CHARACTERISATION OF THE INSERTIONS IN Drosophila melanogaster RIBOSOMAL DNA

Heli Roiha

Cancer Research Campaign

Eukaryotic Molecular Genetics Research Group

Department of Biochemistry

Imperial College of Science and Technology

London SW7 2AZ

Submitted for the degree of Ph.D. February 1982

A B S T R A C T

The characterisation of five cloned segments of ribosomal DNA of Drosophila melanogaster (cDm264, cDm249, cDm207, 14C9, 4G3) which have EcoRI restrictable (Type II) insertions in their 28S ribosomal RNA (rRNA) genes is described. The restriction endonuclease maps of these clones were determined and the regions coding for rRNAs were located by blotting with in vivo labeled rRNA and by S1-mapping. The Type II insertions in these clones are homologous in DNA sequence and have three characteristic EcoRI sites. The sequences of the Type I and Type II insertions are not homologous. The length heterogeneity and distribution of the Type II insertions within D. melanogaster genome were examined by DNA blotting. The lengths of the Type II insertion sequences vary between 0.5 and 3.6 kb and this length heterogeneity is limited to the left-hand part of the insertion sequence. In contrast to the Type I sequences, the Type II insertions seem to be located only within the ribosomal DNA sequences of the X and Y chromosomes.

The junctions between the 28S rRNA gene and the Type II and short 0.5 and 1 kb Type I insertions were sequenced by the Maxam-Gilbert - technique. The Type I and Type II insertions occur at different sites separated by 51 nucleotides. The Type I insertions of different lengths are located at the same site. A thirteen nucleotide sequence present only once at the site of the insertion in the uninterrupted gene is repeated on both sides of the 0.5 kb Type I insertion, while only the left-most part of this sequence is duplicated on both sides of the 1 kb Type I insertion. These structures contrast with the previously determined junction of the major 5 kb Type I insertion which is not associated with sequence duplications and has a deletion of nine nucleotides of coding sequence at its left-hand junction (Roiha et al., 1981). There is no duplication or deletion of sequences at the site of the Type II insertion. These results are discussed in the context of possible transposition and recombination events that appear to have occured at these junctions.

A C K N O W L E D G E M E N T S

I want to thank David Glover for his patience and goodwill over these years. My thanks also to Simon Kidd for his support and criticism at all stages of this work, to Debbie Smith for her encouragement and help with my future plans and to Ross Miller for introducing me to the addictive pleasures of DNA sequencing.

C O N T E N T S

ABSTRACT 1

ACKNOWLEDGEMENTS 2

CONTENTS 3

LIST OF ABBREVIATIONS 6

CHAPTER I INTRODUCTION

I.1. LOCATION OF THE rRNA GENES IN Drosophila melanogaster 7

I.2. THE NUMBER OF rRNA GENES IN Drosophila melanogaster 9

I.3. ORGANISATION OF THE rRNA CODING SEQUENCES 9

I.4. THE rDNA INSERTIONS 11

I.5. TRANSCRIPTION OF THE rDNA SEQUENCES 13

I.6. VARIATION IN rDNA CONTENT 17

I.7. REPLICATION OF rDNA AND DOSAGE COMPENSATION 19

I.8. CONCERTED EVOLUTION OF TANDEMLY ARRANGED GENES 22

I.9. rDNA ON THE X AND Y CHROMOSOMES IN D. melanogaster 24

I.10. rDNA INSERTIONS OF OTHER DIPTERAN SPECIES 26

I.11. PURPOSE OF THIS WORK 29

CHAPTER II MATERIALS AND METHODS

II.1. DNA PREPARATION 30

II.2. IN VITRO RECOMBINANTS 31

II.3. rRNA PREPARATION 32

II.4. PHENOL EXTRACTION AND PRECIPITATION OF NUCLEIC ACIDS 32

II.5. ENZYMES 33

II.6. RESTRICTION ENDONUCLEASE DIGESTIONS 34

II.7. GEL ELECTROPHORESIS 34

II.8. LENGTH STANDARDS FOR GEL ELECTROPHORESIS 35

II.9. PURIFICATION OF RESTRICTION FRAGMENTS FROM AGAROSE GELS 36

II.10. IN VITRO LABELING OF DNA 36

II.11. IN VIVO LABELING OF rRNA 37

II.12. TRANSFER OF DNA FROM AGAROSE GELS TO NITROCELLULOSE AND HYBRIDISATION WITH ^{32}-P LABELED PROBE 37

II.13. SUBCLONING OF RESTRICTION FRAGMENTS 39

II.14. THE CLONING OF rDNA REPEATS CONTAINING SHORT TYPE I INSERTIONS 40

II.15. COLONY HYBRIDISATION 40

II.16. S1 NUCLEASE MAPPING 41

II.17. DNA SEQUENCING 41

CHAPTER III RESULTS

III.1. INTRODUCTION 44

III.2. RESTRICTION MAPPING 44

III.3. S1 NUCLEASE MAPPING OF THE CODING REGIONS 49

III.4. HOMOLOGY BETWEEN THE CLONED TYPE II INSERTIONS 51

III.5. HOMOLOGY BETWEEN THE INTERRUPTION IN THE 28S rRNA GENE IN cDm207 AND THE TYPE I INSERTION SEQUENCES 52

III.6. DISTRIBUTION AND LENGTH VARIATION OF THE TYPE II INSERTION SEQUENCES IN GENOMIC DNA 54

III.7. THE DNA SEQUENCES FLANKING THE TYPE II INSERTIONS 57

III.8. CHARACTERISATION OF D. melanogaster rDNA CLONES WITH SHORT TYPE I INSERTIONS 62

CHAPTER IV DISCUSSION

IV.1. CHARACTERISATION OF THE TYPE II INSERTION SEQUENCES 69

IV.2. THE NUCLEOTIDE SEQUENCES FLANKING THE TYPE II INSERTIONS 70

IV.3. THE NUCLEOTIDE SEQUENCES FLANKING THE SHORT TYPE I INSERTIONS 73

IV.4. rDNA INSERTIONS AS TRANSPOSABLE ELEMENTS 75

FIGURES 79

APPENDIX I THE RESTRICTION MAPS OF THE rDNA CLONES 124

APPENDIX II THE SUBCLONES OF THE rDNA CLONES 143

REFERENCES 152

L I S T O F A B B R E V I A T I O N S

ETS	external transcribed spacer
kb	thousand base pairs, one thousand nucleotides or nucleotide pairs of single or double stranded nucleid acids
LA-plates	Luria agar (10g tryptone, 5 g yeast extract, 5 g NaCl, 15 g agar / 1 liter) - plates
NO	nucleolus organiser
NTS	non-transcribed spacer
rDNA	DNA sequences coding for rRNA
rRNA	ribosomal 18S and 28S RNAs

C H A P T E R I

INTRODUCTION

I.1. LOCATION OF THE rRNA GENES IN DROSOPHILA MELANOGASTER

The haploid genome of D. melanogaster contains approximately 1.65 x 10^5 kb of DNA (Rudkin, 1965, Rasch et al., 1971). About 30% of the DNA is found in darkly staining heterochromatin in the chromosomes of mitotic nuclei (Cooper, 1959, Peacock et al., 1977). The heterochromatic regions include all the centromeres and the entire Y chromosome (Williamson, 1976). Although the Y chromosome contains about the same amount of DNA as the X, it cannot be seen in salivary gland polytene nuclei, because the heterochromatic regions have fused to form the chromocentre. The density of coding sequences per unit of DNA within heterochromatin has been estimated to be about two orders of magnitude lower than in lightly staining euchromatin (Spradling and Rubin, 1981). The satellite sequences which comprise about 25% of the D. melanogaster genome are located entirely in the heterochromatin (Peacock et al., 1977). Satellite sequences have been extensively characterised by Peacock et al. (1977) and Brutlag (1979). Superimposed on the satellite sequences are members of families of middle repetitive sequence elements (Wensink et al., 1974, Glover et al., 1975, Finnegan et al., 1978, Young, 1979, Rubin et al., 1981, Kidd and Glover, 1980, Dawid et al., 1981a).

Cooper (1959) subdivided the heterochromatic DNA into blocks separated by secondary constrictions. The major constrictions on the X and Y chromosomes were identified as sites of nucleolus organisers. On the X chromosome the nucleolus organiser is located in the long arm between the "paranuclear bodies" hC and hB (Figure 1) (Cooper, 1959, Ritossa, 1976). On

the Y the nucleolus organiser is located in the short arm proximal to the male fertility factors (Figure 1) (Cooper, 1959, Ritossa, 1976).

The nucleolar constriction on the X chromosome of D. melanogaster is located between the right-hand breakage points of the sc^4 and sc^8 inversions (Figure 1) (Cooper, 1959). Ritossa and Spiegelman (1965) showed that this region of the X chromosome contained all the genes that code for the 18S and 28S rRNAs in D. melanogaster. Recombination between X chromosomes carrying these inversions results in the region between the proximal breakpoints either getting duplicated (the sc^8- sc^4 chromosome) or deleted (the sc^4- sc^8 chromosome). Ritossa and Spiegelman used these chromosomes to construct fly strains containing between one and four nucleolus organiser regions. Saturation hybridisation of rRNA to DNA extracted from these strains showed that the amount of rDNA correlated with the number of nucleolus organisers present in the genome, as the flies with one (Xsc^4- sc^8/Y), three (Xsc^8- sc^4/Y) or four (Xsc^8- sc^4/ Xsc^8- sc^4) nucleolus organisers contained about half, 1.5 and two times the normal diploid amount of rDNA. The localisation of the 18S and 28S rRNAs in the nucleolus organiser was confirmed by the in situ hybridisation of radioactive rRNA to the nucleolus of salivary gland polytene chromosomes (Pardue et al., 1970, Glover et al., 1975).

The region of the X chromosome between the sc^4 and sc^8 right-hand breakage points is about ten times larger than the size estimated for rRNA genes (Ritossa and Spiegelman, 1976). This region includes the bobbed (bb) locus (see Section I.6.) (Cooper, 1959). Ritossa et al. (1966) investigated the role of the bb locus in rRNA synthesis using saturation hybridisation of rRNA to DNA from bobbed flies. They observed that the rDNA content in these flies is about half of that in the wild-type flies. Thus, the bobbed mutations result from partial deletions of rDNA sequences. Hence it can be concluded that the nucleolus organiser is the cytological and the bb locus the genetical manifestation for the rRNA genes in D. melanogaster.

I.2. THE NUMBER OF rRNA GENES IN D. melanogaster

The number of rRNA genes in the genome can be estimated using rRNA/DNA hybridisation. These experiments had previously indicated that the numbers of the 18S and 28S rRNA genes were equal (Ritossa et al., 1966, Quagliarotti and Ritossa, 1968). This was confirmed by the fact that both genes are transcribed as a single precursor molecule (Tartof and Perry, 1970). Using both rRNAs the proportion of rDNA was found to be around 0.46% in Oregon-R females (Spear and Gall. 1973, Tartof, 1973, 1979). Knowing the sizes of the rRNAs and the D. melanogaster genome, rRNA gene numbers could be calculated to be about 500 in a diploid female and hence about 250 per the X chromosome nucleolus organiser (Tartof, 1973). Spear (1974) examined the rDNA content of the Y chromosome nucleolus organiser in X_{NO}-/ Y flies, which only contain the nucleolus organiser of the Y chromosome. He found that the Y chromosome has little more than half the number of rRNA genes found on the X chromosome. Thus it can be estimated that there is less than 400 copies of rRNA genes in a male fly.

Similar experiments with Urbana-S strain showed that this strain contains only about 100 copies of rRNA genes per the X chromosome (Ritossa and Spiegelman, 1965). This indicates that rDNA content is not conserved and varies widely between different strains.

I.3. ORGANISATION OF THE rRNA CODING SEQUENCES

The examination of actively transcribing rRNA genes in electron microscope by Miller-spreads (Miller and Beatty, 1968) revealed an alternating array of genes and spacers in head to tail tandem arrangement (Hamkalo et al., 1973). Because the 18S and 28S rRNAs are found on the same precursor molecule (Tartof and Perry, 1970), each of the visualised units must contain both the genes. When EcoRI cleaved genomic DNA is hybridised to radioactive rRNA by the method of Southern (1975), four major fragments are seen: the two most strongly hybridising fragments are 17 and 12 kb and the two minor

fragments are 7.4 and 5.4 kb (Tartof and Dawid, 1976, White and Hogness, 1977). The 17 kb fragment is characteristic of the nucleolus organiser of the X chromosome (Tartof and Dawid, 1976).

The organisation of rDNA sequences was examined by both cloning and electron microscopy on uncloned long segments of genomic DNA enriched for rDNA (Glover and Hogness, 1977, White and Hogness, 1977, Wellauer and Dawid, 1977, Pellegrini et al., 1977). This analysis showed that the organisation of the 12 kb unit is similar to the rDNA unit in Xenopus laevis: 18S gene - short spacer - 28S gene - long spacer (Wellauer et al., 1976a). The organisation of the 17 kb unit differed fron the 12 kb unit only by the presence of a 5 kb insertion within the 28S rRNA gene. The structure and organisation of the major rDNA units are shown in Figure 2.

The short internal spacer is included in the primary transcript (Greenberg, 1969, Wellauer and Dawid, 1974). Its length is conserved and it contains the coding sequences for the 5.8S and 2S rRNAs (Pellegrini et al., 1977, Jordan and Glover, 1977). The long external spacer is heterogenous in size. Its average length is 5 kb, but longer ones up to 20 kb can be found (Wellauer and Dawid, 1977, Wellauer et al., 1978, Indik and Tartof, 1980). Homoduplexes and heteroduplexes between different spacers contain small deletion loops in varying positions within the spacer indicating that at least a part of it is internally repetitive. The deletion loops do not extend into a region adjacent to the 5' end of the 18S gene (Wellauer and Dawid, 1977). R-looping experiments with the 34S rRNA precursor showed that this region of 0.92 kb from the 5' end of the 18S gene is included in the primary transcript (Dawid et al., 1978). The rest of the spacer is not transcribed. Restriction analysis of the spacer with AluI has shown that it contains a 250 base pair repeating sequence. The number of these repeats varies between five and twelve per spacer. This repetitive region is located in the central part of the spacer 1.9 kb from the 5' end of the 18S gene (Long and Dawid ,1979a)

and accounts for the size variation previously observed within this spacer.

I.4. THE rDNA INSERTIONS

The 5 kb insertion in the 17 kb major class of rDNA repeats divides the 28Sβ gene into two parts, 28Sβ1 and 28Sβ2, which are 0.74 and 1.13 kb long, respectively (White and Hogness, 1977, Wellauer and Dawid, 1977, Pellegrini et al., 1977). Wellauer and Dawid (1977) showed using R-looping and heteroduplex techniques that insertions of different length classes exist in the rDNA of D. melanogaster. They found that insertions range in size from about 0.5 kb to 6 kb and that rDNA repeats with insertion constitute about two thirds of all the repeating units in the X chromosomal rDNA. Additional electron microscopic and DNA blotting experiments revealed that there are two types of insertion (Glover, 1977, Wellauer et al., 1978). Most insertions between 1.5 and 4 kb do not have sequence homology to the 5 kb insertion and thus form a separate minor class, called Type II. The major Type I insertions include the most abundant 5 kb insertion and homologous small 1 and 0.5 kb insertions. Dawid et al. (1978) characterised cloned rDNA repeats containing Type I insertions of different lengths and found that insertions of 1 and 0.5 kb form a coterminal subset from the right-hand end of the major 5 kb insertion. In the X chromosomal rDNA about 50% of all rDNA repeats have Type I insertions of which two thirds are 5 kb long. 16% of the X chromosomal rDNA repeats have Type II insertions (Wellauer et al., 1978). Using males carrying the sc^4- sc^8 X chromosome, which is deleted for the nucleolus organiser, Wellauer et al. (1978) found that only 16% of all rDNA repeats on the Y chromosome are interrupted and that most if not all the insertions on the Y are of Type II.

Using long genomic segments of rDNA, which contained more than one rDNA repeat, for R-loop mapping of the rDNA sequences Wellauer and Dawid (1977) found that rDNA repeats interrupted by the Type I insertions are

randomly interspersed on the X chromosome. In contrast, the repeats containing the Type II insertions are significantly clustered on the Y chromosome and possibly also on the X chromosome (Wellauer and Dawid, 1977, Wellauer et al., 1978).

The junctions between the 28S rRNA gene and the long 5 kb Type I insertions have been sequenced in three independently derived clones (Roiha et al., 1981, Rae, 1981). When compared with the sequence of the uninterrupted gene, it was found that there is a nine nucleotide deletion of rDNA sequence to the left of the insertion (Figure 16 and 18). The structures of the junctions of the short Type I insertions and the Type II insertions are described in Chapter III and further discussed in Chapter IV.

Peacock and Finnegan (cited by White and Hogness, 1977) used the cloned 17 kb rDNA repeat containing the Type I insertion as the probe in in situ hybridisation and found hybridisation with the chromocenter and several euchromatic bands in addition to the nucleolus organiser region. This implied that the Type I insertion sequences can be located at extranucleolar sites. This was subsequently confirmed by Dawid and Botchan (1977), who found that in caesium chloride gradients containing DNA binding dyes such non-rDNA insertion sequences can be fractionated from those in rDNA. The amount of this non-rDNA component is about 400 kb. The non-rDNA insertion sequences appear to be arranged in blocks of less than 20 kb but greater than 5 kb which are flanked by DNA sequences of different compositions. Such sequences were characterised by Kidd and Glover (1980) who found that the majority of the non-rDNA Type I elements were found in tandem organisation. These elements are closely related to those in rDNA but have undergone several kinds of deletions and substitutions compared to the rDNA insertion sequences. Similar clones were later described by Dawid et al. (1981b). The non-rDNA Type I elements are flanked ~~with~~ and/or interspersed ~~by~~ with non-homologous sequences. Some of

these sequences are middle repetitive and occur in families whose members have diverged. It has been estimated that there are about 100 such non-rDNA insertion-like elements of around 4-5 kb in the genome of D. melanogaster (Kidd and Glover, 1980, Dawid et al., 1981b).

Roiha et al. (1981) determined the junctions of two tandemly arranged Type I elements located outside the rDNA. They found that each unit is flanked on the left by a twenty nucleotide segment and on the right by a thirteen nucleotide segment of rDNA identical to those flanking the 5 kb Type I insertion in the rDNA (Figures 16 and 18). This strongly suggests that these sequences were originally located within the rDNA and have subsequently transposed to the other heterochromatic sites.

The in situ experiments indicated that the majority of the non-rDNA Type I elements are located in the heterochromatic chromocenters with a few copies in a euchromatic locus on the fourth chromosome (Kidd and Glover, 1980, Peacock et al., 1981, Dawid et al., 1981b, Pardue and Dawid, 1981). Using mitotic chromosomes Peacock et al. (1981) showed that the majority of the elements located in the chromocenter are in the X chromosomal heterochromatin. DNA blotting experiments with Type I insertion sequences on several X chromosomes with heterochromatic deficiencies has localised these sequences to the heterochromatic blocks hC and hB on either side of the nucleolus organiser (Figure 1) (deCicco, unpublished results).

The two clusters of rDNA on the X and Y chromosomes comprise about 2% of the genome and nearly 20% of its moderately repetitive DNA.

I.5. TRANSCRIPTION OF THE rDNA SEQUENCES

The cytoplasmic ribosomes of D. melanogaster contain six RNA molecules: 5S, 18S and a 28S complex comprised of two 18S, one 5.8S and one 2S RNA molecules (Jordan et al., 1978). The 5S RNA genes are not linked to the rDNA. They are located on the right arm of the second chromosome at the

position 56F (Prensky et al., 1973). There are about 160 5S genes arranged in a single cluster (Tartof and Perry, 1970), where the genes of 120 nucleotides are separated by spacers of 250 nucleotides (Artavanis-Tsakonas et al., 1977).

Uninterrupted rDNA repeats

The largest rRNA precursor detected in D. melanogaster is 8 kb (34S) which is likely to be the primary transcript as more than 80% of the 5' termini contain pppA or ppA (Levis and Penamn, 1978). A single active rRNA transcription unit carries about 140 RNA polymerase molecules (McKnight and Miller, 1976). By using quantitative RNA/DNA hybridisation experiments Long and Dawid (1979b) determined that there are about 1300 full-length rRNA precursor molecules per embryo nucleus. This means that the total number of nascent and complete rRNA precursor molecules is about 4000 per embryo nucleus (McKnight and Miller, 1976, Long and Dawid, 1979b).

These precursor molecules are processed within the nucleus to the two molecules of 18S and 28S by endonucleolytic cleavages (Jordan et al., 1976, Long and Dawid, 1980a). The 3' end of the precursor exactly corresponds to the 3' end of the 28S gene (Mandel and Dawid, 1981). The spacers within the rDNA repeat are AT-rich. The primary transcript also contains nearly pure AU-streches which flank the genes and single out those regions which will be removed during processing (Jordan et al., 1980). The further processing of the 28S molecule within the nucleus involves the introduction of a "hidden break" of about 120 nucleotides close to the centre of the molecule (Jordan, 1975, Jordan et al., 1976, Wellauer and Dawid, 1977, White and Hogness, 1977). This region is located within a prominent hairpin structure in the 28S rRNA (Wollenzien et al., 1978). It is very AT-rich (Roiha, unpublished results) and thus forms an additional "spacer" which is characteristic for Drosophila rRNA. The two half molecules of about 18S are held together by hydrogen bonds and are transported to the cytoplasm. Later processing steps generate 5.8S and 2S rRNAs which are

120 and 30 nucleotides long, respectively. Both these genes are located within the transcribed spacer separated by a 28 nucleotides long AT-rich region (Jordan and Glover, 1977, Pavlakis et al., 1979). Together these short RNAs correspond to the approximately 160 nucleotides long 5.8S rRNA found in other eukaryotes. They are hydrogen bonded within the 28S rRNA complex (Jordan and Glover, 1977, Pavlakis et al., 1979).

Interrupted rDNA repeats

RNA blotting (Alwine et al., 1977, 1979, Wahl et al., 1979) and kinetic experiments showed that high molecular weight transcripts homologous to rDNA and full-length Type I insertion sequences represent less than one RNA molecule per embryo, larval or adult nucleus. This is more than three orders of magnitude lower than the concentration of the primary rRNA precursor molecules (Long and Dawid, 1979b). Essentially similar results were obtained by Jolly and Thomas (1980) using tissue culture cells. The transcripts from the short Type I insertions are more abundant, but none are of correct size to contain a complete transcription unit (Long and Dawid, 1979b, Kidd, unpublished results). The deletion of rDNA sequences left ~~from~~ of the 5 kb Type I insertion would exclude these rRNA repeats from acting as templates for functional rRNA (Roiha et al., 1981, Rae, 1981). Abortive transcription of these units might be expected, however, as they have identical sequences to the uninterrupted units in the 5' promoter region (Long et al., 1981a). Furthermore, the chromatin in different types of rDNA units has similar sensitivities to DNase I in their rRNA coding sequences, spacers and insertions in the nuclei of D. melanogaster tissue culture cells (C. Wu, cited by Long et al., 1981a). It is possible that transcription of interrupted rRNA genes initiates but that a strong stop signal prevents the transcription into the Type I insertion. The resulting abortive precursor would be 6.8 kb long, but the accumulation of such molecules has not been detected in D. melanogaster embryos (Long et al., 1979b). Alternatively there could be an internal

control region within the transcription unit which is modified in the interrupted units. Such a region could be the segment where the insertions occur. This region is very well conserved between species widely separated by evolution which implies a functional role (Sinclair and Brown, 1971, Gourse and Gerbi, 1980). It seems probable that rRNA genes interrupted by the Type I insertions are transcribed extremely rarely and are not used to produce rRNA. This interpretation is consistent with the observation by electron microscopy of long silent regions interspersed with active transcription units in nucleolar DNA from D. melanogaster embryos (McKnight and Miller, 1976).

Transcription of the Type I insertions does not even occur at a significant level in fly strains which exhibit the bobbed phenotype (Long et al., 1981b). In the bobbed flies 50% or more of the rDNA sequences have been deleted (see Section I.6.) and such flies must be under strong pressure to use all the available rDNA units (Ritossa, 1976). In D. hydei the severity of the similar bobbed phenotype has been directly correlated with the number of uninterrupted rDNA units and not with the total number of rDNA units present in the mutant flies (Franz and Kunz, 1981).

Although rRNA genes interrupted by the Type II insertions are not used efficiently, they are transcribed at a low level (Long et al., 1981c, Kidd and Glover, 1981). The Type II transcripts are present only in the nucleus. The most abundant RNA molecule homologous to the Type II insertion is 3.4 kb long, which corresponds to the length of the most common Type II insertion (see Chapters III and IV). There are in addition a set of higher molecular weight transcripts linked mainly to rDNA sequences upstream from the insertion in the rDNA unit (Kidd and Glover, 1981). The concentration of such molecules is less than one copy per nucleus. These transcripts seem to have one end point at the dA tract at the right-hand junction between the Type II insertions and the 28S β 2 gene (see Chapter III Section 7). This suggests that trans-

cription terminates within this region or that nascent RNA is rapidly cleaved at this site. Kidd and Glover (1981) also observed great variation in the concentration of Type II transcripts between different strains and tissues.

Similarly to the Type I insertions, the rRNA genes containing Type II insertions are not efficiently transcribed in the bobbed mutants (Long et al., 1981b).

Based on the number of rDNA repeats per the X and Y chromosome and the percentage of these repeats carrying insertions it can be estimated that in female and male flies two thirds to a half, respectively, of the total rDNA repeats are split and thus essentially inactive.

I.6. VARIATION IN rDNA CONTENT

As a consequence of tandem organisation of rDNA, the number of units per genome is subject to frequent variations by unequal crossing-over between sister chromatids and homologous chromosomes (Ritossa, 1976). The result of these changes is the gain of extra rDNA units by one chromosome or chromatid at the expense of the other. If the genome contains less than half the normal number of the rDNA units, rRNA synthesis is inadequate and a mutant, bobbed (bb) fly develops. As a corollary to the mechanism that produces the exchanges of rDNA units the number of rRNA genes between different bb alleles is not constant. Flies with rDNA deficiencies show shorter bristles, slow development, lowered fertility and other pleiotropic effects which generally decrease their viability. In severely bobbed flies up to 85% of the rDNA has been deleted. Larger deletions are usually lethal when homozygous. Since bb loci have an additive effect, the minimum number of genes for a non-bobbed phenotype can still be obtained when partially deleted loci are combined.

Another mechanism capable of changing the rDNA content in bobbed males is called rDNA magnification (Ritossa, 1968, Ritossa et al., 1971). This

increase of rDNA is inheritable and occurs in bobbed male flies when an rDNA deficient X chromosome is maintained alongside an rDNA deficient Y chromosome. For example, this phenomenon has been studied on the bb locus carried by the X chromosome of males in the Xbb/Ybb^{+} stocks when it is crossed with the females of $\hat{X}\hat{X}/O$ or $\hat{X}\hat{X}/Ybb^{-}$ (Ritossa et al., 1971, 1973, deCicco and Glover, unpublished results). The "premagnified" males obtained in these crosses, Xbb/O (sterile) or Xbb/Ybb^{-}, are phenotypically extremely bobbed, but show significant increase in their rDNA content. This increase is not stable. When the "premagnified" Xbb/Ybb^{-} males are crossed again $\hat{X}\hat{X}/Ybb^{-}$ females, males are obtained that have the same X and Y chromosomes as their fathers but which are phenotypically only slightly bobbed. When the cross is repeated, the third generation males are phenotypically wild-type and the magnified condition of the locus is stable. The rDNA content of these second and third generation flies returns to wild-type levels which are approximately twice those of the original bobbed X chromosome.

The molecular mechanism(s) of rDNA magnification is not known. Tartof (1973) has proposed that the occurence of unequal non-meiotic crossing-over between sister chromatids generates cells with increased numbers of rDNA units. In such unequal exchanges, bobbed lethal mutations would arise reciprocally with magnified chromosomes. Ritossa (1973) and Ritossa et al. (1973) have proposed the synthesis of transcriptionally inactive extrachromosomal copies of rDNA to explain the sudden increase of rDNA without phenotypic effect in the severely bobbed "premagnified" flies. deCicco and Glover (unpublished results) have studied the rDNA patterns of single flies from different stages of magnification. They observed that all types of rDNA units undergo the amplification and that there is also a concomitant amplification of the Type I sequences outside the nucleolus. These have previously been located within the blocks of heterochromatin flanking the nucleolus organiser region (deCicco, unpublished results).

The rDNA magnification therefore seems to involve large blocks of the X chromosomal heterochromatin. It is possible that the non-rDNA Type I elements could serve as endpoints in the recombinational excision of large blocks of rDNA into extrachromosomal circles.

I.7. REPLICATION OF rDNA AND DOSAGE COMPENSATION

In the polytene chromosomes of salivary gland cells the euchromatin, where most genetic loci are located, is replicated to give on the average 1024 copies per chromosome (Rasch, 1970). In contrast, the heterochromatin is replicated only a few times, if at all (Gall et al., 1971). As a consequence the amount of heterochromatic DNA in polytene nuclei is about the same as in diploid cell (Spear and Gall, 1973). Although the nucleolus organisers are located within heterochromatin, the rDNA is transcriptionally active and might be expected to have different replication properties to the surrounding DNA.

Hennig and Meer (1971) examined the amounts of rDNA in adults, which have both diploid and polytene nuclei (Ashburner, 1970), in eggs, which are considered to have only diploid nuclei, and in purified salivary glands. They found that the proportion of rDNA in salivary gland DNA was much lower than that in diploid cells. They estimated that rDNA on both the X and Y chromosomes undergoes about three times fewer rounds of replication than the surrounding euchromatin. This places rDNA at an intermediate level of polytenisation between that of heterochromatin and euchromatin. Also, while in diploid cells there is a correlation between the amount of rDNA and the number of nucleolus organisers, Spear and Gall (1973) found that in polytene tissue the amount of rDNA was independent of the number of the nucleolus organisers present in the genome. A similar situation was found in pupal and adult polytene ovary tissue (Endow and Gall, 1975, Renkawitz and Kunz, 1975).

The heterogeneity of the insertion sequences and also of the non-

transcribed spacer region gives a complex restriction pattern which is specific for the X or the Y chromosome. This pattern can be used to follow the inheritance of rDNA. Endow and Glover (1979) compared the rDNA restriction patterns of the diploid and polytene tissues of X/ Y and X_{NO}^{-} / Y males and observed that the rDNA pattern in X/ Y polytene DNA is characteristic of the Y nucleolus organiser but not of the X nucleolus organiser. This suggests that the rRNA genes from only the Y nucleolus organiser are replicated during polytenisation. By using interstrain hybrid females containing two X chromosomes with different rDNA patterns, it was possible to determine that also in females the rRNA genes from only one of the two nucleolus organisers are replicated in polytene tissue (Endow, 1980). This replication is independent of whether the X chromosome is derived from the female or male parent. This explains the earlier finding that although the diploid tissue of X/ O flies contained half of the amount of rDNA found in the diploid tissue of X/ X flies, rDNA levels are the same in X/ X and X/ O polytene cells (Spear and Gall, 1973) without involving additional rounds of replication to compensate for the lower rDNA content in the genome of the X/ O fly. It is possible that these results explain the observations of Tartof (1971) who found that the nucleolus organiser on the X chromosome can increase its rDNA content in adult flies, such as X/ O males and X/ X_{NO}^{-} females, which only have one nucleolus organiser. He called this disproportionate increase rDNA compensation and found that it is a somatic event and occured during a single generation.

Procunier and Tartof (1978) mapped the region responsible for dosage compensation, the compensatory response (cr) locus to the heterochromatic block hC, distal of the rDNA by using various deficiencies of the X chromosome. The cr locus is capable of acting in trans to detect the presence of its partner locus in the opposite homologue, and if that homologue is cr^{-}, the cr locus acts in cis to drive the disproportionate

replication of those rDNA units that are contiguous with it. If the cr locus is separated from its rDNA region by an inversion of the X chromosome, no dosage compensation occurs. The Y chromosome contains a cr locus that seems to be located in such a way that it is not capable of compensation. The cr locus on the X chromosome can, however, detect Ycr^- chromosomes and cause compensation. Procunier and Tartof (1975) observed that the 5S RNA genes are also under similar compensatory control.

Sequences in this same region of the X heterochromatin have also been found to interact with the autosomal gene abnormal oocyte (abo), which is located on the second chromosome. Sandler (1970) and Parry and Sandler (1974) showed that abo/abo females exhibit a maternal effect which results in the production of defective eggs. These eggs can be rescued with increasing numbers of either the Y chromosome or the X chromosomal heterochromatin. By testing the rescue capabilities of X chromosomes deficient for various regions of heterochromatin, Parry and Sandler (1974) located the region responsible for the rescue to the hC block, which is distal to the rDNA (designated as Xh^{abo}, Figure 1). Krider and Levine (1975) showed that when flies are kept homozygous for abo for several generations, they cease to exhibit the abo phenotype and, at the same time, increase their rDNA content by several folds. It seems likely that this increased amount of rDNA is related to the loss of the abo phenotype and that the hC region is directly involved in this increase.

Using flies carrying various inversions of the X chromosome, Yedvobnick et al. (1980) showed that Xh^{abo} and cr are not identical and that in contrast to the cr locus, Xh^{abo} functions when separated from adjacent rDNA by an inversion. abo is only one of several genes which are known to interact with and/or regulate heterochromatic functions (Mange and Sandler, 1973, Krider et al. (1979).

The molecular mechanism(s) of dosage compensation and the apparent reversion of the abo mutation together with an increase of the rDNA

content are completely unknown. Graziani et al. (1981) observed that certain rDNA units were selectively replicated during the reversion of the abo/abo phenotype. Similar selective replication has been previously observed during rDNA polytenisation (Endow and Glover, 1979, Endow, 1980). The observation that non-rDNA Type I elements are located in the same region of the X heterochromatin as cr and Xh^{abo} (deCicco, unpublished results) suggests that these elements may serve as intermediaries between the cr and Xh^{abo} and functional rDNA sequences.

I.8. CONCERTED EVOLUTION OF TANDEMLY ARRANGED GENES

The tandemly duplicated gene repeats have evolved in a concerted fashion, which means that the repeats within and between separate clusters are far more similar to each other than would be expected if they had evolved independently following an ancestral duplication event. While keeping tandemly arranged gene units homologous within the clusters in a species, concerted evolution permits their divergence, at several different levels, between related species. Smith (1973, 1976) has proposed that the major mechanisms that drive concerted evolution are selection and exchange of genetic material within and between the clusters. This exchange takes place either by unequal crossing-over between sister chromatids or non-homologous chromosomes or by various mechanisms of disproportionate replication. Concerted evolution has been used to account for the sequence organisation of Xenopus laevis rDNA (Brown et al, 1972). In Xenopus laevis the rDNA units are located in a single cluster, where the coding sequences are separated by non-transcribed spacers. The coding sequences are homogenous and closely related in different Xenopus species (Forsheit et al., 1974). In contrast, the non-coding spacers consist of repetitive DNA and are widely heterogenous in length in related Xenopus species (Brown et al., 1972, Wellauer and Reeder, 1975, Wellauer et al., 1976b, Forsheit et al., 1974).

The gene and spacer regions differ in their rates of acceptance of mutations. Selection against variants of the rRNA genes must be quite strong considering the importance of the rRNA in the ribosome structure and thus in the translational processes in the cell. This is reflected by the remarkable conservation of the rRNA gene sequences throughout evolution (Sinclair and Brown, 1971, Gourse and Gerbi, 1980, Rae et al., 1980). On the other hand, many mutations in the spacer regions are more or less selectively neutral and would not be eliminated. Such mutations would therefore accumulate in time. Spacers with internal repetitive segments seem to be a common feature of tandemly arranged genes (Long and Dawid, 1980b). The overall length of the spacers must be more important than their exact sequence. Selection on the non-transcribed spacer length variation might ensure the correct phasing of nucleosomes and the transcription initiation sites of the units (Gottesfield and Bloomer, 1980, Samal et al., 1981). Selection would also preserve the number of rDNA units at a constant level as flies with too few units would not survive as a result of defective rRNA synthesis.

Unequal crossing-over can take place on several structural levels among the repeated units. At the lowest level there would be the replacement and reorganisation of the internally repetitive units within the non-transcribed spacer. This would account for the generally observed length variation within this spacer (Wellauer et al., 1976a, Wellauer and Dawid, 1977). At a higher level, unequal crossing-over involving complete rDNA units would lead to the replacement and scrambling of the units within the cluster. This would also explain the variability in rRNA gene number per chromosome and between individuals. By this mechanism any mutation which arises in the cluster will either be eliminated by deletion or, if not selected against, spread through the entire cluster by continual duplications. This relatively frequent event would tend to cause the conversion of all units of a single cluster to the

same form. Genetic exchange between rDNA clusters located on different chromosomes can be either a direct transposition or dependent on meiotic or mitotic chromosome exchanges by unequal crossing-over. This event would relatively infrequently introduce blocks of rDNA from other chromosomes into the cluster temporarily restoring heterogeneity and correcting the clusters against each other (Smith, 1973). Such exchanges have been observed between the nucleolus organisers within primate species (Arnheim et al., 1980, Krystal et al., 1981).

When unequal crossing-over and interchromosomal exchange are frequent relative to mutation, this spreading of variants will be so rapid both within and between individuals that the repeats will not be significantly polymorphic and remain horizontally homologous (Smith, 1973, 1976). Arnheim et al. (1980) observed, however, that not all regions of the human rDNA unit participate in concerted evolution and hence show polymorphism. These regions could be either non-transcribed or transcribed parts of the rDNA repeat.

Magnification and other disproportionate replication processes could spread a certain unit through the adjacent units by making many tandem copies of this unit (Ritossa and Scala, 1969, Henderson and Ritossa, 1970, Tartof, 1971). The production of extrachromosomal molecules containing rDNA suggested by Ritossa (1973) and their reinsertion in new random positions within the clusters on the same or different chromosome could in a single generation result in great reorganisation and exchange of the rDNA units.

I.9. rDNA ON THE X AND Y CHROMOSOMES IN D. melanogaster

The organisation of D. melanogaster rDNA is complex and it is necessary to propose additional mechanisms of concerted evolution to explain the fact that the X and Y rDNA clusters are both internally non-homologous and different from each other.

The internal heterogeneity of the rDNA repeats on each nucleolus organiser is mainly due to the different distribution of the two types of insertion sequences found in the 28S rRNA genes. The Type I sequences are located only on the X chromosome, where they interrupt 50% of the rDNA repeats, while the Type II insertions are found in 16% of the repeats on both the X and Y chromosomes (Wellauer et al., 1978). If unequal crossing-over has resulted in the scrambling of units with and without the Type I insertions on the X chromosome, it has also at the same time allowed the clustering of the units containing the Type II insertions on the Y chromosome and possibly also on the X chromosome. There are also differences in both the transcribed and non-transcribed parts of the rDNA repeats between the X and Y chromosomes. The 18S and 28S rRNAs on the X and Y rDNA are very similar (Maden and Tartof, 1974), and this similarity is likely to extend to the whole of the transcribed region. However, Yagura et al. (1979) have found a single nucleotide difference between the X and Y 18S rRNAs in an Oregon-R stock, while the 28S rRNAs were found to be identical. The non-transcribed spacers within each cluster are very similar, but show length heterogeneity. When the spacers on the X and Y rDNA were compared, significant differences were found in the size classes and their frequencies among the spacers (Endow and Glover, 1979, Indik and Tartof, 1980).

These differences suggest that genetic exchange between the X and Y rDNA clusters is infrequent enough to allow at least some differences to be maintained. Additionally, some mechanism either prevents the exchange of the Type I insertions or eliminates them from the Y chromosome. There is genetic evidence that exchange of genetic material at low levels occurs between the nucleolus organisers of the X and Y chromosomes (Palumbo et al., 1973, Maddern, 1981). In normal diploid flies such exchanges have to be double cross-overs to maintain the viability of the fly. This data also suggests that the rDNA clusters on

one or both of the nucleolus organisers do not have uniform polarity. High amounts of rDNA in reverse polarities would lower the possibility of crossing-over to occur and keep the two clusters separate. However, it is difficult to explain the conservation of rDNA sequences on both nucleolus organisers in this kind of orientation. Moreover, there is no molecular evidence for this mixed polarity on any one of the rDNA clusters.

I.10. rDNA INSERTIONS OF OTHER DIPTERAN SPECIES

The species within the Drosophila subgroup have been found to contain sequences homologous to the Type I insertion, which, however, show sequence divergency (Dover and Coen, 1981, Coen, Strachan and Dover, unpublished results). The rDNA in some of the sibling species have the 17 and 12 kb fragments characteristic for rDNA repeats with and without the 5 kb Type I insertion (Tartof, 1979). The proportion of these possibly interrupted rDNA units decreases relative to the uninterrupted units in the sibling species with low amounts of rDNA units. This suggests that any interrupted units carried by these species are inactive and also that there is a relationship between decreasing rRNA gene number and the percentage of interrupted units. It is possible that a finite number of rRNA genes is essential and as this number is approached, those genes which have insertion are selectively lost. Similar observations have been made in Sciara coprophila and Sarcophaga bullata which contain very few 28S genes (Renkawitz et al., 1979, French et al., 1981).

rDNA insertionshave been found in D. hydei and in D. virilis, which belong to a different subgenus of dipterans to D. melanogaster. In D. hydei many rDNA repeats are interrupted by a 6 kb insertion (Renkawitz-Pohl et al., 1980) which shows sequence homology to the right-hand part of the Type I insertion of D. melanogaster (Barnett and Rae, 1979, Renkawitz-Pohl et al., 1980). D. hydei has three clusters of rRNA genes, one on the X and two on the Y chromosome (Hennig et al., 1975, Schafer

and Kunz, 1975). The rDNA insertion sequences are located only within the cluster on the X chromosome (Kunz et al., 1981). Similarly to the Type I insertion of D. melanogaster, the D. hydei insertion does seem to be transcribed (Glatzer, 1979). In D. virilis most of the rDNA units are interrupted by a 10 kb insertion. This insertion appears to be a result of a duplication of an ancestral sequence of around 5 kb (Rae, cited by Spradling and Rubin, 1981). Its restriction map is different to that of the 5 kb Type I insertion in D. melanogaster, but hybridisation and thermal denaturation experiments showed that the two insertions contain closely related sequences (Barnett and Rae, 1979). The D. virilis insertion seems to be transcribed rarely, if at all (Rae, cited by Spradling and Rubin, 1981). This insertion is located at exactly the same position as the Type I insertion in D. melanogaster. The right-hand junctions of these insertions are identical except that there is no homology between the sequences within the two insertions at the junction. The left-hand junctions are quite different, since there is no deletion of coding sequences at the junction in D. virilis. Moreover, the D. virilis insertion is flanked by direct repeat of a fourteen nucleotide segment of rDNA from the right of the insertion (Figures 15 and 18) (Rae et al., 1980). As with the Type I insertions of D. melanogaster, sequences homologous to the 10 kb D. virilis insertion are also located outside the nucleolus (Barnett and Rae, 1979).

Barnett and Rae (1979) have analysed several higher dipteran insects for sequences homologous to the Type I insertion of D. melanogaster. They found sequence homology to the right-hand part of the insertion in genomic DNA of Sarcophaga bullata and Musca domestica. The rDNA of S. bullata and also of Calliphora erythrocephala, which belong to the same infraorder as D. melanogaster, have been characterised in detail. French et al. (1981) found that about 20% of the 72 28S genes in S. bullata contain an insertion. These insertions are heterogenous in length and

vary in size from 1.3 to 12 kb with three abundant size classes around 2. 6 and 10 kb. In C. erythrocephala some of the rDNA units are interrupted by a 6 kb insertion (Beckingham and White, 1980, Beckingham, 1981). The interrupted rDNA units seem to be completely clustered. The non-transcribed spacers of C.erythrocephala are homogenous in length but contain repetitive sequences (Schafer et al., 1981). It is likely that all these insertions which contain homologous sequences occur in the same position within the 28Sβ gene. This suggests that they have been well conserved since the superfamilies where these flies belong to diverged 130 to 180 million years ago (Rohlendorf, 1974).

Barnett and Rae (1979) observed no homology between the Type I insertion of D. melanogaster and genomic DNA from lower dipteran species, Sciara coprophila, crane fly or robber fly, which belong to different infraorders from each other and D. melanogaster. Their analytic approach of DNA blotting with Type I insertion probes has its limitations. Insertions in the rDNA of other species may have little or no homology with those of D. melanogaster and furthermore may be present only in a few copies and so might not be detected. The rDNA of some of these species has been subsequently cloned and analysed. In S. coprophila 10% of the rDNA repeats were found to contain one of the two types of insertions. These insertions of around 1 kb have no homology to the Type I insertion (Renkawitz et al., 1979, Renkawitz-Pohl et al., 1981). Chironomus tentans has a low number of rRNA genes which do not seem to contain insertions (Hollenburg, 1976, Degelman et al., 1979). There may, however, be some additional rDNA units containing insertions, as C. tentans rRNA hybridises to large genomic fragments.

The interrupted rDNA units seem to be inactive in all the dipteran species analysed so far (Long and Dawid, 1979b, Kidd and Glover, 1981, Long et al., 1981c, Rae, cited by Spradling and Rubin ,1981, Glatzer, 1979). Such units have, however, been maintained for over long

evolutionary periods. In flies with low amounts of rDNA, there are only a few interrupted units, and there seems to be selection pressure to control and remove split rRNA genes in these species (Tartof, 1979). This contrasts with their presence in large numbers in other flies with more rDNA units. Such non-functional units may be maintained if they are totally inactive and do not produce any rRNA. There would not be any selection pressure against such units which would be maintained as "silent" genes (Fedoroff, 1979). However, if the interrupted units were non-functional "silent" genes, they should have undergone sequence changes in the same way as the non-transcribed spacer. In contrast to this expectation, the coding and spacer sequences appear to be identical in both types of repeats in all these species. This suggests that either the insertions are capable of transposing into uninterrupted units or that rDNA units containing insertions are corrected against each other and uninterrupted units which combined with rapid spreading along the cluster maintains the homogeneity among the different types of units.

I.11. PURPOSE OF THIS WORK

I have already referred in this Introduction to several results which stem from this thesis work. At the outset of this investigation my aims were to characterise the heterogeneity of the Type II insertions and their chromosomal distribution. This analysis has been carried to the level of the nucleotide sequence between the insertion and the 28S rRNA gene. The significant features of the Type II junctions are described in detail in this thesis. Subsequently I performed a similar analysis on the junctions of certain Type I insertions and the 28S rRNA gene. These results are described and discussed in the context of recombination events that appear to have occured at these junctions.

C H A P T E R II

MATERIALS AND METHODS

II.1. DNA PREPARATION

Bacteria carrying cloned segments of D. melanogaster DNA were propagated under Category II Containment conditions as specified by G.M.A.G. - Plasmid DNA was prepared as described by Clewell and Helinski (1969) using 0.2% Triton to lyse the cells. The lysed debris was pelleted in a Sorvall SS34 rotor at 15 000 rpm for 60 minutes. In some cases the cleared lysate was deproteinised by heating to 65 C for 20 minutes and the protein pelleted by centrifugation in the Sorvall SS34 rotor at 8 000 rpm for 10 minutes. 1.085 g of caesium chloride and 1 mg of ethidium bromide per ml of the resulting supernatant were added and the mixture centrifuged to equilibrium in the Beckman preparative ultracentrifuge 60 Ti rotor at 37 000 rpm for 48 hours. The band containing plasmid DNA was collected into a syringe using a 25 g needle and ethidium bromide was removed either by a passage through a small column of Dowex AG 50 (Wensink et al., 1974) followed by dialysis against TE_2 (10 mM Tris, pH 7.5, 2 mM EDTA) or by phenol or isopropanol extraction after a brief dialysis against TE_2. Deproteinisation of the cleared lysate was also carried out by phenol extraction followed by ethanol precipitation of the nucleic acids. The resulting pellet was resuspended in TE_2 containing caesium chloride and ethidium bromide and further treated as described above.

Alternatively the cells were first resuspended in 50 mM glucose, 25 mM Tris, pH 8, 10 mM EDTA, pH 7.5, 2 mg/ml lysozyme, lysed with cold 0.2 N NaOH, 1 % SDS and the cell debris and protein precipitated with the addition of cold 5 M K-acetate, pH 6. Following centrifugation at 8 000 rpm for 5 minutes, the nucleic acids in the supernatant were precipitated by the addition of an equal

volume of isopropanol. The pellet was dissolved in TE_2 and prepared for equilibrium centrifugation as described above (Birnboim and Doly, 1979).

Both these methods were also applied for small scale preparations of partially purified plasmid DNAs from 1 ml cultures. The cells were lysed and pelleted in eppendorf tubes using an Eppendorf microfuge. Proteins were removed from the cleared lysate by phenol extraction. This was followed by repeated ether extraction and the nucleic acids were precipitated by the addition of 1/10 volume 2 M Na-acetate and 2 volumes of ethanol. Plasmid DNA prepared in this way was treated with RNase (1 mg/ml) before the digestion with restriction endonucleases.

Nuclear DNA from D. melanogaster Oregon R strain embryos was the gift of David Glover and Diane de Cicco. Phage λ strain C_I857S7 DNA was the gift of Christine Clayton and Mike Scott and SV40 DNA was the gift of Bill Chia and Val Lewis.

II.2. IN VITRO RECOMBINANTS

The 17 kb EcoRI fragment containing an rDNA repeat interrupted by the 5 kb Type I insertion, pDm103 was cloned as desribed by Glover et al. (1975) by ligating EcoRI cleaved size selected D. melanogaster DNA fragments with pSC101 (Cohen et al., 1973). The 17 kb EcoRI fragment was later transferred into the EcoRI site of ColE1 (Hershfield et al., 1974) (cDm103) by David Glover and pBR322 (Bolivar, 1977) (pcDm103) by Simon Kidd.

cDm207, cDm264, cDm249 and cDm238 were isolated from a pool of recombinant DNA clones constructed by Gerry Rubin by inserting randomly sheared D. melanogaster DNA fragments into the EcoRI site of ColEI by the dA/dT connector method (Wensink et al., 1974). The 12 kb EcoRI fragment of cDm238 containing an uninterrupted rDNA repeat was transferred into the EcoRI site of pBR322 by Simon Kidd. 14C9 and 4G3 were constructed in a similar way but with pSF2124 (So et al., 1975) as the cloning vector by Spyros Artavanis-Tsakonas and Paul Schedl. MB27 was constructed by Mick Browne as described

by Roiha and Glover (1981).

II.3. rRNA PREPARATION

Embryonic rRNA was prepared essentially as described by Rubin and Hogness (1975). Four to 19 hours embryos were collected and extensively rinsed with 0.7% NaCl on fine nylon gauze, dechorionated by soaking in 60% chloros for three minutes and rinsed again with 0.7% NaCl. The embryos were homogenised in cold 50 mM Na-acetate, pH 5, 1 mM EDTA, 0.5% diethylbicarbonate (DEP) using approximately 0.5 ml of this buffer per 0.5 g of enbryos (Levis and Penman, 1977). The homogenate was transferred into a centrifuge tube and mixed with an equal volume of 50 mM Na-acetate, pH 5, 1 mM EDTA, 1% SDS. The mixture was extracted twice with water saturated phenol and the aqueous phase layered onto a sucrose stepgradient containing equal volumes of 30, 25, 20, 15 and 10% sucrose in 50 mM Na-acetate, pH 5, 1 mM EDTA, 0.1% DEP. The gradient was spun in Sorvall SW27 rotor at 25 000 rpm for 15 hours. Fractions were collected in clean vials and 18S and 28S rRNA peaks were identified by UV illumination of small aliquots of each fraction spotted onto an agarose gel containing ethidium bromide. These fractions were pooled and the rRNA precipitated with ethanol. rRNA preparations were stored in a small volume of ethanol at -20 C.

rRNA from tissue culture cells was prepared from Eschalier's K_c-cells (Eschalier and Ohanessian, 1970). The cells were grown as described by Rubin and Hogness (1975). Cells grown for 4-5 days in 10 ml flasks were collected by centrifugation for 10 minutes at 2 000 rpm at room temperature and rRNA prepared as described above.

II.4. PHENOL EXTRACTION AND PRECIPITATION OF NUCLEIC ACIDS

Phenol extractions of DNA solutions were done by adding an equal volume of redistilled phenol saturated with TE_2, gently vortexing and separating the phases by centrifugation at 8 000 rpm for 10 minutes in the Sorvall SS34 rotor

or for 30 seconds in an Eppendorf microfuge. The phenol phase was back extracted with TE_2 and the aqueous phases pooled. The aqueous phase was extracted three times with ether to remove phenol.

Ethanol precipitation of DNA was done by adding 1/10 volume of 3 M NaCl or 1/10 volume of 0.3 M Na-acetate, pH 7 and 2.5 volumes of ethanol or an equal volume of isopropanol followed by storage overnight at -20 C or placing for 5-15 minutes in a dry ice-ethanol bath. Precipitates were recovered by centrifugation at 8 000 rpm for 20 minutes at -4 C in the Sorvall SS34 rotor or for 5 minutes in an Eppendorf microfuge.

Phenol extractions of RNA were performed in a similar way except that water saturated phenol was used.

RNA was precipitated by adding 1/10 volume of 2 M Na-acetate, pH 5 and 2.5 volumes of ethanol.

II.5. ENZYMES

Restriction endonucleases AluI, BglI, BglII, BclI, BstI, CauI, CauII, EcoRI, HaeIII, HhaI, HindIII, HinfI, KpnI, MspI, PstI, SacI, SalI, SmaI, SstI, TaqI and XmaI were used in this work. These enzymes were prepared by members of the Department or purchased from Bethesda Research Laboratories, Boehringer Mannheim AG, New England Biolabs Ltd or Miles Ltd. References for these enzymes can be found in the reviews by Roberts (1978, 1980).

T4 DNA ligase was prepared according to the method of Murray et al. (1979) by Lesley Woods. E.coli DNA polymerase I (Jovin et al., 1969) was the gift of Alan Fersht. Calf intestinal phosphatase was purchased from Boehringer Mannheim AG and further purified by Diane de Cicco as described by Efstratiadis et al. (1979). Endonuclease S1 (Vogt, 1973, 1980) was purchased from Sigma. Reverse transcriptase from avian myoblastosis virus was obtained from J. Beard, Life Sciences, Inc., St. Petersburg, Florida. Terminal deoxynucleotidyltransferase was purchased from Ratcliff Biochemicals, Los Alamos, New Mexico. Klenow subfragment of E.coli DNA polymerase I was

purchased from New England Biolabs Ltd.

II.6. RESTRICTION ENDONUCLEASE DIGESTIONS

The following buffers were used:

- BglI, BglII, BstI, EcoRI, HaeIII, HindIII, HinfI, SalI:
  100 mM Tris, pH 7.6, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM β-mercaptoethanol
- AluI, BclI, CauI, CauII, KpnI, MspI, SacI, SstI, TaqI, XmaI:
  6 mM Tris, pH 7.6, 6 mM $MgCl_2$, 6 mM β-mercaptoethanol
- SmaI, PstI:
  10 mM Tris, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 100 mg/ml bovine serum albumine

All the digestions were carried out for one hour at 37 C. In the case of multiple digestions involving enzymes requiring high and low salt concentrations, the low salt digest was done first, then salt concentration was increased and the second enzyme was added. The digestions were stopped either by freezing the samples at -20 C or by the addition of 1/10 volume of gel loading buffer (20% Ficoll, 25 mM EDTA, pH 7.5 and Orange G tracking dye) and electrophoresis.

II.7. GEL ELECTROPHORESIS

Electrophoresis of DNA was carried out on horizontal slab gels (20 x 18 x 0.4 cm) of 0.6 - 2.25% agarose in either TB buffer (89 mM Tris, pH 8, 89 mM boric acid, 2.5 mM EDTA) or TA buffer (40 mM Tris-acetate, pH 7.8, 5 mM Na-acetate, 1 mM EDTA). The agarose gels contained 1 ug/ml ethidium bromide. The gels were connected to the buffer by paper wicks. Agarose gels of 0.6 - 1% were electrophoresed at 1.5 V/cm for 13 hours or 10 V/cm for four hours. Gels of higher than 1% agarose were run at 10 V/cm for two hours. Samples were loaded in 20-200 ul of a buffer containing 2% Ficoll, 2 mM EDTA and Orange G tracking dye.

Single stranded DNA for S1 nuclease mapping gels was electrophoresed

essentially as described by Favaloro et al. (1980). 1.5% agarose gels were formed in 50 mM NaCl, 2 mM EDTA and run submerged in 30 mM NaOH, 2 mM EDTA at 2-5 V/cm for 8-16 hours. Samples were loaded in 10 ul of a buffer containing 2% Ficoll, 30 mM NaOH, 2 mM EDTA, 0.5% xylene cyanol and 0.5% Orange G.

Acrylamide gels (20 x 38 x 0.003 cm) of 8% or 20% were run in Tris-borate buffer (Peacock and Dingman, 1967). All sequencing gels contained 7 M urea and were run hot (about 28 mA). The polymerisation mixtures used for 8% and 20% acrylamide gels contained 7.8% (w/v) acrylamide, 0.2% (w/v) bis-acrylamide or 19.3% (w/v) acrylamide, 0.7% (w/v) bis-acrylamide, respectively, and 7M urea, 0.007% (w/v) ammonium persulphate, 0.007% (v/v) TEMED catalyst in Tris-borate buffer (Peacock and Dingman, 1967, Maniatis et al., 1975). Acrylamide stocks prepared in distilled water were previously deionised by stirring with amberlite MB-1. This solution was poured into a mold formed by sealing the glass plates with adhesive tape (UTF/1/AT, Universal Scientific Ltd.). Thin (0.003 cm) gels (Sanger and Coulson, 1978) were used. The gels were usually poured on the day before the run. The gels were run as described by Maxam and Gilbert (1980). After the run, one of the glass plates was removed, the gel transferred onto an exposed sheet of X-ray film, covered with cling film and placed in a light tight cassette. The gel was exposed to preflashed Fuji RX medical X-ray film at -70 C for 2-4 days (Laskey and Mills, 1976).

II.8. LENGTH STANDARDS FOR GEL ELECTROPHORESIS

DNA length standards were generated by appropriate restriction digestions of SV40, phage λ , pBR322 or pAT153 DNA. The sizes of SV40 fragments were calculated from the nucleotide sequence (Fiers et al., 1978). The sizes of phage λ DNA cleaved by EcoRI or HindIII were taken from the electrophoretic measurements made by Philippsen et al. (1978). The sizes of pBR322 and its derivative pAT153 were calculated from the nucleotide sequence (Sutcliffe,

1979).

A plot of electrophoretic mobility against the length of the standard restriction fragments (in kb) was used to calibrate the gel and so to determine the lengths (in kb) of unknown DNA fragments.

II.9. PURIFICATION OF RESTRICTION FRAGMENTS FROM AGAROSE GELS

Gel slices containing DNA fragments stained with ethidium bromide were cut out, passed through a syringe into an eppendorf or a centrifuge tube depending on the volume of the slice. The syringe was rinsed with an equal volume of 0.1 M Tris, pH 7.6, which was added to the agarose. Two volumes of phenol saturated with TE_2 were added and the mixture was incubated at 65 C for 10-20 minutes with frequent shaking till the agarose was melted. The phases were separated by centrifugation for 20 minutes at 9 000 rpm at 4 C in the Eppendorf microfuge or the Sorvall HB4 rotor. The aqueous phase was removed and the phenol and interphases were back extracted with fresh TE_2 buffer. The DNA in the pooled aqueous phases was re-extracted with an equal volume of phenol, the phenol phase back extracted with buffer, and the aqueous phases pooled and ether extracted. The DNA was ethanol precipitated twice, washed with cold ethanol and stored at -20 C.

II.10. IN VITRO LABELING OF DNA

DNA was labeled by nicktranslation to specific activities of 10^7-10^8 cpm/ug as described by Rigby et al. (1977). Depending on the required specific activity of the DNA either one or two ^{32}P α-deoxytriphosphates (Amersham, 350 Ci/mmol) were used in the reaction at a specific activity of 200 Ci/mmole. Reactions were terminated and unincorporated nucleotides removed by passing the reaction mixture over a Sephadex G-50 column equilibrated with TE_2.

The recessed 3' ends of double stranded DNA fragments (as created by HindIII) were labeled using reverse transcriptase essentially as described

by Wahl et al. (1979) or by Klenow subfragment of E.coli DNA polymerase I in 6 mM Tris, pH 7.5, 6 mM $MgCl_2$, 6 mM β-mercaptoethanol (Ross Miller, personal communication). 20-30 ul of α-^{32}P dTTP or α-^{32}P dGTP (Amersham, 2000-3000 Ci/mmole) was used in each reaction depending on the sequence of the 5' end template. The reactions were stopped by phenol extraction and ethanol precipitation of the DNA.

The 3' overhangs of double stranded DNA fragments (as created by HhaI) were labeled using terminal deoxynucleotidyltransferase as described by Maxam and Gilbert (1977). 100 uCi of α-^{32}P ATP (Amersham, 2000-3000 Ci/mmole) was used in each reaction. Alkaline hydrolysis was carried out overnight. Alternatively, ^{32}P cordycepin triphosphate (Amersham, 3000 Ci/mmole) was used and the reaction with no hydrolysis was performed as above.

II.11. IN VIVO LABELING OF rRNA

2.5 mCi of neutralised ~~phosphorus~~ phosphate (Amersham) per 100 ml culture was added ~~o~~to the cells twelve hours after innoculation and the cells were grown for further 4-5 days. rRNA was prepared as described above. 18S and 28S rRNA peaks were identified by Cerenkov counting of the fractions, the appropriate fractions were pooled and ethanol precipitated. rRNA was stored in ethanol at -20 C.

II.12. TRANSFER OF DNA FROM AGAROSE GELS TO NITROCELLULOSE AND HYBRIDISATION WITH ^{32}P - LABELED PROBE

DNA was transferred from agarose gels to nitrocellulose essentially as described by Southern (1975). After electrophoresis the DNA fragments in the gel were denatured in 0.5 M NaOH, 1.5 M NaCl for 1-2 hours, and the gel then neutralised in 1.5 M Tris, pH 7.4, 3 M NaCl for one hour. The nitrocellulose filter was rinsed several times in distilled water before the transfer which was carried out in 20 x SSC (SSC is 0.15 M NaCl, 0.015 M Na-citrate) for 12-48 hours. After the tranfer the nitrocellulose filter was rinsed in 2 x

SSC and baked at 80 C for a minimum of two hours.

A modification of the procedure described by Jeffreys and Flavell (1977) was used for the hybridisation of nicktranslated ^{32}P DNA to restriction fragments transferred to nitrocellulose filters. The filters were preincubated before hybridisation in 3 x SSC, 10 x Denhardt's solution (1 x solution contains 0.02% Ficoll, 0.02% bovine serum albumine, 0.02% polyvinyl pyrrolidine; Denhardt, 1966), 0.1% SDS at 65 C for three hours. The solution was then made to 100 ug/ml denatured sonicated salmon sperm DNA and the preincubation was continued for a further hour. The hybridisation was carried out at 65 C for 12-15 hours in the above solution (except that the SSC was at 5 x concentration) containing 10-20 ng/ml ^{32}P DNA labeled by nick translation. The nitrocellulose filters were washed 2 x 30 minutes in 2 x SSC, 5 x Denhardt's solution, 0.1% SDS, 0.1% tetrasodium pyrophosphate and 100 ug/ml denatured sonicated salmon sperm DNA, then 6 x 30 minutes in 0.2 x SSC, 0.1% SDS and finally 2 x 20 minutes at room temperature in 3 mM Tris base.

The hybridisation of DNA transferred to nitrocellulose filter with ^{32}P labeled RNA was carried out in 5 x SSC, 50% formamide at 45 C for 12 hours with ^{32}P in vivo labeled rRNA probe at 10 000 cpm/ml of hybridisation solution. The filters were washed 4 x 15 minutes in 5 x SSC, 50% formamide and 6 x 15 minutes in 2 x SSC at room temperature.

After drying the filters were autoradiographed at -70C using preflashed Fuji RX medical X-ray film.

The probe hybridised to DNA fixed on nitrocellulose filter could be removed, if the filter was not dried by baking at 80 C after the washes. The probe was removed by either extensively rinsing the filter in water at 65 C or by soaking in 30 mM NaOH at room temperature for one hour followed by neutralisation in 1.5 M Tris, pH 7.4, 3 M NaCl. After drying the filter could be preincubated and hybridised as described above.

II.13. SUBCLONING OF RESTRICTION FRAGMENTS

The vectors used for the subcloning of restriction fragments were pBR322 (Bolivar, 1977) and its shorter derivative pAT153 (Twigg and Sherratt, 1980) which contain genes for ampicillin and tetracycline resistancies. The vector and target plasmid were digested with the appropriate restriction endonucleases and the reactions were terminated by heating at 65 C for 10 minutes. The vector DNA was dephosphorylated to prevent recircularisation in the ligation reaction. The DNA was ethanol precipitated and resuspended in 10 mM Tris, pH 7.5, 10 mM $MgCl_2$ and calf intestinal phosphatase was added. Following incubation at 37 C for 30 minutes the reaction was terminated by phenol extraction and the DNA was ethanol precipitated. For the ligation, 0.1 ug of dephosphorylated vector DNA and 5 ug of plasmid DNA were mixed, heated at 65 C for 5 minutes, quickly cooled on ice and made 100 mM Tris, pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 15 mM dithio treitol and 1 mM spermine. DNA concentrations were between 50 and 100 ug/ml. T4 ligase was added and the reaction mixture was incubated for one hour at 15 C. Afterwards the reaction mixture was diluted with the above buffer to a DNA concentration of 5-10 ug/ml, more T4 ligase was added and the incubation was continued for a further 15-20 hours. The DNA was ethanol precipitated, resuspended in 10 mM Tris, pH 7.6, 1 mM EDTA and used to transform E.coli HB101 to ampicillin resistance as described by Mandel and Higa (1970). The ampicillin resistant colonies were grided on LA-plates containing 100 ug/ml ampicillin and the colonies transformed by the plasmid containing the wanted restriction fragment were identified by a modification of the colony hybridisation procedure of Grunstein and Hogness (1975) using gel purified restriction fragments labeled by nicktranslation as probes. Alternatively, small scale DNA preparations were made of some the ampicillin resistant colonies and their restriction pattern was compared to the original plasmid and the wanted subclone identifed.

II.14. THE CLONING OF rDNA REPEATS CONTAINING SHORT TYPE I INSERTIONS

D. melanogaster genomic DNA (100 ug) was cleaved by EcoRI and electrophoresed on 0.7% agarose gel. The region of the gel containing 12-17 kb long EcoRI fragments was cut out and the DNA purified as described before. Approximately 4.5 ug of such DNA was ligated with 1 ug of EcoRI cleaved and dephosphorylated pAT153 and used to transform E.coli HB101 as described above. The transformation efficiency was 7×10^3 colonies per 1 ug of vector and about 8 000 colonies were subsequently screened. Colonies transformed by a plasmid containing a short Type I insertion within rDNA were identified by a modification of the colony hybridisation procedure of Grunstein and Hogness (1975) using three gel purified fragments as probes. Colonies were selected which hybridised with the 12 kb EcoRI fragment containing an uninterrupted rDNA repeat (purified from pDm238 (Kidd and Glover, 1980) and with the 0.8 kb BstI fragment containing the right-hand end of a long Type I insertion (purified from pc25 (Kidd and Glover, 1980), but not with the 4.3 kb HindIII/BstI fragment containing the remaining left-hand part of the long Type I insertion (purified from pc2, Kidd and Glover, 1980).

II.15. COLONY HYBRIDISATION

A modification of the procedure described by Grunstein and Hogness (1975) was used. The colonies either spread or grided on LA-plates were replica plated on a nitrocellulose disk or a disk of Whatman 541 paper by placing the disk briefly onto the colonies. The replica colonies on the filters were grown on a fresh LA-plate (containing 100 ug/ml ampicillin) overnight or 2-4 hours depending on the strength of innoculum. The colonies on the filters were lysed by placing them on Whatman 3 MM paper soaked in 0.5 M NaOH, 1.5 M NaCl for 5 minutes and afterwards neutralised for two minutes on 1 M Tris, pH 7.4, 3 M NaCl. The disks were extensively rinsed in 2 x SSC to remove any agar or lysed debris and the colonies fixed onto the

disk by baking at 80 C for at least two hours. The procedure used for hybridisation of nicktranslated ^{32}P DNA to colonies bound on either nitrocellulose or Whatman 541 paper was the same as described above.

II.16. S1 NUCLEASE MAPPING

The S1 nuclease mapping procedure of Berk and Sharp (1977) was used with modifications as described by Favaloro et al. (1980). 100 ug of DNA and 90-10 ng of 28S rRNA and/or 50-5 ng of 18S rRNA were ethanol precipitated in a 1.5 ml eppendorf tube with 25 ug of E.coli tRNA carrier and resuspended in 10 ul of hybridisation buffer (80% formamide, 0.4 M NaCl, 0.04 M Pipes, pH 6.4, 0.001 M EDTA) (Casey and Davidson, 1977). The tubes were capped tightly and incubated submerged in a waterbath at 85 C for 15 minutes to denature the DNA. They were then rapidly transferred into an adjacent waterbath at 52 C, where the incubation was continued for two hours. Annealings were terminated by the addition of 200 ul of cold S1 digestion buffer (0.25 M NaCl, 0.03 M Na-acetate, pH 4.6, 4.5 mM $ZnSO_4$, 20 ug/ml denatured salmon sperm DNA) containing 200 units per ml S1 nuclease. S1 digestion was carried out for 30 minutes at 37 C and terminated by the addition of 75 ul of 2.5 M ammonium acetate, 50 mM EDTA. The DNA-rRNA hybrids were precipitated with 20 ug E.coli tRNA and an equal volume of isopropanol, and later resuspended in 10 ul of loading buffer (50 mM NaOH, 5 mM EDTA, 6.6% sucrose with bromphenol blue and xylene cyanol). The samples were then electrophoresed on alkaline gels, the DNA transferred to nitrocellulose filter and hybridised as described before.

II.17. DNA SEQUENCING

DNA fragments endlabeled and gel purified as described before were sequenced essentially as described by Maxam and Gilbert (1980). The reactions and base modifications chosen were C + T with hydrazine, C with hydrazine and salt, A over C with NaOH, and G with dimethylsulphate. 5 ug of alkaline

degraded salmon sperm DNA purified by phenol extraction was used as carrier in each reaction. Two time points were used for A and G reactions to recover reaction products for the entire size range (aliquots were added into stop mixture and stored on ice). The initial 90 C step of the A reaction was carried out in sealed capillaries submerged in water bath over a bunsen flame. 1 M Na-acetate was used in the initial precipitation of the A cleavages. The supernatants were removed with a pointed capillary using a water suction pump. After the precipitations and washes the DNA was dissolved in 25 ul 0.5 M piperdine and the strand scission was performed in sealed capillaries submerged in a water bath over a bunsen flame. After the reaction the contents of the capillary were rinsed into 400 ul of water and lyophilised overnight. Finally, the DNA was dissolved in de-ionised formamide containing 0.02 M EDTA, 0.02% xylene cyanol and bromphenol blue to give about 2×10^3 Cerenkov counts per ul and stored at -20 C. Before loading onto a sequencing gel the samples were immersed in boiling water for five minutes.

C H A P T E R III

RESULTS

III.1. INTRODUCTION

When rRNA was hybridised to Eco RI cleaved D. melanogaster DNA, many fragments shorter than the major 17 and 12 kb, which represent the complete interrupted and uninterrupted units were found to contain rDNA segments (David Glover, personal communication). As both the major classes of rDNA units only contain one EcoRI site in the 18S gene, it was postulated that these additional rDNA units contained insertions with EcoRI sites and would thus form a third class of rDNA repeats found in D. melanogaster. To obtain all these arrangements of rDNA in their genomic organisation it was advantageous not to use a bank of EcoRI digested DNA and a new library of randomly sheared fragments of D. melanogaster DNA constructed by linking the D. melanogaster fragments to the cloning vector by the dA/dT method was used instead. This library was screened for rDNA sequences, the positive clones were digested with EcoRI and the ones that contained rDNA fragments shorter than 12 kb were selected. The characterisation of five of these clones, cDm264, cDm207, cDm249, 14C9 and 4G3, and the insertions they contain is described in this chapter.

III.2. RESTRICTION MAPPING

The characterisation of these clones was initiated by mapping the cleavage sites of the restriction endonucleases EcoRI, HindIII and XmaI. The initial mapping of the clones cDm103 and cDm207 were previously described by Glover and Hogness (1977) and Glover (1977). Other restriction endonucleases which recognise hexameric sites and thus have few or no sites

in the vectors were chosen for the further mapping. The sites for these enzymes, BglII, SalI, KpnI, BstI, SstI and PstI are shown on the complete physical maps in Figure 3. Only the BglII and PstI maps of cDm103 were determined in this work (for the details of the KpnI and SstI maps, see Kidd and Glover, 1980, Kidd, 1980). The detailed maps for each enzyme which show sizes of restriction fragments are shown in Appendix I. The restriction endonuclease cleavage maps were derived from the lengths of fragments produced by complete digestions with a single enzyme or by double or triple digestions with more enzymes. Ambiguities were resolved using partial digestions or using subcloned or gel purified fragments as starting material for second cleavages.

These maps revealed new restriction fragments resulting from a second class of insertion sequences (the Type II insertions) in the 28S rRNA gene in these clones. The most characteristic of these seem to be the two small EcoRI fragments of 0.6 and 0.7 kb which are represented in each clone. Another characteristic feature is the cluster of restriction sites of SalI, BglII, KpnI and PstI near these EcoRI sites. Each of these clones has a similar restriction pattern in this region, the one exception being cDm207 in which the insertion is much longer (Glover, 1977). The restriction map which I have determined for the right-hand part of cDm207 resembles closely the Type I insertion in cDm103. This similarity was investigated further (see Section III.5.).

The rDNA sequences in the clones cDm207 and 1409 are interrupted in the transcribed spacer region by sequences which in their restriction patterns do not conform to rDNA segments (Figure 3). The interruptions are at different sites just to the left of the PstI site in the transcribed spacer region. The non-rDNA sequences at these sites in the two clones are not similar as they have completely different restriction site patterns. DNA blotting experiments (Southern, 1975) showed that the sequences which interrupt cDm207 are repetitive (data not shown). These sequences have not

been further characterised.

The restriction maps also show patterns of restriction sites characteristic of regions from the previously mapped rDNA clone, cDm103 (Glover and Hogness, 1977, Glover, 1977). The clones cDm264, cDm249 and 4G3 each have BglII, XmaI, EcoRI and HindIII sites which are characteristic of 18S rRNA gene. Each of the clones with the exception of cDm207 has a 0.9 kb HindIII fragment characteristic of the 3' end of the 28S rRNA gene.

In order to confirm that the familiar restriction fragments code for rRNA and to determine the extent of coding sequences present in the novel restiction fragments, ^{32}P labeled 18S or 28S rRNA was hybridised to double digestions of these clones using the method of Southern (1975). These results, for double digestions with EcoRI/BglII and EcoRI/HindIII are shown in Figure 4. The physical map on the top of the Figure represents the rDNA insert in cDm264. In order to facilitate the discussion of the results, restriction fragments, whose ends fall in coding or spacer sequences and thus represent restriction fragments characteristic of uninterrupted rDNA sequences are marked with the letters a to d and the fragments with an end in the Type II insertion with the letters w, x and z. The prime marks indicate a significant deviation on the length of corresponding fragments in the other plasmids when compared to cDm264. The corresponding fragments in the other plasmids are indicated on the autoradiograms. Because BglII and HindIII do not cleave the vectors (ColE1 and pSF2124) and the only EcoRI site in the vectors has been replaced by dA/dT bridges, the largest fragment in each digestion contains the vector and the flanking D. melanogaster sequences from the insert.

18S rRNA gene

These experiments show that the clones cDm207 and 14C9 do not contain any 18S rRNA gene sequences, since there is no hybridisation of 18S rRNA to any of their restriction fragments (Figure 4, panels B and E, tracks 2 and 5). The 18S gene contains sites for the three enzymes used in this

experiment. In the EcoRI/BglII double digestion of cDm103 three fragments of 6.2 kb, 3.5 kb (a) and 1.5 kb (b) contain the left, right and middle parts of the gene, respectively (Figure 4, panel B, track 1; for the map of cDm103, see Figure 3). As the 6.2 kb fragment spans the distance from the BglII site within the 18S gene to the BglII site in the Type I insertion (cDm103, Figure 3), it is not found in the clones which contain Type II insertions. The 1.5 kb fragment b which contains the main part of the 18S gene can be found in cDm249, cDm264 and 4G3 (Figure 4, panel B, tracks 3, 4 and 6). The 3.5 kb fragment a is found in cDm249 and cDm264 (Figure 4, panel B, tracks 3 and 4). The fragment w of cDm264 and w' of 4G3 map leftwards from the 1.5 kb fragment b. The difference in the lengths of these fragments from two separate clones reflect the length variations of the non-transcribed spacer regions. The vector fragment of cDm264 (Figure 4, panel B, track 4) hybridises to 18S rRNA at both its ends and thus shows stronger signal compared to the vector fragments of cDm249 and 4G3.

In the EcoRI/HindIII double digestion of cDm103 three fragments hybridise to 18S rRNA (Figure 4, panel E, track 1). Fragment c contains the non-transcribed spacer and the main part of the 18S gene. The slightly shorter fragment contains the sequence from the left-hand EcoRI site to the HindIII site at the left-hand part of the Type I insertion and thus is not seen in the other clones. The smallest fragment contains a short segment of the 3' end of the gene and is represented also in cDm249, cDm264 and 4G3 (Figure 4, panel E, tracks 3, 4 and 6). The non-transcribed spacer sequences of cDm264 and 4G3 map within fragments c and c', which show the same length difference as the fragments w and w' on the EcoRI/BglII double digestion. The fragments z and z' from cDm264 and cDm249, respectively, contain the region between the EcoRI sites in the 18S gene and the insertion. Their length difference of 1.4 kb is discussed in the connection with the 28S rRNA gene.

28S rRNA gene

These sets of double digestions contain fragments characteristic of uninterrupted parts of the 28S gene which are present on all the clones. One such fragment is the EcoRI/BglII fragment a which contains sequences from the left end of the 28S gene and the conserved transcribed spacer. It is found in cDm103, cDm249 and cDm264 (Figure 4, panel C, tracks 1, 3 and 4). The 0.9 kb EcoRI/HindIII fragment d from the 3' end of the 28S gene occurs in all the clones except cDm207 (Figure 4, panel F, tracks 1, 3, 4, 5 and 6). This conforms with the previous data which showed that cDm207 is interrupted by a long sequence and only contains the gene sequences leftwards from the insertion (Glover, 1977). Accordingly in the EcoRI/HindIII digestion, all the 28S rRNA conding sequences of cDm207 are included in the 7.5 kb EcoRI fragment (Figure 4, panel F, track 2). The large fragments containing the vector in EcoRI/HindIII double digestions of cDm264 and 4G3 both hybridise to 28S rRNA. This indicates the presence of a short segment of the gene on the right-hand junction of the vector in these clones.

There is a characteristic set of restriction fragments containing sequences from the 28S gene and Type II insertion. The 0.7 kb EcoRI/HindIII fragment y from the right-hand junction of the Type II insertion and the 28S gene is present on all the clones except cDm207 (Figure 4, panel F, tracks 3, 4, 5 and 6). The 2.3 kb BglII fragment x which contains the left-hand junction of the 28S gene and the Type II insertion can be found in cDm207, cDm264 and 14C9 (Figure 4, panel C, tracks 2, 4 and 5). The corresponding BglII fragment x' in cDm249 is 1.4 kb shorter, but it contains 28S rRNA sequences. The same length difference can be seen between the EcoRI/HindIII fragments z and z' in panel F, tracks 3 and 4. These fragments contain the bulk of the coding sequence. It also contains the 3' end of the 18S sequence and can also be seen in panel E, tracks 3 and 4. Altogether, this suggests that in cDm249 some sequences have been deleted

in the region of the 28S β1 gene and the left-hand part of the insertion.

III.3. S1 NUCLEASE MAPPING OF THE CODING REGIONS

The experiments in the previous Section only localise the coding sequences into certain restriction fragments. In order to accurately map the ends of both the genes and the site of the interruption within the 28S rRNA gene relatively to restriction sites within these genes, the S1 endonuclease mapping technique of Berk and Sharp (1977) was used. Mature 18S or 28S rRNA was annealed to EcoRI or BglII cleaved plasmid DNAs. Each of these enzymes cleaves within the 18S gene and BglII has a site in the 28S gene. The results of these experiments on the clones cDm264, cDm249, 14C9 and 4G3 are shown in Figure 5. Similar mapping was also performed in the previously characterised clones cDm103 (Figure 5) and cDm207 (data not shown).

18S rRNA gene

The 18S rRNA protects a 1.68 kb fragment when annealed to EcoRI cleaved cDm264, cDm249 and 4G3 (Figure 5, tracks 1). This determines the length of the main part of the 18S gene. A fragment of similar size is also protected by 18S rRNA in cDm103 (Figure 5) in good agreement with previous measurements of 1.68 kb R-loop structures of 18S rRNA annealed to cDm103 (White and Hogness, 1977).

In BglII cleaved plasmids a fragment of 1.58 kb is protected in each case (Figure 5, tracks 2). This locates the BglII site 1.58 kb from the 3' end of the 18S rRNA gene. BglII cleaved cDm264 annealed to 18S rRNA protected an additional fragment of 0.8 kb (Figure 5, cDm264, track 2). The restriction sites shown in Figure 3 indicate that this fragmental gene is located by the dA/dT bridge at the left-hand junction of the vector. Following EcoRI digestion the S1 protected fragment is 0.61 kb long (Figure 5, cDm264, track 1), which indicates that the EcoRI site is 0.19 kb from the 3' end of the gene. This short protected fragment is not shown on this

gel. These experiments confirm the previous observation that the clones cDm207 (data not shown) and 14C9 (Figure 5, 14C9, tracks 1 and 2) do not contain 18S rRNA sequences. The measurements on the other clones can be used to accurately locate the 18S rRNA coding sequence around the BglII and EcoRI sites as follows: 5'-0.29 kb-BglII site-1.39 kb-EcoRI site-0.19 kb-3'. The total length of the gene is 1.87 kb.

28S rRNA gene

Mature 28S rRNA consists of four polynucleotide chains: 28S α, 28S β, 2S and 5.8S which are held together by hydrogen bonds (Jordan et al., 1976, Jordan and Glover, 1977). The clones cDm264 and 14C9 show the same pattern of three protected fragments in hybridisation to purified 28S rRNA (Figure 5, cDm264 and 14C9, tracks 3 and 4). An identical pattern is obtained with cDm103 (Figure 5, cDm103, tracks 3 and 4). The protected fragments in the EcoRI cleaved plasmids are 1.73 kb, 1.26 kb and 0.78 kb. The 1.73 kb fragment corresponds to the 28S α chain which in electron microscopic R-loop experiments was previously estimated to be 1.75 kb (White and Hogness, 1977). The 1.26 kb and 0.78 kb fragments correspond to the 28S β chain which is interrupted by the insertion in these two plasmids. These lengths agree well with the previous measurements of 1.21 kb and 0.74 kb (White and Hogness, 1977). The total length of the 28S gene is 3.79 kb.

The BglII site within the 28S gene can be located within the 28S α sequence, since the S1 protected fragment in the BglII cleaved material is only 1.58 kb long (Figure 5, tracks 4). This places the BglII site 1.58 kb from the 5' end of the 28S α coding sequence and 0.14 kb left from the processing site between the 28S α and 28S β chains. The small fragment of 0.14 kb is not shown on this gel.

The EcoRI cleaved cDm249 protects only two of the three 28S rRNA chains, the 28S α and 28S β 2 chains (Figure 5, cDm249, tracks 3 and 4). This indicates that the 1.4 kb deletion localised to this region by DNA blots (Southern, 1975) must consist of the sequences coding for the whole

28Sβ1 chain. Since this chain is 0.78 kb long, the deletion must also include at least 0.6 kb of Type II insertion sequence.

The only 28S rRNA fragment protected by the clone 4G3 is the 28Sβ2 chain of 1.24 kb (Figure 5, 4G3, tracks 3 and 4). This was already suggested by the restriction mapping which showed that this clone is interrupted by the vector sequences at the left-hand part of insertion (Figure 3). The other 28S gene located to the right-hand part of the clones cDm264 and 4G3 in the experiments described in the previous Section must be less than 0.4 kb long as they are not represented on this gel.

III.4. HOMOLOGY BETWEEN THE CLONED TYPE II INSERTIONS

The Type II insertions in the four clones characterised in this work range between 2.7 and 3.4 kb. The sequence which interrupts the 28S rRNA gene in cDm207 is at least 8 kb long, but the restriction pattern in the left-hand part of the interruption is similar to the other clones which contain Type II sequences. Most of the restriction enzyme cleavage sites within the insertion are at identical positions in these clones (Figure 3), notably the three EcoRI sites and the cluster of sites for SalI, BglII, KpnI and PstI. Another characteristic fragment is the 0.8 kb EcoRI/HindIII fragment which spans over the right-hand junction of the Type II insertion and the 28S gene. This fragment is found in all the clones except cDm207, where the corresponding fragment is 0.7 kb long and does not contain any coding sequences (Figure 3). The 1.1 kb SalI and 1 kb SstI/EcoRI fragments from the left-hand part of the insertion are found in cDm264 and 14C9, and the SstI/EcoRI fragment also in 4G3, where the vector sequences interrupt the insertion between the SalI and SstI sites (Figure 3). In cDm207 these fragments are 0.2 kb longer. In the clones cDm264 and 14C9 the insertions are identical in length and also in the pattern of the restriction sites. In cDm249 there is a deletion of both insertion and coding sequences at the left-hand junction. In 4G3 the cloned part of the insertion is identical

to the one in cDm264.

In order to investigate the sequence homology within the Type II insertions in these clones, two fragments of the insertion sequence, 1.3 kb SalI fragment from cDm851 (see Appendix II) and the two small EcoRI fragments from cDm207, were gel purified, nicktranslated and used as probes in hybridisation to EcoRI/SalI digested plasmid DNAs (Figure 6). Altogether, these probes cover about 85% of the total length of the Type II insertions found in these plasmids. Panel A in Figure 6 shows the result in the hybridisation of the SalI fragment and Panel B of the EcoRI fragments. With both the probes the only hybridisation observed was to the corresponding fragments in the other plasmids: in cDm264 and 14C9 the SalI fragment is 0.2 kb shorter than in cDm207 as already observed when comparing the restriction maps, and the relevant sequences in cDm249 and 4G3 are included in the 4.3 kb and 21 kb fragments, respectively (Figure 3). This indicates that these parts of the Type II insertions have strong sequence homology in these clones.

III.5. HOMOLOGY BETWEEN THE INTERRUPTION IN THE 28S rRNA GENE IN cDm207 AND THE TYPE I INSERTION SEQUENCES

EcoRI cleavage of cDm207 results in four fragments of 0.6 kb, 0.7 kb, 7.5 kb and 12 kb. The sequence which interrupts the 28S gene in cDm207 is at least 8.1 kb long as it is interrupted by the vector sequences at its right-hand end (Glover, 1977). This interruption has a similar restriction pattern as the Type II insertions at its left-hand end and the sequences included in these fragments were found to be homologous to the Type II insertion (Figure 6). The restriction enzyme cleavage map of the right-hand part of this interruption resembles a Type I insertion (Figure 3, cDm103 and cDm207). The degree of sequence homology was tested in a similar experiment to that described in the previous Section. A HindIII/BstI fragment of 4.3 kb from pc225 (a subclone of cDm103 which contains

Type I sequences, see Appendix II) was gel purified, nicktranslated and used as aprobe in hybridisation to cDm207 which was cleaved by several restriction endonucleases as indicated in Figure 7. The restriction map of cDm207 shows the fragments used in this experiment and the letters on the map refer to the indicated fragments on the autoradiogram. As this autoradiogram shows, all the restriction fragments from the right-hand part of cDm207 show sequence homology to the Type I insertion. This homology extends leftwards up to the HindIII site, but the 2.1 kb BglII/ HindIII fragment which spans over the three EcoRI sites in the Type II insertion does not hybridise to the Type I sequences. Although the smallest SstI fragment of 0.4 kb is not shown on this gel, it is very likely to be homologous to the Type I sequences, as this arrangement of three SstI sites at the left-hand part of the sequence is characteristic of all the Type I segments cloned and investigated (Kidd and Glover, 1980). The sequence that interrupts the 28S rRNA gene in cDm207 thus consists of Type II insertion sequences linked to the Type I insertion sequences. Judging from the restriction sites, the change from one type of insertion to the other is direct and there are no other sequences between the two insertions detectable by restriction mapping.

The detailed restriction mapping shown in Figure 3 revealed that the Type I sequence in cDm207 is diverged from the insertion in cDm103 and more resembles the structure of some of the units in the clone cDm219. This clone contains several Type I sequence elements in tandem array and originates outside rDNA sequences in the heterochromatin (Kidd and Glover, 1980).

This experiment also shows that there is no cross-hybridisation between the restriction fragments from the Type II and the Type I insertions. As the Type I probe covers about 80% of the 5 kb insertion, this indicates that the whole of the Type I and Type II insertions do not have enough homology to be detected by the hybridisation technique used.

III.6. DISTRIBUTION AND LENGTH VARIATION OF THE TYPE II INSERTION SEQUENCES IN GENOMIC DNA

The Type I insertion sequences within rDNA show considerable length heterogeneity. The majority of these insertions are 5 kb long and 1 kb and 0.5 kb insertions homologous to the right-hand end of the long 5 kb insertions form a minor class (Wellauer and Dawid, 1977, Wellauer et al., 1978). The clones characterised on this work indicate that the Type II insertions show some length heterogeneity: in cDm264 and 14C9 the insertion in 3.4 kb long and cDm249 2.4 kb. The part of the insertion in cDm207 which contains the Type II insertion sequences is about 3.6 kb long. I have investigated the extent of this heterogeneity and the distribution of the Type II sequences within D. melanogaster genome using DNA blotting (Southern, 1975).

In order to examine the general length heterogeneity of the Type II insertion, HindIII cleaved D. melanogaster DNA was hybridised with the nicktranslated EcoRI fragments from the middle part of the Type II insertion (pcDm207cd, see Appendix II). The major band of hybridisation is identical in size to the 8.3 kb fragment from cDm264 (Figure 8). This HindIII fragment of cDm264 contains the entire Type II insertion with the conserved flanking sequences between the HindIII sites in the 18S and 28Sβ2 genes. This result suggests that the insertion in the clone cDm264 represents the most abundant size class of the Type II insertions. Additional genomic HindIII fragments are also seen to hybridise to the probe in the long exposure shown in Figure 8. These include fragments slightly larger than 8.3 kb and smaller fragments down to 3.7 kb in length. The corresponding 6.8 kb fragment in the clone cDm249 could be included in the band of strong hybridisation between 5 kb and 8.3 kb, but the shorter discrete bands at 5.3 kb, 4.3 kb and 3.7 kb are not represented in the other clones. These fragments can only be seen in a heavily exposed autoradiogram and thus represent minor size classes of the Type II sequences. The

weakness of the signal suggests that they may be present only in a few copies per genome.

The length heterogeneities within the different parts of the Type II insertion were investigated by hybridising gel purified probes from within these parts of the insertion to appropriate digestions of D. melanogaster DNA using cDm264 digests as size markers. The insertion was divided into three parts using BglII, EcoRI and EcoRI/HindIII digestions.

BglII cleaves cDm264 within the 28S gene and the Type II insertion to generate a 2.3 kb fragment. Since the BglII site is in the 28Sα gene, probes included in the insertion would only show the length variation within the insertion sequence in this fragment. This 2.3 kb fragment is the only one detected in a short exposure of the filter, when a probe from the left-hand part of the insertion (the 1.2 kb SstI/BglII fragment purified from 264 R12, see Appendix II) is hybridised to total genomic DNA (data not shown). This again suggests that this region of the Type II insertion as represented in cDm264 is the major arrangement. In a longer exposure (Figure 9 A, track 2) 2.5 kb fragment and shorter fragments of 2.1 kb, 1.8 kb and 1.5 kb are seen. This suggests length heterogeneity within the left-hand part of the Type II insertion. The 2.5 kb fragment corresponds to the fragment found in cDm207. The length of the shortest of these fragments corresponds to a BglII fragment which contains only about 0.5 kb of insertion sequence leftwards from the BglII site in the insertion. The presumably rare rDNA units with deletions of the coding sequence such as observed in the clone cDm249 would also be included in these shorter BglII fragments.

The length heterogeneity of the left-hand part of the Type II insertion was also examined in SalI/EcoRI digests of D. melanogaster DNA probed with the SalI fragment of 1.3 kb purified from cDm851 (see Appendix II). The most prominent band of hybridisation is at 1.1 kb corresponding to the fragment of identical size in cDm264 (Figure 9 B,

track 2). Much fainter bands of 4.5 kb, 4.3 kb and 1.3 kb are also seen. The 1.3 kb fragment corresponds to the fragment found in cDm207. The long fragments could represent insertions where some of the left-hand sequences have been deleted as already described in the previous experiment. When these deletions remove the left-hand SalI site, only long EcoRI/SalI fragments which include the segments of rDNA unit to the EcoRI site in the 18S gene would hybridise with the probe used. However, these bands hybridise faintly and thus represent a very minor class of insertion.

The sequence variation within the two small EcoRI fragments was investigated using as probes these fragments purified from pcDm207cd (see Appendix II). In a short exposure only these two bands could be detected (data not shown). On a longer exposure (Figure 9 C, track 2) three other fragments of 7.3 kb, 5.5 kb and 4.3 kb can be detected. Some of these may include insertions with long deletions of the left-hand sequences including one or even two of the EcoRI sites. Alternatively, these fragments may represent hybridisation of the probe to genomic sequences other than rDNA. When the hybridisation intensities between these long fragments and the short pair are compared, it is obvious that the long fragments are very much less frequent and again may represent only a few copies per genome.

In order to examine the length divergence within the right-hand part of the Type II insertion. D. melanogaster DNA was cleaved with EcoRI and HindIII. In cDm264, this cleavage results in a 0.76 kb fragment which consists of 0.4 kb of the Type II insertion sequence and 0.34 kb of the 28Sβ2 gene (for the mapping of this fragment, see Section III.7.). This digestion was probed with a 0.32 kb EcoRI/HinfI fragment from p264 RH23 (see Appendix II) which includes only insertion sequences. As the HindIII site is in the 28Sβ2 gene, this experiment monitors the length variation within the 0.4 kb of insertion sequence included in the EcoRI/HindIII fragment. As shown in Figure 9 D, track 2, the probe hybridises only with

with the 0.7 kb EcoRI/HindIII fragment in D. melanogaster DNA.

III.7. THE DNA SEQUENCES FLANKING THE TYPE II INSERTION

Introduction

The nucleotide sequences at the junctions between the gene and the intervening sequence have been determined for the 28S genes from several organisms (Wild and Sommer, 1980, Allet and Rochaix, 1979, Dujon, 1980, Rae et al., 1980). It has not been possible in any of these cases to recognise a consensus sequence which might serve as a recognition site for processing enzymes such as has been postulatedfor genes transcribed by RNA polymerase II (Breatnach et al., 1978). It was of interest to determine the nucleotide sequence at the junctions between the insertions and the 28S gene in D. melanogaster, since there are major differences between D. melanogaster and the other systems with split 28S rRNA genes. In D. melanogaster there are both uninterrupted genes and genes split by one of the two types of insertions (Wellauer et al., 1978) and furthermore genes containing insertions are probably not transcribed (Long and Dawid, 1979, Long et al., 1981c, Kidd and Glover, 1981). Another interesting feature was the finding that in D. virilis, a fourteen nucleotide sequence present once at the site of the insertion in the uninterrupted rDNA unit is repeated at the both sides of the insertion (Rae et al., 1980), a structure which has so far been associated only with eukaryotic transposing elements and proviruses (Dunsmuir et al., 1980, Farabaugh and Fink, 1980, Calos and Miller, 1980).

The clones and restriction mapping

I have sequenced the junctions between rDNA and the major Type II insertion in cDm264 and between the Type II insertion and the Type I insertion in cDm207. The restriction fragments which contain the junctions were first subcloned in pBR322 to facilitate the additional restriction mapping necessary for sequencing by the Maxam-Gilbert technique (1980).

The left-hand junction in both cDm264 and cDm207 is included in 1.1 kb BglII/SalI fragments. These were subcloned into Bst. I/SalI cleaved pBR322 (p264 BS21 and p207 BS2, see Appendix II). It can be estimated that the SalI site is about 100 nucleotides from the junction, since the coding sequence within the 1.1 kb subcloned fragments is about 1 kb long (the BglII site was mapped about 140 nucleotides from the right-hand end of the 28Sα chain, the "gap" between 28Sα and 28Sβ chains is about 120 nucleotides long (White and Hogness, 1977) and the 28Sβ1 gene is about 780 nucleotides long).

The right-hand junction in cDm264 is included in the 0.76 kb EcoRI/HindIII fragment which was subcloned into EcoRI/HindIII cleaved pBR322 (p264 RH23, see Appendix II). The right-hand junction between the Type II insertion and the 28Sβ2 gene in p 264 RH23 can be placed about 0.3 kb from the HindIII site using measurements from the corresponding region in cDm103. The junction between the Type I insertion and the gene in cDm103 is included in 0.3 kb BstI/HindIII fragment, pc4 (see Appendix II), which was known to contain only a very short segment of the insertion sequence (Simon Kidd, personal communication). As the average nucleotide sequence ladder in thin 8% acrylamide gels can be accurately read only up to 250-300 nucleotides, it was necessary to determine additional restriction sites within the subcloned fragment. This was performed by purifying the 0.7 kb EcoRI/HindIII fragment and cleaving it with restriction enzymes not previously used (AluI, BclI, BglI, CauI, CauII, HaeIII, HhaI, HinfI, MspI and TaqI). All these enzymes, except BclI, BglI and HaeIII, cleaved within the fragment, but only AluI, HhaI and HinfI were chosen for further mapping. The map of p264 RH23 is shown in Appendix I.

The junction of the Type II and Type I insertions in cDm207 is included in 0.66 kb EcoRI/HindIII fragment which was subcloned into the similarly cleaved pBR322 (p207 RH16, see Appendix II). It was possible to estimate the transition from one type of insertion to the other to be

around 0.2 kb from the HindIII site. This corresponds to the position of the transition from the gene to the left-hand part of the Type I insertion in cDm103. A direct transition between the two types of insertion within p207 RH16 was suggested by the observation that the estimated lengths for the insertion sequences involved, 0.46 kb and 0.2 kb, accounted for the total length of the junction fragment. Additional restriction mapping was performed as with p264 RH23 using several restriction enzymes (AluI, BclI, HaeIII, HhaI, HinfI, MspI, XbaI and XhoI). All these enzymes, except BclI and HaeIII, cleaved within the fragment. The AluI, HhaI and HinfI sites were accurately mapped (see Appendix I).

When the restriction endonuclease maps of these subclones were compared, it was found that their left-hand regions were similar up to the HhaI site (see Appendix I). It was concluded that the HhaI site is very near the junction and so sequencing was performed from the HhaI sites and also from the HindIII sites in both these subclones. Additionally, DNA around the HinfI site of p207 RH16 was sequenced as shown in Figure 10.

The end labeling, gel purification of DNA fragments was performed as described in Chapter II. Most of the fragments were sequenced twice using independently labeled material. To minimise errors, each sequencing ladder was read only in the well separated part and only sets of sequencing chemistry which did not result in any ambiguous signals were used. Gaps in the nucleotide ladder at modified bases were not observed in this work.

The nucleotide sequence flanking the Type II insertion in cDm264

The general strategy for sequencing these junctions is shown in Figure 10. A total of 276 nucleotides were sequenced from the SalI site in p264 BS21. The part of this sequence which contains the junction is shown in Figure 11a. A total of 290 nucleotides were sequenced leftwards from the HindIII site of p264 RH23 together with 126 nucleotides leftwards and 40 nucleotides rightwards from the HhaI site. These sequences over-

lap and together give a total of 435 nucleotides. The sequence around the right-hand junction is shown in Figure 11b.

The sequence of the 28S β gene in the vicinity of the insertion in an uninterrupted rDNA unit has been determined by Ross Miller and Lesley Woods (Roiha et al., 1981). This sequence is shown in Figure 12. The sites of the Type I and Type II insertions were positioned by comparing this sequence with that of the comparable regions from the subcloned left and right junctions. It was found that the Type I and Type II insertions occur at different sites separated by 51 nucleotides within the 28S β gene. In Figure 12 the sites of the insertions are marked with arrow heads. The Type II insertion is located five nucleotides from the position of the intervening sequence in the 28S rRNA gene in some strains of Tetrahymena (Wild and Sommer, 1980). The division of the 28S β gene into the units shown in boxes in Figure 12 is to facilitate comparison between the structures of the two types of insertion in different plasmids and does not imply any functional significance of these segments.

In Figure 11 the gene sequences are shown in boxes. At the left-hand junction of the Type II insertion and the gene the only conspicious structural element found is a direct repeat of six nucleotides, GGGGAG (Figure 11a). These repeats are contiguous and are formed partly from rRNA coding sequence and partly from the insertion sequence. The sequencing gel of this region is shown in Figure 13. There are no such repeats at the right-hand junction, which instead shows a tract of 22 dA residues in the insertion sequence (Figure 11b). These are directly followed by the coding sequence. The sequencing gel showing the right-hand junction is shown in Figure 14. The arrangement of sequences of the junctions of the Type II insertion as found in cDm264 are shown in a schematic form in Figure 15.

The nucleotide sequence flanking the Type II insertion and the 28S β1 gene in cDm207

A total of 270 nucleotides were sequenced leftwards from the SalI site in p207 BS2. The relevant part of this sequence is shown in Figure 11a. When this sequence was compared to the corresponding sequence in p264 BS21, it was found that the Type II insertion occurs at exactly the same site in both these plasmids. The structure at the junction of the Type II insertion and 28S β 1 gene in cDm207 differs from the one in cDm264 by an additional 23 nucleotides which occur between the two repeated hexameric sequences. The sequencing gel of this region is shown in Figure 13. There are two additional single base differences in the sequenced segments of cDm207 and cDm264 as indicated in Figure 11a.

The nucleotide sequence at the transition from the Type II to the Type I insertion in cDm207

In p207 RH16 a total of 255 nucleotides were sequenced leftwards from the HindIII site, 160 nucleotides leftwards from the HhaI site and 230 nucleotides rightwards from the HinfI site following the strategy shown in Figure 10. These sequences form an overlapping segment of 340 nucleotides leftwards from the HindIII site. The part of this sequence which contains sequences homologous to the Type II insertion are shown in Figure 11a and the Type I insertion in Figure 16. The structure of the right-hand junction of the Type II insertion and the 28S gene segment β' in cDm207 is very similar to the one found in cDm264 (Figure 11b). A tract of 19 dA residues occurs within the insertion sequence at the junction and there is a single base change within the insertion sequence. In cDm207 the Type II insertion sequences are linked to the Type I insertion sequences with 51 nucleotides of rDNA. These nucleotides correspond to the segments β' and β'' which occur between the sites of the two types of insertion (Figure 12). The junction between the β'' segment of the 28S gene and the Type I insertion in cDm207 is identical to that found at the left-hand end of

the insertion in cDm103 and the units of cDm219 (Figure 16). The segment of the Type I insertion in cDm207 which I have sequenced has several single base changes and an internal deletion of 22 nucleotides compared to the insertion in cDm103. These differences are indicated in Figure 16. The sequencing gel of this region is shown in Figure 17.

The structure of cDm207 is shown schematically in Figure 15 alongside the structures of the Type II insertion in cDm264, the Type I insertion in cDm103 and the Type I segments in cDm219. The corresponding region flanking the insertion in D. virilis is also shown (Rae et al., 1980).

III.8. CHARACTERISATION OF D. melanogaster rDNA CLONES WITH SHORT TYPE I INSERTIONS

Introduction

The nucleotide sequence at the junctions of long Type I sequences with rDNA in two different configurations was previously determined by Ross Miller and Lesley Woods (Roiha et al., 1981). The cloned DNAs which they sequenced were an rDNA unit containing a 5 kb Type I insertion cDm103 (Glover and Hogness, 1977) and a fragment from the chromocentral heterochromatin consisting of five tandemly repeated Type I elements of about 3.5 kb cDm219 (Kidd and Glover, 1980). In addition, the left-hand junction of the Type I element with rDNA has been determined in cDm207 (see Section III.7.). In each of these cases the junctions are identical. They differ from the arrangements of sequences around the insertion in D. virilis rDNA (Rae et al., 1980). In both D. melanogaster and D. virilis the insertions are located at an identical site in rDNA, but the long Type I insertions in D. melanogaster are not flanked by sequence duplications and at their left-hand sides there is a deletion of nine nucleotides of rDNA coding sequence (Figure 18). This segment has been termed β''' (Figures 12 and 15). The arrangements of sequences at the right-hand junction are, however, similar in both D. melanogaster and D. virilis

(Figures 15 and 18). As the majority of the Type I segments is found in one of these arrangements, the identity of the sequences at the left-hand junctions in these clones suggests that they represent the principal sequence arrangement in the genome. This is surprising since it would seem that an rDNA unit which is incapable of producing functional rRNA because of the deletion of nine nucleotides of rDNA (β^{111}) has been spread around in the genome.

Besides the major 3.5 to 5 kb long Type I insertions there is a set of shorter Type I insertion (Wellauer and Dawid, 1977, Wellauer et al., 1978). The majority of these are either 0.5 kb or 1 kb in length. They have sequence homology to the extreme right-hand part of the long insertion (Wellauer et al., 1978, Dawid et al., 1978). I have determined the sequences at the junctions of these short Type I insertions in order to find out whether the previously characterised structure of junctions was the only one in the genome and to see to what extent the ends of these short Type I insertions resemble those of transposable elements. The entire lengths of these short insertions were sequenced together with the corresponding region in the long 5 kb Type I insertion in cDm103 to determine the degree of sequence homology between these regions.

The clones and restriction mapping

The clones which contain rDNA sequences interrupted with short Type I insertions were selected from libraries constructed as described in Chapter II and by Roiha and Glover (1981). These libraries were screened for hybridisation to rDNA and sequences from the right-hand, but not to the sequences from the left-hand part of the long Type I insertions. The plasmid DNA preparations were digested with EcoRI, HindIII and BstI, since the restriction patterns of these enzymes were sufficient to characterise an rDNA clone with Type I insertion from the right-hand part of the long 5 kb insertion. The maps of the region around the site of the insertion in the 28S β1 - 28S β2 gene in these clones and the

corresponding region in cDm103 are shown in Figure 19.

The clone RI10 was defined to be an rDNA clone containing an insertion of approximately 0.5 kb as it has only one BstI site (BstI sites are found only in the insertions in rDNA) and three BglII sites (two of which are located in the 28S∝ and 18S genes). It also contains the expected BglII fragment of approximately 1.2 kb between the sites in 28S∝ gene and the insertion. A fragment of this size is not found in cloneswhich are uninterrupted or which contain a long Type I insertion.

The clones RI9 and MB27 were both originally thought to contain single insertions of slightly longer than 1 kb. These clones have two BstI sites 1.2 kb and 0.8 kb apart, respectively. Both clones have two BglII sites 0.46 kb apart in the insertion within the short BstI fragments (two additional BglII sites are located in the 28S∝ and 18S genes). New BglII fragments of 1.3 kb and 1.5 kb, respectively, are present in both these clones. Both of these new fragments are cleaved by BstI. The left-hand BstI/BglII fragment in the insertion of MB27 is 0.6 kb long. The corresponding fragment in RI9 is 0.44 kb and it was found to contain a HindIII site. DNA sequencing eventually revealed that RI9 was a rearranged segment of rDNA which contains two Type I insertions.

Each of these clones and cDm103 have two characteristic KpnI sites within their insertions. It was necessary to map further sites within the central BglII fragment in order to have convenient sites for the second restriction cleavages and to be able to sequence the whole BglII fragment. The single HaeIII site and the HinfI sites leftwards from it were mapped using the purified 0.44 kb BglII fragment of MB27, RI9 and cDm103 and the 1.2 kb BglII fragment of RI10. The complete maps of this region are shown in Figure 19.

The nucleotide sequences flanking the short Type I insertions

The sequencing was performed according to the strategy shown in Figure 19. The arrows indicate the extent of the nucleotides sequenced

in each end labeled fragment. The right-hand 1122 nucleotides of the 5 kb Type I insertion in cDm103 are shown in Figure 20 together with the flanking 28S gene sequence which is underlined. The insertion sequence is given negative numbers from the junction and the rDNA is numbered with respect to the position of the Type II insertion (Figure 11). The junctions were identified by comparing these sequences to the sequence of the uninterrupted 28S gene (Figure 11).

Each of the short Type I insertions has an identical right-hand junction with rDNA to the one found in cDm103 and also in D. virilis (Roiha et al., 1981, Rae et al., 1980). The sequences of these junctions are shown in Figure 21. In this Figure the rDNA sequence is underlined and the insertion sequence overlined. A sequencing gel of the right-hand junction in the clone MB27 is shown in Figure 22.

The structures of the left-hand junctions differ from the one determined for the long insertion in cDm103 or in the Type I insertion in cDm207. There is no deletion of the coding sequence in any of these clones and three of the short insertions are flanked on both sides by an rDNA segment which is present as a single copy in the uninterrupted rDNA unit. However, there is a variety of arrangements of the rDNA sequences at the left-hand junctions of the short insertions and the gene. The sequences of the left-hand junctions are shown in alignment in Figure 21 to facilitate comparison. The sequence of the uninterrupted rDNA is shown for comparison in Figure 21c. This sequence and the sequence of the left-hand junction in cDm103 were determined by Ross Miller and Lesley Woods (Roiha et al., 1981). In this Figure the duplicated segments are indicated by the asterisks. The length of the duplicated sequence varies from seven to fifteen nucleotides. In Figure 22 the dotted lines placed against the sequence of the MB27 right-hand junction indicate the variation in the length of the duplication. The left end points of these three short insertions are shown in Figure 20 by the solid arrows at the

positions -1008, -528 and -50 for the 1 kb, 0.5 kb and 0.047 kb insertions, respectively, below the 1122 nucleotides from the right-hand end of the long Type I insertion.

The clone RI10 contains an insertion of approximately 0.53 kb (Figure 20). It is flanked by a 15 nucleotide repeated sequence (Figure 21b). This sequence repeat is entirely contained within the rDNA at the right-hand junction. At the left-hand side the junction cannot be unambiguously placed. The left-hand ten nucleotides of the repeated segment must be included in the rDNA coding sequence, but the remaining five could be contiguous with either rDNA (Figure 21c) or the insertion sequence (Figure 20), since common nucleotides occur at these positions in the two sequences. The sequence of the left-hand junction in the clone RI10 is shown in Figure 22 where the dotted lines show the extent of the duplication.

The clone MB27 has a 1 kb insertion which is flanked by a repeated segment of seven nucleotides (Figure 21b). There is an ambiguity whether the last nucleotide of the repeat at the left-hand side is contributed by the Type I sequence (Figure 20) or the gene (Figure 21c). The left-hand junction sequence of MB27 is shown in Figure 22.

The clone RI9 contains a rearranged rDNA segment (see Chapter IV) with two Type I insertions 380 nucleotides apart (Figure 19). The shorter of these is only 47 nucleotides long. The dotted lines in Figure 19 indicate the displacement of this insertion to bring its right-hand junction into alignment with the other clones. It is flanked on both sides by an unambiguous fourteen nucleotide repeat contiguous with the rDNA coding sequence (Figure 21b). The larger insertion of 747 nucleotides is located downstream in the rDNA unit (Figure 19). Its right-hand junction is similar to the ones found in the other clones (Figure 21a), but its left-hand junction is at an unusual site in the 28S $\beta 2$ gene. To locate this sequence rearrangement within the 28S β 2 gene, I determined the

sequence rightwards from the HindIII site in the 28S β2 gene in the clone MB27. This site is indicated by an open arrow at the position +440 below the rDNA sequence in Figure 20. The other open arrow at the position -748 in Figure 20 locates the junction in the insertion sequence. This insertion is not flanked by a short duplicated segment and there is a direct transition from the insertion sequence to the 28S β 2 sequence at the left-hand junction (Figure 21a). The nature of the rearrangement, however, means that the insertion is flanked by a direct repeating sequence of 380 nucleotides. It is not clear whether this has the same significance as the short repeats observed flanking the other insertions.

The sequence homogeneity between the short and long Type I insertions

It was known from the previous hybridisation and heteroduplex experiments that the Type I insertions of different lengths are related in sequence; the 0.5 kb insertion is homologous with the right-hand half of the 1 kb insertions which themselves are homologous with the right-hand part of the 5 kb insertions (Wellauer et al., 1978, Dawid et al., 1978). When the entire sequences of these shorter insertions were compared to that of the right-hand part of the 5 kb insertion in cDm103 (Figure 20), it was found that each of the shorter insertions is an extremely homologous segment of this sequence. There are only a few scattered single base differences between the long and short insertion sequences. These are listed in the legend to Figure 20. The sequence divergence is 0.28% in RI10, 0.39% in MB27 and 0.40% in the 747 nucleotide insertion in RI9. The short insertion in RI9 is homologous to the corresponding sequence in cDm103. The flanking coding sequences are entirely homologous with the exception of one base difference in RI9 (see legend to Figure 20).

There was an ambiguity about two nucleotides at the positions -193 and -191 within the insertion sequences (Figure 20) in all the clones, when plasmid DNA cleaved with BglII and labeled at the 3' end (Figure 19) was used. When the corresponding sequence was read in the other strand

following 3' labeling of BstI cleaved DNA, no ambiguities were found. The following corrections should be made for the sequence shown in Figure 20: -193: A, -191, no nucleotide.

C H A P T E R IV

DISCUSSION

IV.1. CHARACTERISATION OF THE TYPE II INSERTION SEQUENCES

This thesis work is primarily conserned with the characterisation of the Type II sequences found in D. melanogaster rDNA. The Type II insertions have been defined as a class of insertion which contain EcoRI sites. The clones characterised in this work contain complete insertions with three characteristic EcoRI sites. Their Type II insertions are internally homologous but have no sequence homology with the Type I insertions. The lengths of these Type II insertions vary between 2.4 and 3.6 kb.

Since the Type II insertion is not cleaved by HindIII and since the flanking HindIII sites are contained within conserved sequences in the transcription unit, DNA blots of HindIII cleaved genomic DNA hybridised with Type II sequences could be used to analyse the size variation of the Type II insertions in the D. melanogaster genome. This analysis shows that the Type II insertions are less heterogenous than the Type I insertions and range in length from about 3.6 to 0.5 kb. The 3.4 kb insertion present in cDm264 is the most abundant size class. The shorter insertions seem to be represented only by a few copies in the genome. It is not certain that the 4.3 and 3.7 kb HindIII fragments which hybridise to Type II sequences are derived from rDNA as they are too short to contain a full complement of flanking rDNA sequences. They may represent rare arrangements of Type II insertion sequences containing HindIII sites or they may share only short regions of homology with the Type II insertion sequences and originate elsewhere in the genome. For

the main part the length heterogeneity within the Type II insertions is limited to the left-hand part of the insertion and, similarly to the Type I insertions, the right-hand sequences are present in the shorter insertions. Similar results were obtained by Long et al. (1981c).

Genomic DNA blots reveal a major difference in the genomic distribution of the two types of insertions. DNA blots of EcoRI cleaved genomic DNA hybridised with Type I insertion sequences show a complex pattern of homologous fragments (Dawid and Botchan, 1977, Kidd and Glover, 1980), indicating that roughly half of these sequences are not inserted within rDNA but are located elsewhere in the genome. In the genomic DNA blots hybridised with different parts of the Type II insertion sequences the major hybridisation occurs with a fragment of identical size to the comparable fragment in cDm264, while the minor hybridisation results from fragments representing the rare shorter Type II insertions. This indicated that, unlike the Type I insertion sequences, the Type II sequences are to a large degree only found inserted within rDNA sequences.

IV.2. THE NUCLEOTIDE SEQUENCES FLANKING THE TYPE II INSERTIONS

When the junctions of both types of rDNA insertion were compared with the sequence of the uninterrupted gene, it was found that the insertions occur at different sites separated by 51 nucleotides. Furthermore, the structures of the junctions of the Type II insertion sequences are quite different from those of the 5 kb Type I insertion in D. melanogaster and the 10 kb insertion in D. virilis (Roiha et al., 1981, Rae et al., 1980). In the two clones characterised in this work there is neither a deletion of adjacent rDNA sequences as found at the left-hand junction of the 5 kb Type I insertion in D. melanogaster nor a duplication of a short segment of the coding sequence from the right of the insertion as found in D. virilis.

A characteristic feature of both the right-hand junctions between

the 28S β2 gene and the Type II insertion characterised in this work is the presence of a strech of dA residues in the insertion sequence. In both the clones characterised in this work and by Dawid and Rebbert (1981) the Type II insertion interrupts the 28S β2 gene at an identical site. The length of the strech of dA residues found in all these clones varies between 13 and 22 residues. There is perfect homology between all the Type II insertion sequences left from the strech of dAs in spite of the different number of dAs present. An AT-rich region seems to be a common feature of the right-hand junction of the insertions in the rDNA of Tetrahymena, Chlamydomonas chloroplasts and yeast mitochondria for which the AT-content of the right-most 25 nucleotide pairs is about 75% (Wild and Sommer, 1980, Allet and Rochaix, 1979, Dujon, 1980). Similar strèches have also been observed in the ends of some classes of transposable sequences (Dawid et al., 1981, Pan et al., 1981). It has been postulated that AT-rich regions are an important feature at processing sites of the ribosomal transcripts (Jordan et al., 1980). Subsequently it was found that the most abundant RNA molecule homologous to the Type II insertion sequence seems to have an end point at the dA tract at the right-hand junction (Kidd and Glover, 1981). Alternatively, this region could act as a termination signal for transcription.

The left-hand junction between the 28S β1 gene and the insertion of about 3.5 kb in cDm264 and cDm207 are at an identical site. A hexamer repeat, GGGGAG, formed partly from rDNA sequence and partly from insertion sequence is found at the junction in both these clones. The insertion sequences are identical, except that 23 nucleotides of extra sequence occurs between the two repeats in cDm207. In the clones characterised by Dawid and Rebbert (1981) the left-hand junctions are different from each other and also from those in cDm264 and cDm207. In a clone carrying a Type II insertion of 1.4 kb characterised by Dawid and Rebbert (1981) the breakpoint is at an identical site with cDm264 and cDm207, but

insertion sequences, as one would expect, have no homology. In a clone containing an insertion of 3.5 kb, which is similar to the one in cDm264, the breakpoint occurs two base pairs upstream and thus creates a deletion of coding sequence. The first eight nucleotides in the insertion sequence right from the junction in this clone have diverged from those in cDm264 and cDm207 and the hexamer repeat at the junction is not present in this clone. Interestingly, the 23 nucleotides observed in cDm207 but not in cDm264 are present in this clone. The rest of the insertion sequence in these three clones is, however, identical. In a third clone characterised by Dawid and Rebbert (1981) four base pairs of the coding sequence are deleted at the junction. Thus it appears that similarly to the 5 kb Type I insertion, some of the Type II insertions are associated with deletions of gene sequences at the left-hand junction. This was observed on a larger scale in the clone cDm249, where this region contains a long deletion of the 28S gene sequences combined with a deletion of the Type II sequences.

In cDm207 the Type II insertion is followed by a Type I insertion. These insertion sequences are separated by 51 nucleotides of rDNA corresponding to the segments β' and β'' which occur between the sites of the two types of insertion (Figure 11). Since the insertions occur at different sites it is possible that cDm207 arose by recombination between rDNA units containing different types of insertion. Similar clones have been isolated by Long et al. (1981c). The fact that both the sequence and the restriction map of the cDm207 Type I insertion more resemble those found in the Type I elements located outside the rDNA raises the alternative possibility that recombination has occured between an rDNA unit containing a Type II insertion and a Type I elements elsewhere in the genome. Moreover, since the rDNA repeat in cDm207 is interrupted in the transcribed spacer by a repetitive sequence, this clone could be the result of a double crossing-over event. Alternatively, interruptions within the

transcribed spacer, which are also seen in the clone 14C9, could represent insertions of mobile elements into rDNA sequences.

IV.3. THE NUCLEOTIDE SEQUENCES FLANKING THE SHORT TYPE I INSERTIONS

Unlike the long 5 kb Type I insertions, the 1, 0.5 and 0.047 kb insertions are flanked by duplicated segments and do not result in deletions of the coding sequence. The 0.5 and 0.047 kb insertions are flanked by the same fourteen nucleotide segment that has been found at the ends of the 10 kb insertion in D. virilis (Rae et al., 1980). The 1 kb insertion is flanked by a duplication of the seven left-most nucleotides of these fourteen nucleotides, but if one mismatch is allowed, the duplication extends over eleven nucleotides (Figure 21). Similar structures in two clones containing 0.5 and 1 kb Type I insertions were characterised by Dawid and Rebbert (1981). It is possible that the 1 kb insertions were originally flanked by complete duplications of the fourteen nucleotide segment, but have subsequently lost three nucleotides by deletion. Similar events at the left-hand junction of the 5 kb insertion could account for the more extensive deletion of the whole duplicated segment and an additional nine nucleotides of the adjacent coding sequence. Interestingly, all the Type I insertions of varying lengths in D. melanogaster and the insertion in D. virilis (Rae et al., 1981) interrupt the 28S β2 gene at an identical site.

All the sites where left-hand junctions of the short insertions occur within the 5 kb insertion sequence have short segments of homology to the segment of rDNA coding sequence included in the duplication. These homologies involve nucleotides -1008 to -1004, -530 to -526 and -51 to -49 in the insertion sequence and nucleotides +71 to +75 in the coding region (Figure 20). Similar homologies are found in the clones characterised by Dawid and Rebbert (1981), even though the left-hand junctions of the two 0.5 kb insertions are not at an identical site, but

occur 29 nucleotides apart. The short insertions could have originated as a result of an unequal crossing-over between an rDNA repeat carrying a long insertion andan uninterrupted repeat. This model assumes that as few as three to five nucleotide homology is sufficient to allow recombination. Alternatively, the short insertions could have arisen by an intramolecular recombination between these homologous segments within an rDNA unit containing a 5 kb insertion. This would generate an extra-chromosomal 0.5 and 1 kb insertion element with a short segment of homology to rDNA. This element could be capable of integration at the site of the homology in another rDNA unit and could therefore generate a short flanking duplication upon reinsertion. Aberrant excision could result in a variety of sequence rearrangements around the site of excision and such events could explain the deletions observed at some of the junctions. The duplication of a long part of the 28S $\beta 2$ sequence in the clone RI9 could be explained by the reinsertion of such an aberrantly excised element into another rDNA unit containing an insertion. Alternatively, this clone could be the result of a large deletion extending from the left of the 747 nucleotide insertion into the adjacent rDNA unit. These events need not be frequent, since a single new rDNA unit generated by these mechanisms could spread throughout the nucleolus organiser by unequal crossing-over. The rDNA units containing short Type I insertions represent about 15% of the repeats in the nucleolus organiser of the X chromosome (Wellauer et al., 1978). In contrast to the expectations of these models, the reciprocal products of recombination, that is insertions containing sequences only from the left-hand part of the 5 kb insertion sequences, have not been observed and are either rare or absent. However, the alternative of short rDNA insertions giving rise to longer ones is unlikely as it requires a complicated scheme of sequental insertion of various elements at a single site in some of the rDNA repeats.

IV.4. rDNA INSERTIONS AS TRANSPOSABLE ELEMENTS

The target site duplications observed in the short Type I insertions in D. melanogaster and the insertions in D. virilis suggest that these structures arose by the insertion of a transposable sequence into rDNA. (Calos and Miller, 1980, Dunsmuir et al., 1980, Will et al., 1981, Farabaugh and Fink, 1980). If the Type I insertions are transposable sequences, it is striking that they differ substantially from some well characterised transposable elements. These elements in prokaryotes and eukaryotes frequently contain long terminal repeats (Calos and Miller, 1980, Cameron et al., 1979, Farabaugh and Fink, 1980, Levis et al., 1980, Potter et al., 1980). Such end repeats were not observed within the insertion in any of the short Type I insertions characterised in this work or by Dawid and Rebbert (1981) or within the sequenced parts of the 5 kb Type I insertion by Roiha et al. (1981) or by Rae (1981). Homologous sequences, which may be widely diverged, between the left- and right-hand ends of the different insertions were searched for using the computer program of Brutlag et al. (1982). Only short scattered areas of homology were found and none in the regions near the junctions in any of the insertions. In addition, searches were carried out for regions of symmetry or dyad symmetry within each insertion and the flanking gene sequences, but no structures of statistical significance were observed. Similar negative results were obtained by Dawid and Rebbert (1981). Hence, it can be concluded that end repeats have probably never been a feature of the Type I insertions. Dawid et al. (1981) have characterised a class of transposable elements which do not carry long terminal repeats, but which are flanked by short duplications of the site of insertion. It is possible that the Type I insertions belong to such a class. Although the 5 kb insertion is not flanked by a target site duplication, it nevertheless seems likely to be the basic unit of Type I elements, because of its abundance in rDNA and its presence in non-rDNA sequences in D. melano-

gaster and because it is also present in D. virilis. The sequences related to the Type I element of D. melanogaster are found in species which diverged over one hundred million years ago (Barnett and Rae, 1979). Although the interrupted rDNA units are inactive their coding and spacer sequences are well conserved, at least on the level of electron microscopic resolution (Wellauer and Dawid, 1977, Wellauer et al., 1978, Dawid et al., 1978). DNA sequencing of short segments of the coding sequences from all types of repeats has confirmed the sequence conservation down to a few scattered single nucleotide changes (Long et al., 1981a, Roiha et al., 1981, Dawid et al., 1981, Dawid and Rebbert, 1981, Rae, 1981). This further suggests that the Type I elements are capable of transposition, since such inactive genes should have undergone sequence changes in the same way as the non-transcribed spacer. The many sequence rearrangements around the target site of the Type I insertions suggest that there is a mechanism of site specific recombination at this site. This is, however, speculative and there is no direct evidence for continual transposition.

The non-rDNA Type I elements are flanked by short segments of the rDNA coding sequence, which suggests that the progenitor of these sequences was originally located within the nucleolus organiser (Roiha et al., 1981). After its excision by a mechanism possibly related to the normal transposition of these sequences, the progenitor element could have become tandemly amplified either by autonomous replication or by unequal crossing-over following its reinsertion at a new chromosomal site. This event may be relatively recent since the sequences of these non-rDNA Type I elements are still closely related to the rDNA insertion sequences. Similar non-rDNA insertion elements have been observed in D. virilis, which suggests that these events may be frequent and occur in several species (Barnett and Rae, 1979). The differences between the non-rDNA Type I elements and the rDNA insertion suggest that exchange does not occur between the two types of sequence (Kidd and Glover, 1980). It is

possible that only the rDNA Type I insertions are transposable in which case they would have a very specific integration process limited to a sequence in the 28S rRNA gene.

A computer analysis confirmed that the Type II insertions do not have repeated end sequences. The Type II insertions are, moreover, not associated with target site duplications and thus have little similarity with transposable elements. Another difference to the Type I insertions is the genomic distribution of these sequences, as the Type II insertions are found only among rDNA where they are evenly distributed between the X and Y nucleolus organisers (Wellauer et al., 1978). Since they have not so far been observed in the rDNA of other flies, these sequences may represent a relatively recent insertion typical of D. melanogaster and/ or its sibling species. This proposal is supported by the still significant clustering of these sequences among the rDNA units. It is possible that the end sequences of the Type II insertions have undergone deletions and become defective. Some mechanism of transposition or correction must, however, be operating among the rDNA units containing Type II sequences, since the sequences of such units are well conserved.

It is likely that the insertions of D. melanogaster rDNA arose as a result of two independent insertions of two extraneous transposable elements which have subsequently undergone varying structural modifications. The fact that rDNA repeats containing either type of insertion are not used to produce rRNAs and that both types of insertion are located very near each other within a well conserved region of rDNA suggests that their insertion rendered the rDNA unit incapable of transcription. There has therefore been no selection pressure against such units. The sequence conservation of the interrupted units can be explained by various correction mechanisms or by the continual transposition of the insertion sequences. A limited exchange of genetic material between the nucleolus organisers of the X and Y chromosomes would maintain the similarity of

the separate rDNA clusters. An additional mechanism must be postulated, which restricts the Type I sequences to the nucleolus organiser of the X chromosome. It can be suggested that the number of the uninterrupted units has stayed constant and the different amounts of rDNA units on the X and Y chromosomes reflect the independent spreading of the interrupted non-active rDNA repeats.

F I G U R E S

FIGURE 1.

The diagrammatic representation of the X and Y chromosomes of D. melanogaster.

The euchromatic region of the X chromosome is shown in white and the heterochromatic regions are shown in black. C is the centromere; NO is the nucleolus organiser; IH is the junction of the euchromatic and heterochromatic regions on the X chromosome; hA - hD are the main segments of the X chromosome divided by constrictions which are indicated by the arrows below the diagrams. The approximate positions of sc^4 and sc^8 inversions and the locations of cr and Xh^{abo} are also indicated on the X chromosome (redrawn from Cooper, 1959, Procunier and Tartof, 1978, Parry and Sandler, 1974).

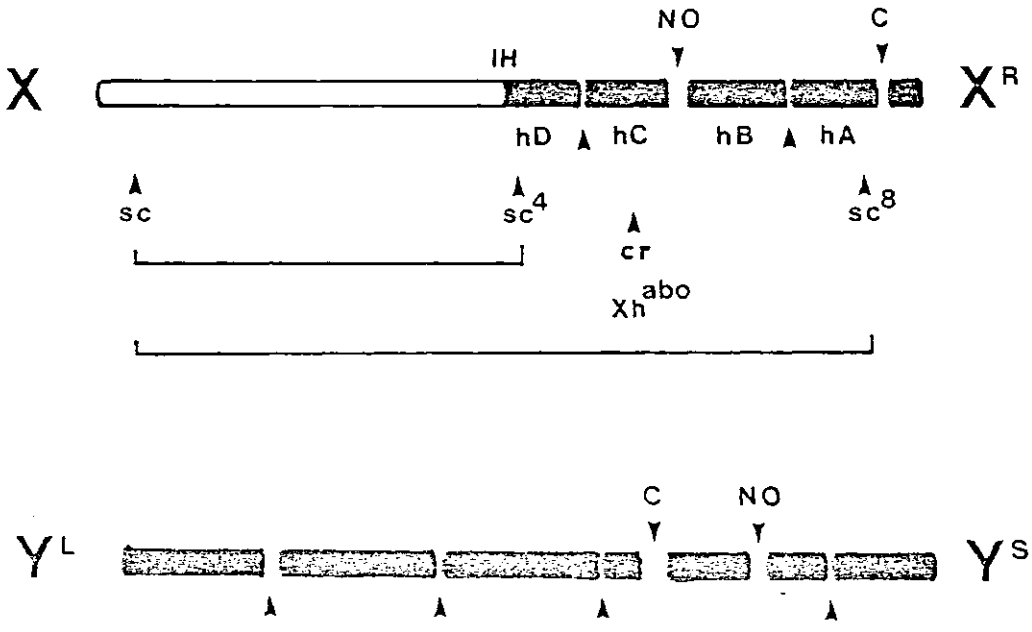
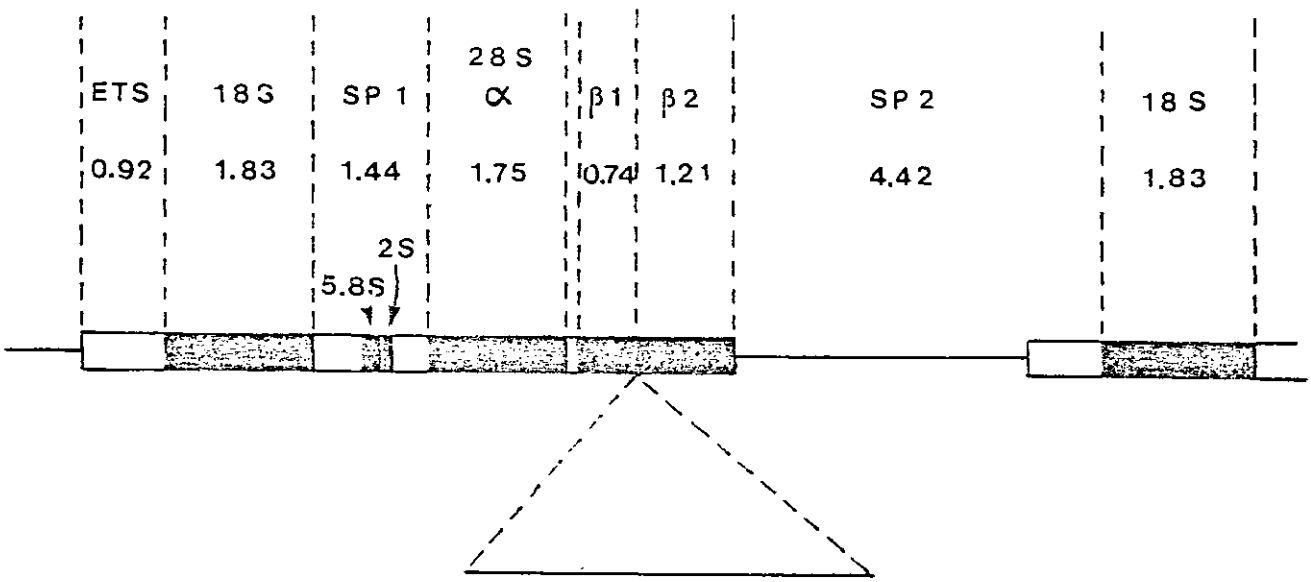

X
IH
NO
C
X^R
hD
hC
hB
hA
sc
sc^4
sc^8
cr
Xh^{abo}
Y^L
C
NO
Y^S

FIGURE 2.

The organisation of the major rDNA units of D. melanogaster.

The non-transcribed regions of the rDNA unit are indicated by the thin line. The open boxes represent transcribed sequences which are removed during processing. The position of the external transcribed spacer (ETS) is from Long and Dawid (1979a). The position of the mature rRNAs in the rDNA unit are indicated by the black boxes. The position of the 5 kb insertion found in the 17 kb rDNA unit is indicated within the 28S rRNA gene (redrawn from White and Hogness, 1977).



5.32
INSERTION

FIGURE 3.

Physical maps of the Type II rDNA clones and cDml03.

The vector is indicated by the stippled box and the D. melanogaster DNA by the solid line. The thickest line represents DNA complementary to the mature ribosomal RNAs. The medium thick line represents those rDNA sequences also contained in rDNA units that lack any type of insertion. The thin lines correspond to either insertions in the 28S gene or some other sequence that interrupts the contiguity of the rDNA. Genes for the mature rRNAs are given the nomenclature used by White and Hogness (1977). NTS refers to the non-transcribed spacer sequences together with the adjacent external transcribed spacer sequences. The internal transcribed spacer (TS) contains the genes for 5.8S and 2S rRNA (Jordan and Glover, 1977).

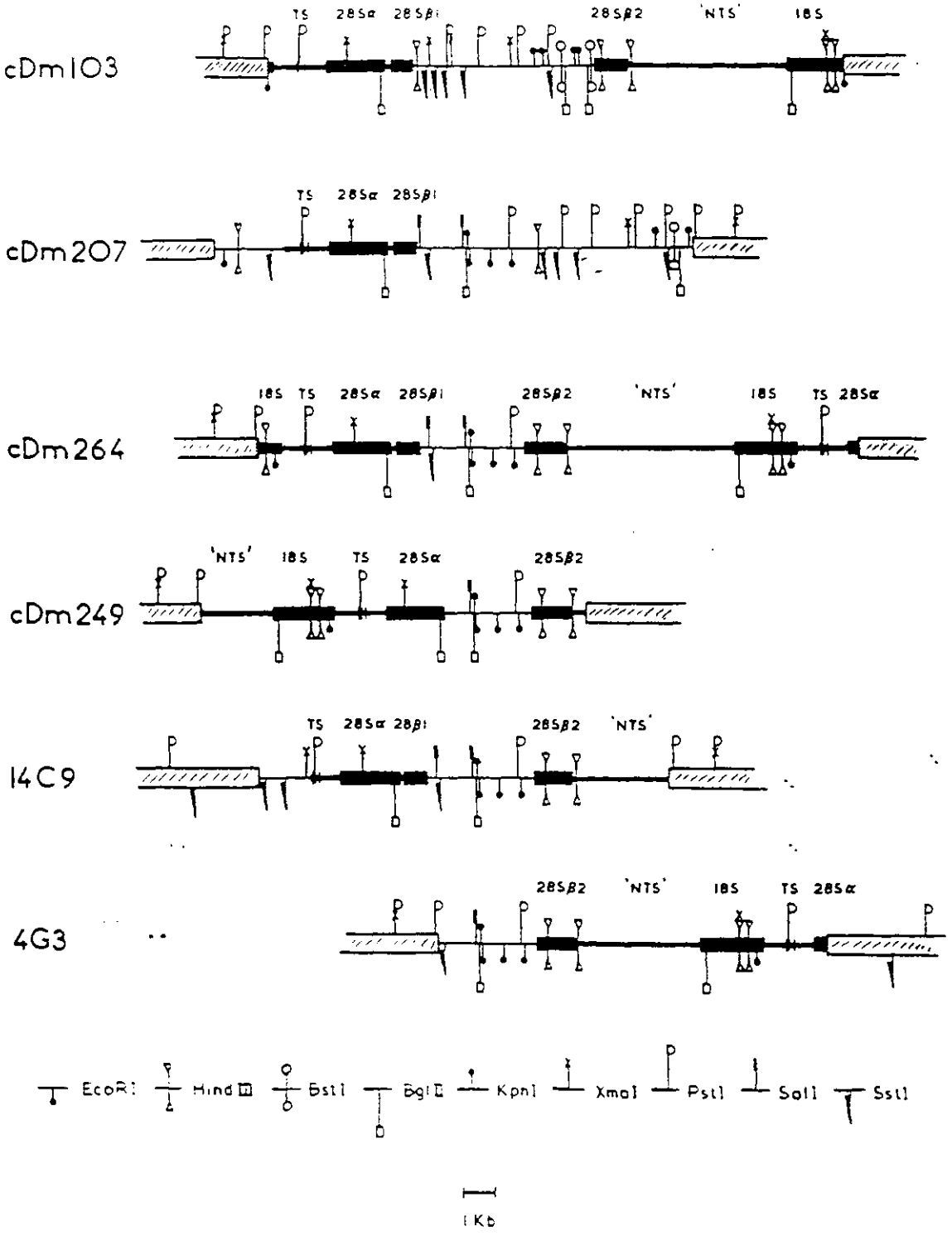

cDm103
TS
28Sα
28Sβ1
28Sβ2
'NTS'
18S
cDm207
TS
28Sα
28Sβ1
cDm264
18S
TS
28Sα
28Sβ1
28Sβ2
'NTS'
18S
TS
28Sα
cDm249
'NTS'
18S
TS
28Sα
28Sβ2
14C9
TS
28Sα
28β1
28Sβ2
'NTS'
4G3
28Sβ2
'NTS'
18S
TS
28Sα
EcoRI
Hind III
BstI
BglII
KpnI
XmaI
PstI
SalI
SstI
1 Kb

FIGURE 4.

Hybridisation of ^{32}P labeled 18S and 28S rRNAs to the clones containing Type II insertions.

The physical map represents the rDNA segment in the clone cDm264 with symbols for cleavage sites as in Figure 3. The fragments a - d marked above the map are characteristic of uninterrupted rDNA sequences and the fragments x, y and z below the map of sequences bordering the Type II insertion. The prime mark indicates a significant deviation in the length of the corresponding fragments in the other plasmids. In each panel the tracks contain cDm103 (1), cDm207 (2), cDm249 (3), cDm264 (4), 14C9 (5) and 4G3 (6). Panels A and D show ethidium bromide stained gels of EcoRI/BglII and EcoRI/HindIII digestions of these plasmids, respectively, and the panels B and C, E and F are autoradiograms of these gels following transfer to nitrocellulose and hybridisation with ^{32}P labeled rRNAs: EcoRI/BglII, 18S rRNA (B) and 28S rRNA (C), EcoRI/HindIII, 18S rRNA (E) and 28S rRNA (F).

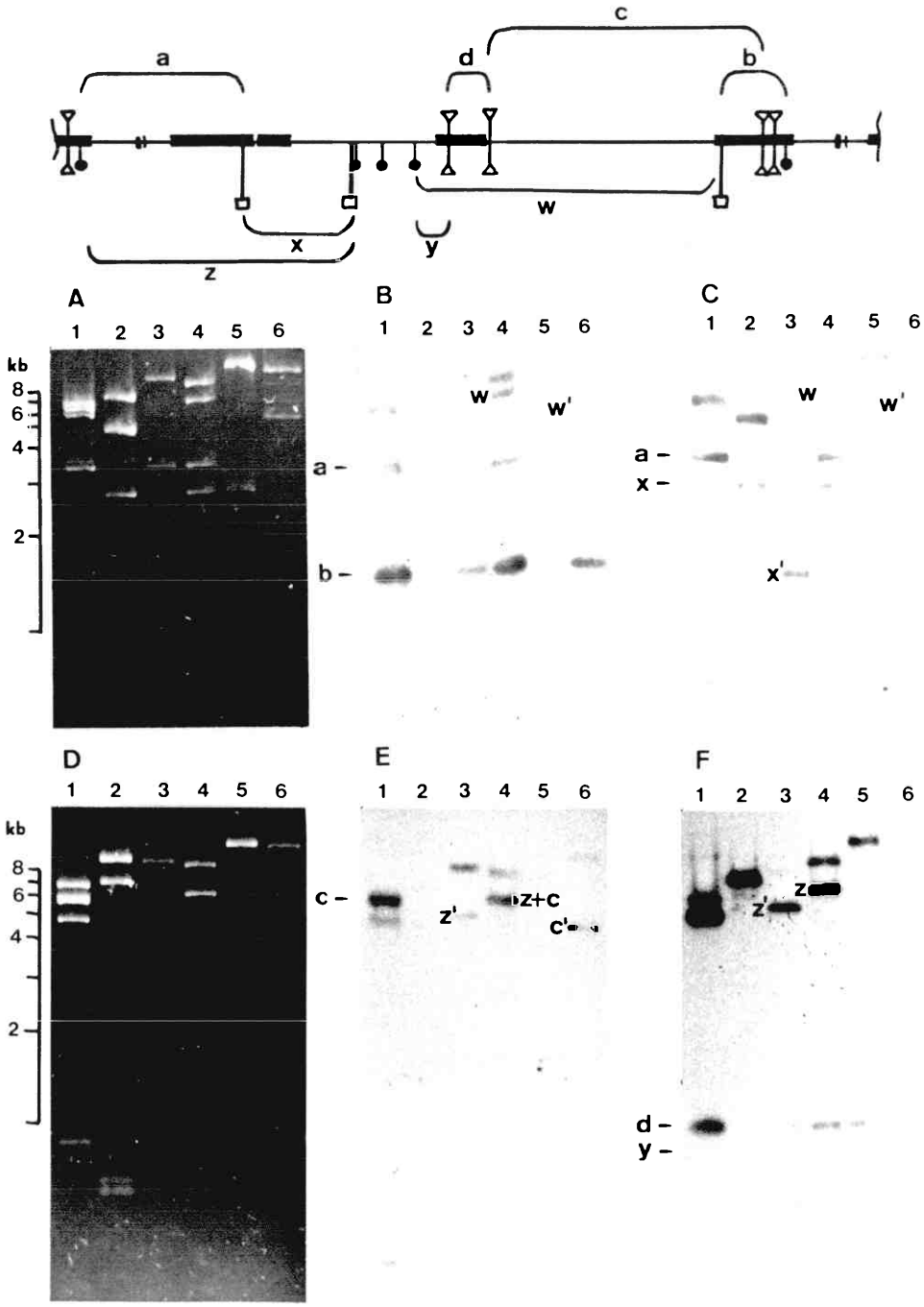

a
b
c
d
w
x
y
z
A
B
C
D
E
F
1 2 3 4 5 6
kb
8
6
4
2
a –
b –
w
w'
x'
c –
z'
z+c
c'
z
d –
y –

FIGURE 5.

Protection of plasmid DNA from S1 endonuclease by 18S and 28S rRNA.

Each plasmid DNA was cleaved with EcoRI (tracks 1 and 3) or BglII (tracks 2 and 4) and then annealed with rRNA as described in Chapter II. For each plasmid, tracks 1 and 2 show DNA protected by 18S rRNA and tracks 3 and 4 show DNA protected by 28S rRNA.

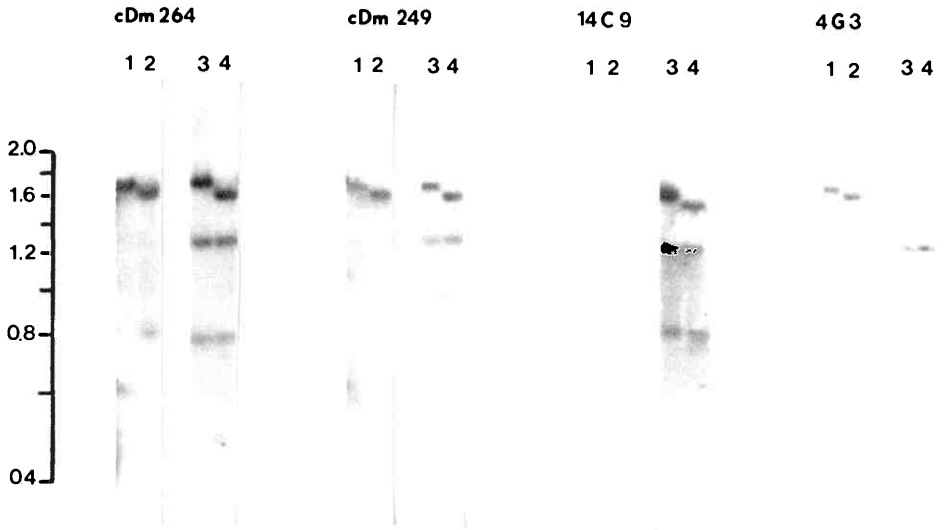

cDm 264
1 2 3 4
cDm 249
1 2 3 4
14C9
1 2 3 4
4G3
1 2 3 4
2.0
1.6
1.2
0.8
04

FIGURE 6.

Sequence homogeneity within the Type II insertions in the cloned segments of D. melanogaster rDNA.

Each plasmid DNA was digested with EcoRI and SalI (A) or EcoRI (B), fractionated by gel electrophoresis and transferred to nitrocellulose for hybridisation with the nick translated probe. The probe in panel A was the 1.3 kb SalI fragment from cDm851 (see Appendix II) and in panel B the two EcoRI fragments from the Type II insertion in cDm207. In each panel the tracks contain cDm207 (1), cDm249 (2), cDm264 (3), 14C9 (4) and 4G3 (5).

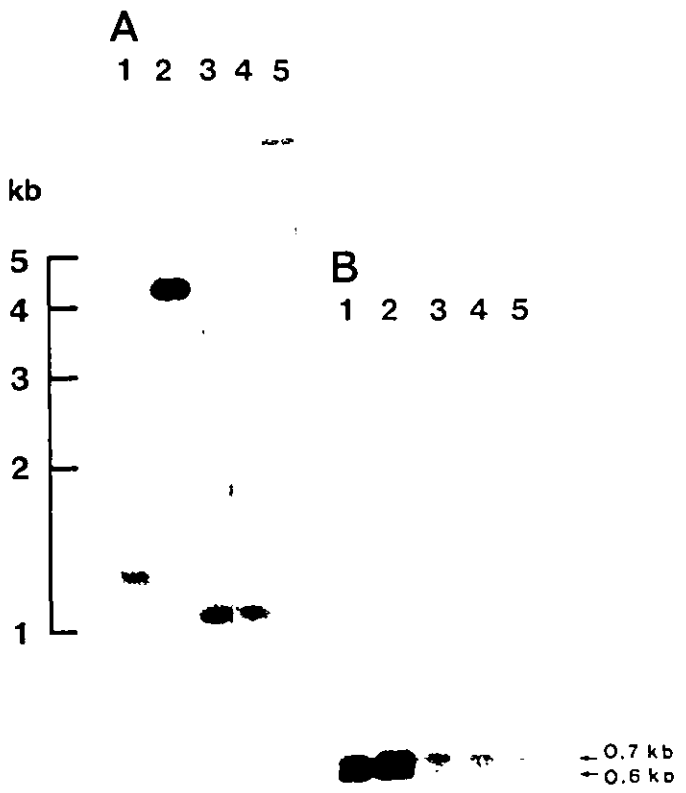
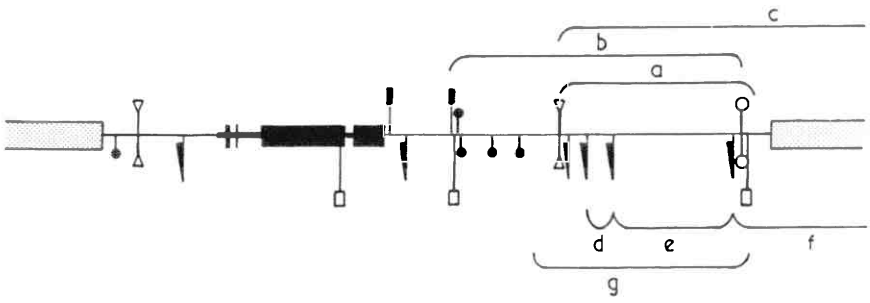

A
1 2 3 4 5
kb
5
4
3
2
1
B
1 2 3 4 5
0.7 kb
0.6 kb

FIGURE 7.

Hybridisation of Type I insertion sequences to cDm207.

The physical map shows cDm207 with the sites for the restriction endonucleases used in this experiment. The digested plasmid DNA was fractionated by gel electrophoresis, transferred to nitrocellulose and hybridised with ^{32}P labeled pc225 (see Appendix II). The letters on the map refer to the indicated bands on the autoradiogram. The cDm207 was digested with HindIII and BglII (1), HindIII (2), EcoRI and BglII (3), SstI (5) and SalI and BstI (6). Track 4 is a EcoRI digest of cDm103.


c
b
a
d
e
f
g

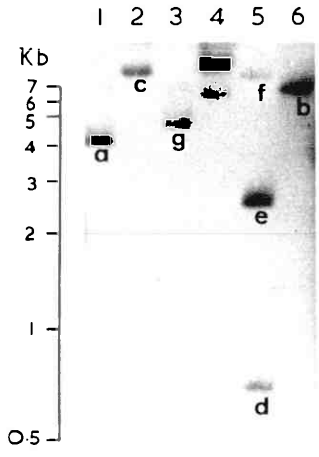

1 2 3 4 5 6
Kb
7
6
5
4
3
2
1
0.5
a
c
g
f
b
e
d

FIGURE 8.

Length heterogeneity of the Type II insertions.

cDm264 (A) and D. melanogaster DNA (B) were digested with HindIII, fractionated by gel electrophoresis, transferred to nitrocellulose and hybridised with ^{32}P labeled pcDm207cd (see Appendix II). Sizes in kb are indicated to the right of the tracks.

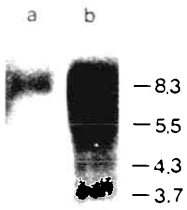

a
b
8.3
5.5
4.3
3.7

FIGURE 9.

Length heterogeneity within the Type II insertions.

cDm264 (1) and D. melanogaster DNA (2) were digested with BglII (A), SalI and EcoRI (B), EcoRI (C) and EcoRI and HindIII (D), fractionated by gel electrophoresis, transferred to nitrocellulose and hybridised with ^{32}P labeled probe. The probes were the 1.2 kb SstI/BglII fragment from p264 SR12 (A), the 1.3 kb SalI fragment from cDm851 (B), the 0.7 kb and 0.6 kb EcoRI fragments from pcDm207cd (C) and the 0.32 kb EcoRI/HinfI fragment from p264 RH23 (D). For the maps of these subclones, see Appendix II.

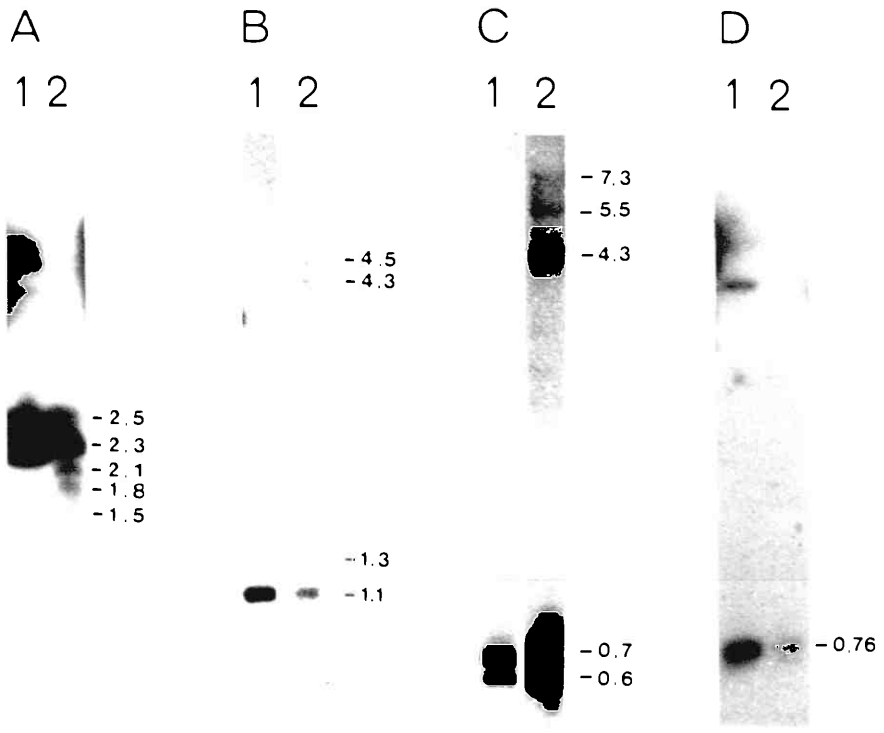

A
1 2
- 2.5
- 2.3
- 2.1
- 1.8
- 1.5
B
1 2
- 4.5
- 4.3
- 1.3
- 1.1
C
1 2
- 7.3
- 5.5
- 4.3
- 0.7
- 0.6
D
1 2
- 0.76

FIGURE 10.

Sequencing strategies.

The schematical physical maps present the clones cDm264 and cDm207 with the subclones containing the junctions between the insertion sequences and the gene beneath each map. Only those restriction endonuclease cleavage sites used for either subcloning or sequencing are shown. The arrows below the maps of the subclones indicate the extent and direction of the sequences read from various restriction sites. In each of the maps of the original clones, the ribosomal genes are indicated by filled blocks, the Type II insertions by dotted shading and the Type I insertions by horizontal hatching. These regions are not indicated in the subclones. The plasmid vectors are indicated by cross-hatching.

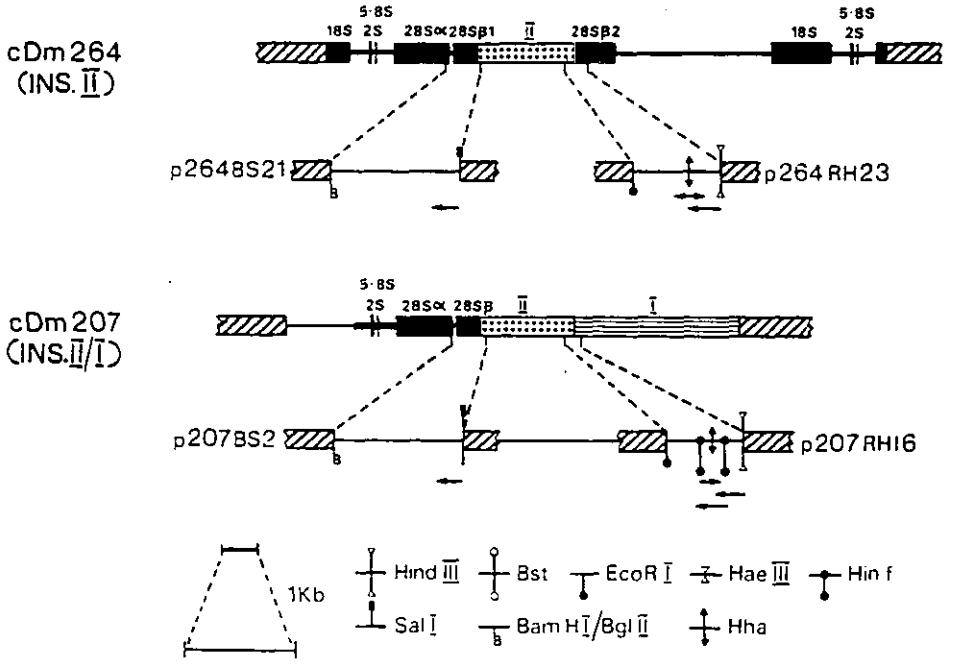

cDm 264
(INS. II)
5·8S
18S
2S
28Sα
28Sβ1
II
28Sβ2
18S
5·8S
2S
p2648S21
p264 RH23
cDm 207
(INS. II/I)
5·8S
2S
28Sα
28Sβ
II
I
p207BS2
p207RHI6
1Kb
Hind III
Bst
EcoR I
Hae III
Hin f
Sal I
Bam HI/Bgl II
Hha

FIGURE 11.

The sequences of the junctions of the Type II insertions within rDNA.

a. The sequence of the left-hand junction of the Type II sequences in cDm264 and cDm207 comes from the subclones p264 BS21 and p207 BS2, respectively. Only the regions in which the cDm207 sequence differs from the cDm264 are shown.

b. The sequence of the right-hand junction of the Type II sequences in cDm264 and cDm207 comes from the subclones p264 RH23 and p207 RH16, respectively. The single base of the cDm207 sequence which differs from cDm264 is indicated. Δ signifies a deletion.

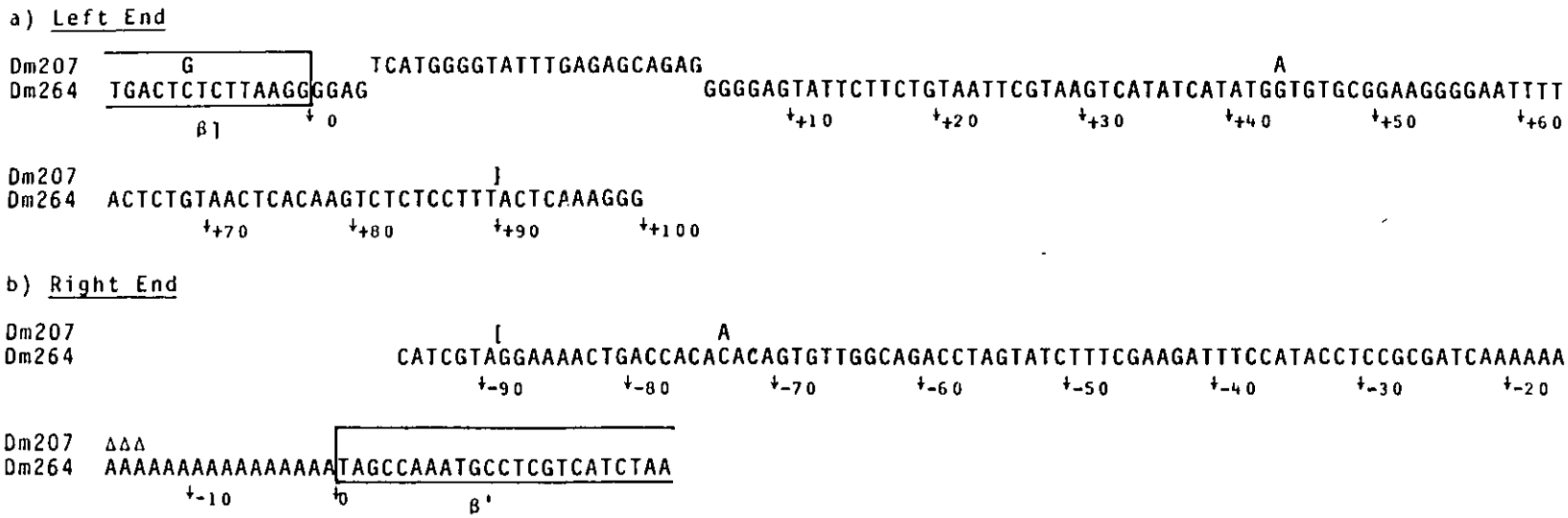

a) Left End
Dm207 G TCATGGGGTATTTGAGAGCAGAG A
Dm264 TGACTCTCTTAAGGGGAG GGGGAGTATTCTTCTGTAATTCGTAAGTCATATCATATGGTGTGCGGAAGGGGAATTTT
β] 0 +10 +20 +30 +40 +50 +60
Dm207]
Dm264 ACTCTGTAACTCACAAGTCTCTCCTTTACTCAAAGGG
+70 +80 +90 +100
b) Right End
Dm207 [A
Dm264 CATCGTAGGAAAACTGACCACACACAGTGTTGGCAGACCTAGTATCTTTCGAAGATTTCCATACCTCCGCGATCAAAAAA
-90 -80 -70 -60 -50 -40 -30 -20
Dm207 ΔΔΔ
Dm264 AAAAAAAAAAAAAAAATAGCCAAATGCCTCGTCATCTAA
-10 0 β'

FIGURE 12.

The sequences of a 497 nucleotide segment of the uninterrupted 28S gene. (Roiha et al., 1981).

The two "halves" of the 28S β gene on either side of the insertion are termed $\beta 1$ and $\beta 2$ (White and Hogness, 1977). The numbering of the sequence has been arbitrarily begun from the position of the Type II insertion which is marked with arrow heads. In the 28S β genes containing the Type I insertion, similarly marked with arrow heads, the β''' segment is substituted by the insertion. The arrow heads marked Tp indicate the position of the intervening sequence in a homologous region of the 28S rRNA gene of Tetrahymena pigmentosa (Wild and Sommer, 1980).

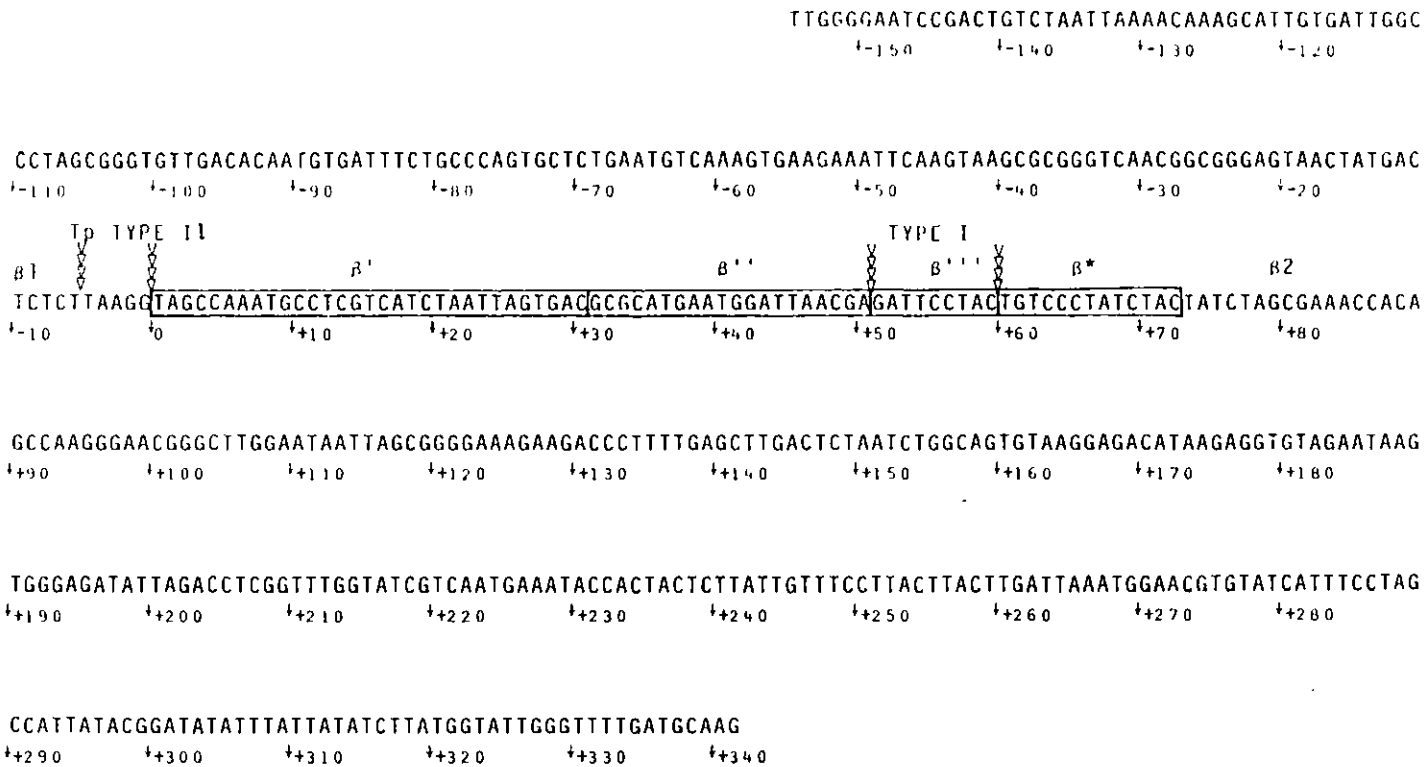

TTGGGGAATCCGACTGTCTAATTAAAACAAAGCATTGTGATTGGC
-150 -140 -130 -120
CCTAGCGGGTGTTGACACAATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAAGTAAGCGCGGGTCAACGGCGGGAGTAACTATGAC
-110 -100 -90 -80 -70 -60 -50 -40 -30 -20
Tp TYPE II
TYPE I
β1 β' β'' β''' β* β2
TCTCTTAAGGTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATTAACGAGATTCCTACTGTCCCTATCTACTATCTAGCGAAACCACA
-10 0 +10 +20 +30 +40 +50 +60 +70 +80
GCCAAGGGAACGGGCTTGGAATAATTAGCGGGGAAAGAAGACCCTTTTGAGCTTGACTCTAATCTGGCAGTGTAAGGAGACATAAGAGGTGTAGAATAAG
+90 +100 +110 +120 +130 +140 +150 +160 +170 +180
TGGGAGATATTAGACCTCGGTTTGGTATCGTCAATGAAATACCACTACTCTTATTGTTTCCTTACTTACTTGATTAAATGGAACGTGTATCATTTCCTAG
+190 +200 +210 +220 +230 +240 +250 +260 +270 +280
CCATTATACGGATATATTTATTATATCTTATGGTATTGGGTTTTGATGCAAG
+290 +300 +310 +320 +330 +340

FIGURE 13.

Sequencing gels showing the left-hand junction of the Type II insertion. The sequences in cDm207 (A) and cDm264 (B) were obtained from the subclones p207 BS2 and p264 BS21, respectively. In both cases the plasmid DNA was cleaved with SalI and EcoRI and 3' labeled at the SalI site. The sequences shown are from the non-coding strand. The rDNA and insertion sequences are indicated by lines placed against the sequence. The 23 nucleotides which occur between the two hexameric sequences in cDm207 are indicated separately.

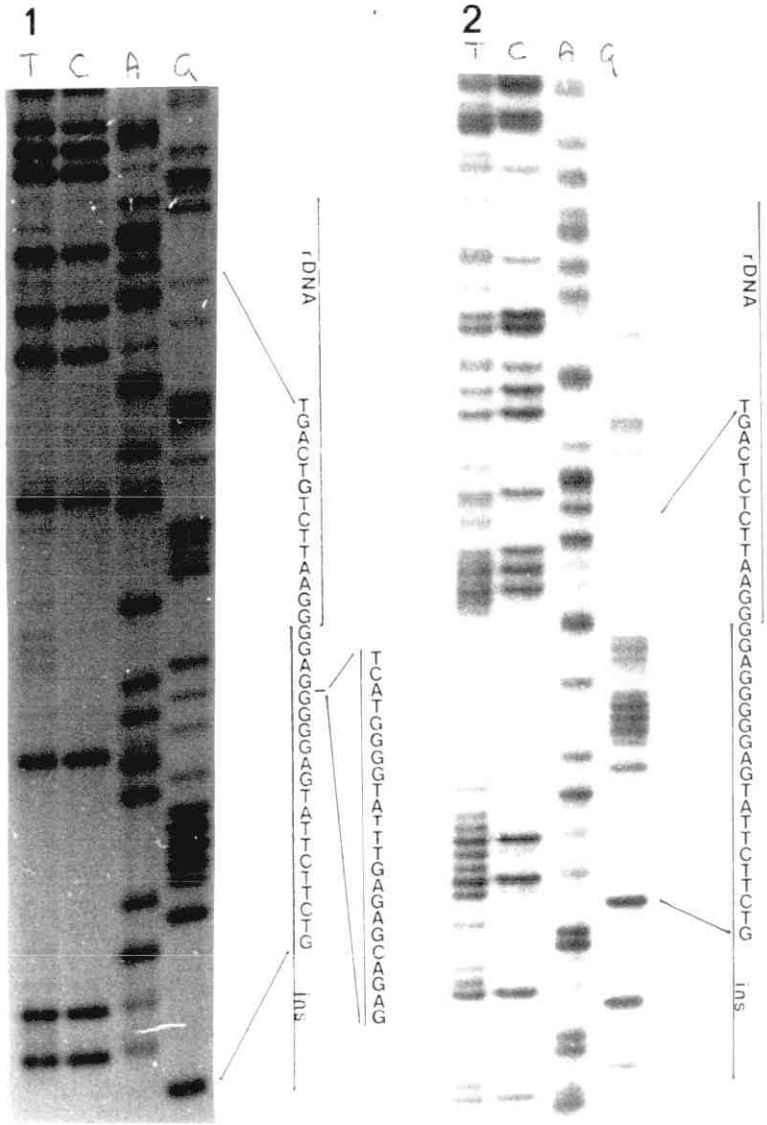

1
T C A G
rDNA
TGACTGTCTTAAGGGGAGGGGGAGTATTCTTCTG
ins
TCATGGGGTATTTGAGAGCAGAG
2
T C A G
rDNA
TGACTCTCTTAAGGGGAGGGGGAGTATTCTTCTG
ins

FIGURE 14.

Sequencing gel showing the right-hand junction of the Type II insertion in cDm264.

The subclone p264 RH23 of cDm264 was cleaved with HhaI and EcoRI and 3' labeled at the HhaI site. The sequence shown is from the non-coding strand. The rDNA and insertion sequences are indicated by lines placed against the sequence.

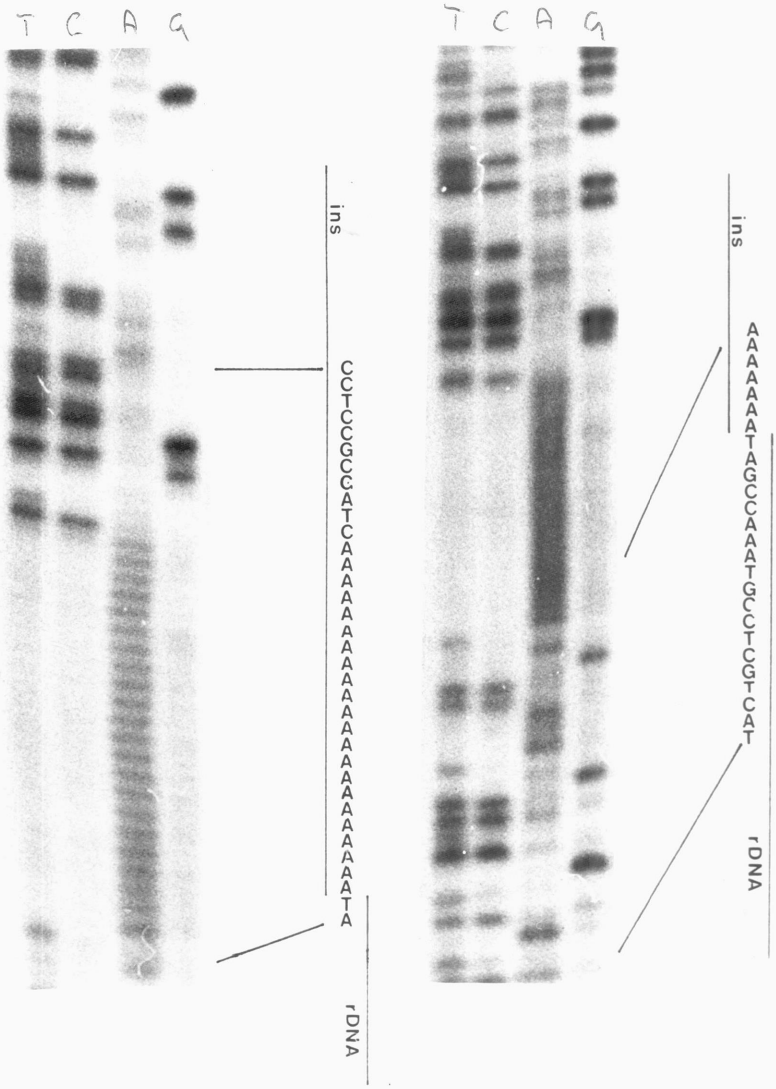
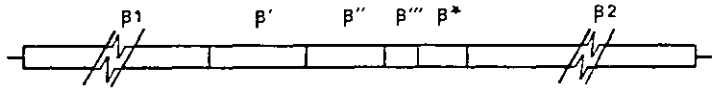

T C A G
ins
CCTCCGCCATCAAAAAAAAAAAAAAAAAAAAAAAAAAAATA
rDNA
T C A G
ins
AAAAAAAATAGCCAAATGCCTCGTCAT
rDNA

FIGURE 15.

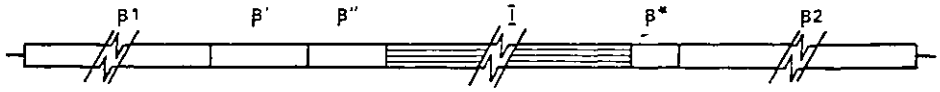
A schematic comparison of the arrangements of segments of the 28S rRNA gene which flank the insertions in D. melanogaster and D. virilis. The genes are indicated by boxes and the insertion sequences as in the legend to Figure 10.

D. melanogaster – INS I

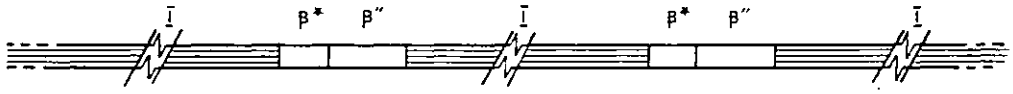
pDm 238



cDm103



cDm219



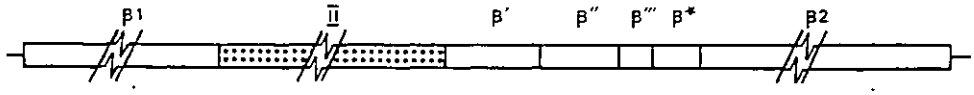
D. virilis

pDvL3-8



D. melanogaster – INS II

cDm264



cDm207

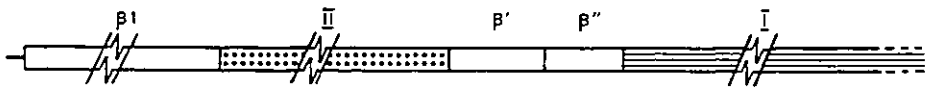


FIGURE 16.

The sequences of the Type I insertions at the junctions with 28S β gene.

The continuous sequence from the junction of the units 3 and 4 from cDm219 and the sequences of the left- and right-hand junctions of cDm103 are taken from Roiha et al. (1981). Only the differences between this sequence and the corresponding junction sequences in cDm103 and the left-hand junction sequence in cDm207 are indicated below. Δ indicates a deleted nucleotide. The extent of the sequence determined for these clones is indicated by brackets.

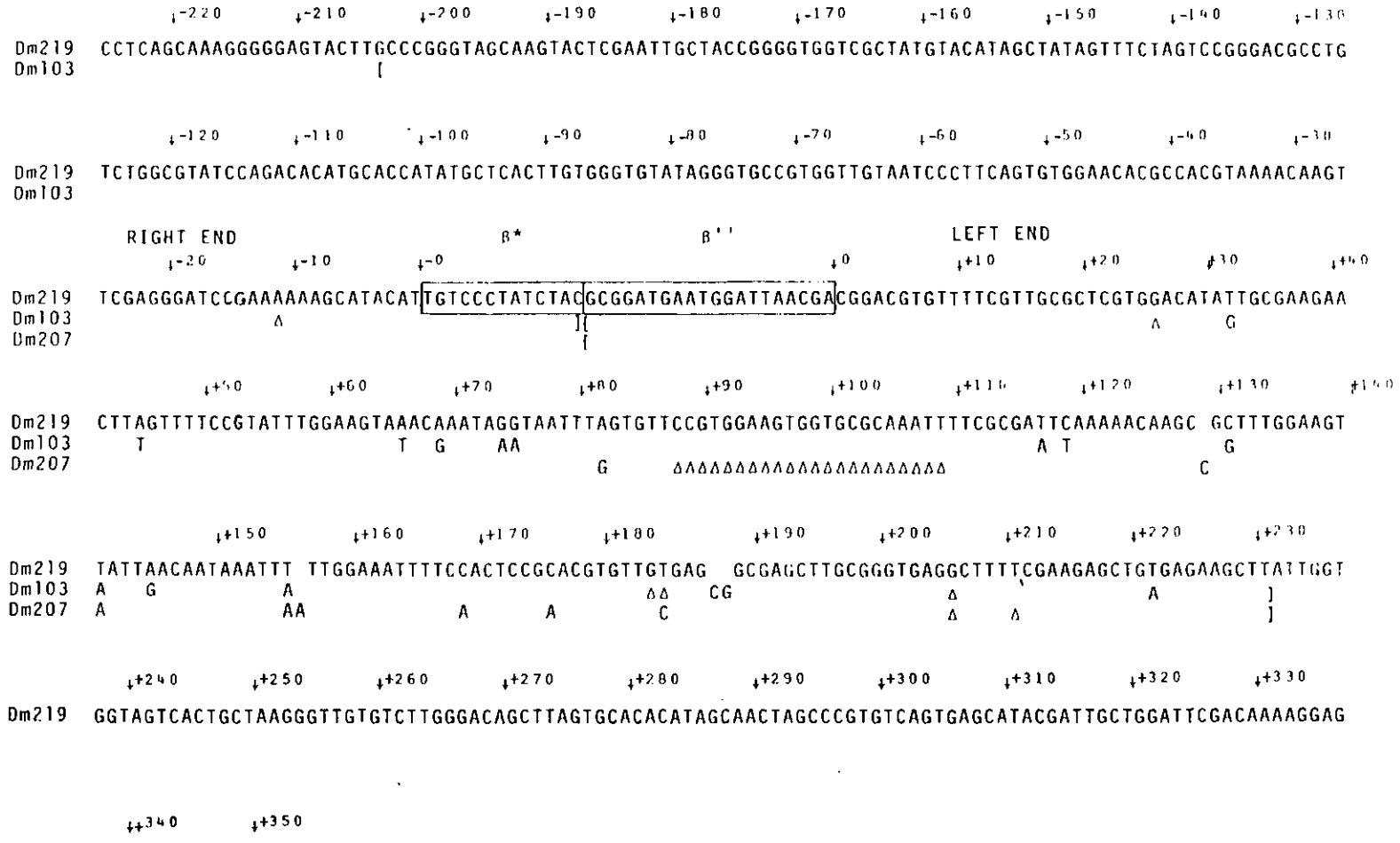

-220 -210 -200 -190 -180 -170 -160 -150 -140 -130
Dm219 CCTCAGCAAAGGGGGAGTACTTGCCCGGGTAGCAAGTACTCGAATTGCTACCGGGGTGGTCGCTATGTACATAGCTATAGTTTCTAGTCCGGGACGCCTG
Dm103
-120 -110 -100 -90 -80 -70 -60 -50 -40 -30
Dm219 TCTGGCGTATCCAGACACATGCACCATATGCTCACTTGTGGGTGTATAGGGTGCCGTGGTTGTAATCCCTTCAGTGTGGAACACGCCACGTAAAACAAGT
Dm103
RIGHT END
β*
β''
LEFT END
-20 -10 -0 0 +10 +20 +30 +40
Dm219 TCGAGGGATCCGAAAAAAGCATACATTGTCCCTATCTACGCGGATGAATGGATTAACGACGGACGTGTTTTCGTTGCGCTCGTGGACATATTGCGAAGAA
Dm103 A A G
Dm207
+50 +60 +70 +80 +90 +100 +110 +120 +130 +140
Dm219 CTTAGTTTTCCGTATTTGGAAGTAAACAAATAGGTAATTTAGTGTTCCGTGGAAGTGGTGCGCAAATTTTCGCGATTCAAAAACAAGC GCTTTGGAAGT
Dm103 T T G AA A T G
Dm207 G ΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔ C
+150 +160 +170 +180 +190 +200 +210 +220 +230
Dm219 TATTAACAATAAATTT TTGGAAATTTTCCACTCCGCACGTGTTGTGAG GCGAGCTTGCGGGTGAGGCTTTTCGAAGAGCTGTGAGAAGCTTATTGGT
Dm103 A G A ΔΔ CG Δ A]
Dm207 A AA A A C Δ Δ]
+240 +250 +260 +270 +280 +290 +300 +310 +320 +330
Dm219 GGTAGTCACTGCTAAGGGTTGTGTCTTGGGACAGCTTAGTGCACACATAGCAACTAGCCCGTGTCAGTGAGCATACGATTGCTGGATTCGACAAAAGGAG
+340 +350
Dm219 GACCACCATTGGACAA

FIGURE 17.

Sequencing gel showing the right-hand junction of the Type II insertion and the left-hand junction of the Type I insertion in cDm207.

The sequence was obtained from the subclone p207 RH16 which was cleaved with HinfI and HhaI and 3' labeled at the HinfI site. The sequence shown is from the coding strand. The Type I, Type II and rDNA sequences are indicated by lines placed against the sequence.

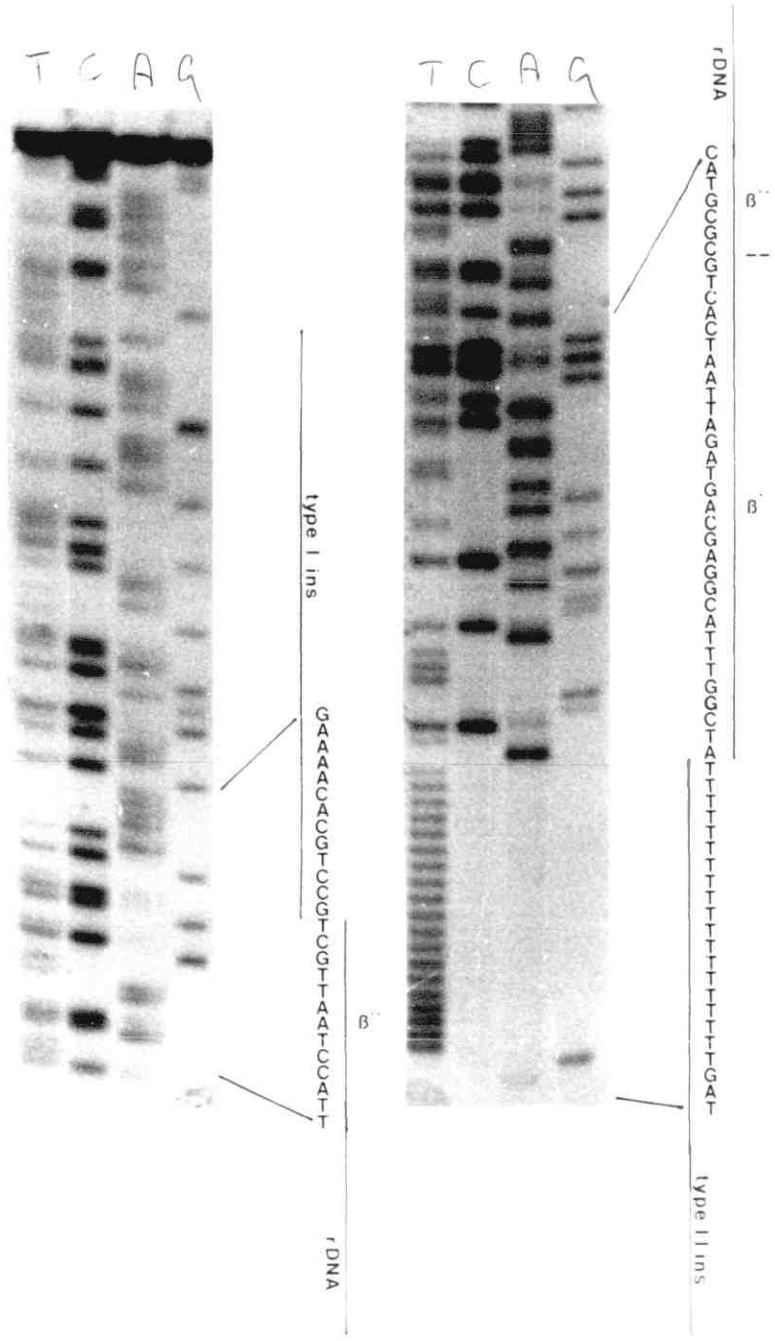

T C A G
rDNA
CATGCGCGTCACTAATTAGATGACGAGGCATTTGGCTATTTTTTTTTTTTTTTTTTTTTTGAT
type I ins
type I ins
GAAAACACGTCCGTCGTTAATCCATT
rDNA

FIGURE 18.

The comparison of the junctions of the Type I insertion in D. melanogaster and the insertion in D. virilis rDNA.

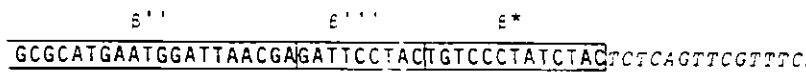
The D. melanogaster sequence is taken from Roiha et al. (1981) and the D. virilis sequence from Rae et al. (1980). Segments of the 28S β gene are indicated above the boxed-in sequence and the insertion sequence is italicized.

A) Left End

D. melanogaster (Dm103)

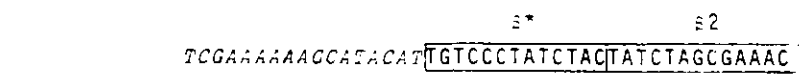


D. virilis (DvL3-8)



B) Right End

D. melanogaster (Dm103)



D. virilis (DvL3-8)

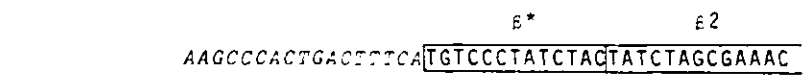


FIGURE 19.

Sequencing strategies.

The physical maps show the regions of the rDNA segments which were sequenced. The arrows show the directions and extent to which sequences were read from the terminally labeled nucleotide. The shaded blocks represent rDNA and the solid lines the insertion sequences. The dotted lines indicate the displacement of the left-hand insertion in the clone RI9 to bring its right-hand junction into alignment with the other clones. Bgl = BglII, Bst = BstI, H3 = HaeIII, HF = HinfI, Kpn = KpnI, Sst = SstI.

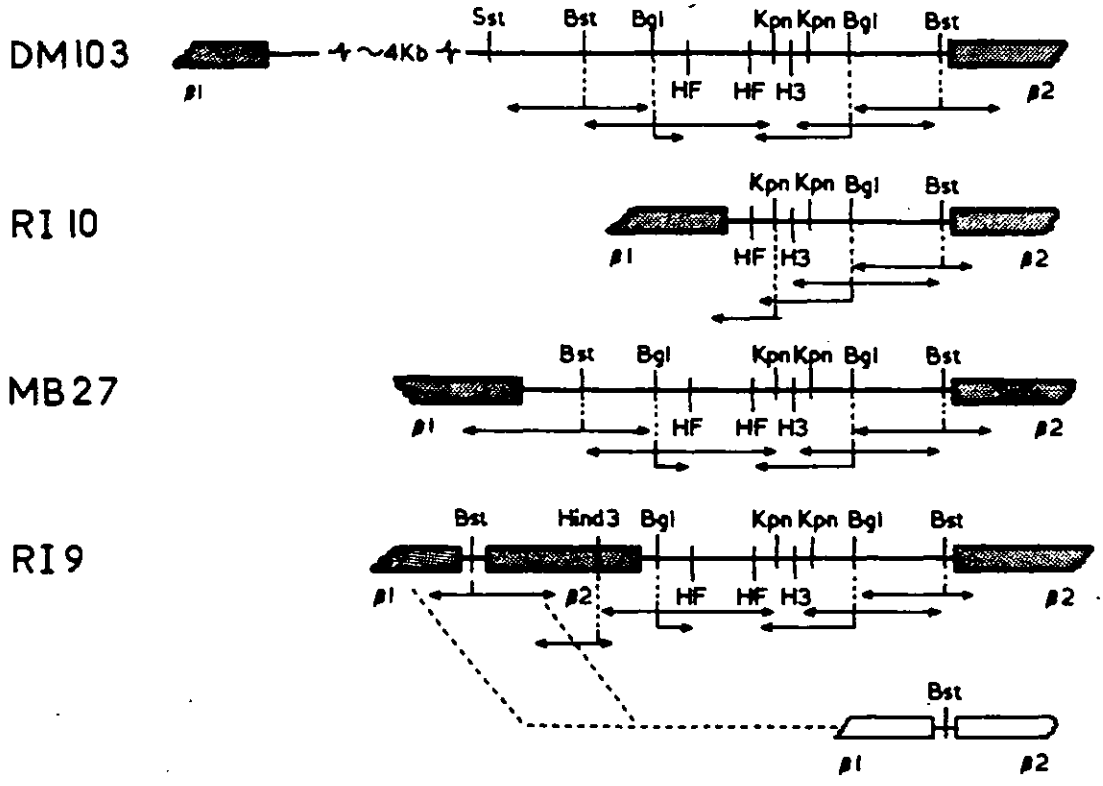

DM103
~4Kb
Sst
Bst
Bgl
Kpn Kpn
Bgl
Bst
HF
HF H3
β1
β2
RI 10
MB27
RI 9
Hind3
Bst

FIGURE 20.

A contiguous sequence of the 1122 right-hand nucleotides of the Type I insertion and its junction with rDNA in cDm103.

The right-hand side of the insertion is given negative numbers from the point at which it meets rDNA and the rDNA is numbered with respect to the position of the Type II insertion (Figure 11). The rDNA sequence is underlined (nucleotides +61 - +537). The solid arrows indicate the left-hand end points of the 1 kb insertion (-1008), 0.5 kb insertion (-528) and the 0.047 kb insertion (-50). The open arrows indicate the points at which the left-hand of the 747 nucleotide insertion in RI9 joins with the 28S β2 gene (-748 to +441). The sequences of the shorter insertions are identical with that of the cDm103 except at the following positions: clone RI10: -604 C, -123 G; clone MB27: -933 C, -888 C, -82 C; clone RI9: (747 nucleotide insertion) -603 C, -286 G, -264 A.

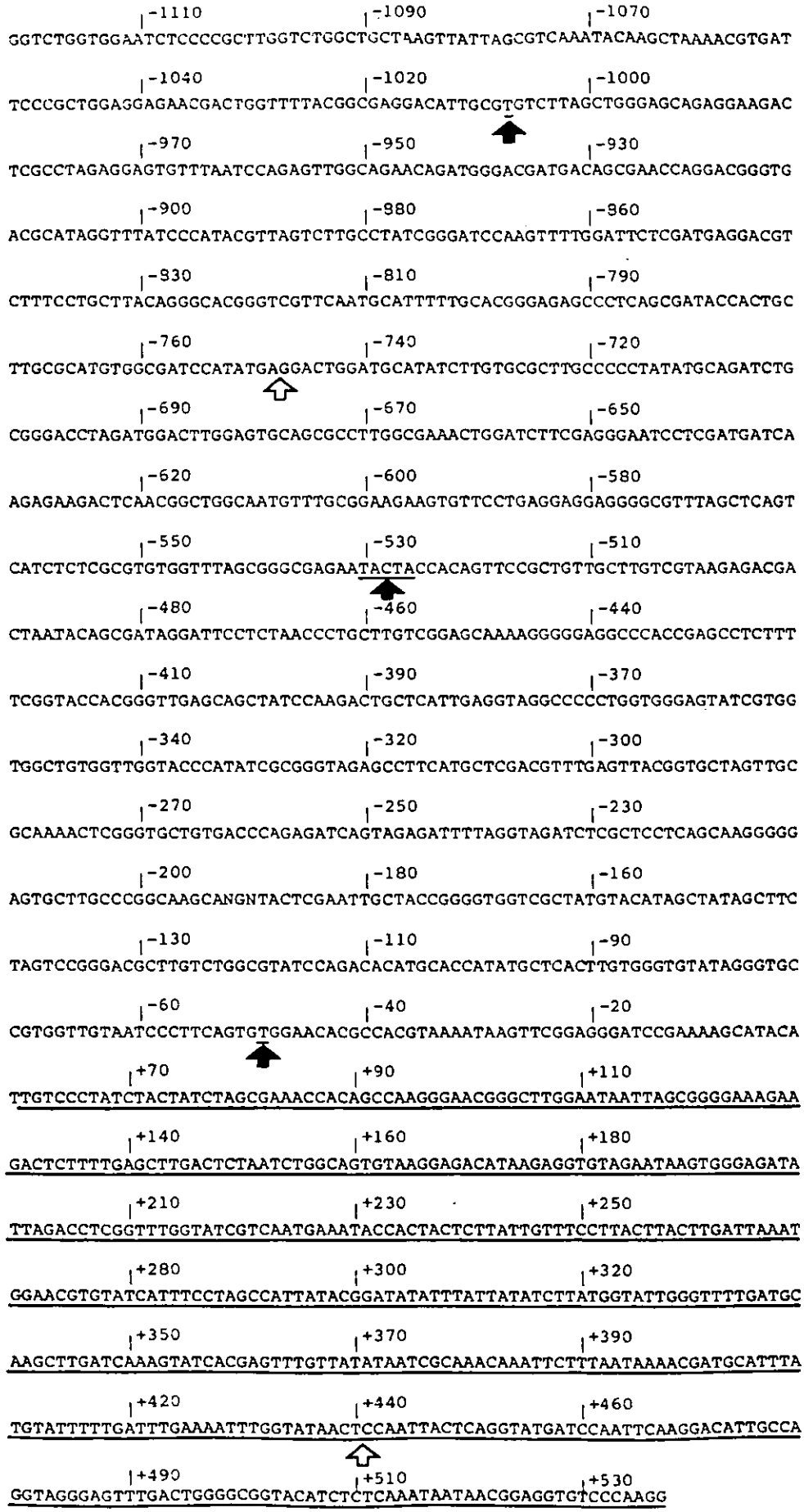

-1110
-1090
-1070
GGTCTGGTGGAATCTCCCCGCTTGGTCTGGCTGCTAAGTTATTAGCGTCAAATACAAGCTAAAACGTGAT
-1040
-1020
-1000
TCCCGCTGGAGGAGAACGACTGGTTTTACGGCGAGGACATTGCGTGTCTTAGCTGGGAGCAGAGGAAGAC
-970
-950
-930
TCGCCTAGAGGAGTGTTTAATCCAGAGTTGGCAGAACAGATGGGACGATGACAGCGAACCAGGACGGGTG
-900
-880
-860
ACGCATAGGTTTATCCCATACGTTAGTCTTGCCTATCGGGATCCAAGTTTTGGATTCTCGATGAGGACGT
-830
-810
-790
CTTTCCTGCTTACAGGGCACGGGTCGTTCAATGCATTTTTGCACGGGAGAGCCCTCAGCGATACCACTGC
-760
-740
-720
TTGCGCATGTGGCGATCCATATGAGGACTGGATGCATATCTTGTGCGCTTGCCCCCTATATGCAGATCTG
-690
-670
-650
CGGGACCTAGATGGACTTGGAGTGCAGCGCCTTGGCGAAACTGGATCTTCGAGGGAATCCTCGATGATCA
-620
-600
-580
AGAGAAGACTCAACGGCTGGCAATGTTTGCGGAAGAAGTGTTCCTGAGGAGGAGGGGCGTTTAGCTCAGT
-550
-530
-510
CATCTCTCGCGTGTGGTTTAGCGGGCGAGAATACTACCACAGTTCCGCTGTTGCTTGTCGTAAGAGACGA
-480
-460
-440
CTAATACAGCGATAGGATTCCTCTAACCCTGCTTGTCGGAGCAAAAGGGGGAGGCCCACCGAGCCTCTTT
-410
-390
-370
TCGGTACCACGGGTTGAGCAGCTATCCAAGACTGCTCATTGAGGTAGGCCCCCTGGTGGGAGTATCGTGG
-340
-320
-300
TGGCTGTGGTTGGTACCCATATCGCGGGTAGAGCCTTCATGCTCGACGTTTGAGTTACGGTGCTAGTTGC
-270
-250
-230
GCAAAACTCGGGTGCTGTGACCCAGAGATCAGTAGAGATTTTAGGTAGATCTCGCTCCTCAGCAAGGGGG
-200
-180
-160
AGTGCTTGCCCGGCAAGCANGNTACTCGAATTGCTACCGGGGTGGTCGCTATGTACATAGCTATAGCTTC
-130
-110
-90
TAGTCCGGGACGCTTGTCTGGCGTATCCAGACACATGCACCATATGCTCACTTGTGGGTGTATAGGGTGC
-60
-40
-20
CGTGGTTGTAATCCCTTCAGTGTGGAACACGCCACGTAAAATAAGTTCGGAGGGATCCGAAAAGCATACA
+70
+90
+110
TTGTCCCTATCTACTATCTAGCGAAACCACAGCCAAGGGAACGGGCTTGGAATAATTAGCGGGGAAAGAA
+140
+160
+180
GACTCTTTTGAGCTTGACTCTAATCTGGCAGTGTAAGGAGACATAAGAGGTGTAGAATAAGTGGGAGATA
+210
+230
+250
TTAGACCTCGGTTTGGTATCGTCAATGAAATACCACTACTCTTATTGTTTCCTTACTTACTTGATTAAAT
+280
+300
+320
GGAACGTGTATCATTTCCTAGCCATTATACGGATATATTTATTATATCTTATGGTATTGGGTTTTGATGC
+350
+370
+390
AAGCTTGATCAAAGTATCACGAGTTTGTTATATAATCGCAAACAAATTCTTTAATAAAACGATGCATTTA
+420
+440
+460
TGTATTTTTGATTTGAAAATTTGGTATAACTCCAATTACTCAGGTATGATCCAATTCAAGGACATTGCCA
+490
+510
+530
GGTAGGGAGTTTGACTGGGGCGGTACATCTCTCAAATAATAACGGAGGTGTCCCAAGG

FIGURE 21.

A comparison of insertion junctions.

The rDNA sequences to the left and right of insertion are shown in alignment to facilitate comparison. The sequence at the left-hand junction in cDm103 is from Roiha et al. (1981). The rDNA sequences are underlined and the insertion sequences overlined. The duplicated sequence is indicated by the asterisks. The sequence of uninterrupted rDNA is shown for comparison (Roiha et al., 1981).

a) Insertions not flanked by duplications:

LEFT JUNCTION RIGHT JUNCTION

RI9 insertion 0.747kb

GAAAATTTGGTATAACTGGACTGGATGCATATCTT----------ATCCGAAAAGCATACATTGTCCCTATCTACTATCTAGCAA

Dml03 insertion approximately 5kb

GAATGGATTAACGACGGACGTGTTTTCGTTGCGCT----------ATCCGAAAAGCATACATTGTCCCTATCTACTATCTAGCAA

b) Insertions flanked by duplications:

LEFT JUNCTION RIGHT JUNCTION

RI10 insertion 0.525(0.530)kb

GAATGGATTAACGAGATTCCTACTGTCCCTATCTACTACCACAGTTCCGCTG-------ATCCGAAAAGCATACATTGTCCCTATCTACTATCTAGCAA

********** **********

MB27 insertion 1.006(1.007)kb

GAATGGATTAACGAGATTCCTACTGTCCCTGTCTTAGCTGGGAGCAG----------ATCCGAAAAGCATACATTGTCCCTATCTACTATCTAGCAA

******* *******

RI9 insertion 0.047kb

GAATGGATTAACGAGATTCCTACTGTCCCTATCTACTTGGAACACGCCACGTAAAAT--ATCCGAAAAGCATACATTGTCCCTATCTACTATCTAGCAA

********** **********

c) Uninterrupted rDNA:

GAATGGATTAACGAGATTCCTACTGTCCCTATCTACTATCTAGCGAA

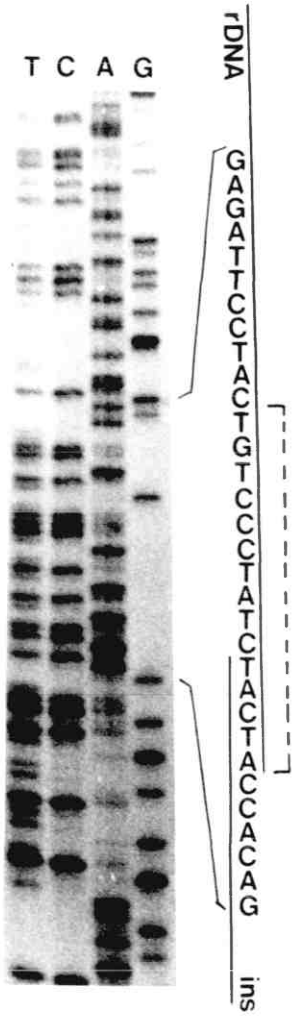
FIGURE 22.

Sequencing gels showing the left-hand junctions of the insertions in RI10 and MB27 and the right-hand junction in MB27.

The clone RI10 was cleaved with KpnI and BglII and 3' labeled at the KpnI site. The clone MB27 was cleaved with BstI and BglII (left junction) or BstI and EcoRI (right junction) and 3' labeled at the BstI site. The sequences from the left- and right-hand junctions are from the non-coding and coding strands, respectively. The rDNA and insertion sequences are indicated by solid lines placed against the sequence. The dashed line indicated the segment of rDNA duplicated on either side of the insertion. The variations seen in the length of the duplication are indicated against the sequence of the MB27 right-hand junction.

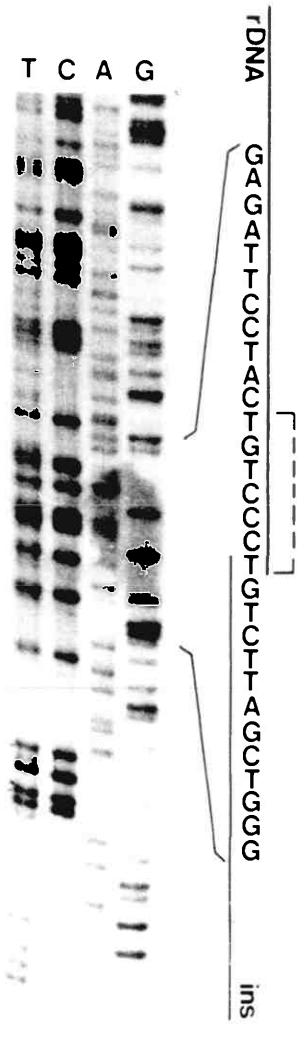
RI 10

Left Junction



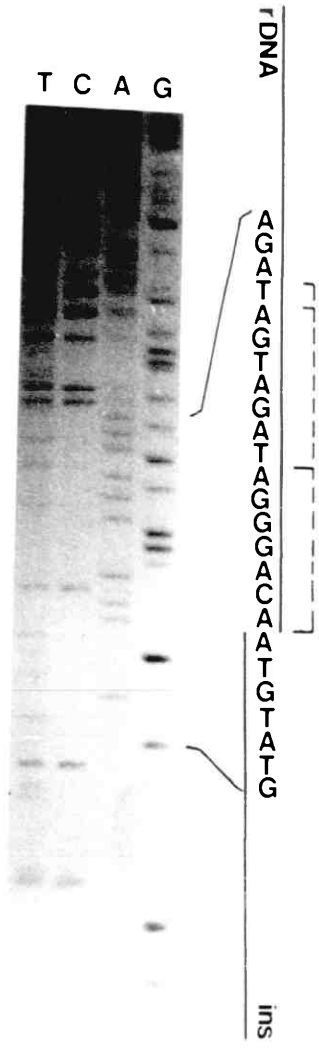
MB 27

Left Junction



MB 27

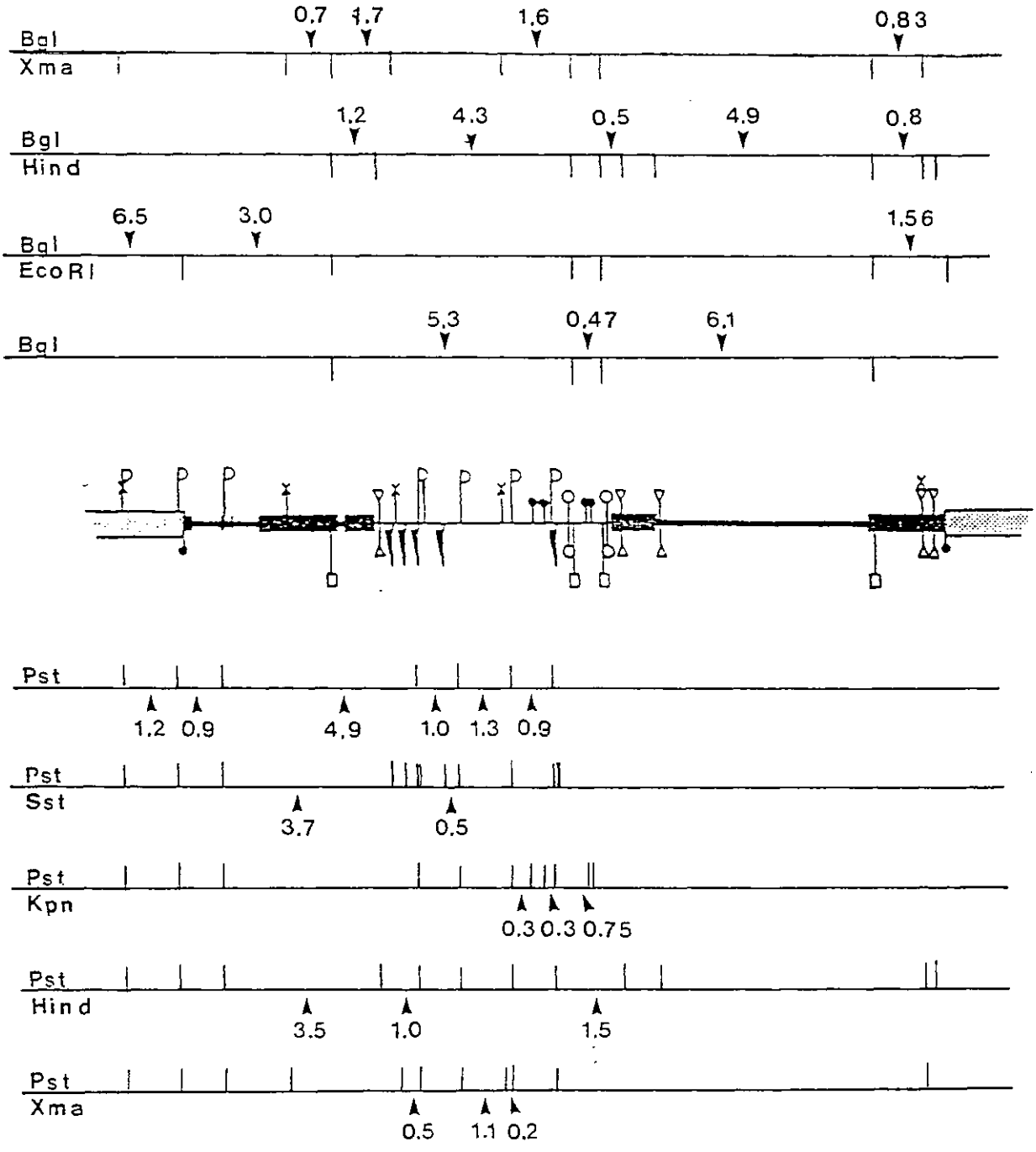
Right Junction

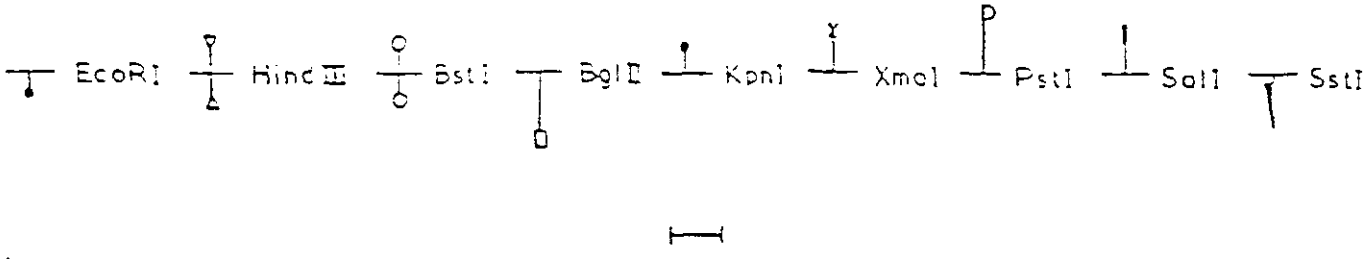


A P P E N D I X I

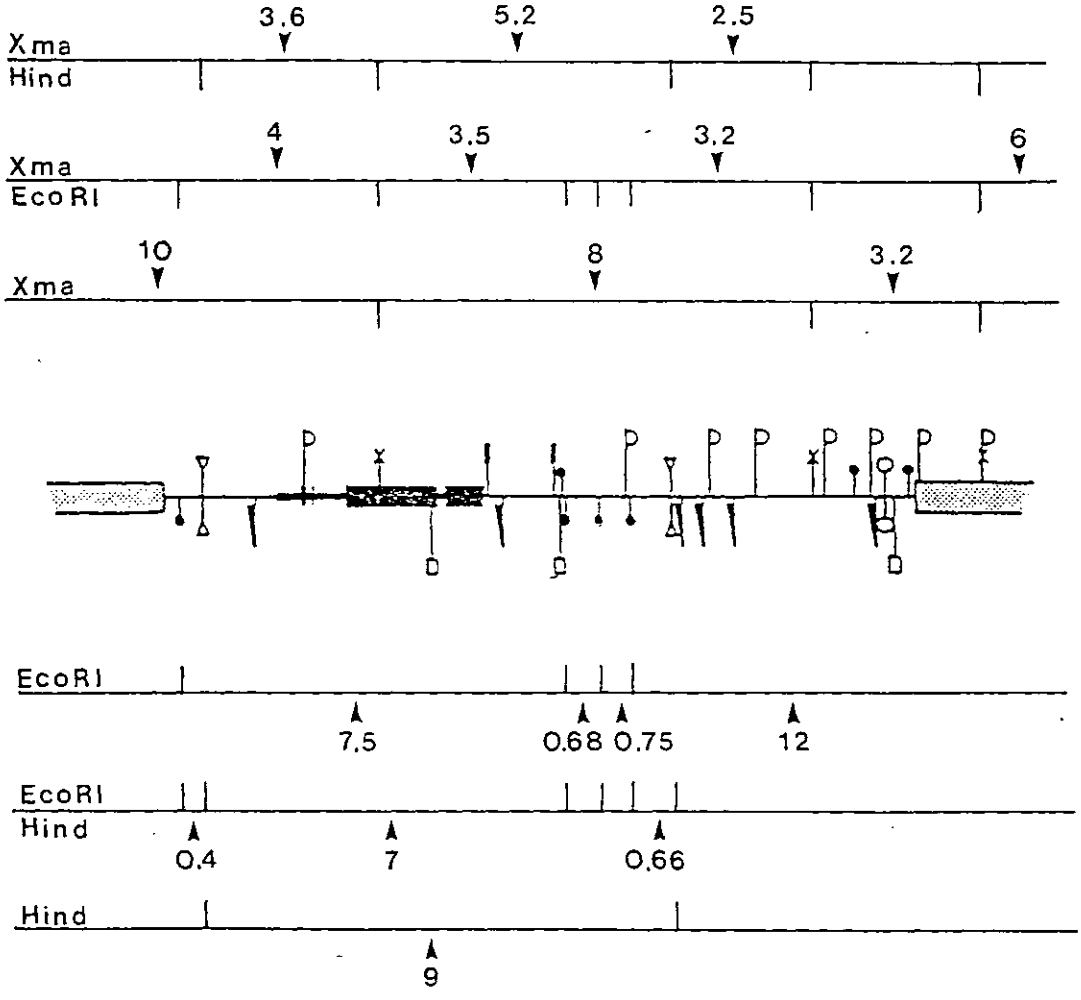
THE RESTRICTION ENDONUCLEASE MAPS OF THE rDNA CLONES

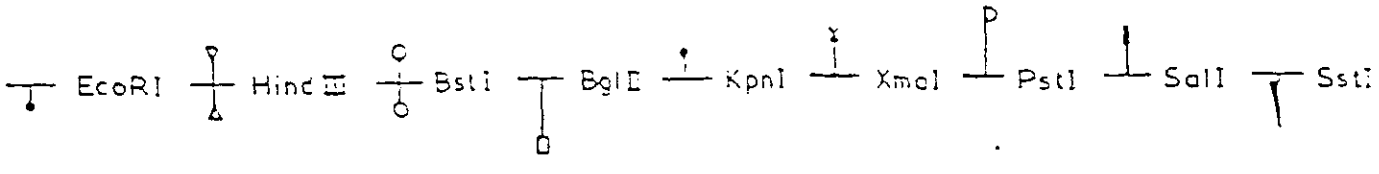
cDm103; Bgl II, Pst I



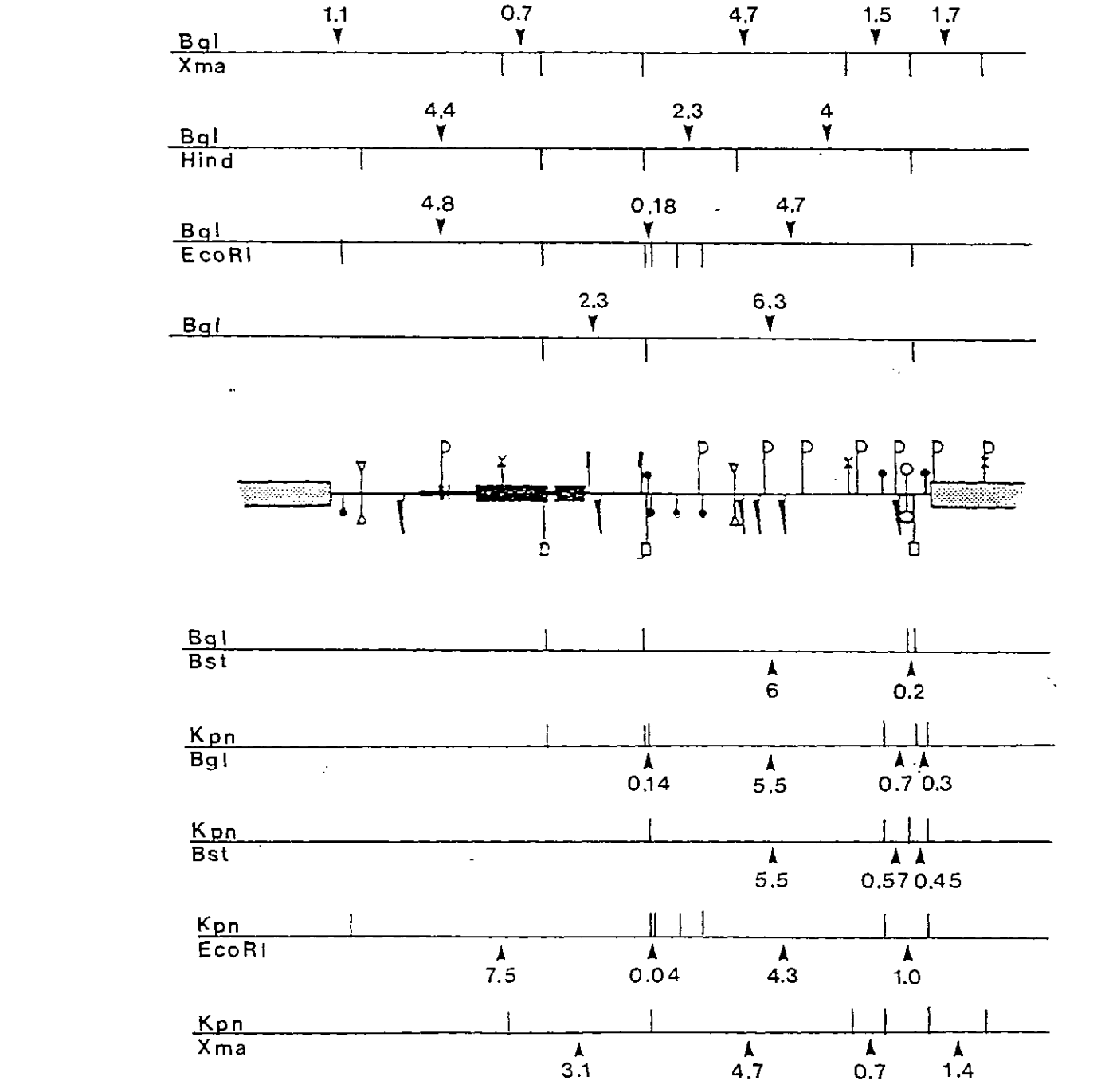


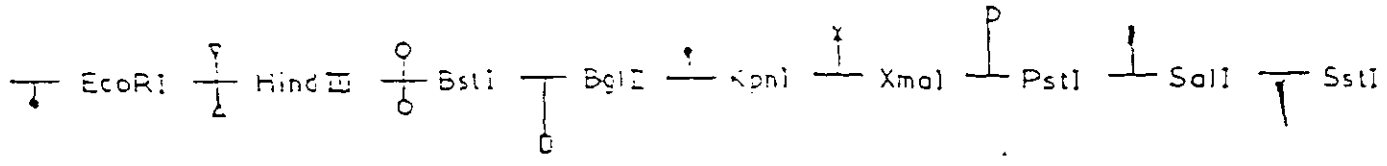
cDm207;EcoRI, HindIII, XmaI



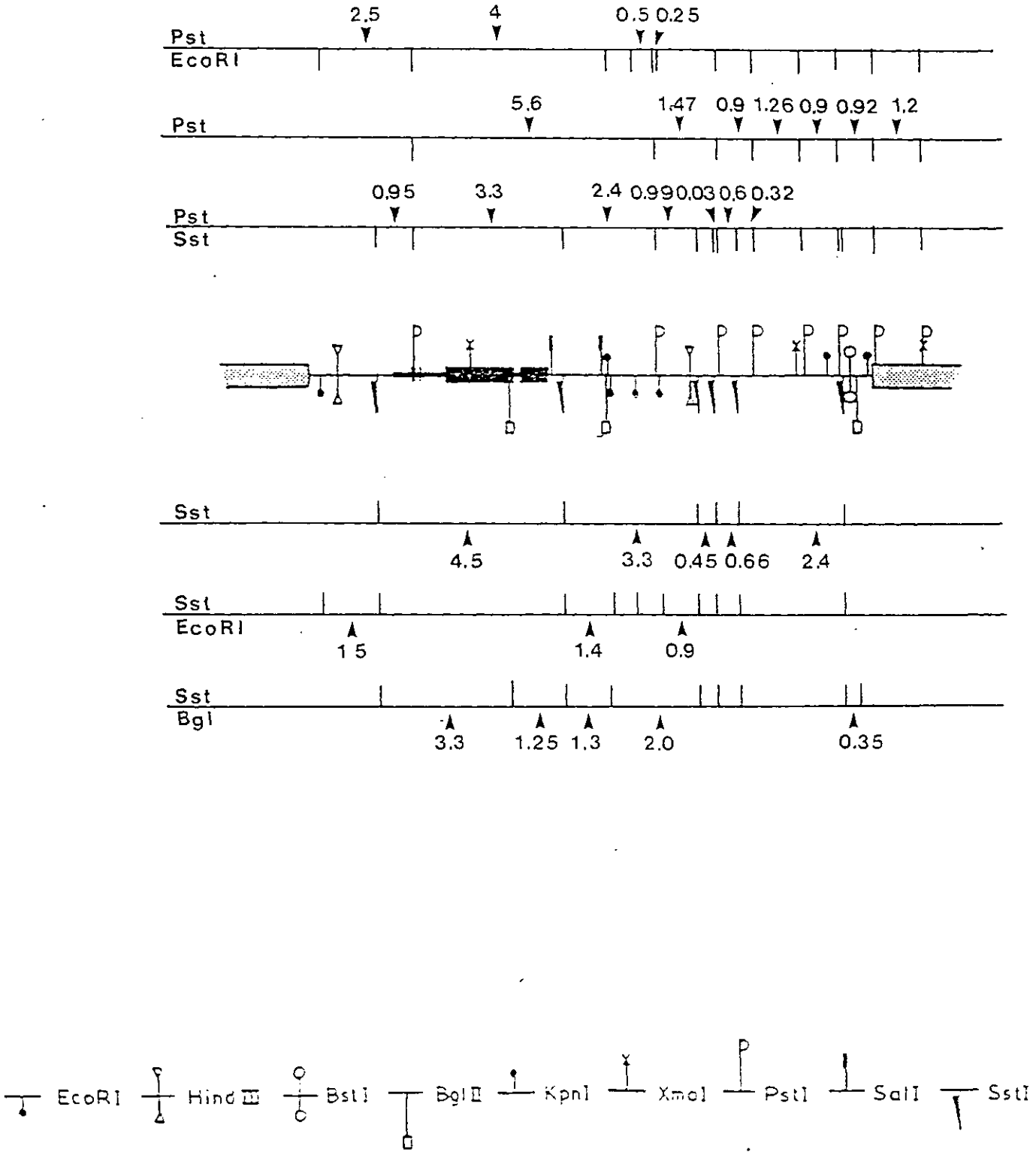


cDm207; BglII, Kpn I

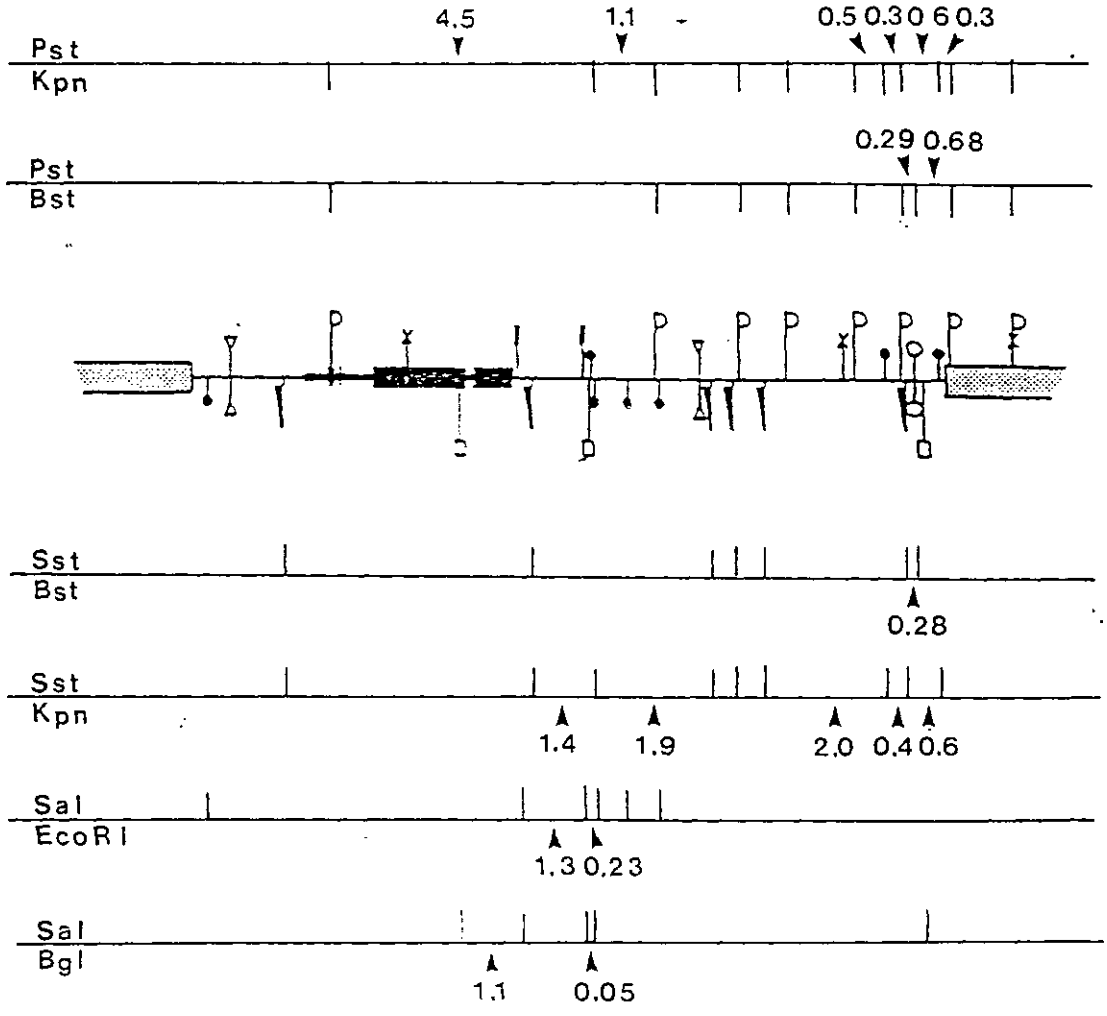


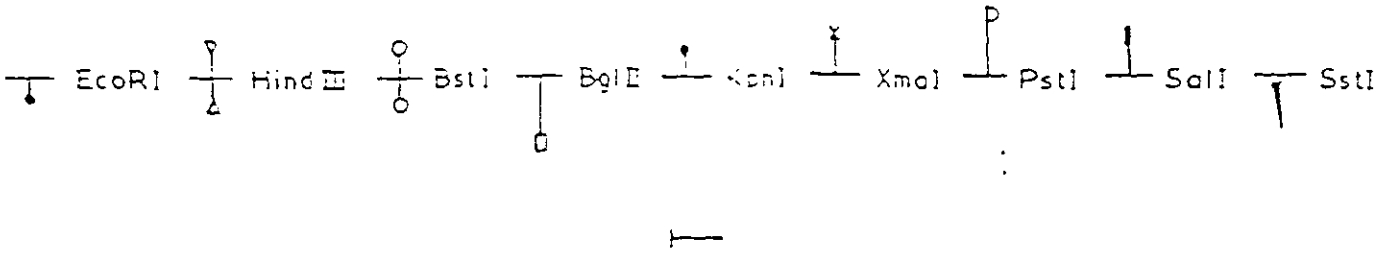


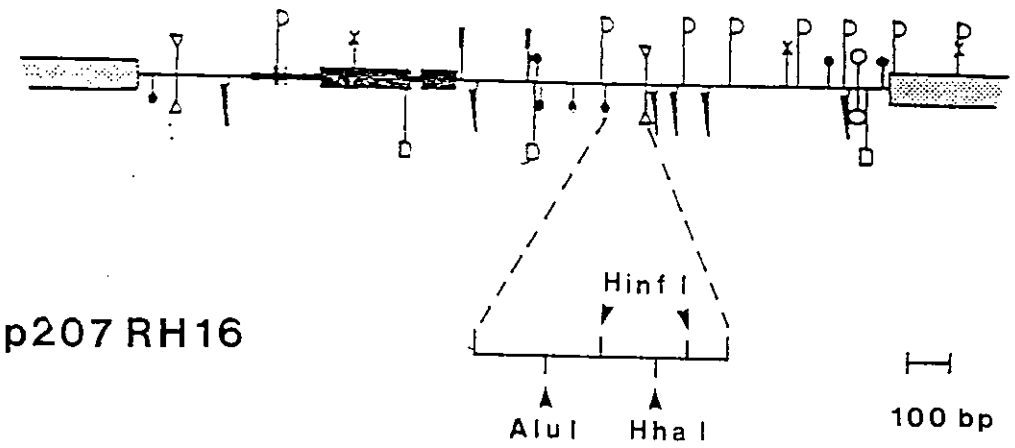
cDm207 ; Pst I, Sst I

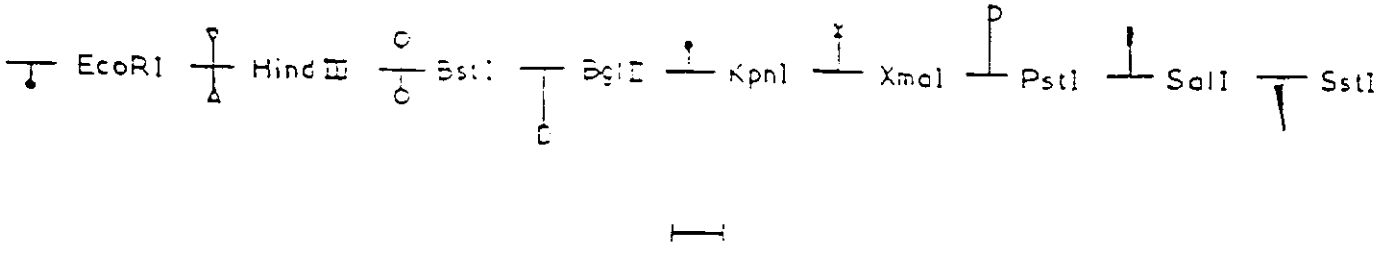


cDm207; Bst I, Kpn I, Sal I

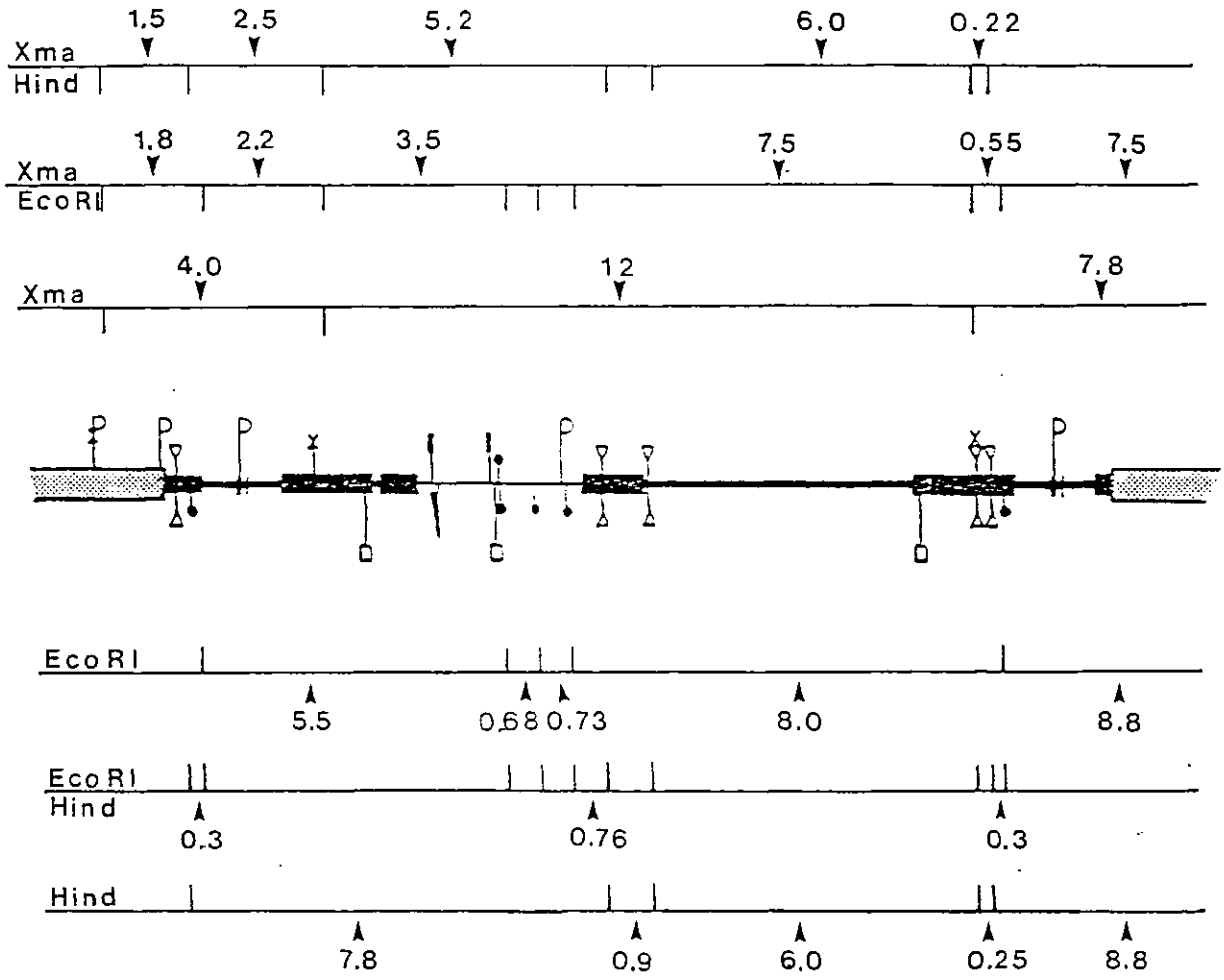


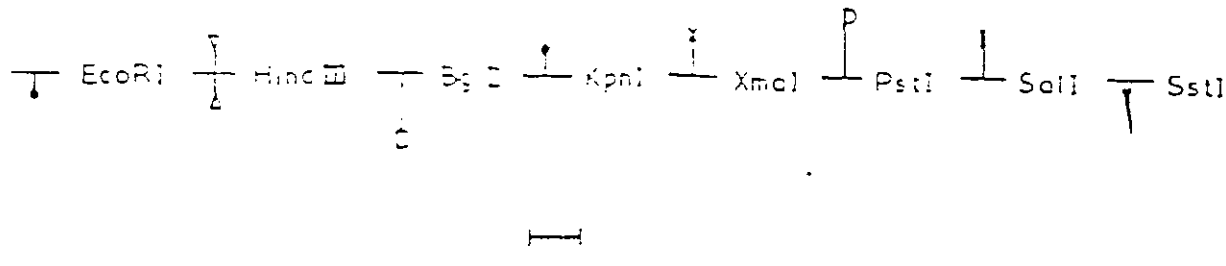



p207 RH16
Hinf I
AluI
Hha I
100 bp

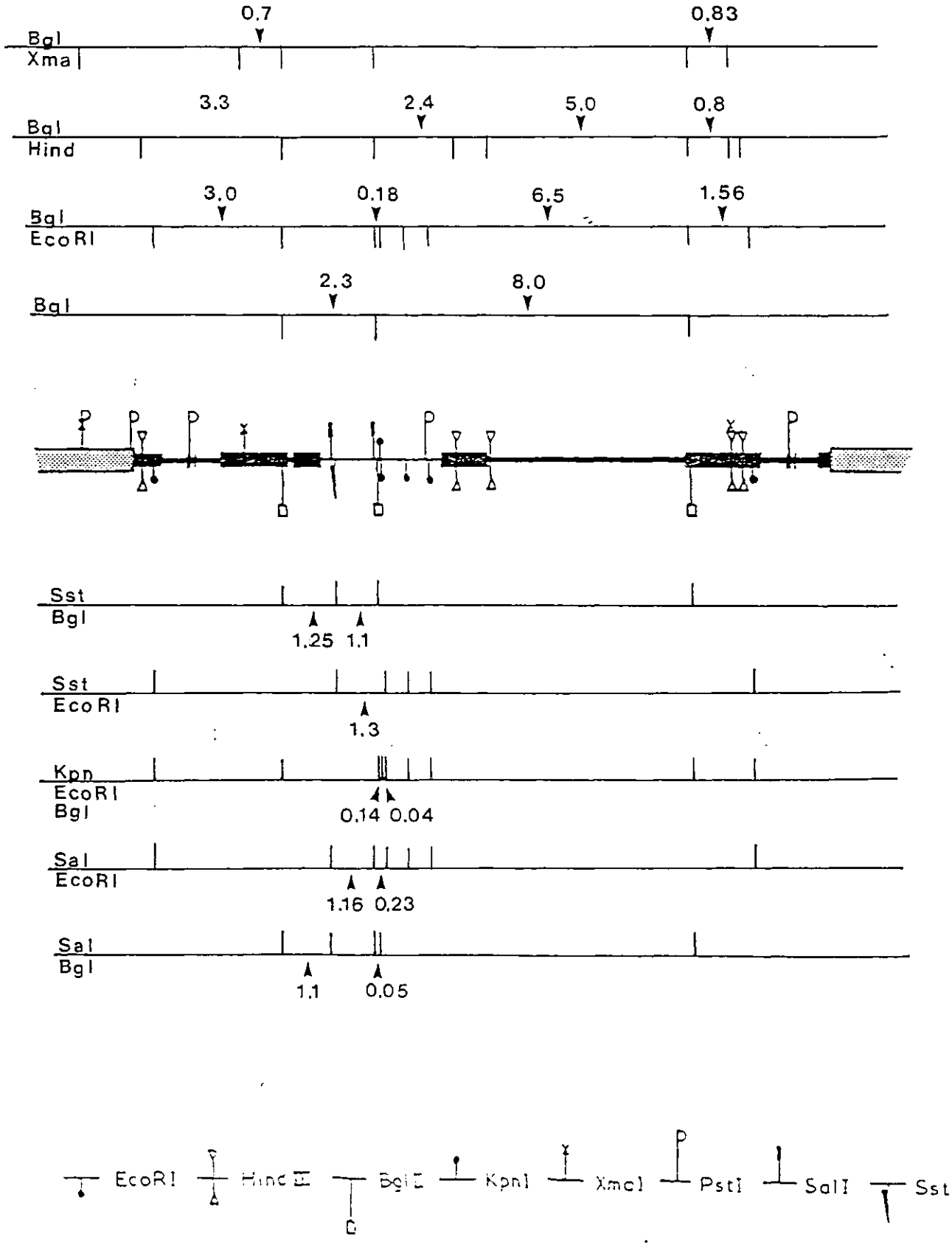

EcoRI
Hind III
Bst I
Bgl II
Kpn I
Xma I
Pst I
Sal I
Sst I

cDm 264 ; EcoRI , Hind III , Xma I

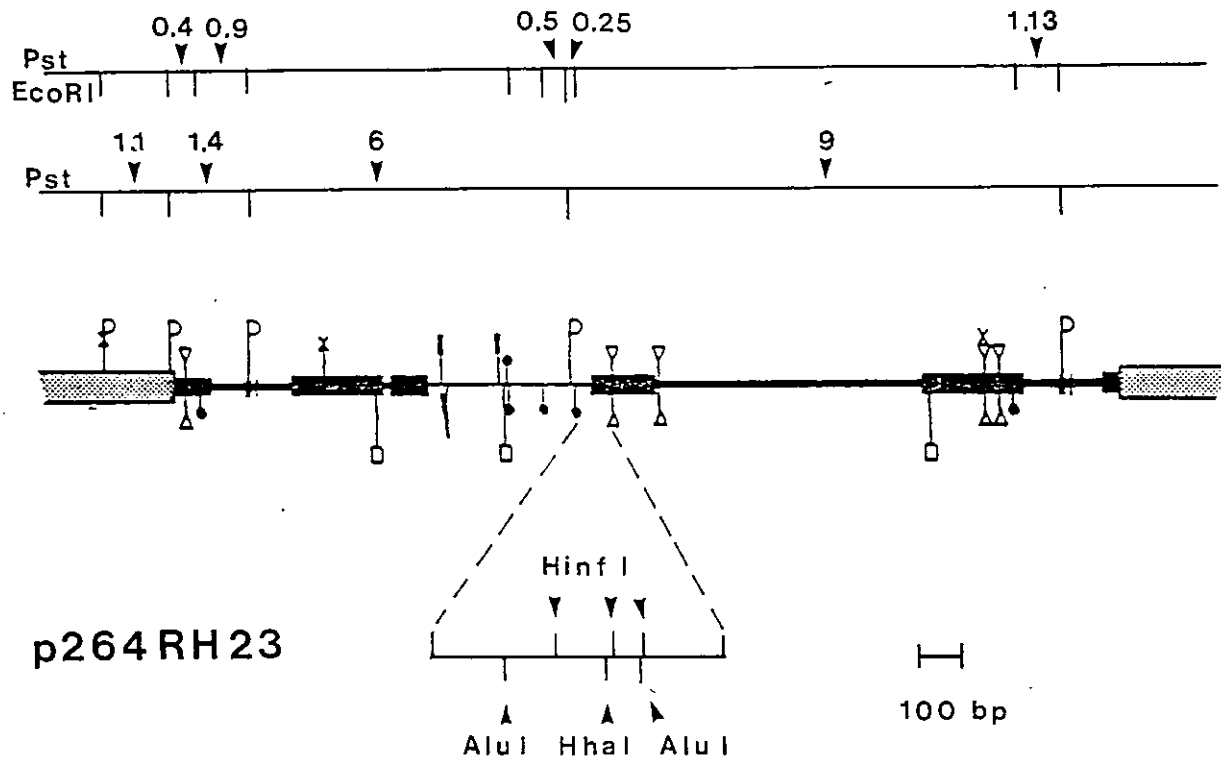


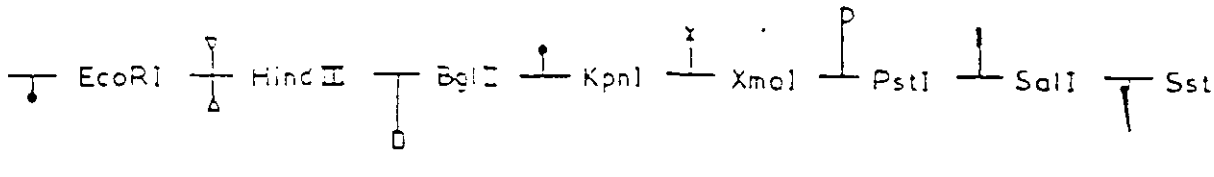


cDm 264; Bgl II, Kpn I, Sal I, Sst I

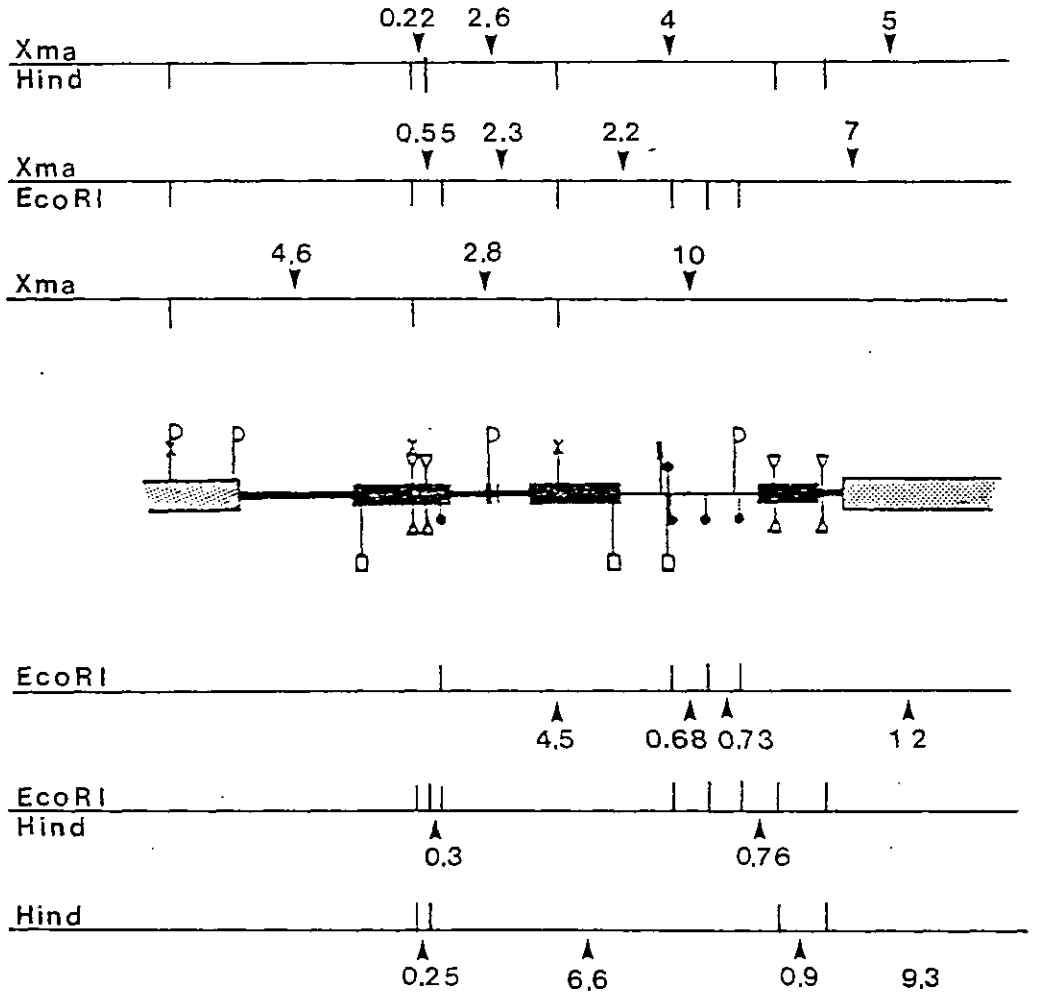


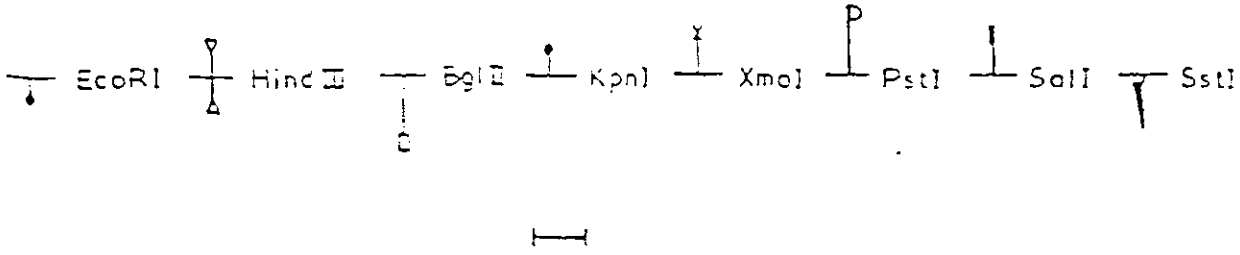
cDm 264 ; Pst I


0.4 0.9
0.5 0.25
1.13
Pst
EcoRI
1.1 1.4
6
9
Pst
Hinf I
p264RH23
100 bp
AluI HhaI AluI

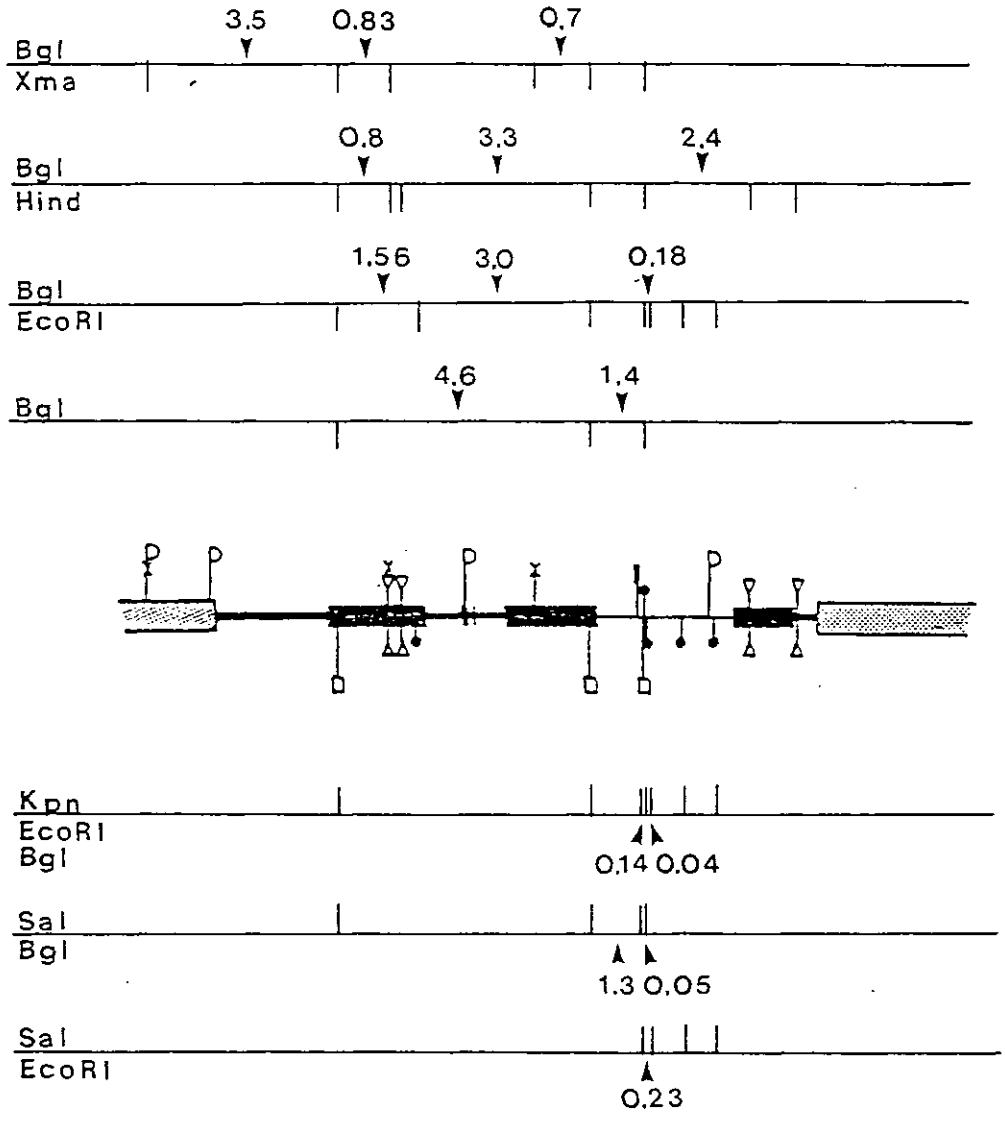

EcoRI
Hind III
Bgl II
KpnI
XmaI
PstI
SalI
SstI

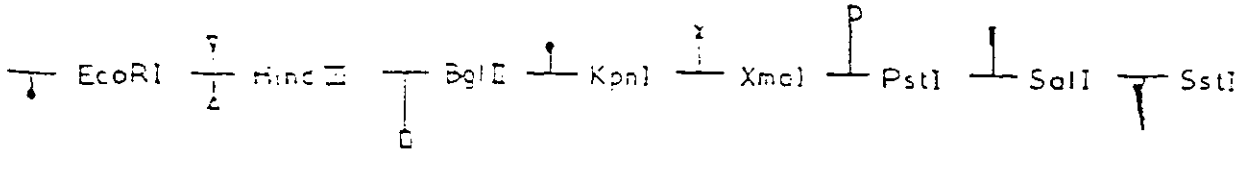
cDm 249 ; Eco RI , Hind III ; Xma I



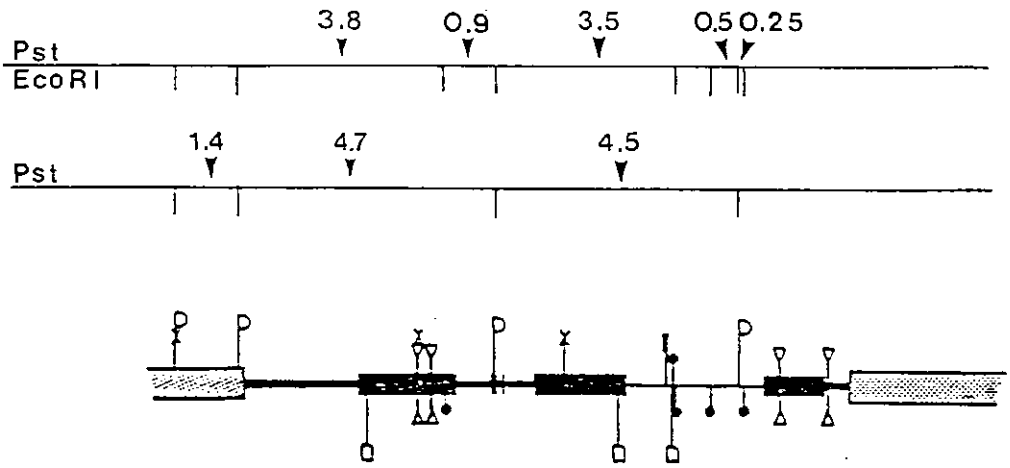


cDm 249 ; Bgl II , Kpn I , Sal I



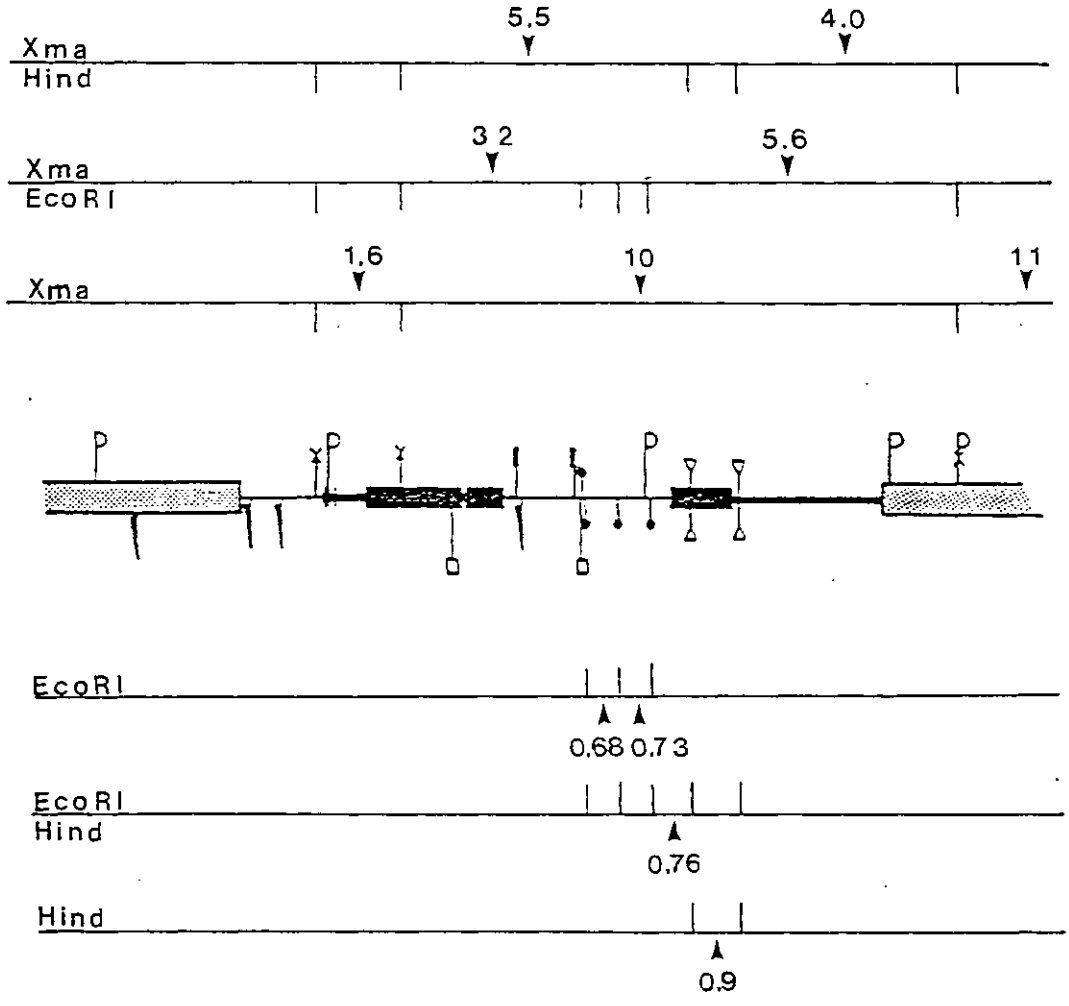


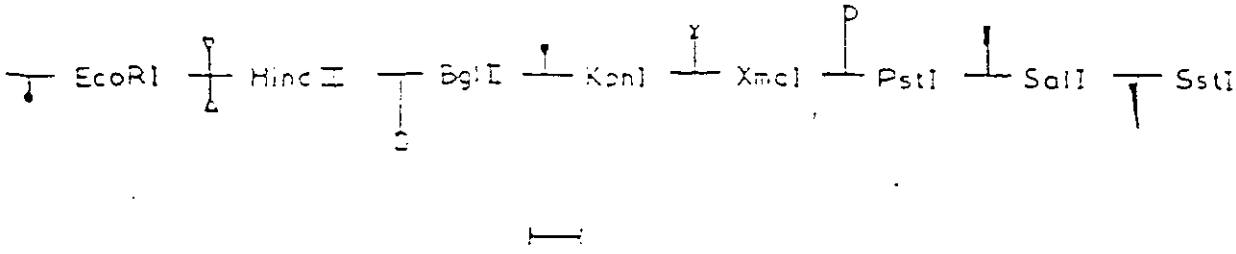
cDm 249 ; Pst I



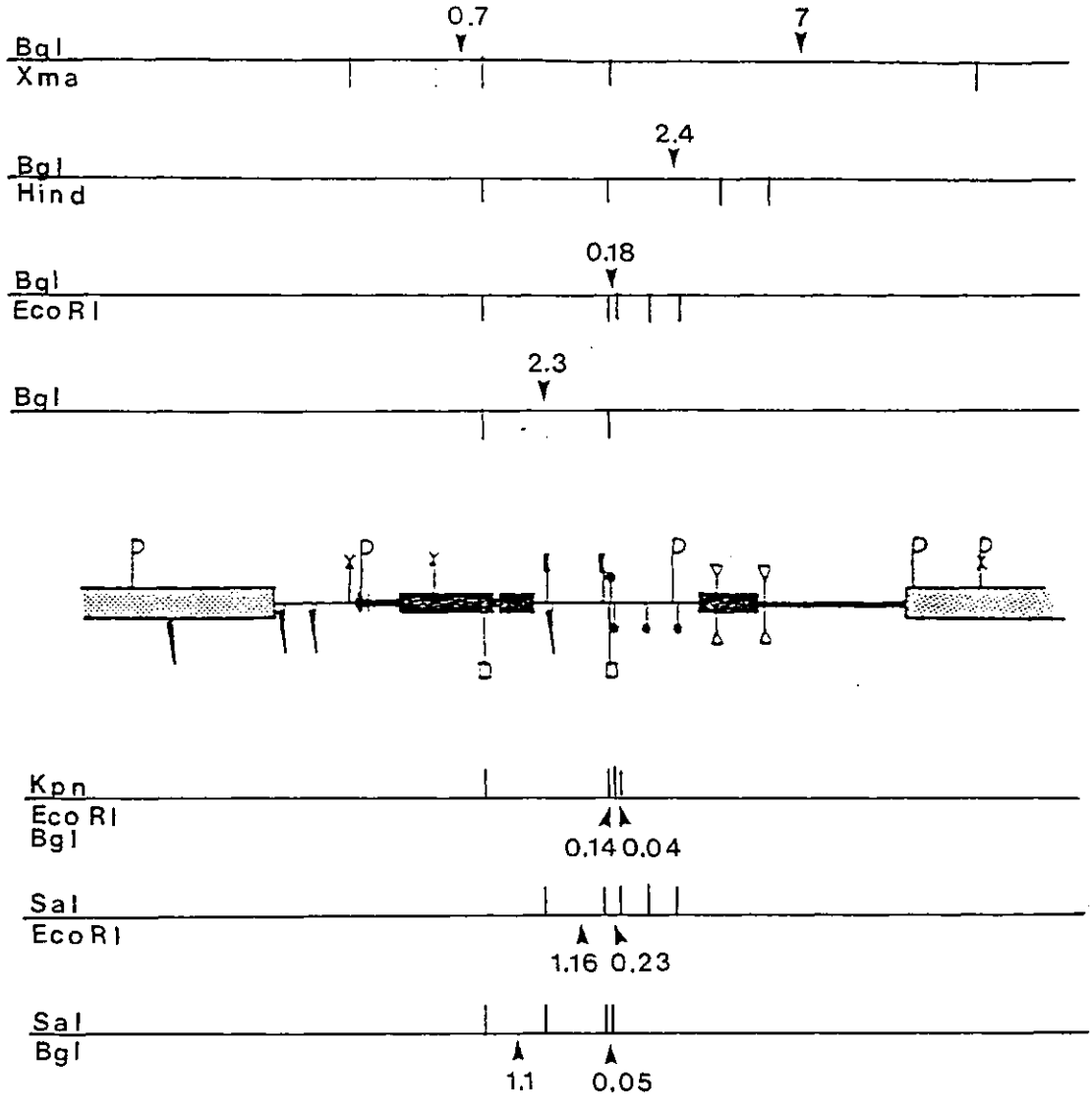
EcoRI HincII BglI KpnI XmaI PstI SalI SstI

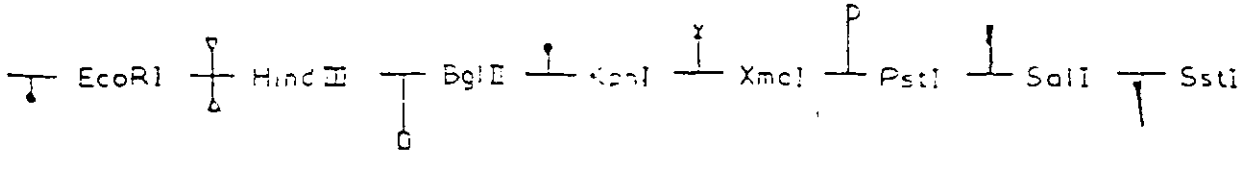
14C9 ; Eco RI , Hind III , Xma I



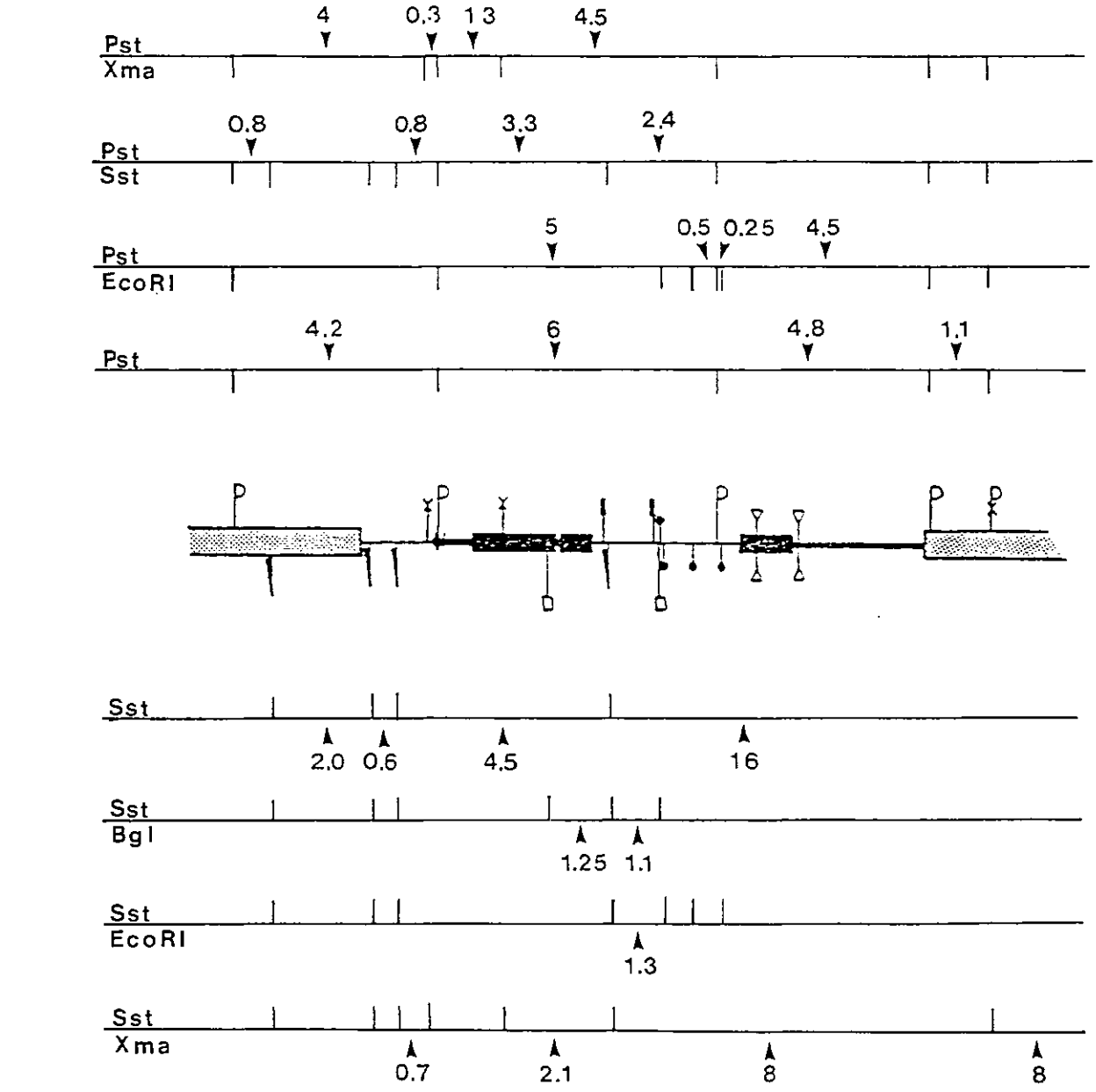


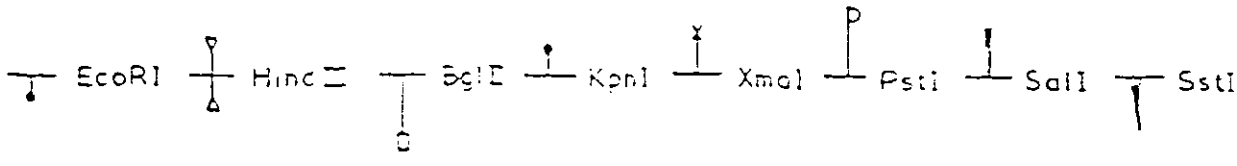
14C9 ; Bgl III, Kpn I, Sal I



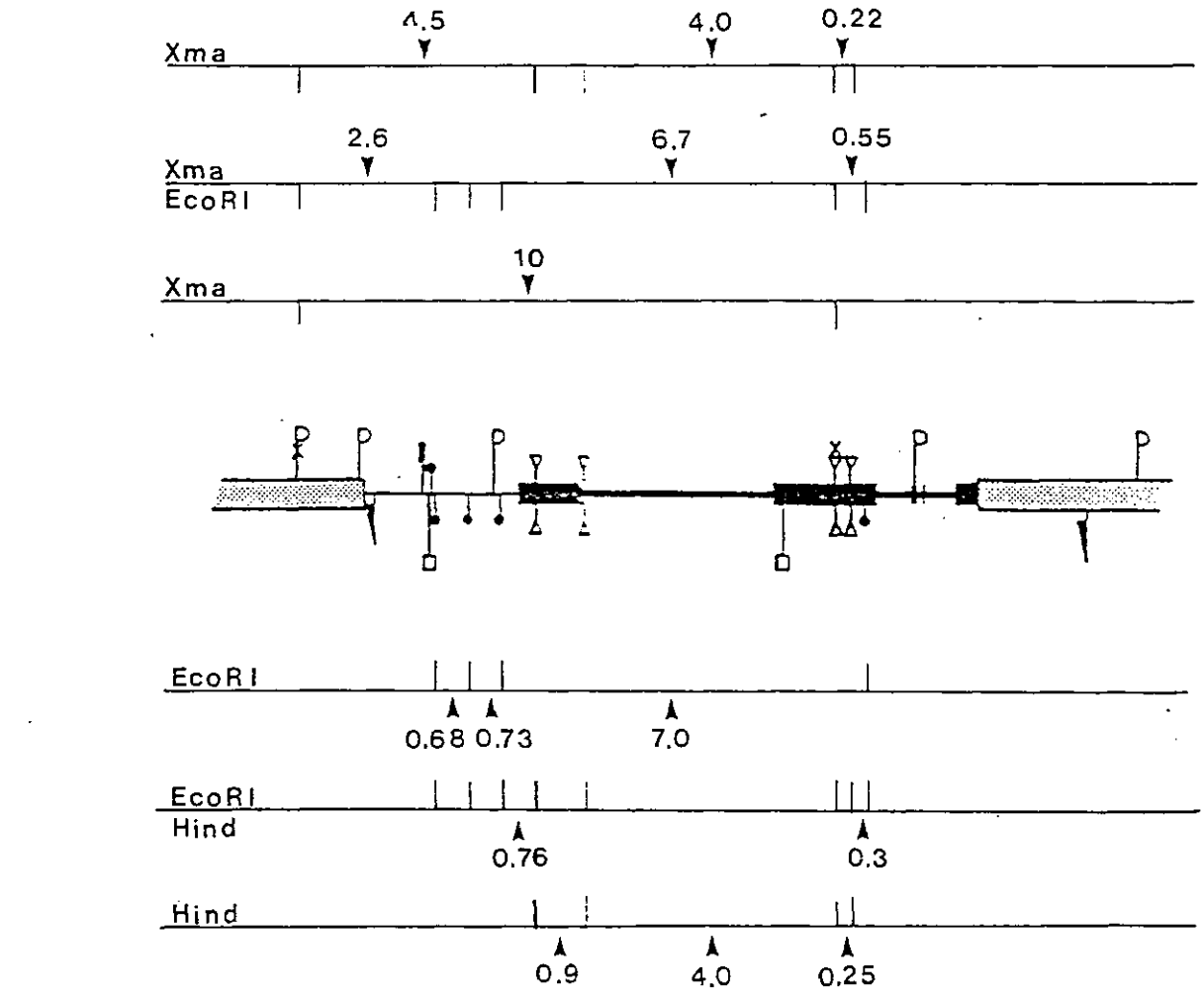


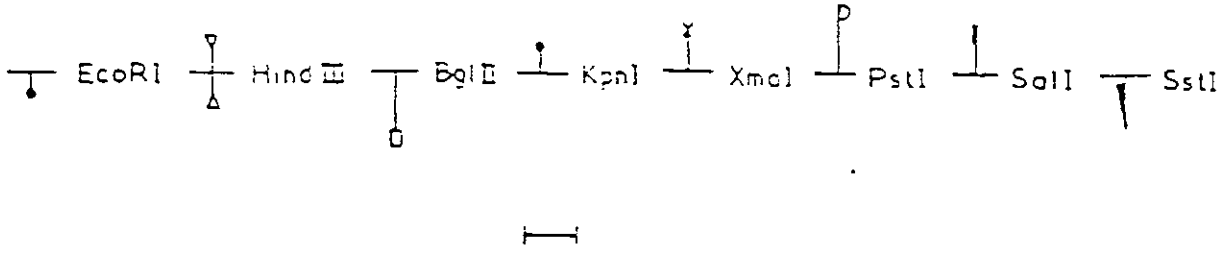
14C9 ; Pst I , Sst I



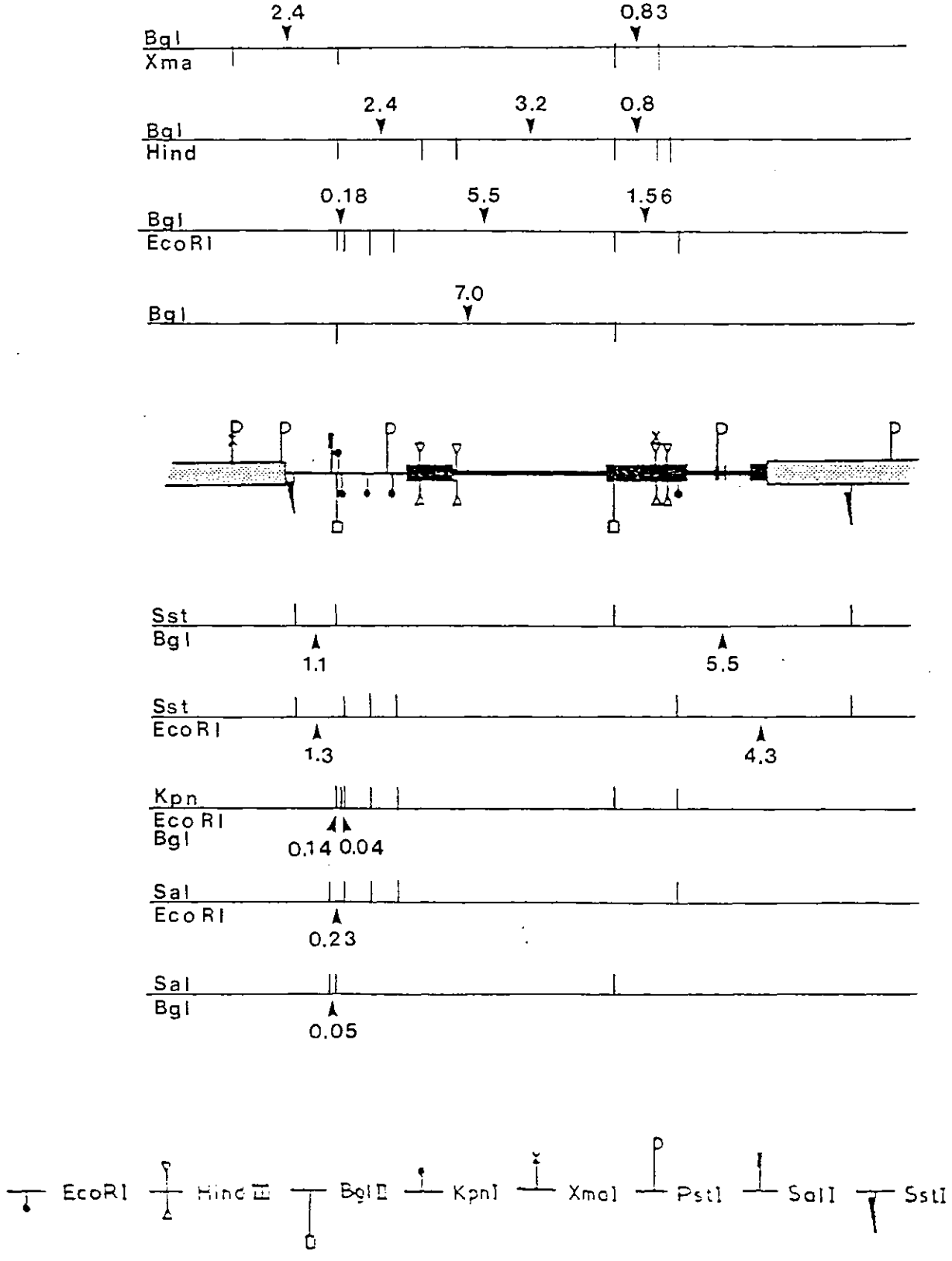


4G3 ; Eco RI, Hind III, Xma I

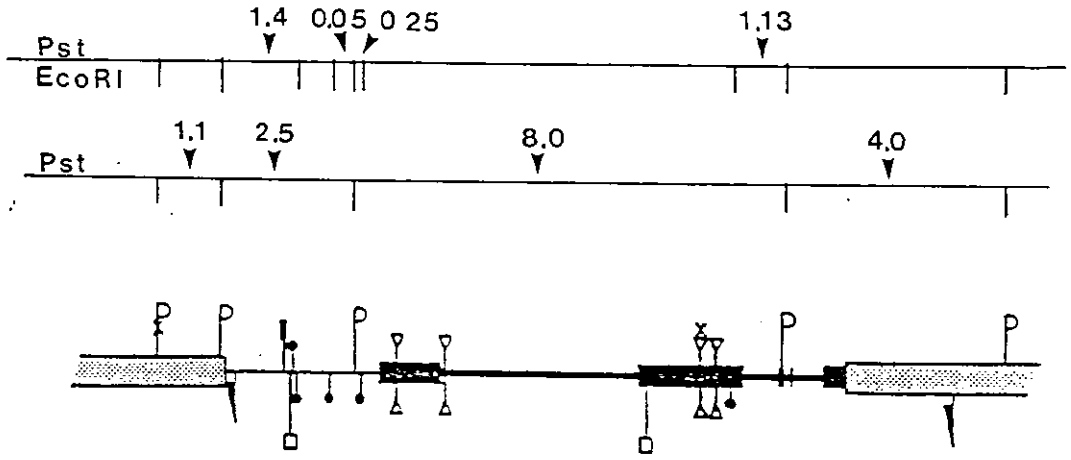


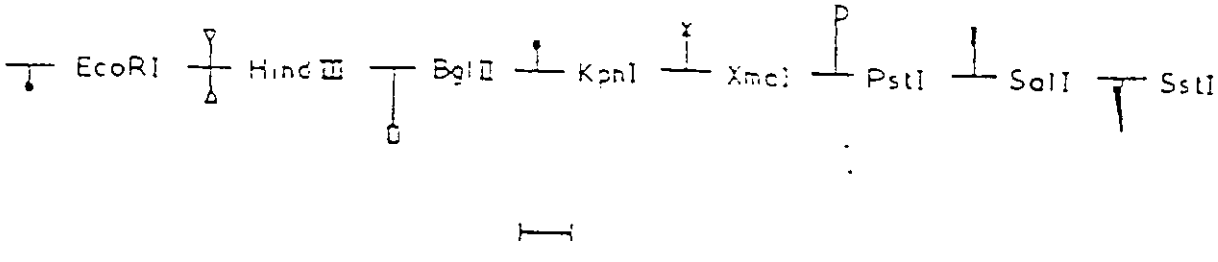


4G3 ; Bgl II, Kpn I, Sal I, Sst I



4G3 ; Pst I


1.4
0.05
0.25
1.13
Pst
EcoRI
1.1
2.5
8.0
4.0
Pst

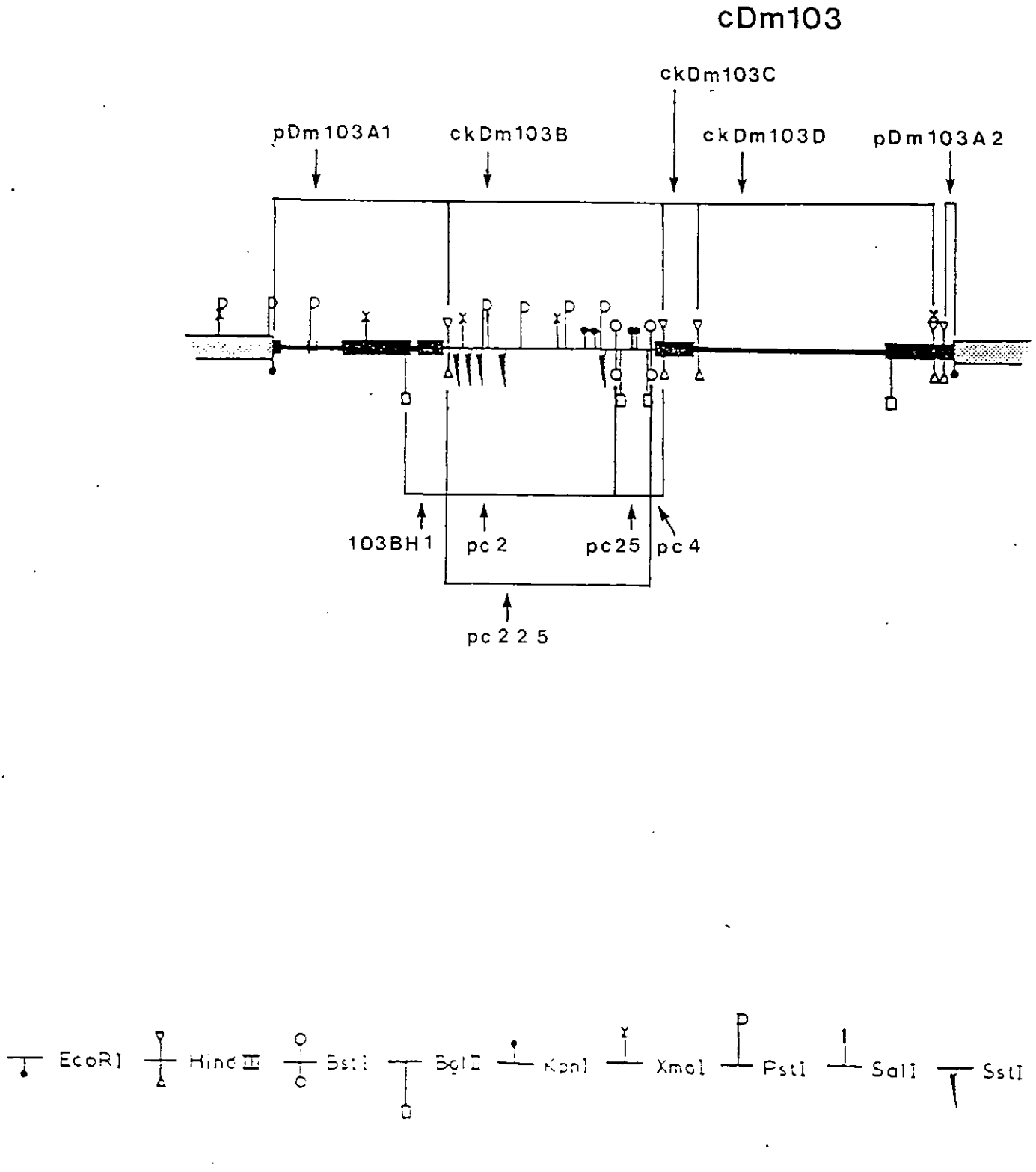

EcoRI
Hind III
Bgl II
KpnI
XmaI
PstI
SalI
SstI

A P P E N D I X II

THE SUBCLONES OF THE rDNA CLONES

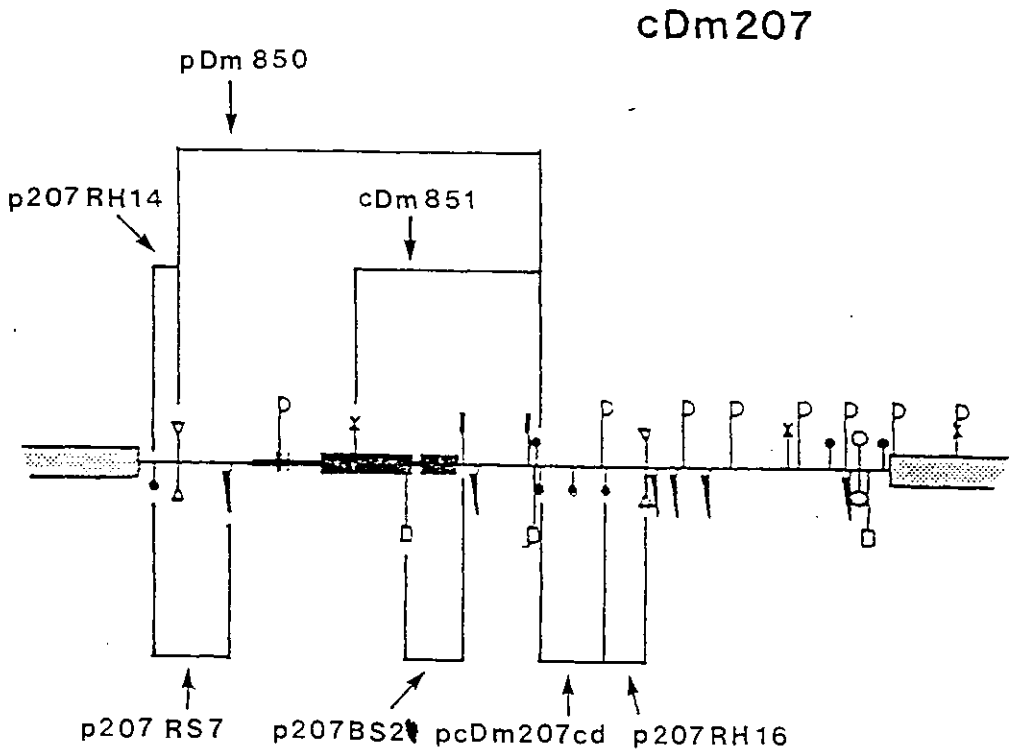
The subclones of cDm103.

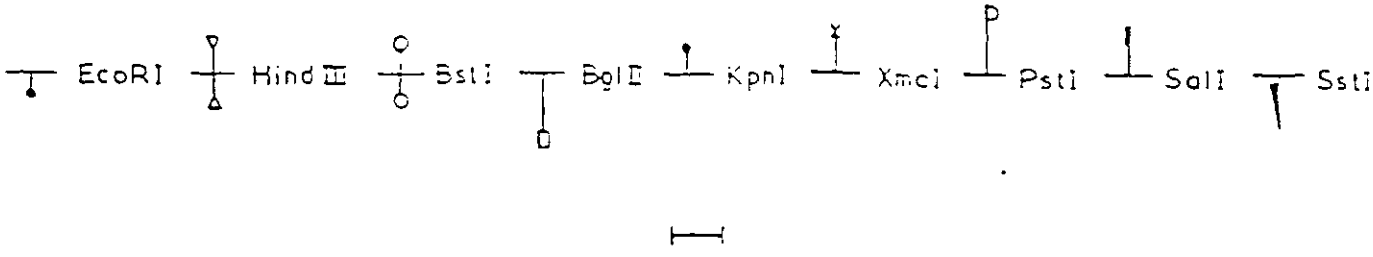
- pDm103A1, EcoRI/HindIII fragment in pSC101 (Glover and Hogness, 1977)
- ckDm103B, HindIII fragment in pML2 (Glover and Hogness, 1977)
- ckDm103C, " " " " " " " "
- ckDm103D, " " " " " " " "
- pDm103A2, HindIII/EcoRI fragment in pSC101 (Glover and Hogness, 1977)
- 103 BH1, BglII/HindIII fragment in pAT153 (Roiha et al., 1981)
- pc2, HindIII/BstI fragment in pBR322 (Kidd and Glover, 1980)
- pc25, BstI fragment in pBR322 (S. Kidd)
- pc225, HindIII/BstI and BstI fragments in pBR322 (S. Kidd)
- pc4, BstI/HindIII fragment in pBR322 (S. Kidd)


cDm103
ckDm103C
pDm103A1
ckDm103B
ckDm103D
pDm103A2
103BH1
pc 2
pc25
pc 4
pc 2 2 5
EcoR1
Hind III
Sst I
Bgl II
Kpn I
Xma I
Pst I
Sal I
Sst I

The subclones of cDm207.

- p207 RH14, EcoRI/HindIII fragment in pBR322
- pDm850, HindIII/EcoRI fragment in pSC101 (Glover, 1977)
- cDm851, XmaI/EcoRI fragment in ColE1 (Glover, 1977)
- h207 RS7, EcoRI/SstI fragment in Homer II (Chia, Scott and Rigby, Nucl. Acids Res., submitted) (S. Kidd)
- p207 BS21, BglII/SalI fragment in pBR322
- pc207cd, the two EcoRI fragments in pBR322 (S. Kidd)
- p207 RH16, EcoRI/HindIII fragment in pBR322


cDm207
pDm 850
p207RH14
cDm 851
p207 RS7
p207BS2
pcDm207cd
p207RH16


EcoRI
Hind III
BstI
BglII
KpnI
XmcI
PstI
SalI
SstI

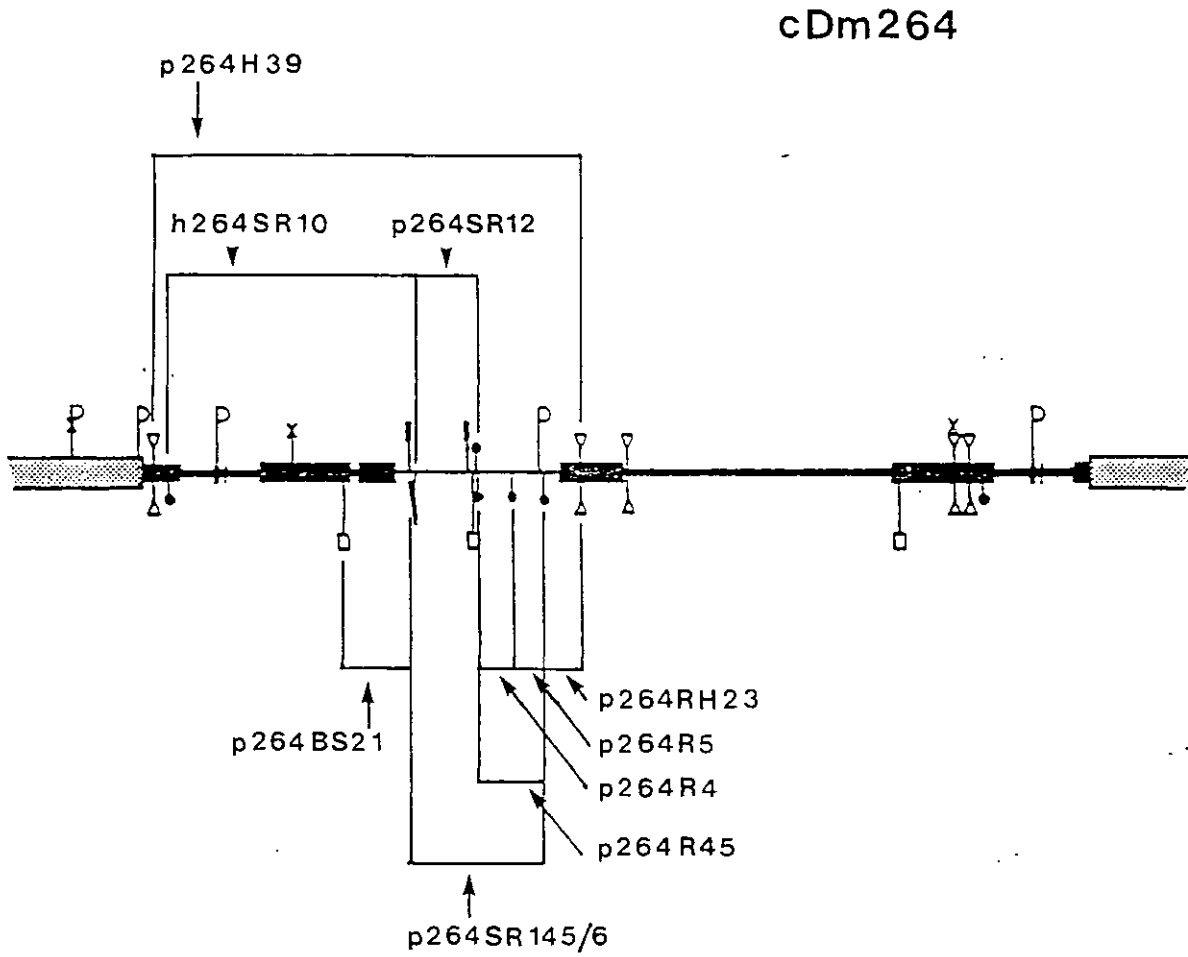
The subclones of cDm264.

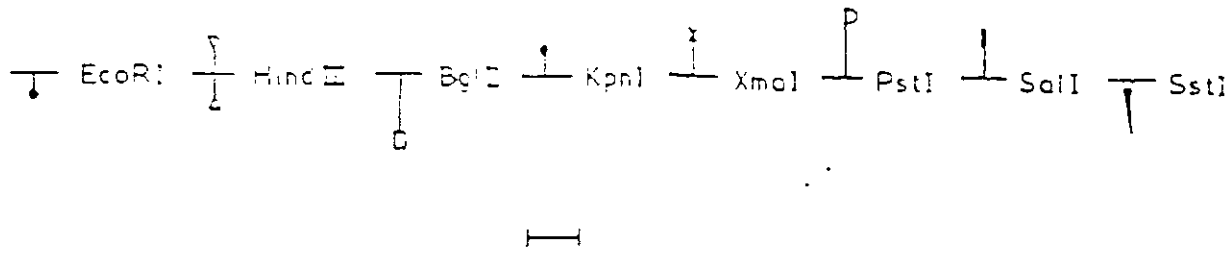
- p264 H39, HindIII fragment in pBR322
- h264 SR10, EcoRI/SstI fragment in Homer II (S. Kidd)
- p264 SR12, SstI/EcoRI fragment in Homer II' (Homer II' is Homer II with the cos site deleted) (S. Kidd)

  (h264 SR1, the same fragment in Homer II)
- p264 BS21, BglII/SalI fragment in pBR322
- p264 R4, the left-hand EcoRI fragment in pAT153 (S. Kidd)
- p264 R5, the right-hand EcoRI fragment in pAT153 (S. Kidd)
- p264 R45, the two EcoRI fragments in pAT153 (S. Kidd)
- p264 SR145/6, SstI/EcoRI and the two EcoRI fragments in Homer II (S. Kidd)
- p264 RH23, EcoRI/HindIII fragment in pBR322

- p2643/5, the full-length reorganised 264 in pBR322 (not shown) (S. Kidd)

  (p2383/12, the similarly reorganised 238 in pBR322) (S. Kidd)

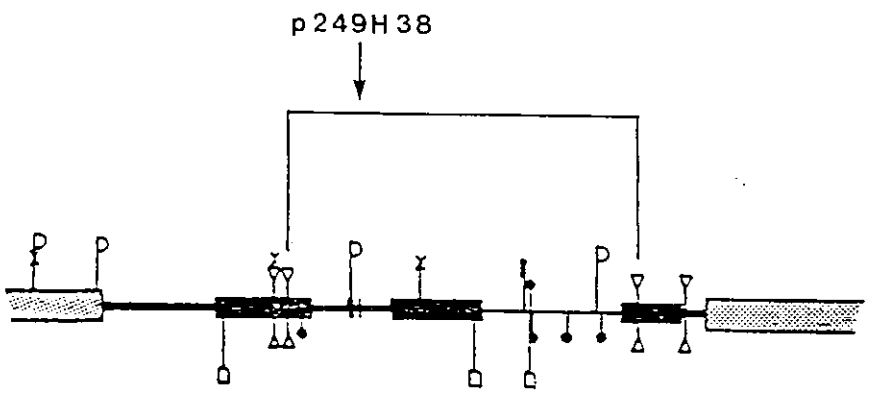

cDm264
p264H39
h264SR10
p264SR12
p264BS21
p264RH23
p264R5
p264R4
p264R45
p264SR145/6

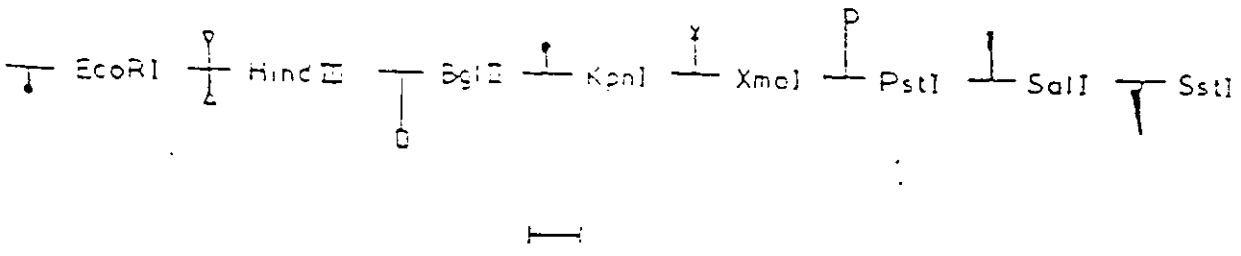

EcoRI
HincII
BglII
KpnI
XmaI
PstI
SalI
SstI

The subclone of cDm249.

- p249 H38, HindIII fragment in pBR322

cDm249





R E F E R E N C E S

Allet, B., Rochaix, J-D. (1979) Cell, 18, 55 - 60

Alwine, J.C., Kemp, D.J., Stark, G.R. (1977) Proc. Nat. Acad. Sci. U.S.A., 74, 5350 - 5354

Alwine, J.C., Kemp, D.J. (1979) Methods in Enzymol., 68, 220 - 244 (Wu, R., Ed.) Academic Press, N.Y.

Arnheim, N., Krystal, M., Scmickel, R., Wilson, G., Ryder, O., Zimmer, E. (1980) Proc. Nat. Acad. Sci. U.S.A., 77, 7323 - 7327

Artavanis - Tsakonas, S., Schedl, P., Tschudi, C., Pirrotta, V., Steward, R., Gehring, W.J. (1977) Cell, 12, 1057 - 1067

Ashburner, M. (1970) Advanc. Insect Physiol., 7, 1 - 196

Barnett, T., Rae, P.M.M. (1979) Cell, 16, 763 - 775

Beckingham, K., White, R. (1980) J. Mol. Biol., 137, 349 - 373

Beckingham, K. (1981) J. Mol. Biol., 149, 141 - 169

Berk,A.J., Sharp,P.A. (1977) Cell, 12, 721 - 732

Birnboim, H.C., Doly, J. (1979) Nucl. Acids Res., 7, 1513 - 1523

Bolivar, F., Rodrigues, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., Falkow, S. (1977) Gene, 2, 45 - 113

Breatnach, R., Benoist, C., O'Hare, K., Gannon, F., Chambon, P. (1978) Proc. Nat. Acad. Sci. U.S.A., 75, 4853 - 4857

Brown, D.D., Wensink, P.C., Jordan, E. (1972) J. Mol. Biol., 63, 57 - 73

Brutlag, D., Peacock, W.J. (1979) J. Mol. Biol., 135, 565 - 580

Brutlag, D.L., Clayton, J., Friedland, P., Kedes, L.H. (1982) Nucl. Acids Res., 10, 279 - 294

Cameron, J.R., Lohe, Y., Davis, R.W. (1979) Cell, 16, 739 - 751

Calos, M.P., Miller, J.H. (1980) Cell, 20, 579 - 595

Casey, J., Davidson, N. (1977) Nucl. Acids Res., 4, 1539 - 1552

Clewell, D.B., Helinski, D.R. (1969) Proc. Nat. Acad. Sci. U.S.A., 62, 1159 - 1166

Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Helling, R.B. (1973) Proc. Nat. Acad. Sci. U.S.A., 70, 3240 - 3244

Cooper, K.W. (1959) Chromosoma, 10, 535 - 588

Dawid, I.B., Botchan, P. (1977) Proc. Nat. Acad. Sci. U.S.A., 74, 4233 - 4237

Dawid, I.B., Wellauer, P.K., Long, E.O. (1978) J. Mol. Biol., 126, 749 - 768

Dawid, I.B., Long, E.O., DiNocera, P.P.,Pardue, M.L. (1981) Cell, 25, 399 - 408

Dawid, I.B., Rebbert,M.L. (1981) Nucl. Acids Res., 9, 5011 - 5020

Degelmann, A., Royer, H-D., Hollenburg, C.P. (1979) Chromosoma, 71, 263 - 281

Denhardt, D.T. (1966) Biochem. Biophys. Res. Commun., 23, 641 - 646

Dover, G., Coen, E. (1981) Nature, 290, 731 - 732

Dujon, B. (1980) Cell, 20, 185 - 197

Dunsmuir, P., Brorein, W.J., Simon, M.A., Rubin, G.M. (1980) Cell, 21, 575 - 580

Efstratiadis, A., Vournakis, J.N., Donis-Keller, H., Chaconas, G., Dougall, D.K. (1979) Nucl. Acids Res., 4, 4165 - 4174

Endow, S.A., Gall, J.G. (1975) Chromosoma, 50, 175 - 192

Endow, S.A., Glover, D.M. (1979) Cell, 17, 597 - 605

Endow, S.A. (1980) Cell, 22, 149 - 155

Eschalier, G., Ohanessian, A. (1970) In Vitro, 6, 162 - 172

Farabaugh, P.J., Fink, G.R. (1980) Nature, 286, 352 - 356

Favaloro, J., Treisman, R., Kamen, R. (1980) Methods in Enzymol., 65, 709 - 717 (Maldave, K. and Grossman, L., Eds.) Academic Press, N.Y.

Fedoroff, N.V. (1979) Cell, 16, 647 - 710

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., van de Voorde, A., van Heuverswyn, H., van Heereweghe, J., Volckaertand, G., Ysebaer, M. (1978) Nature, 273, 113 - 120

Finnegan, D.J., Rubin, G.M., Young, M.W., Hogness, D.S. (1978) Cold Spring Harbor Symp. Quant. Biol., 42, 1053 - 1063

Forsheit, A.B., Davidson, N., Brown, D.D. (1974) J. Mol. Biol., 90, 301 - 314

Franz, G., Kunz, W. (1981) Nature, 292, 638 - 640

French, C.K., Fouts, D.L., Manning, J.E. (1981) Nucl. Acids Res., 9, 2563 - 2576

Gall, J.G., Cohen, E.H., Polan, M.L. (1971) Chromosoma, 33, 319 - 344

Glatzer, K.H. (1975) Chromosoma, 53, 371 - 379

Glatzer, K.H. (1979) Chromosoma, 75, 161 - 175

Glover, D.M., White, R.L., Finnegan, D.J., Hogness, D.S. (1975) Cell, 5, 149 - 157

Glover, D.M., Hogness, D.S. (1977) Cell, 10, 167 - 176

Glover, D.M. (1977) Proc. Nat. Acad. Sci. U.S.A., 74, 4932 - 4436

Gottesfield, J.M., Bloomer, L.S. (1980) Cell, 21, 751 - 760

Gourse, R.L., Gerbi, S.A. (1980) Nucl. Acids Res., 8, 3623 - 3637

Graziani, F., Vicari, L., Boncinelli, E., Malva, C., Manzi, A., Mariani, C. (1981) Proc. Nat. Acad. Sci. U.S.A., 78, 7662 - 7664

Greenberg, J.R. (1969) J. Mol. Biol., 46, 85 - 98

Grunstein, M., Hogness, D.S. (1975) Proc. Nat. Acad. Sci U.S.A., 72, 3961 - 3965

Hamkalo, B.A., Miller, O.L., Bakken, A.H. (1973) Cold Spring Harbor Symp. Quant. Biol., 38, 915 - 919

Henderson, A., Ritossa, F. (1970) Genetics, 66, 463 - 473

Hennig, W., Meer, B. (1971) Nature New Biol., 233, 70 - 72

Hennig, W., Link, B., Leoncin, I.O. (1975) Chromosoma, 51, 57 - 63

Hershfield, V., Boyer, H.W., Yanovsky, C., Lovett, M.A., Helinski, D.R. (1974) Proc. Nat. Acad. Sci U.S.A., 71, 3455 - 3459

Hollenberg, C.P. (1976) Chromosoma, 57, 185 - 197

Indik, Z.K., Tartof, K.D. (1980) Nature, 284, 477 - 479

Jeffreys, A.J., Flavell, R.A. (1977) Cell, 12, 1097 - 1108

Jolly, D.I., Thomas, C.A. (1980) Nucl. Acids Res., 8, 67 - 84

Jordan, B.R. (1975) J. Mol. Biol., 98, 277 - 280

Jordan, B.R., Jourdan, R., Jacq, B. (1976) J. Mol. Biol., 101, 85 - 105

Jordan, B.R., Glover, D.M. (1977) FEBS Letters, 78, 271 - 274

Jordan, B.R., Latil-Lamotte, M., Jourdan, R. (1980) FEBS Letters, 117, 1227 - 1231

Jovin, T.M., Englund, D.T., Bertsch, L.L. (1969) J. Biol. Chem., 244, 2996 - 3008

Kidd, S.J., Glover, D.M. (1980) Cell, 19, 103 - 119

Kidd, S.J. (1980) Ph.D. Thesis, University of London

Kidd, S.J., Glover, D.M. (1981) J. Mol. Biol., 151, 645 - 662

Krider, H.M., Levine, B.I. (1975) Genetics, 81, 501 - 513

Krider, H.M., Yedvobnick, B., Levine, B.I. (1979) Genetics, 92, 879 - 889

Krystal, M., D'Eustachio, P., Ruddle, F.H., Arnheim, N. (1981) Proc. Nat. Acad. Sci. U.S.A., 78, 5744 - 5748

Kunz, W., Petersen, G., Renkawitz-Pohl, R., Glatzer, K.H., Schafer, M. (1981) Chromosoma, 83, 145 - 158

Laskey, R.A., Mills, R.D. (1976) FEBS Letters, 82, 314 - 316

Levis, R., Penman, S. (1978) J. Mol. Biol., 121, 219 - 238

Levis, R., Dunsmuir, P., Rubin, G.M. (1980) Cell, 21, 581 - 588

Long, E.O., Dawid, I.B. (1979a) Nucl. Acids Res., 7, 205 - 215

Long, E.O., Dawid, I.B. (1979b) Cell, 18, 1185 - 1196

Long, E.O., Dawid, I.B. (1980a) J. Mol. Biol., 138, 873 - 878

Long, E.O., Dawid, I.B. (1980b) Ann. Rev. Biochem., 49, 727 - 764

Long, E.O., Rebbert, M.L., Dawid, I.B. (1981a) Proc. Nat. Acad. Sci. U.S.A., 78, 1513 - 1517

Long, E.O., Collins, M., Kiefer, B.I., Dawid, I.B. (1981b) Mol. Gen. Genet., 182, 377 - 384

Long, E.O., Rebbert, M.L., Dawid, I.B. (1981c) Cold Spring Harbor Symp.

Quant. Biol., 45, 667 - 672

Maddern, R.H. (1981) Genet. Res., Camb., 38, 1 - 7

Maden, B.E.H., Tartof, K. (1974) J. Mol. Biol., 90, 51 - 64

Mandal, R.K., Dawid, I.B. (1981) Nucl. Acids Res., 9, 1801 - 1811

Mandel, M., Higa, A. (1970) J. Mol. Biol., 53, 159 - 162

Mange, A.D., Sandler, L. (1973) Genetics, 73, 73 - 86

Maniatis, T., Jeffrey, A., van de Sande, H. (1975) Biochemistry, 14, 3787 - 3794

Maxam, A.M., Gilbert, W. (1977) Proc. Nat. Acad. Sci. U.S.A., 74, 560 - 564

Maxam,A.M., Gilbert, W. (1980) Methods in Enzymol., 65, 499 - 530 (Maldave, K. and Grossman, L., Eds.) Academic Press, N.Y.

McKnight, S.L., Miller,O.L. (1976) Cell, 8, 305 - 314

Meyer, G.F., Hennig, W. (1974) Chromosoma, 46, 121 - 144

Miller, O.L., Beatty, B.R. (1968) Science, 164, 955 - 957

Murray, N.E., Bruce, S.A., Murray, K. (1979) J. Mol. Biol., 132, 493 - 505

Palumbo, G., Caizzi, R., Ritossa, F. (1973) Proc. Nat. Acad. Sci. U.S.A., 70, 1883 - 1885

Pan, J., Elder, J.T., Duncan, C.H., Weissman, S.M. (1981) Nucl. Acids Res., 9, 1151 - 1170

Pardue, M.L., Gerbi, S.A., Echardt, R.A. (1970) Chromosoma, 29, 268 - 290

Pardue, M.L., Dawid, I.B. (1981) Chromosoma, 83, 29 - 44

Parry, D.M., Sandler, L. (1974) Genetics, 77, 535 - 539

Pavlakis, G.N., Jordan, B.R., Wurst, R.M., Vournakis, J.N. (1979) Nucl. Acids Res., 7, 2213 - 2238

Peacock, A.C., Dingman, C.W. (1967) Biochemistry, 6, 1818 - 1827

Peacock, W.J., Lohe, A.R., Gerlach, W.L., Dunsmuir, P., Dennis, E.S., Appels, R. (1977) Cold Spring Harbor Symp. Quant. Biol., 42, 1121 - 1135

Peacock, W.J., Lohe, A.R. (1980) Insect Cytogenetics, Symposia of Royal Entomological Society of London, 10, 1 - 12 (Blackham, R.L., Hewitt, G.M., Ashburner, M., Eds.), Blackwell Publications, London

Peacock, W.J., Appels, R., Endow, S., Glover, D.M. (1981) Genet. Res., Camb., 37, 209 -214

Pellegrini, M., Manning, J., Davidson, N. (1977) Cell, 10, 213 - 224

Philippsen, P., Kramer, R.A., Davis, R.W. (1978) J. Mol. Biol., 123, 371 - 386

Potter, S., Truett, M., Phillips, M., Maher, A. (1980) Cell, 20, 649 - 647

Prensky, W., Steffensen, D.M., Hughes, W.L. (1973) Proc. Nat. Acad. Sci. U.S.A., 70, 1860 - 1864

Procunier, J.D., Tartof, K.D. (1975) Genetics, 81, 515 - 523

Procunier, J.D., Tartof, K.D. (1978) Genetics, 88, 67 - 79

Quagliarotti, G., Ritossa, F.M. (1968) J. Mol. Biol., 36, 57 - 69

Rae, P.M.M., Kohorn, B.D., Wade, R.P. (1980) Nucl. Acids Res., 8, 3491 - 3504

Rae, P.M.M. (1981) Nucl. Acids Res., 9, 4997 - 5010

Rasch, E. (1970) Introduction to Quantitative Cytochemistry, 2, 357 - 397 (Wied, G.L., Bahr, G.F., Eds.) Academic Press, N.Y.

Rasch, E.M., Barr, H.J., Rasch, P.W. (1971) Chromosoma, 33, 1 - 18

Renkawitz, R., Kunz, W. (1975) Chromosoma, 53, 131 - 140

Renkawitz, R., Gerbi, S.A., Glatzer, K.H. (1979) Mol. Gen. Genet., 173, 1 - 13

Renkawitz-Pohl, R., Glatzer, K.H., Kunz, W. (1980) Nucl. Acids Res., 8, 4593 - 4611

Renkawitz-Pohl, R., Matsumoto, L., Gerbi, S.A. (1981) Nucl. Acids Res., 9, 3747 - 3762

Rigby, P.W.J., Dieckmann, M., Rhodes, C., Berg, P. (1977) J. Mol. Biol., 113, 237 - 251

Ritossa, F.M., Spiegelman, S. (1965) Proc. Nat. Acad. Sci. U.S.A., 53, 737 - 745

Ritossa, F.M., Atwood, K.C., Spiegelman, S. (1966) Genetics, 54, 819 - 834

Ritossa, F.M. (1968) Proc. Nat. Acad. Sci. U.S.A., 60, 509 - 516

Ritossa, F.M., Scala, G. (1969) Genetics Suppl., 61, 305 - 317

Ritossa, F., Malva, C., Boncinelli, E., Graziani, F., Polito, L. (1971) Proc. Nat. Acad. Sci. U.S.A., 68, 1580 - 1584

Ritossa, F., Scalenghe, F., Dituri, N., Contini, A.M. (1973) Cold Spring Harbor Symp. Quant. Biol., 38, 483 - 490

Ritossa, F. (1973) Proc. Nat. Acad. Sci. U.S.A., 70, 1950 - 1954

Ritossa, F. (1976) Genetics and Biology of Drosophila melanogaster, 1b, 801 - 846 (Ashburner, M., Novitski, E., Eds.) Academic Press, N.Y.

Roberts, R.J. (1978) Gene, 4, 183 - 193

Roberts, R.J. (1980) Methods in Enzymol., 65, 1 - 18 (Maldave, K., Grossman, L., Eds.) Academic Press, N.Y.

Rohlendorf, B. (1974) The Historical Development of Diptera, 567 - 577 Edmonton, University of Alberta Press

Roiha, H., Miller, J.R., Woods, L.C., Glover, D.M. (1981) Nature, 290, 749 - 753

Roiha, H., Glover, D.M. (1981) Nucl. Acids Res. 9, 5521-5532

Rubin, G.M., Hogness, D.S. (1975) Cell, 6, 207 - 213

Rubin, G.M., Brorein, W.J., Dunsmuir, P., Flavell, A.J., Levis, R., Strobel, E., Toole, J.J., Young,E. (1981) Cold Spring Harbor Symp. Quant. Biol., 45, 619 - 628

Rudkin, G.F. (1965) Genetics, 53, 665 - 681

Samal, B., Worcel.A., Louis, C., Schedl, P. (1981) Cell, 23, 401 - 409

Sandler, L. (1970) Genetics, 64, 481 - 443

Sanger, F., Coulson, A.R. (1978) FEBS Letters, 87, 107 - 110

Schafer,M., Wyman, A.R., White, R. (1981) J. Mol. Biol., 146, 179 - 200

Schafer, U., Kunz, W. (1975) Mol. Gen. Genet., 137, 365 - 368

Sinclair, J., Brown, D.D. (1971) Biochemistry, 10, 2761 - 2769

Smith, G.P. (1973) Cold Spring Harbor Symp. Quant. Biol., 38, 507 - 514

Smith, G.P. (1976) Science, 191, 528 - 535

So, M., Gill, R., Falkow, S. (1975) Mol. Gen. Genet., 142, 239 - 249

Southern, E.M. (1975) J. Mol. Biol., 98, 503 - 517

Spear, B.B., Gall, J.G. (1973) Proc. Nat. Acad. Sci. U.S.A., 70, 1359 - 1363

Spear, B.B. (1974) Chromosoma, 48, 159 - 179

Spradling, A.C., Rubin, G.M. (1981) Ann. Rev. Gen., submittted

Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol., 43, 77 - 90

Tartof, K.D., Perry, R.P. (1970) J. Mol. Biol., 51, 171 - 183

Tartof, K.D. (1971) Science, 171, 294 - 297

Tartof, K.D. (1973) Genetics, 73, 57 - 71

Tartof, K.D., Dawid, I.B. (1976) Nature, 263, 27 - 30

Tartof, K.D. (1979) Cell, 17, 607 - 614

Twigg, A.J., Sherratt, D. (1980) Nature, 283, 216 - 218

Vogt, V.M. (1973) Eur. J. Biochem., 33, 192 - 200

Vogt, V.M. (1980) Methods in Enzymol., 65, 248 - 255 (Maldave, K., Grossman, L., Eds.) Academic Press, N.Y.

Wahl, G.M., Stern, M., Stark, G.R. (1979) Proc. Nat. Acad. Sci. U.S.A., 76, 3683 - 3687

Wellauer, P.K., Dawid, I.B. (1974) J. Mol. Biol., 89, 379 - 395

Wellauer, P.K., Reeder, R.H. (1975) J. Mol. Biol., 94, 151 - 161

Wellauer, P.K., Dawid, I.B., Brown, D.D., Reeder, R.H. (1976 a) J. Mol. Biol., 105, 461 - 486

Wellauer, P.K., Reeder, R.H., Dawid, I.B., Brown, D.D. (1976b) J. Mol. Biol., 105, 487 - 505

Wellauer, P.K., Dawid, I.B. (1977) Cell, 10, 193 - 212

Wellauer, P.K., Dawid, I.B. (1978) J. Mol. Biol., 126, 769 - 782

Wellauer, P.K., Dawid, I.B., Tartof, K.D. (1978) Cell, 14, 269 - 278

Wensink, P.C., Finnegan, D.J., Donelson, J.E., Hogness, D.S. (1974) Cell, 3, 315 - 325

White, R.L., Hogness, D.S. (1977) Cell, 10, 177 - 192

Wild, M.A., Sommer, M. (1980) Nature, 283, 693 - 694

Will, B.M., Bayev, A.A., Finnegan, D.J. (1981) J. Mol. Biol., 153, 897 - 915

Williamson, J. (1976) Genetics and Biology of Drosophila melanogaster, 1b, 667 - 699 (Ashburner, M., Novitski, E., Eds.) Academic Press, N.Y.

Wollenzien, P.L., Youvan, D.C., Hearst, J.E. (1978) Proc. Nat. Acad. Sci. U.S.A., 75, 1642 - 1646

Yagura, T., Yagura, M., Muramatsu, M. (1979) J. Mol. Biol., 133, 533 - 547

Yedvobnick, B., Krider, H.M., Levine, B.I. (1980) Genetics, 95, 661 - 672

Young, M.W. (1979) Proc. Nat. Acad. Sci. U.S.A., 76, 6274 - 6278