J.G. (1981) "In vitro transcription of cloned 5S RNA genes of the newt Notophthalmus." J. Cell. Biol. 90, 323-331

- Kedes, L.H. (1979) "Histone genes and histone messengers." *Ann. Rev. Bioch.* 48, 837-870
- Kedes, L.H. and Birnstiel, M.L. (1971) "Reiteration frequency and clustering of DNA sequences complementary to histone messenger RNAs." *Nature N.B.* 230, 165-179
- Kedes, L.H., Chang, A.C.Y., Houseman, O. and Cohen, S.N. (1975b) "Isolation of histone genes from unfractionated sea urchin DNA by subculture cloning in *E.coli*." *Nature* 255, 533-538
- Kedes, L.H., Cohn, R.H., Lowry, J.C., Chang, A.C.Y., and Cohen, S.N. (1975a) "The organisation of sea urchin histone genes." *Cell* 6, 359-369
- Kedes, L. and Maxson, R. (1981) "Histone gene organisation: paradigm lost." *Nature* 294, 11-12
- Kemp, D.J. (1975) "Unique and repetitive sequences in multiple genes for feather keratins." *Nature* 254, 573-577
- Keshet, E., Shaul, Y., Kaminchik, J. and Aviv, H. (1980) "Heterogeneity of 'virus-like' genes encoding retrovirus-associated 30S RNA and their organisation within the mouse genome." *Cell* 20, 431-439
- Kincaid, J.M.Jr. (1969) "Qualitative species differences and quantitative tissue differences in the distribution of lysine-rich histones." *J. Biol. Chem.* 244, 3375-3386

Kincaid, J.M.Jr. and Cole, R.D. (1966) "A structural comparison of different lysine-rich histones of calf thymus." J. Biol. Chem. 241, 5798-5805

Kirby, K.S. and Cooke, E.A. (1967) "Isolation of DNA from mammalian tissues." Bioch. J. 14,

Kleim, P., Heinrikson, R.L. and Fitch, W. (1981) "An examination of the expected degree of sequence similarity that might arise in proteins that have converged to a similar conformational state." J. Mol. Biol. 151, 179-197

Klien, J. (1978) "H2 mutations: their genetics and effect on immune function." Adv. in Immun. 26, 55-141

Kleinschmidt, A.K. (1968) "Monolayer techniques in electron microscopy of nucleic acid molecules." Meth. in Enz. 12, 361-377

Kornberg, R.D. (1977) "The structure of chromatin." An. Rev. Bioch. 46, 931-954

Kossel, A. (1928) "The protamines and histones."

Kranse, M.O. and Stein, G.S. (1976) "Arginine-rich histone synthesis and acetylation in W138 cells stimulated to proliferate." Exp. Cell. Res. 100, 63-70

Krieg, P.A., Robins, A.J., Gail, M.J., Titmas, R.C. and Wells, J.R.E. (1982) "Chicken histone H5: selection of a cDNA recombinant using an extended primer." Nucl. Acid. Res. 9, 1495-1505

- Krolewski, J.J., Bertelsen, A.H., Humayan, M.Z. and Rush, M.G. (1982) "Members of the Alu family of interspersed repetitive DNA sequences are in the small circular DNA population of monkey cells grown in culture." *J. Mol. Biol.* 154, 399-415
- Krumlauf, R., Jeanspierre, M. and Young, B.D. (1982) "Construction and characterisation of genomic libraries from specific human chromosomes." *P.N.A.S.* 79, 2971-2975
- Kunkel, N.S. and Weinberg, E.S. (1978) "Histone gene transcripts in the cleavage and mesenchyme blastula embryo of the sea urchin S.purpuratus." *Cell* 14, 313-326
- Kurtz, D.T. (1981) "Rat alpha-2u globulin is encoded by a multigene family." *J. Mol. Appl. Gen.* 1, 29-38
- Laemmli, U.K. (1970) "Cleavage of structural proteins during assembly of the head of bacteriophage T4." *Nature* 227, 680-685
- Lai, E.C., Woo, S.L.C., Dugaiczyk, A. and O'Malley, B.W. (1979) "The ovalbumin gene: alleles created by mutations in the intervening sequences." *Cell* 16, 201-211
- Lamb, M.M. and Daneholt, B. (1979) "Characterisation of active transcription units in Balbiani rings of Chironomus tentans." *Cell* 17, 835-848
- Langan, J.A., Rall, C.C. and Cole, R.D. (1971) "Variation in the primary structure at a phosphorylation site in lysine-rich histone." *J. Biol. Chem.* 246, 1942-1948

Lebkowski, J.S. and Laemmli, U.K. (1982a) "Evidence for two levels of DNA folding in histone-depleted Hela interphase nuclei." J. Mol. Biol. 156, 309-324

Lebkowski, J.S. and Laemmli, U.K. (1982b) "Non-histone proteins and long-range organisation of Hela interphase DNA." J. Mol. Biol. 156, 325-344

Lee, F., Mulligan, R., Berg, P. and Ringwold, G. (1981) "Glucocorticoids regulate expression of DHFR cDNA in mouse mammary tumour virus chimaeric plasmids." Nature 294, 228-232

Lefevre, G. (1981) "The distribution of randomly recovered X-ray-induced, sex-linked genetic effects in Drosophila." Genetics 99, 261-280

Levin, J.M. and Cooke, P.R. (1981) "Conformational changes induced by salt in complexes of histones and superhelical nuclear DNA." J. Cell. Sci. 50, 199-208

Lewin, B. (1974) "Lethal mutants and gene numbers." Nature 251, 373-375

Lewin, B. (1975) "Units of transcription and translation: sequence compartments of heterogenous nuclear and mRNA." Cell 4, 77-93

Lewin, B. (1980) "Gene Expression: vol2 eukaryotic genomes." 2nd ed.

- Liao, L.W. and Cole, R.D. (1981) "Condensation of dinucleosomes by individual subfractions of H1 histone." J. Biol. Chem. 256, 10124-10128
- Lichtler, A.C., Sierra, F., Clarke, S., Wells, J.R.E., Stein, J.L. and Stein, G.S. (1982) "Multiple H4 histone mRNAs of HeLa cells are encoded in different histone genes." Nature 298, 195-198
- Lifschytz, E. and Falk, R. (1968) "Fine structure analysis of a chromosome segment in Drosophila melanogaster. Analysis of X-ray induced lethals." Mutat. Res. 6, 235-244
- Lifschytz, E. and Falk, R. (1969) "Fine structure analysis of a chromosome segment in Drosophila. Analysis of EMS-induced lethals." Mutat. Res. 8, 147-155
- Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) "The organisation of the histone genes in Drosophila melanogaster. Functional and evolutionary implications." C.S.H.S.Q.B. 42, 1047-1051
- Lima-de-Faria, (1977) "Prediction of gene location and classification of genes according to the chromosome field." in 'Specific Eukaryotic genes.' Engberg, Klenow and Leith ed. pp25-37
- Livingston, V.W-C. and Alexander-Jackson, E. (1970) "A specific type of organism cultivated from malignancy: bacteriology and proposed classification." Anal. Nat. Acad. Sci. 174, 636-654
- Livingston, V.W-C. and Livingston, A.M. (1974) "Some cultural, immunological and biochemical properties of Progenitor Cryptocides." Trans. Nat. Acad. Sci. 34, 569-582

References

- Lorimer, W.S.III, Stone, P.R. and Kidwell, W.R. (1977) "Control of histone H1 dimer-poly(ADP-ribose) complex formation by poly(ADP-ribose) glycohydrolase." *Exp. Cell. Res.* 106, 261-266
- Mandel, J.L. and Chambon, P. (1979) "DNA methylation: organ-specific variation in the methylation pattern within and around ovalbumin and other chicken genes." *Nucl. Acid. Res.* 7, 2081-2103
- Mandel, M. and Higa, A. (1970) "Ca-dependant bacteriophage DNA infection." *J. Mol. Biol.* 53, 159-162
- Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstradiatis, A. (1978) "The isolation of structural genes from libraries of eukaryotic DNA." *Cell* 15, 687-701
- Maniatis, T., Jeffrey, A. and van de Sande, H. (1975) "Chain length determination of small double and single stranded DNA molecules by polyacrylamide gele electrophoresis." *Eur. J. Bioch.* 14, 3787-3794
- Mannherz, H.G. and Goody, R.S. (1976) "Proteins of contractile systems." *Ann. Rev. Bioch.* 45, 427-465
- Manning, J.E., Schmid, C.W. and Davidson, N. (1975) "Interspersion of repetitive and not-repetitive DNA sequences in the Drosophila melanogaster genome." *Cell* 4, 141-155
- Marsden, M.P.F. and Laemmli, U.K. (1979) "Metaphase chromosome structure: evidence for a radial loop model." *Cell* 17, 849-858

- Martin, J.P.Jr and Fridovich, I (1981) "Evidence for a natural gene transfer from the ponyfish to its bioluminescent bacterial symbiont *Photobacter leiognathi*." J. Biol. Chem. 256, 6080-6089
- Marushige, Y. and Marushige, K. (1975) "Transformation of sperm histone during formation and maturation of rat spermatozoa." J. Biol. Chem. 250, 39-45
- Marx, J.L. (1981) "Genes that control development." Science 213, 1485-1488
- Mason, P.J., Torok, I., Kiss, I., Karch, F. and Udvardy, A. (1982) "Evolutionary implications of a complex pattern of DNA sequence homology extending far upstream of the hsp70 genes at loci 87A7 and 87C1 in *Drosophila melanogaster*." J. Mol. Biol 156, 21-35
- Matthews, B.W., Remington, S.J., Grutter, M.G. and Anderson, W.F. (1981) "Relation between hen egg white lysozyme and bacteriophage T4 lysozyme: evolutionary implications." J. Mol. Biol. 147, 545-558
- Mauron, A., Levy, S., Childs, G. and Kedes, L. (1981) "Monocistronic transcription is the physiological mechanism of sea urchin embryonic histone gene expression." Mol. Cell. Biol. 1, 661-671
- Maxam, A.M. and Gilbert, W. (1978) "Sequencing end-labelled DNA with base-specific chemical cleavages." Meth. in Enz. 65, 499-559
- Maxson, L.R., Sarich, V.M. and Wilson, A.C. (1974) "Continental drift and the use of albumin as an evolutionary clock." Nature 255, 397-400

References

- Maxson, R.E.Jr. and Wilt, F.H. (1981) "The rate of synthesis of histone mRNA during the development of sea urchin embryos." *Dev. Biol.* 83, 380-386
- May, R.M. (1982) "The cause and control of infectious disease." *Nature* 297, 539-540
- May, F.E.B., Weber, R. and Westley, B.R. (1982) "Isolation and characterisation of the Xenopus laevis albumin genes: loss of 74k albumin gene sequences by library amplification." *Nucl. Acid. Res.* 9, 2791-2807
- Mæer, A. and Champagne, M. (1976) "Histones from nucleated erythrocytes of the marine invertebrate Sipunculus nudus." *Exp. Cell. Res.* 103, 119-125
- McClelland, A., Smith, D.F. and Glover, D.M. (1981) "Short intervening sequences close to the 5' ends of the three Drosophila larval serum protein 1 genes." *J. Mol Biol* 153, 257-272
- Macleod, D. and Bird, A. (1982) "DNase sensitivity and methylation of active versus inactive rRNA genes in Xenopus species hybrids." *Cell* 29, 211-218
- McCready, S.J., Cox, B.S. and McLaughlin, C.S. (1977) "Superhelical DNA in yeast chromosomes." *Exp. Cell. Res.* 108, 473-478
- McCready, S.J., Godwin, J., Mason, D.W., Braæll, I.A. and Cook, P.R. (1980) "DNA is replicated at the nuclear cage." *J. Cell. Sci.* 46, 365-386

References

- McDonald, J.A. and Kelley, W.N. (1971) "Lesch-Nyhan syndrome: altered kinetic properties of a mutant enzyme." *Science* 171, 689-691
- McGhee, J.D. and Felsenfeld, G. (1980) "Nucleosome structure." *Ann. Rev. Bioch.* 49, 1115-1156
- McGrogan, M., Spector, D.J., Goldenberg, C.J., Halbot, D. and Raskas, H.J. (1979) "Purification of specific adenovirus 2 RNAs by preparative hybridisation and selective thermal elution." *Nucl. Acid. Res.* 6, 593-607
- McKeown, M., Taylor, W., Kindle, K.L., Firtel, R.A., Bender, W. and Davidson, N. (1978) "Multiple heterogenous actin genes in Dictyostelium." *Cell* 15, 789-800
- McKnight, S.L. and Miller, O.L. (1976) "Ultrastructural patterns of RNA synthesis during early embryogenesis of Drosophila melanogaster." *Cell* 8, 305-319
- Meistrich, M.L., Reid, B.O. and Barcellona, W.J. (1976) "Changes in sperm nuclei during spermiogenesis and epididymal maturation." *Exp. Cell. Res.* 99, 72-78
- Mertin, J.E. and Gurdon, J.B. (1977) "Purified DNSs are transcribed after microinjection into Xenopus oocytes." *P.N.A.S.* 74, 1502-1506
- Meunier-Rotival, M., Soriano, P., Cuny, G., Straus, F. and Bernardi, G. (1982) "Sequence organisation and genomic distribution of the major family of interspersed repeats of mouse DNA." *P.N.A.S.* 79, 355-359

References

- Michelson, A. and Orkin, S.H. (1980) "The 3' untranslated regions of the duplicated human alpha globin genes are unexpectedly divergent." Cell 22, 371-378
- Miki, B.L.A. and Needin, J.M. (1975) "The histones of rainbow trout erythrocytes include an erythrocyte-specific histone." Can. J. Bioch ,1158-1169
- Miller, T.E., Huang, C-Y. and Pogo, A.O. (1978) "Rat liver nuclear skeleton and ribonucleoprotein complexes containing HnRNA." J.Cell. Biol. 76, 675-691
- Mirault, M.-E., Goldschmidt-Clermont, M., Artavanis, T.S. and Shedl, P. (1979) "Organisation of the multiple genes for the 70k dalton heat-shock protein in Drosophila melanogaster." P.N.A.S. 76, 5254-5258
- Mir zebekov, A.D. (1981) "Mobile genetic elements: new Soviet results." Nature 291, 12
- Mische,D. and Pardue, M.L. (1982) "Organisation and expression of alpha-tubulin genes in Drosophila melanogaster." J. Mol. Biol. 156, 449-466
- Moorman, A.F.M., de Boer, P.A.J., de Laaf, R.T.M., van Dongen, W.M.A.M. and Destree, O.J.H. (1981) "Primary structure of the histone H3 and H4 genes and their flanking sequences in a minor histone gene of Xenopus laevis." F.E.B.S. letts. 136, 45-54

References

- Moorman, A.F.M., de Laaf, R.T.M., Destree, O.H.J., Telford, J. and Birnstiel, M.L.(1980) "Histone genes from X.laevis: molecular cloning and initial characterisation." *Gene* 10, 185-193
- Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981) "The SV40 72 bp repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants." *Nucl. Acid Res.* 9, 6047-6068
- Moysis, R.K., Bonnet, J., Li, D.W. and Ts'o, P.O.P. (1981a) "An alternative view of mammalian sequence organisation" *J. Mol. Biol.* 153, 841-870
- Moysis, R.K., Bonnet, J., Li, D.W. and Ts'o, P.O.P. (1982b) "An alternative view of mammalian sequence organisation: short repetitive sequences are organised into scrambled tandem clusters in syrian hamster DNA." *J. Mol. Biol* 153, 871-896
- Mulvihill, E.R., LePennec, J-P. and Chambon, P. (1982) "Chicken oviduct progesterone receptor: location of specific regions of high-affinity binding in cloned DNA fragments of hormone-responsive genes." *Cell* 24, 621-632
- Murray, N.E. and Murray, K. (1974) "Manipulation of restriction targets in phage lambda to form receptor chromosomes for DNA fragments." *Nature* 251, 476-481
- Naveh-Many, J. and Cedar, H. (1982) "Topographical distribution of 5-methylcytosine on animal and plant DNA." *Mol. Cell Biol.* 2, 758-762

Newcombe, H.B. (1971) "The genetic effects of ionising radiation." Adv. Genet. 16, 239-303

Newrock, K.M., Cohen, L.H., Hendricks, M.B., Donnelly, R.J. and Weinberg, E.S. (1978) "Stage-specific mRNAs coding for subtypes of H2A and H2B histones in the sea urchin embryo." Cell 14, 327-336

Newsholme, E.A. (1976) "The role of the Fructose-6-phosphate/ Fructose-1,6-diphosphate cycle in metabolic regulation and heat generation." Bioch. Soc. Trans. 4, 978-984

Nieukoop, P.D. and Faber, J. (ed) (1975) "Normal table of Xenopus laevis (Daudin)." North Holland pub.

Nishioka, Y., Leder, A. and Leder, P. (1980) "Unusual alpha-globin-like gene that has cleanly lost both globin intervening sequences." P.N.A.S. 77, 2806-2809

Niva, O. and Sugahara, T. (1981) "5-azacytidine induction of mouse endogenous type C virus and suppression of DNA methylation." P.N.A.S. 78, 6290-6294

Noble, G.K. (1954) "The biology of the Amphibia." (Dover edn.)

Nudel, U., Katcoff, D., Carmon, Y., Zevin-Sonkin, D., Levi, Z., Shaul, Y., Shani, M. and Yaffe, D. (1980) "Identification of recombinant phages containing sequences from different rat myosin heavy chain genes." Nucl. Acid. Res. 8, 2133-2146

References

Nudel, U., Katcoff, D., Zakut, R., Shani, M., Carmon, Y., Finer, M., Cosnek, H., Ginsberg, I. and Yaffe, D. (1982) "Isolation and characterisation of rat skeletal muscle and cytoplasmic actin genes." P.N.A.S. 79, 2763-2767

Nussinov, R. (1981) "Eukaryotic dinucleotide preference rules and their implication for degenerate codon usage." J. Mol. Biol 149, 125-131

O'Brien, S.J. (1973) "On estimating functional gene numbers in eukaryotes." Nature N.B. 242, 52-54

Okada, T.A. and Comings, D.E. (1979) "Higher order structure of chromosomes." Chromosoma 72, 1-14

Okayama, H., Edson, C.M., Fukushima, M., Ueda, K. and Hayaishi, O. (1978) "Purification and properties of poly(ADP-ribose)synthetase from rat." J. Biol. Chem 252, 7000-7005

Old, R.W. Callan, H.G. and Gross, K.W. (1977) "Localisation of histone gene transcripts in newt lampbrush chromosomes by in situ hybridisation." J. Cell. Sci, 27, 57-79

Orgel, L.E. and Crick, F.H.C. (1980) "Selfish DNA: the ultimate parasite." Nature 284, 604-607

Orkin, S.H., Kazanian, H.H.Jr., Antonarkis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G. and Giardina, P.J.V. (1982) "Linkage of beta-thalassemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta globin gene

cluster." Nature 296, 627-631

Overton, G.C. and Weinberg, E.S. (1978) "Length and sequence heterogeneity of the histone gene repeat unit of the sea urchin S.purpuratus." Cell 14, 247-257

Owerbach, D., Rutter, W.J., Shows, T.B., Gray, P., Goeddel, D.V. and Lawn, R.M. (1981) "Leukocyte and fibroblast interferon genes are located on human chromosome 9." P.N.A.S. 78, 3123-3127

Panyin, S., Bilek, D. and Chalkley, R. (1971) "An electrophoretic comparison of vertebrate histones." J. Biol. Chem. 246, 4206-4215

Pardoll, D.M., Vogelstein, B. and Coffey, D.S. (1980) "A fixed site of DNA replication in eukaryotic cells." Cell 19, 527-536

Pardue, M.L., Kedes, L.H., Weinberg, E.S. and Birnstiel, M.L. (1977) "Localisation of sequences coding for histone mRNA in the chromosomes of Drosophila melanogaster." Chromosoma 63, 135-151

Parker, M., Needham, M. and White, P. (1982) "Prostatic steroid binding protein: gene duplication and steroid binding." Nature 298, 92-94

Parker, D.R. and Williamson, J.H. (1974) "Some radiation effects on segregation in Drosophila." Genetics 78, 163-171

Patient, R.K., Elkington, F.A., Kay, R.M. and Williams, J.G. (1980) "Internal organisation of the major adult alpha- and beta-globin

genes of X.laevis." Cell 21, 565-573

Paulson, J.R. and Laemmli, U.K. (1977) "The structure of histone depleted metaphase chromosomes." Cell 12, 817-828

Perutz, M.F. and Lehmann, H. (1968) "Molecular pathology of haemoglobin." Nature 219, 902-909

Pluciennick, A., Skowronski, J. and Jaworski, J. (1982) "Nucleotide sequence of bovine 1.715 satellite DNA and its relation to other bovine satellite sequences." J. Mol. Biol. 158, 293-304

Poccia, D.L. and Hinegardner, R.T. (1975) "Developmental changes in chromatin proteins of the sea urchin from blastula to mature larva." Dev. Biol. 45, 81-89

Poirier, G.G., de Murcia, G., Jongstra-Bilen, J., Niedergang, C. and Mandel, P. (1982) "Poly(ADP-ribosyl)ation of polynucleosomes cause relaxation of chromatin structure." P.N.A.S. 79, 3423-3427

Portmann, R., Schaffner, W. and Birnstiel, M (1976) "Partial denaturation mapping of cloned histone DNA from the sea urchin Psammechinus miliaris." Nature 264, 31-34

Potter, S.S. (1982) "DNA sequence of a foldback transposable element in Drosophila." Nature 297, 201-204

Proudfoot, N. (1980) "Pseudogenes." Nature 286, 840-841

Rabbits, T.H. and Forster, A. (1978) "Evidence for non-contiguous variable and constant region genes in both germ line and myeloma DNA." Cell 13, 319-327

Radloff, R., Bauer, W. and Vinograd, J. (1967) "A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells." P.N.A.S. 57, 1514-1521

Rao, M.R.S., Rao, B.J. and Ganguly, J. (1982) "Localisation of testis-variant histones in rat testis chromatin." Bioch J. 205, 15-21

Ricciardi, R.P., Miller, J.S. and Roberts, B.E. (1979) "Purification and mapping of specific mRNAs by hybridisation-selection and cell-free translation." P.N.A.S. 76, 4927-4931

Richter, J.D. and Smith, L.D (1981) "Differential capacity for translation and lack of competition between mRNAs that segregate to free and membrane bound polysomes." Cell 27, 183-191

Rigby, P.W. (1982) "The oncogenic circle closes." Nature 297, 451-453

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) "Labelling DNA to high specific activity in vitro by nick translation with DNA polymerase I." J. Mol. Biol. 113, 237-251

Roberts, R.J. (1978) "Restriction and modification enzymes and their recognition sequences." Gene 4, 183-193

References

Roberts, B.E. and Patterson, B.M. (1973) "Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell free system from commercial wheat germ." P.N.A.S. 70, 2330-2334

Roberts, J.H., Stark, P., Giri, C.P. and Smulsen, M. (1975) "Cytoplasmic poly(ADP-ribose) polymerase during the HeLa cell cycle." Arch. Bioch. Bioph. 171, 305-315

Robertson, M. (1981) "Gene families, hopeful monsters and the selfish genetics of DNA." Nature 293, 333-334

Robins, D.M., Paek, I., Seeburg, P.H. and Axel, R. (1982) "Regulated expression of growth hormone genes in mouse cells." Cell 29, 623-631

Robinson, S.I., Nelkin, B.D. and Vogelstein, B (1982) "The ovalbumin gene is associated with the nuclear matrix of chicken oviduct cells." Cell 28, 99-106

Rodakis, G.C., Moschonas, N.K. and Kafatos, F.C. (1982) "Evolution of a multigene family of chorion proteins in silkworms." Mol. Cell. Biol. 2, 554-563

Rodgers, W.H. and Gross, P.R. (1978) "Inhomogeneous distribution of egg RNA sequences in early embryo." Cell 14, 279-288

Rosenberg, H., Singer, M. and Rosenberg, M. (1978) "Highly reiterated sequences of simians." Science 200, 394-402

References

Rowe, W.P. and Ko z ak, C.A. (1980) "Germ-line re-insertions of AKR murine leukemia virus genomes in Akv-1 congenic mice." P.N.A.S. 77, 4871-4874

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J.L., LeMeur, M., Bregegere, F., Gannon, F., LePennec, J.P., Chambon, P. and Kourilsky, P. (1979) "The ovalbumin gene region: common features in the organisation of three genes expressed in the chick oviduct under hormonal control." Nature 279, 125-132

Ruderman, J.V., Baglioni, C. and Gross, P.R. (1974) "Histone mRNA and histone synthesis during embryogenesis." Nature 247, 36-38

Ruderman, J.V. and Gross, R.R. (1974) "Histones and histone synthesis in sea urchin development." Dev. Biol. 36, 286-298

Ruiz-Carrillo, A., Waugh, L.J. and Allfrey, V.G. (1975) "Processing of newly synthesised histone molecules." Science 190, 117-128

Ryffel, G.U., Muellener, D.B., Wyler, T., Wahli, W. and Weber, R. (1981) "Transcription of single-copy vitellogenin gene of Xenopus involves expression of middle-repetitive DNA." Nature 291, 429-431

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) "DNA sequencing with chain terminating inhibitors." P.N.A.S. 74, 5463-5467

Saus, J., Timoneda, J., Herandez-Yago, J. and Grisolia, S. (1982) "Scope of the ATP-ubiquitin system for intracellular protein degradation." F.E.B.S. Letts 143, 225-227

Schafer, M.P., Boyd, C.D., Jolsloshev, P. and Crystal, R.G. (1980) "Structural organisation of a 17kb segment of the alpha(2) collagen gene: evaluation by R-loop mapping." Nucl. Acid. Res. 8, 2241-2253

Scheer, U. (1978) "Changes of nucleosome frequency in nucleolar and non-nucleolar chromatin as a function of transcription: an electron microscopic study." Cell 13, 535-549

Scheller, R.H., Anderson, D.M., Posakony, J.W., McAllister, L.B., Britten, R.J. and Davidson, E.H. (1981) "Repetitive sequences of the sea urchin." J. Mol. Biol. 149, 15-39

Schibler, U., Pittet, A-C., Young, R.A., Hagenbuckle, O., Tosi, M., Gellman, S. and Wellauer, P.K. (1982) "The mouse alpha amylase multigene family. Sequence organisation of members expressed in pancreas, salivary gland and liver." J. Mol. Biol. 155, 247-266

Schlesinger, D.H. Goldstein, G. and Niall, H.D. (1975) "Complete amino acid sequence of Ubiquitin, an adenylate cyclase stimulating polypeptide probably universal in living cells." Biochem. 14, 2214-2218

- Schmid, C.W. and Deininger, P.L. (1975) "Sequence organisation of the human genome." Cell 6, 345-358
- Schmid, C.W. and Jelinek, W.R. (1982) "The Alu family of dispersed repetitive elements." Science 216, 1065-1070
- Schmid, C.W., Manning, J.E. and Davidson, E.H. (1975) "Inverted repeat sequences in the Drosophila genome." Cell 5, 159-172
- Schwartz, R.H. and Rothblum, K.N. (1981) "Gene switching in myogenesis: differential expression of the actin multigene family." Biochem. 20, 4122-4129
- Searle, A.G. (1974) "Nature and consequences of induced chromosomal damage in mammals." Genetics 78, 173-186
- Searle, R.L. (1980) "Rapid turnover of the histone-ubiquitin conjugate protein A24." Nucl. Acid Res. 9, 3151-3158
- Seidman, J.G., Leder, A., Nau, M., Norman, B. and Leder, P. (1978) "Antibody diversity." Science 202, 11-17
- Selby, P.B. (1973) "X-ray-induced specific locus mutation rates in young male mice." Mutat. Res. 18, 77-88
- Setlow, R.B. and Setlow, J.K. (1972) "Effects of radiation on polynucleotides." An. Rev. Biophys. Bioeng. 1, 293-346
- Seyedin, S.M and Kistler, W.S. (1979a) "H1 histone subfractions of mammalian testes 1) Organ specificity in the rat." Biochem 18, 1371-1375

- Seyedin, S.M. and Kistler, W.S. (1979b) "H1 histone subfractions of mammalian testes 2) Organ specificity in mice and rabbits." *Biochem* 18, 1376-1379
- Sgaramella, V. and Khorana, H.G. (1972) "A further study of the T4-ligase-catalysed joining of DNA at base-paired ends." *J. Mol. Biol.* 72, 493-502
- Sharp, P.A., Sugden, B. and Sambrook, J (1973) "Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis." *Bioch. J.* 12, 3055-3063
- Sherod, D., Johnson, G. and Chalkley, R. (1974) "Studies on the heterogeneity of lysine-rich histones in dividing cells." *J. Biol. Chem.* 249, 3923-3931
- Shinohara, T., Robenson, E.A., Appella, E. and Piatigorsky, J. (1982) "Multiple gamma-crystallins of the mouse lens: fractionation of mRNAs by cDNA cloning." *P.N.A.S.* 79, 2783-2787
- Shires, A., Carpenter, M.P. and Chalkley, R. (1975) "New histones found in mature mammalian testis." *P.N.A.S.* 72, 2714-2718
- Shires, A., Carpenter, M.P. and Chalkley, R. (1976) "A cysteine-containing H2B-like histone found in mature mammalian testis." *J. Biol. Chem.* 251, 4155-4158
- Skirbinski, D.O.F. and Ward, R.D (1981) "Relationship between allozyme heterozygosity and rates of divergence." *Genet. Res.* 38, 71-92

References

- Skirbinski, D.O.F. and Ward, R.D. (1982) "Correlations between heterozygosity and evolutionary rate of proteins." *Nature* 298, 490-492
- Sierra, F., Lichtler, A., Marashi, F., Rickles, R., Van Dyke, T., Clarke, S., Wells, J., Stein, G. and Stein, G. (1982) "Organisation of human histone genes." *P.N.A.S.* 79, 1795-1799
- Singer, M. (1982) "SINES and LINES: highly repeated short and long interspersed sequences in mammalian genomes." *Cell* 28, 433-434
- Sittman, D.B., Chiu, P.I-M., Pan, C-J., Cohn, R.H., Kedes, L.H., and Marzluft, W.F. (1981) "Isolation of two clusters of mouse histone genes." *P.N.A.S.* 78, 4078-4082
- Smerdon, M.J. and Isenberg, I. (1976) "Interactions between the subfractions of calf thymus H1 and non-histone chromosomal proteins HMGl and 2." *Biochem* 15, 4242-4247
- Smith, G.P. (1976) "Evolution of repeated DNA sequences by unequal crossover." *Science* 191, 528-535
- Smith, H.O. and Birnstiel, M.L. (1976) "A simple method for DNA restriction site mapping." *Nucl. Acid. Res.* 3, 2387-2398
- Sollner-Webb, B., Camerini-Otero, R.D. and Felsenfeld, G. (1976) "Chromatin structure as probed by nucleases: evidence for the central role of histones H3 and H4." *Cell* 9, 179-193
- Southern, E.M. (1975) "Long range periodicities in mouse satellite DNA." *J. Mol. Biol.* 94, 51-69

References

- Southern, E.M. (1975b) "Detection of specific sequences among DNA fragments separated by gel electrophoresis." J. Mol. Biol. 98, 503-517
- Southern, E.M. (1979) "Measurement of DNA length by gel electrophoresis." Analyt. Biochem. 100, 319-323
- Spalding, J., Kajiwar, K. and Mueller, G.C. (1966) "The metabolism of basic proteins in HeLa cell nuclei." P.N.A.S. 56, 1535-1542
- Spencer, F.A., Hoffman, F.M. and Gelbart, W.M. (1982) "Decapentaplegic: a gene complex affecting morphogenesis in Drosophila melanogaster." Cell 28, 451-461
- Spinelli, G., Melli, M., Arnold, E., Casano, C., Gianguzza, F. and Ciaccio, M. (1980) "High molecular weight RNA containing histone messenger in the sea urchin Paracentrotus lividus ." J. Mol. Biol. ,111-122
- Spohr, G., Reith, W. and Sures, I. (1981) "Organisation and sequence analysis of a cluster of repetitive DNA elements from Xenopus laevis." J. Mol. Biol. 151, 573-592
- Sorti, R.V., Coen, D.M. and Rich, A (1976) "Tissue-specific forms of actin in developing chick." Cell 8, 521-527
- Sotirov, N. and Janes, E.W. (1972) "Quantitative differences in the content of the histone f2c between chicken erythrocytes and erythroblasts." Exp. Cell Res. 73, 13-16

References

Stanfield, S.W. and Lengyel, J.A. (1980) "Small circular DNA of Drosophila melanogaster homologous to transcripts in the nucleus and cytoplasm." Biochem. 19, 3873-3877

Stedman, E. and Stedman, E. (1943) "Probable function of histone as a regulator of mitosis." Nature 152, 556-557

Stedman, E. and Stedman, E. (1950) "Cell specificity of histones." Nature 166, 780-781

Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) "Ovomucoid intervening sequences specify functional domains and generate protein polymorphisms." Cell 21, 681-687

Stein, R., Gruenbaum, Y., Pollak, Y., Razin, A. and Ceder, H. (1982) "Clonal inheritance of the pattern of DNA methylation in mouse cells." P.N.A.S. 79, 61-65

Steinmetz, M., Winoto, A., Minard, K. and Hood, L. (1982) "Clusters of genes encoding mouse transplantation antigens." Cell 29, 489-498

Stellwagen, R.H. and Cole, R.D. (1968) "Comparisons of histones obtained from mammary gland at different stages of development and lactation." J. Biol. Chem. 243, 4456-4462

Stephens, R.E., Pan, C-J., Ajiro, K., Dolby, T.W. and Borun, T.W. (1977) "Studies of human histone mRNA." J. Biol. Chem 252, 166-172

References

Stephenson, E.C., Erba, H.P. and Gall, J.G. (1982a) "Characterisation of a cloned histone gene cluster of the newt Notophthalmus viridescens." Nucl. Acid. Res. 9, 2281-2295

Stephenson, E.C., Erba, H.P. and Gall, J.G. (1981b) "Histone gene clusters of the newt Notophthalmus are separated by long tracts of satellite DNA." Cell 24 639-647

Streek, R.E. (1982) "A multicopy insertion sequence in the bovine genome with structural homology to the Long Terminal Repeats of retroviruses." Nature 298, 767-769

Studier, F.W. (1969) "Effects of the conformation of single-stranded DNA on renaturation and aggregation." J. Mol. Biol 41, 199-209

Subak-Sharp, H., Burk, R.R., Crawford, L.V., Morrison, J.M., Hay, J. and Kier, H.M. (1966) "An approach to evolutionary relationships of mammalian DNA viruses through analysis of the pattern of nearest neighbour base sequence." C.S.H.S.Q.B. 31, 737-748

Sugino, A., Goodman, H.M., Heyneker, H.L., Shine, J., Boyer, H.W. and Cozzarelli, N.R. (1977) "Interaction of bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base-paired ends." J. Biol. Chem. 252, 3987-3994

Sures, I., Levy, S. and Kedes, L.H. (1980) "Leader sequences of Strongylocentrotus purpuratus histone mRNAs start with a unique heptapeptide common to all five histone genes." P.N.A.S. 77, 1265-1269

Sures, I., Lowry, J. and Kedes, L.H. (1978) "The DNA sequence of sea urchin H2A, H2B and H3 histone coding and spacer regions." *Cell* 15, 1033-1044

Sutcliffe, J.G. (1978) "pBR322 restriction map derived from DNA sequence: accurate DNA size markers up to 4361 bp long." *Nucl. Acid Res.* 5, 2721-2731

Szybalski, W., Kubinski, H., Hradecna, Z. and Summers, W.C. (1971) "Analytical and preparative separation of complementary DNA strands." *Meth. in Enz.* 21, 383-413

Tabin, C.J., Hoffman, J.W., Goff, S. and Weinberg, R.A. (1982) "Adaptation of a retrovirus as a eukaryotic vector transmitting the herpes simplex virus thymidine kinase gene." *Molec. Genet.* 2, 426-436

Taylor, S.M. and Jones, P.A. (1979) "Multiple new phenotypes induced in 10T half and 10T3 cells treated with 5- α -methylcytidine." *Cell* 17, 771-779

Temin, H.M. (1980) "Origin of retroviruses from cellular moveable genetic elements." *Cell* 21, 599-600

Templeton, A.R., Salle, R.D. and Walbot, V. (1981) "Speciation and inference of rates of molecular evolution from genetic differences." *Hered.* 47, 439-442

Thiebaud, C.H. and Fischberg, M. (1977) "DNA content in the genus Xenopus." *Chromosoma* 59, 253-257

References

- Thomas, M., Cameron, J.R. and Davis, R.W. (1974) "Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA." P.N.A.S. 71, 4579-4583
- Thompson, J.R.Jr. and Woodruff, R.C. (1981) "A model for spontaneous mutation in Drosophila caused by transposing elements." Heredity 47, 327-335
- Tiemeier, D., Enquist, L. and Leder, P. (1974) "Improved derivative of a phage lambda EK2 vector for cloning recombinant DNA." Nature 263, 526-527
- Tiemeier, D.C., Tilghman, S.M., Polsky, F.I., Seidman, J.G., Leder, A., Edgell, M.H. and Leder, P. (1978) "A comparison of two cloned beta globin genes and their surrounding and intervening sequences." Cell 14, 237-245
- Turner, P.C. and Woodland, H.R. (1982a) "H3 and H4 histone cDNA sequences from Xenopus: a sequence comparison of H4 genes." Nucl. Acid. Res. 10, 3769-3780
- Turner, P.C. and Woodland, H.R. (1982b) "Histone gene number and organisation in Xenopus: Xenopus borealis has a homogenous major cluster." submitted.
- Twigg, A.J. and Sherrat, D. (1980) "Trans-complementable copy-number mutants of plasmid ColE1." Nature 283, 216-218
- Tymowska, J. and Fischberg, M. (1973) "Chromosome complements of the genus Xenopus." Chromosoma 44, 335-342

Ulla, E. and Melli, M. (1982) "Cloning and characterisation of cDNA clones to 7S RNAs of Hela cells." Nucl. Acid Res. 10, 2209-2223

Ullrich, A., Dull, T.J., Gray, A., Phillips, J.A.III and Peter, S. (1982) "Variation in the sequence and modification state of the human insulin gene flanking regions." Nucl. Acid. Res. 10, 2225-2240

Ullrich, A., Gray, A., Goeddel, D.V. and Dull, T.J. (1982) "Nucleotide sequence of a portion of human chromosome 9 containing a leukocyte interferon gene cluster." J. Mol. Biol. ,467-486

Upchurch, K.S., Leyva, A., Arnold, W.J., Holmes, E.W. and Kelley, W.N. (1975) "HPRT defficiency: association of reduced catalytic activity with reduced levels of immunologically detectable enzyme protein." P.N.A.S. 72, 4142-4146

Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. (1981) "Direct repeats flank three small nuclear RNA pseudogenes in the human genome." Cell 26, 11-17

van Dongen, W. (1982) "Histone genes and their expression in early development of Xenopus laevis." PhD thesis University of Amsterdam

van Dongen, W., de Laaf, L., Moorman, A. and Destree (boo), O. (1981) "The organisation of histone genes in X.laevis." Nucl. Acid Res. 9,

References

van der Ploeg, L.H.T. and Flavell, R.A. (1980) "DNA Methylation in the human gamma-delta-beta globin locus in erythroid and non-erythroid tissues." *Cell* 19, 947-958

Vanin, E.F., Goldberg, G.I., Tucker, P.W. and Smithies, O. (1980) "A mouse alpha-globin-related pseudogene lacking intervening sequences." *Nature* 286, 222-226

Vanyushin, B.F., Main, A.L., Vasilyev, V.K. and Belozersky, A.N. (1973) "The content of 5-methyl cytosine in animal DNA: species and tissue specificity." *B.B.A.* 299, 397-403

Vinograd, G. (1963) "Sedimentation equilibrium in a buoyant density gradient." *Meth. in Enz.* 6, 854-870

Wahli, W., Dawid, I., Wyler, T., Jaggi, R.B., Weber, R. and Ryffel, G.U. (1979) "Vitellogenin in X.laevis is encoded in a small family of genes." *Cell* 16, 535-549

Wallis, J.W., Hereford, L. and Grunstein, M. (1980) "Histone H2B genes of yeasts encode two different proteins." *Cell* 22, 799-805

Warren, T.G. and Shields, D. (1982) "Cell-free synthesis of somatostatin precursors." *P.N.A.S.* 79, 3729-3733

Wasylyk, B. and Chambon, P. (1981) "A T to A base substitution and small deletions in the conalbumin TATA box drastically decrease specific in vitro transcription." *Nucl. Acid Res.* 9, 1813-1824

References

- Watson, D.K. and Moudrianakis, E.N. (1982) "Histone-dependant reconstitution and nucleosomal location of a non-histone chromosomal protein: the H2A-specific protease." *Biochem.* 21, 248-256
- Weatherall, D.J. and Clegg, J.B. (1982) "Thalassemia revisited" *Cell* 29, 7-9
- Weinberg, R.A. (1980) "Origins and roles of endogenous retroviruses." *Cell* 22, 643-644
- Weintraub, H. and Groudine, M. (1975) "Chromosomal subunits in active genes have an altered conformation." *Science* 193, 848-856
- Weisbrod, S. (1982) "Active chromatin." *Nature* 297, 289-295
- Weisbrod, S., Groudine, M. and Weintraub, H. (1980) "Interaction of HMG 14 and 17 with actively transcribed genes." *Cell* 19, 289-301
- Weiss, E.H., Cheah, S.E., Grosveld, F.G., Dahl, H.H.M., Solomon, E. and Flavell, R.A. (1982) "Isolation and characterisation of a human alpha-1(1)-like collagen gene from a human cosmid library." *Nucl. Acid Res* 10,
- Wellauer, P.K., Dawid, I.B., Brown, D.D. and Reeder, R.H. (1976) "The molecular basis for length heterogeneity in ribosomal DNA from X.laevis." *J. Mol. Biol* 105, 461-486
- Wells, R.D., Hardies, S.C., Horn, G.T., Klein, B., Larson, J.E., Neuendorf, S.K., Panayatatos, N., Patient, R.K. and Selsing, E. (1980) "RPC-5 column chromatography for the isolation of DNA fragments." *Meth in En* 65, 327-347

- Westley, B. and Knowland, J. (1978) "An Estrogen receptor from Xenopus laevis liver possibly connected with vitellogenin synthesis." Cell 15, 367-374
- Wetmur, J.G. and Davidson, N. (1968) "Kinetics of renaturarion of DNA." J. Mol. Biol. 31, 349-370
- Whitaker, M.J. and Steinhardt, R.A. (1981) "The relation between the increase in reduced nicotinamide nucleotides and the initiation of DNA synthesis in sea urchin eggs." Cell 25, 95-103
- Wickens, M.P., Woo, S., O'Malley, B.W. and Gurdon, J.B. (1980) "Expression of a chicken ovalbumin gene injected into frog oocyte nuclei." Nature 285, 628-634
- Widmer, H.J., Andres, A-C., Niessing, J., Hosbach, H.A. and Weber, R. (1981) "Comparative analysis of cloned larval and adult globin cDNA sequences of Xenopus laevis." Dev. Biol. 88, 325-332
- Wigler, M.H. (1981) "Inheritance of methylation patterns in vertebrates." Cell 24, 285-286
- Wilde, C.D., Crowther, C.E. and Cowan, N.J. (1982) "Diverse mechanisms in the generation of human beta-tubulin pseudogenes." Science 217, 549-552
- Will, B.M., Bayev, A.A. and Finnegan, D.J. (1981) "Nucleotide sequence of terminal repeats of 412 transposable elements of Drosophila melanogaster." J. Mol. Biol. 153, 897-915

References

- Williamn, J.G. and Penman, S. (1975) "The messenger RNA sequences in growing and resting mouse cells." *Cell* 6, 197-206
- Wilkinson, K.D., Urban, M.K. and Haas, A.L. (1980) "Ubiquitin is the ATP-dependant proteolysis factor I of rabbit reticulocytes." *J. Biol Chem.* 255, 7529-7532
- Wilson, A.C., Carlson, S.S. and White, T.J. (1977) "Biochemical evolution." *An. Rev. Bioch.* 46, 573-639
- Wilson, M.C. and Melli, M (1977) "Determination of the number of histone genes in human DNA." *J. Mol. Biol.* 110, 511-535
- Wold, B.J., Klein, W.H., Hough-Evans, B.R., Britten, R.J. and Davidson, E.H. (1978) "Sea urchin embryo mRNA sequences expressed in the nuclear RNA of adult tissues." *Cell* 14, 941-950
- Wolff, S., Abrahamson, S., Bender, M.A. and Conger, A.D. (1974) "The uniformity of normalised radiation-induced mutation rates amobg different species." *Genetics* 78, 133-134
- Woodland, H.R. (1980) "Histone synthesis during the development of Xenopus." *F.E.B.S. Letts* 121, 1-7
- Woodland, H.R. and Adamson, E.D. (1977) "The synthesis and storage of histones during the oogenesis of Xenopus laevis." *Dev. Biol* 57, 118-135
- Woodland, H.R. and Wilt, F.H. (1980a) "The functional stability of sea urchin histone mRNA injected into oocytes of X.laevis." *Dev. Biol.* 75, 199-213

References

- Woodland, H.R. and Wilt, F.H. (1980b) "Stability and translation of sea urchin histone mRNA molecules injected into X.laevis eggs and developing embryos." Dev. Biol. 75, 214-221
- Wu, R.S. and Bonner, W.M. (1981) "Separation of basal histone synthesis from S-phase synthesis in dividing cells." Cell 27, 321-330
- Wu, M. and Davidson, N. (1973) "Secondary structures in transfer RNA genes." J. Mol. Biol. 78, 35-41
- Wu, R.S., Nishioka, D. and Bonner, W.M. (1982) "Differential conservation of histone H2A variants between mammals and sea urchins." J. Cell. Biol. 93, 426-431
- Xin, J-H., Brandhorst, B.P., Britten, R.J. and Davidson, E.H. (1982). "Cloned embryo mRNAs not detectably expressed in adult sea urchin coelomates." Dev. Biol. 89, 527-531
- Yamamoto, K.R., Alberts, B.M., Beninger, R., Lauhorne, L. and Treiber, G. (1970) "Rapid bacteriophage sedimentation in the presence of polyethylene glycol, and its application to large-scale virus preparations." Virol. 40, 734-744
- Yunis, J.J. and Prakesh, O. (1982) "The origin of man: a chromosomal pictorial legacy." Science 215, 1525-1529
- Young, M.W. (1979) "Middle repetitive DNA: a fluid component of the Drosophila genome." P.N.A.S. 76, 6274-6278

References

- Young, M.W. and Judd, B.H. (1978) "Nonessential sequences, genes and polytene chromosome bands of Drosophila." Genetics 88, 723-742
- Yu, L.C., Szabo, P., Borun, T.W. and Prenskey, W. (1977) "The localisation of the genes coding for histone H4 in human chromosomes." C.S.H.S.Q.B. 42, 1101-1105
- Zachau, H.G. and Igo-Kemenes, T. (1981) "Face to Phase with nucleosomes." Cell 24, 597-598
- Zernik, M., Heintz, N., Boime, I., and Roeder, R.G. (1980) "Xenopus laevis histone genes: variant H1 genes are present in different clusters." Cell 22, 807-815