MOLECULAR AND CELLULAR BIOLOGY, July 1985, p. 1591–1600
0270-7306/85/071591-10$02.00/0

Vol. 5, No. 7

The 45-kb Unit of Major Urinary Protein Gene Organization Is a Gigantic Imperfect Palindrome

J. O. BISHOP,* G. G. SELMAN, JANET HICKMAN, LINDA BLACK, R. D. P. SAUNDERS, AND A. J. CLARK

Department of Genetics, University of Edinburgh, Edinburgh, EH9 3JN, United Kingdom

Received 17 January 1985/Accepted 29 March 1985

The multigene family which codes for the mouse major urinary proteins consists of about 35 genes. Most of these are members of two distinct groups, group 1 and group 2. The group 1 and group 2 genes are organized in head-to-head pairs within 12 to 15 remarkably uniform chromosomal units or domains about 45 kilobase pairs (kb) in size. The 45-kb units are located on chromosome 4, and many of them are adjacent to each other. We propose that the 45-kb unit is a unit both of organization and of evolutionary change. In this study the homologies within the unit were observed by examining, in an electron microscope, heteroduplex and foldback structures made from cloned major urinary protein genes. These show that the 45-kb unit is a gigantic imperfect palindrome. Each arm of the palindrome contains two regions of inverted symmetry of 9.5 and 4.5 kb separated by a 3-kb nonsymmetrical region. We argue that the nonsymmetrical regions arose by a series of deletion events in the two arms of the palindrome. The center of the 45-kb unit is an 8-kb sequence without inverted symmetry flanked by the 9.5-kb regions, which contain the 4-kb genes and their immediate 5′ and 3′ flanking regions. The junction between adjacent 45-kb units is a 2- to 4-kb sequence without inverted symmetry flanked by the 4.5-kb regions. Some of the 45-kb units are arranged as direct tandem repeats. Others appear to be in inverted orientation with respect to a neighboring unit. Cloned major urinary protein genes show few incidences of the repetitive elements B1, B2, R, and MIF. Two elements, a B1 and an R, may be a constant feature of the 45-kb units. If so, in those cases in which the units are in tandem array, both of these elements will occur with a 45-kb periodicity. A comparison of corresponding parts of different 45-kb units shows that they differ because of a number of deletion or insertion events, particularly in the regions 3′ to the genes.

Mouse major urinary protein (MUP) genes are a family of about 35 related genes clustered near the brown locus on chromosome 4 (2, 3, 16). Many of the genes are pseudogenes (P. Ghazal, A. J. Clark, and J. O. Bishop, Proc. Natl. Acad. Sci. U.S.A., in press). The active genes that are expressed in the liver specify a family of closely related proteins (10, 12).

Our interest in the MUP genes stems from the fact that different members display different patterns of expression. For example, testosterone stimulates the expression of some genes more than others in the livers of female mice (11). Thyroxine and growth hormone induce the expression of different subsets of MUP genes (15). Different subsets are expressed in submaxillary, lachrymal, and mammary glands (22). The variety of these effects and the underlying similarity of the genes lead us to suppose that it may be possible to relate differences in their expression to sequence differences.

We isolated and characterized a number of MUP genes (3, 8, 9). The transcription unit is 4 kilobase pairs (kb) long and contains seven exons and six introns. The coding region extends from exon 1 through to the 5′ end of exon 6. Exon 7 is entirely noncoding. Some mRNA (short mRNA) contains all of exon 6, and different molecules terminate at two different major polyadenylation [poly(A)] sites. In other mRNA molecules (long mRNA) the noncoding region of exon 6 is spliced in at least two different configurations to exon 7 (8; A. Chave-Cox, unpublished data). The ratio of long to short mRNA found in liver and other tissues is 10 or more.

Most of the MUP genes can be classified in two groups, group 1 and group 2 (3). The genes within each group are quite similar (sequence data show that different group 1 genes differ by about 1%; group 2 genes differ by more) but members of different groups differ by about 10% (A. J. Clark, D. Barrett, R. Bingham, P. Ghazal, and J. O. Bishop, unpublished data). The exon 1 sequences of six group 1 genes are nearly identical. The exon 1 sequences of the four group 2 genes that have been studied share a common consensus that is different from the group 1 consensus, and in particular they all carry the same stop codon. Thus, these group 2 genes are all pseudogenes, descended from the same ancestral pseudogene (Ghazal et al., in press).

The group 1 and group 2 genes are part of large 45-kb-long units of DNA organization (Fig. 1). Each unit contains one group 1 gene and one group 2 gene, about 15 kb apart and in divergent transcriptional orientation (i.e., head-to-head configuration) (9). From the results of DNA-DNA hybridization, it was found that not only the two genes but also extensive 5′ and 3′ flanking regions are homologous. Thus, the 45-kb unit is, in effect, a gigantic imperfect palindrome. Some of the 45-kb units are adjacent to each other. Because there is more conservation within group 1 and within group 2 than between the groups, we suggest that the 45-kb unit is the unit not only of organization but also of evolution of the MUP genes. We propose that the MUP locus contains an array of 45-kb units and that all members of the contemporary array are descended fairly recently from a common ancestral 45-kb unit (Ghazal et al., in press).

In this study we describe an electron microscope analysis of heteroduplexes formed between a number of group 1 and group 2 genes and of two types of foldback structures. One foldback structure is developed by a type of clone which spans the middle of a gigantic imperfect palindromic 45-kb unit and includes the 5′ ends of two MUP genes (BS102 and BS109; Fig. 1). The other is developed by clones isolated from the region between two 45-kb units and contains the

* Corresponding author.

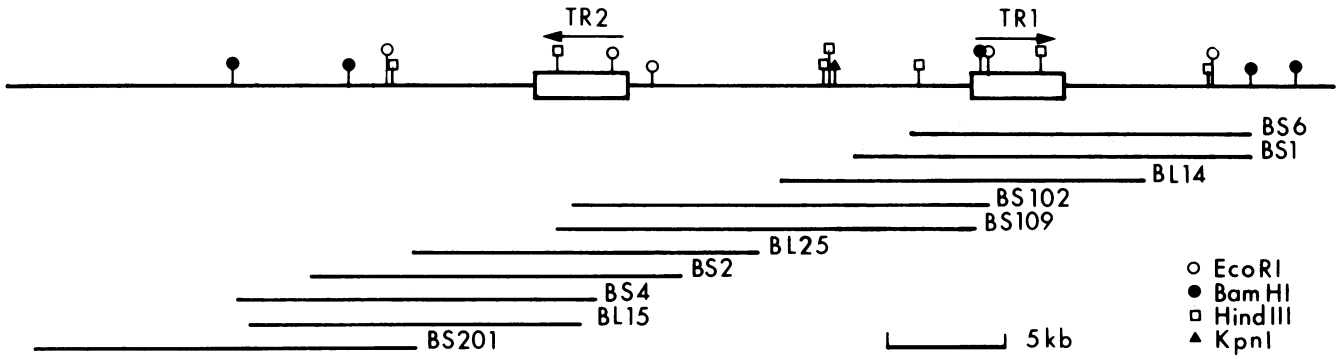

FIG. 1. The 45-kb MUP gene pair and the cloned sequences examined in this study. The top line shows a segment of chromosome 4 containing two MUP transcription units (boxes labeled TR1 and TR2) in head-to-head orientation (7). Only selected restriction sites, useful for orientation, are shown. TR1 and TR2, Transcription units 1 and 2, group 1 and group 2 genes, respectively, transcribed in the directions shown by the arrows. The lines at the bottom of the figure show the relationship of different clones to the chromosome map. Some of the clones (BS1, BS2, BS4) contain deletions (not shown) relative to the map.

inverted 3′ flanking regions of genes in the two adjacent units (BS201; Fig. 1). We also investigated the incidence of four abundant dispersed repetitive elements within the complex.

MATERIALS AND METHODS

Cloned DNA sequences. The isolation of bacteriophage lambda (Charon 4A) clones containing MUP genes and flanking sequences and of the plasmid subclones BS6-3 and BS3-B3 has been described previously (3, 8, 9). Plasmids containing cloned repetitive sequences were those described by Bennett et al. (1). Plasmids pM142, pM134, pM290, and pM225 carry the B2, MIF, R, and B1 elements, respectively.

Electron microscopy. Electron microscopy was carried out as described previously (6).

(i) Heteroduplex formation. The two DNA samples were mixed, denatured in 0.1 N NaOH–20 mM EDTA, neutralized, and annealed, each at a concentration of 3 μg/ml, for 2 h at 27°C in 50% formamide–0.1 M Tris-hydrochloride–50 mM NaCl–10 mM EDTA (pH 8.5).

(ii) Foldback formation. DNA was denatured and neutralized as described above and annealed for 30 min at 27°C at a concentration of 3 μg/ml.

(iii) Spreading. The annealed samples were diluted 22-fold with a hyperphase containing 0.1 M Tris-hydrochloride, 10 mM EDTA (pH 8.5), 50% formamide, and single- and double-stranded markers M13 (6.407 kb) and pAT153 (3.657 kb) and cytochrome *c* was added to 0.2 mg/ml. The hypophase contained 15% formamide, 10 mM Tris-hydrochloride, and 1 mM EDTA. Parlodion-coated grids were stained in uranyl acetate, rotary shadowed with PtPd, and coated with carbon before removal of the Parlodion with ethanol. Strand lengths were measured with a Ferranti Cetec Graphics digitizer.

Localization of repetitive elements. Charon 4A clones were digested with *Hin*dIII-*Bam*HI or *Eco*RI. The fragments were separated on agarose gels, transferred to nitrocellulose filters, and probed with nick-translated plasmid DNA as described previously (7). The final filter wash was with 0.2× SET (0.3 M NaCl, 6 mM Tris-hydrochloride, 0.4 mM EDTA [pH 8.0]).

RESULTS

Heteroduplexes between MUP genes. Heteroduplexes between five pairs of Charon 4A clones were examined (Fig. 2). In all cases the MUP genes in the two clones were orientated in the same direction relative to the arms of the vector DNA. Thus, each heteroduplex molecule contained renatured vector arms of 19.8 and 10.9 kb. Single-stranded loops can be generated in these structures in two ways. (i) Any nonoverlapping region at the end of one of the cloned fragments generates a single-stranded loop at the boundary with the vector arm. (ii) An insertion or deletion in either of the fragments can generate a single-stranded loop within the region of cloned mouse DNA. Single-stranded loops will be formed only if the overlapping regions of the two clones are well matched. When this is not the case, two-stranded loops or D-structures are formed.

Heteroduplexes between two group 1 genes or between two group 2 genes are comparisons between homologous parts of different 45-kb gene-pair units. Heteroduplexes between a group 1 gene and a group 2 gene compare the two inverted 16-kb regions of homology within the 45-kb units.

Heteroduplex data (Fig. 2) were interpreted in conjunction with restriction site mapping data (3, 8, 9). The interpretation is shown in Fig. 3 and the derivation of each interpretation is described in the Appendix.

Group 1 genes (BS1 × BS6). Results of restriction site mapping showed that the group 1 genes and their flanking sequences are very similar. However, insertions or deletions were observed in the nearby flanking regions. MUP gene BS1 appeared to carry a large deletion relative to the other group 1 genes. Heteroduplex mapping (Fig. 2 and 3) confirmed that this is the case. The deletion was 2 ± 0.1 kb in length and was within 0.1 kb of the poly(A) site. Otherwise, BS1 and BS6 formed a duplex over their common length which extended 2.8 kb into the 5′ flanking region and 6.1 kb 3′ to the deletion. The restriction site mapping (8) detected a second small deletion or insertion (about 0.1 kb) which was not detected in the heteroduplexes.

Group 2 genes (BS2 × BL25, BL15 × BS4). Two combinations of group 2 genes were heteroduplexed. BS2 and BS4 have identical restriction site patterns over their common length and are provisionally considered to be different overlapping fragments of the same gene which are however inserted in the vector in the opposite orientation. BS2 was heteroduplexed with BL25, and BS4 was heteroduplexed with BL15. The duplex between BS2 and BL25 extended 2.1 kb into the 5′ flanking region, where it broke down for at least 0.2 kb, the length of the single-stranded region of BS2. In the 3′ flanking region the duplex broke down 1.5 kb from

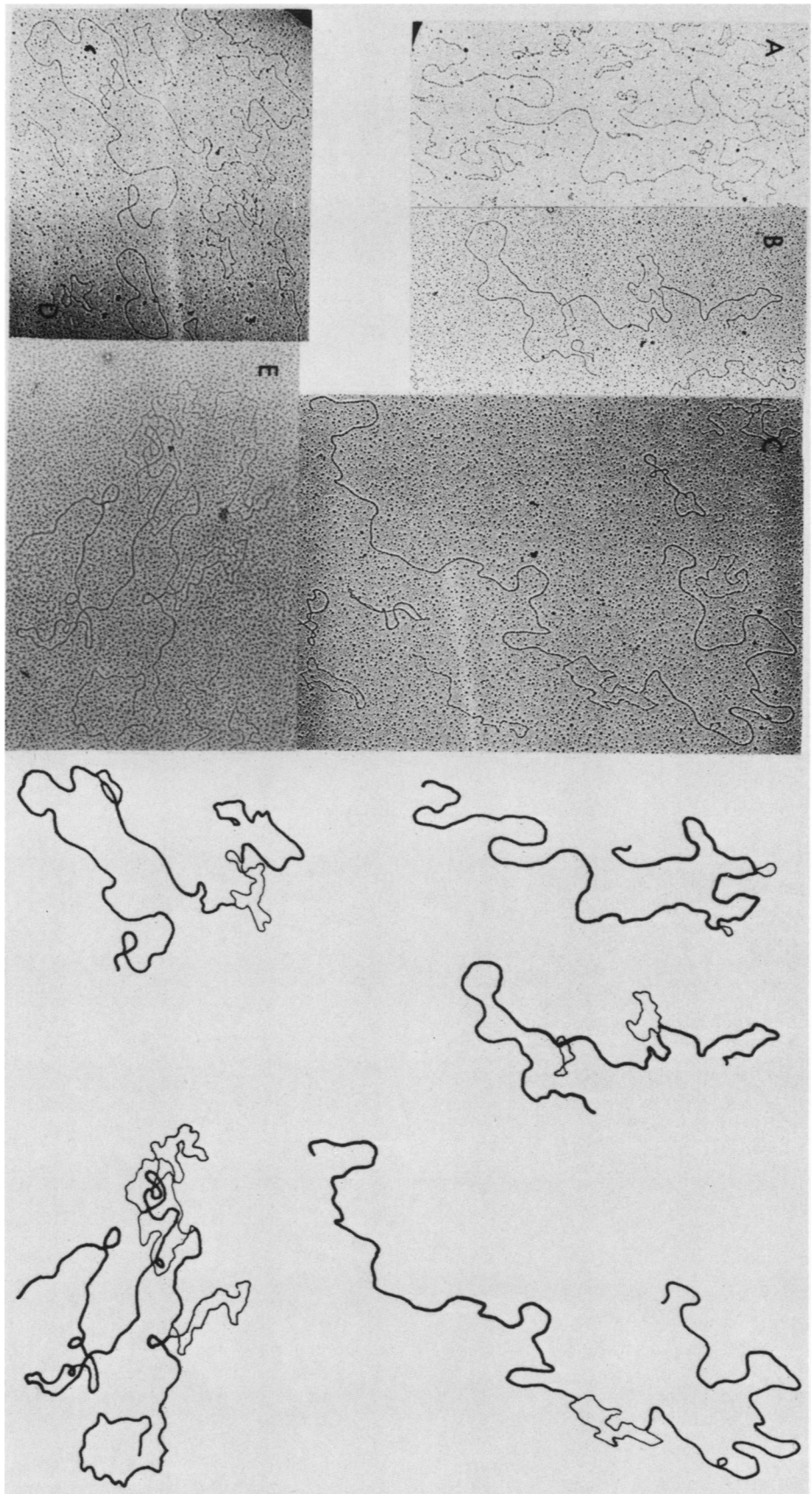

FIG. 2. Heteroduplexes formed by MUP genes and flanking sequences cloned into Charon 4A. (A) BS1 × BS6; (B) BL14 × BS2; (C) BL15 × BS4; (D) BL25 × BS2; (E) BS1 × BS4. Line drawings of the heteroduplexes are provided for clarity.

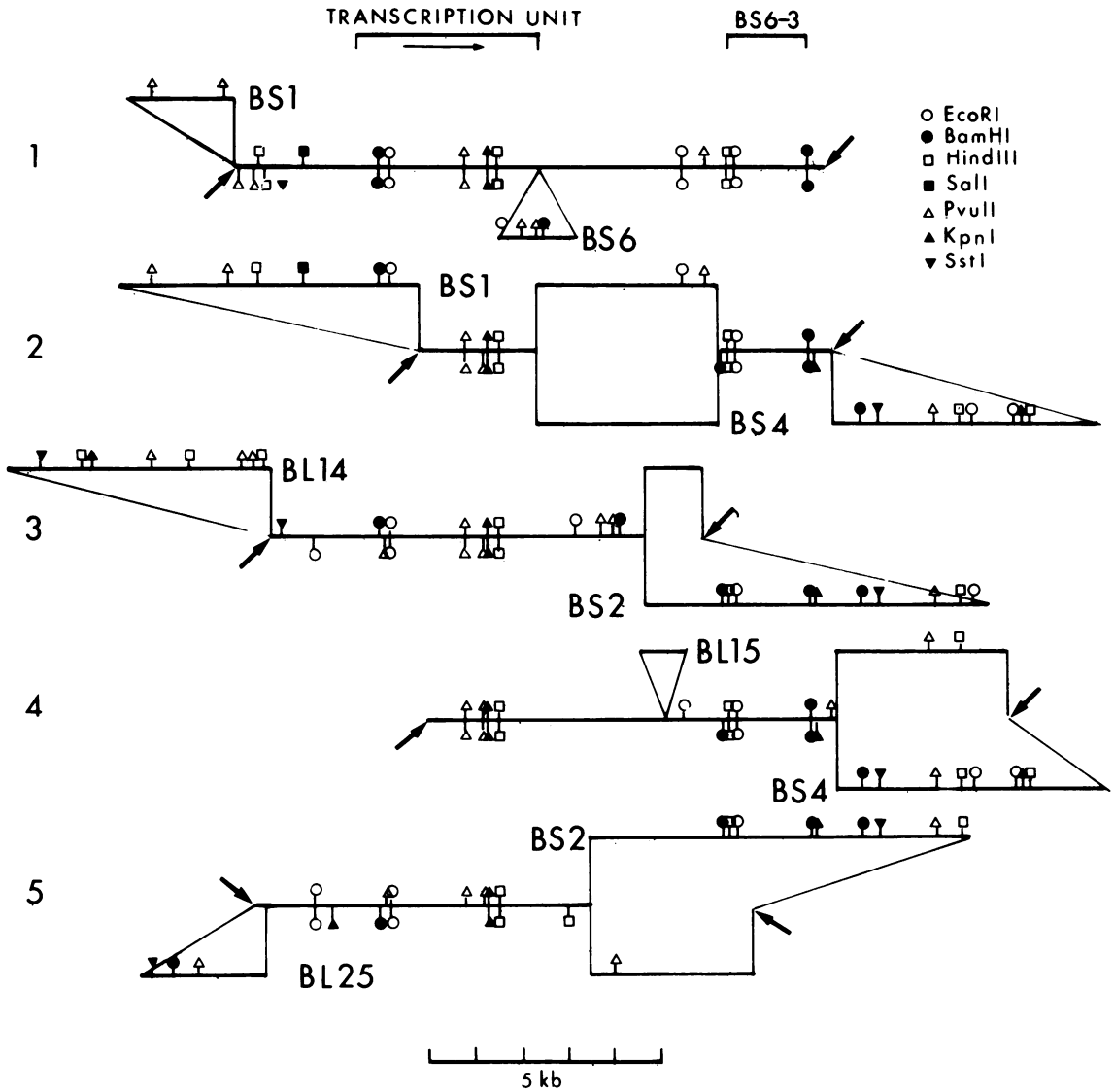

FIG. 3. Interpretation of heteroduplex structures in relation to restriction site maps. DNA is represented by horizontal lines proportional in length to the length of the DNA (see scale). Only the cloned mouse DNA is shown. The lines above and below represent single strands looped out of the duplex, and these are assigned to one or another clone as shown. Diagonal arrows mark the junction points of the mouse DNA with the duplexed arms of the bacteriophage vector. Vertical and diagonal lines join the points of transition between single and double strands. The diagrams are aligned in relation to the ubiquitous *Hin*dIII site 3 kb from the cap site and 1 kb from the poly(A) addition site. BS6-3 is a plasmid subclone of BS6 which hybridizes with all of the *Hin*dIII-*Bam*HI fragments that lie directly below it on the figure. (1) BS1 × BS6 (group 1); (2) BS1 × BS4 (group 1 × group 2); (3) BL14 × BS2 (group 1 × group 2); (4) BL15 × BS4 (group 2); (5) BS2 × BL25 (group 2).

the poly(A) addition site and did not re-form over a further distance of at least 3 kb.

The DNA fragments cloned in BS4 and BL15 terminated at the 5′ end at approximately the same position, close to the middle of the transcription unit. In the 3′ direction the duplex extended 3.9 kb beyond the poly(A) addition site where it was interrupted by a 1-kb deletion or insertion. The duplex then continued for a further 3 kb, followed by at least 3.7 kb of mismatched sequence.

Group 1 × group 2 heteroduplexes (BS1 × BS4, BL14 × BS2). Again, advantage was taken of the opposite orientation of BS2 and BS4 relative to the vector. BS2 (group 2) was heteroduplexed with BL14 (group 1). The duplex extended 2.2 kb into the 5′ flanking region to the boundary between mouse and Charon 4A DNA in BS2. In the 3′ direction, the duplex extended 2.2 kb beyond the poly(A) site and then broke down for at least 1.2 kb.

Because the 5′ end of the DNA cloned in BS4 lies within the transcription unit, the BS1 × BS4 heteroduplex provided information only about 3′ sequences. As in the case of the BS1 × BS6 heteroduplex, homology extended downstream to a point close to the poly(A) site. Homology was then absent for a distance of about 4.2 kb in the immediate 3′ flanking region. The duplex was formed at this point and continued downstream for at least 2.3 kb.

The group 1 × group 1 and group 2 × group 2 results reflect relationships between repeats of the 45-kb unit. On the other hand, the group 1 × group 2 results reflect relationships between the two arms of the unit. They show homologous stretches of DNA extending at least 2.9 kb into

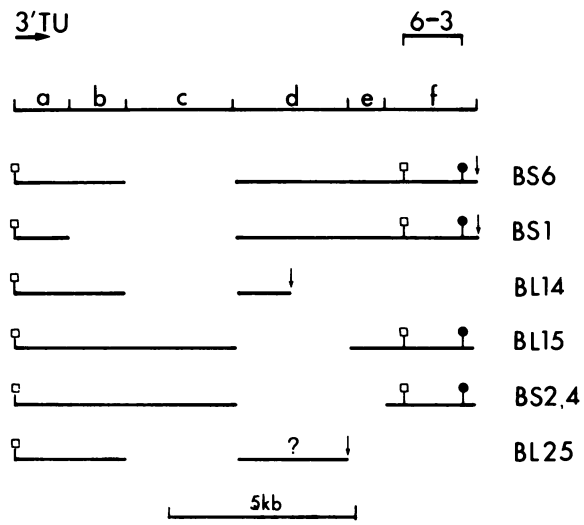

FIG. 4. Heteroduplex and mapping relationships of 3′ flanking regions in terms of deletions or insertions or both. The map extends in the 3′ direction from the *Hin*dIII site which lies 1 kb 5′ to the poly(A) addition site. 3′TU marks the 3′ end of the transcription unit; 6-3 denotes subclone BS6-3 which hybridizes with the conserved *Hin*dIII-*Bam*HI site shown in the figure. The 3′ ends of four clones are marked with arrows. Two clones extend further to the right, but these regions have not been heteroduplexed. The question mark indicates a region of BL25 which has not been shown directly to be homologous to segment d. Restriction site symbols are defined in Fig. 3.

the 5′ flanking regions of symmetrically opposed genes (BL14 × BS2).

The heteroduplexes showed a number of different configurations in the 3′ flanking region. Interpretation of these is made easier because of identities in restriction site patterns. In particular, a 1.7-kb sequence between a *Hin*dIII site and a *Bam*HI site, located 4 to 5.5 kb 3′ to the poly(A) addition site in different genes, was present in many genes of both groups. The conservation of this region was established by probing Southern blots of the cloned sequences with BS6-3, a plasmid subclone of clone BS6 (see Fig. 3).

Coextensive nonduplexed regions may be due to nonoverlapping deletions or insertions, to parallel dissimilar insertions, or to mutational sequence changes. The longest loop due to unmatched single-stranded mouse sequences was observed in the BS1 × BS4 heteroduplex and is known to be at least partly due to deletion or insertion of segments b and e (Fig. 3 and 4). Thus, it is worthwhile to ask whether the available data can be explained solely in terms of deletions or insertions or both.

Figure 4 shows an interpretation of the mapping and heteroduplex data which is based on the assumption that the contemporary genes differ from their common ancestral gene mainly because of a series of deletion or insertion events or both. The figure shows that six group 1 and group 2 genes can, in fact, be related through a series of deletions to a hypothetical ancestral gene. This form of presentation was adopted for reasons of clarity. Some or all of the proposed deletion events may have been insertions instead.

Foldback structures formed by clones containing flanking sequences in opposite orientation. The two MUP genes in the 45-kb unit are about 15 kb apart. As a consequence, it was possible to isolate two clones, BS102 and BS109, that contain the intervening DNA together with parts of the 5′ ends of both transcription units. Because of the head-to-head arrangement of the genes and homologies both between genes and between flanking regions, hairpin-loop foldback structures are formed when these clones are denatured and annealed briefly at a low concentration. Both clones formed structures with 8-kb loops, double-stranded stems, and single-stranded tails (Fig. 5). The stem contained the 5′ ends of the genes and the homologous 5′ flanking sequences. The tails were the vector arms together with the 3′ end of the longer gene fragment. Each loop was a region between the

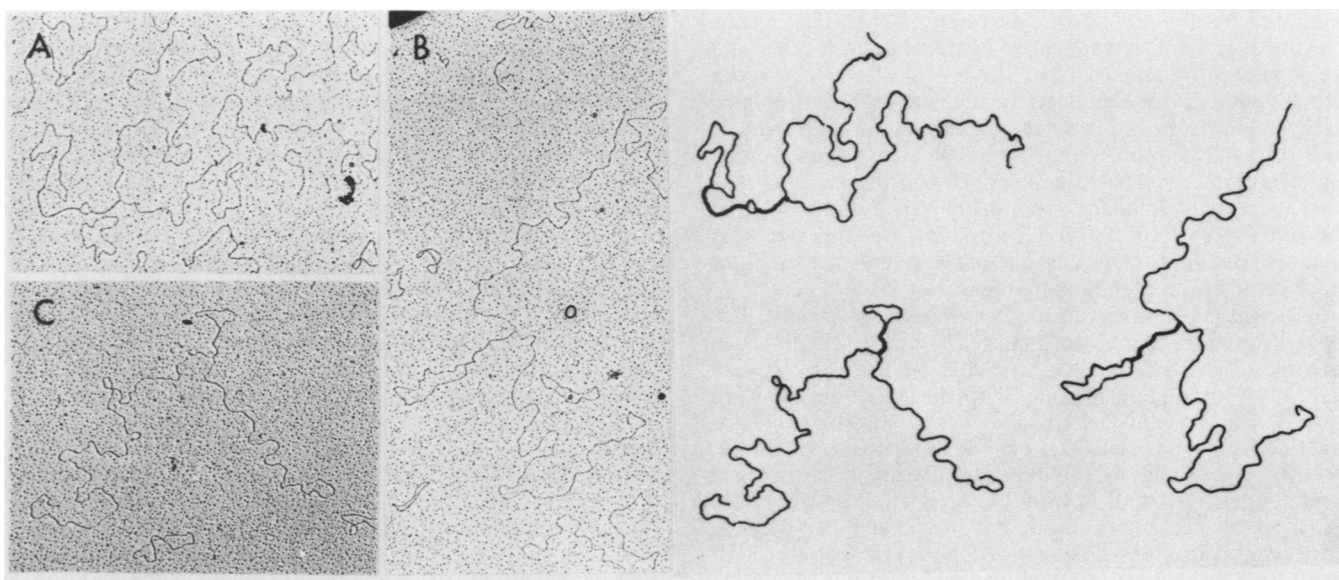

FIG. 5. Foldback structures. (A) BS109 showing a stem with a single bubble; (B) BS109 showing a stem structure without a bubble; (C) BS201. Line drawings of the heteroduplexes are provided for clarity.

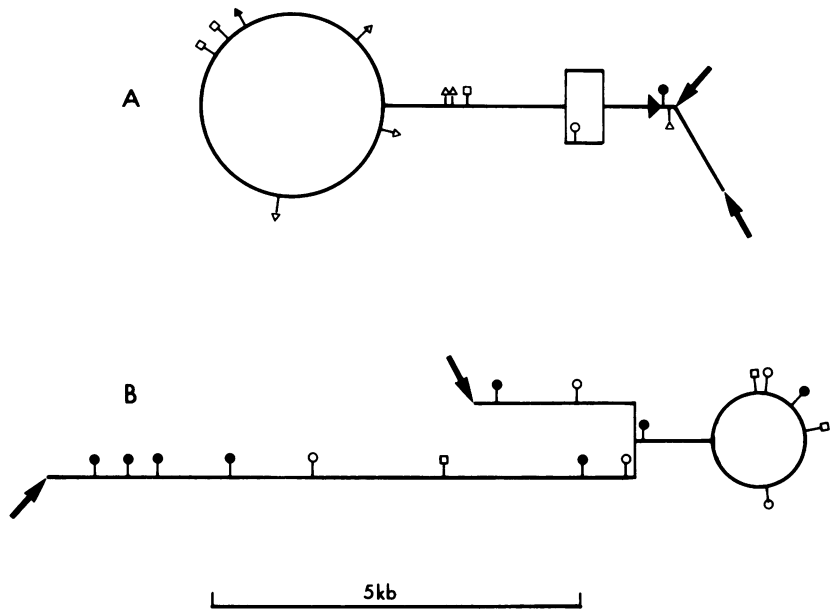

FIG. 6. Interpretation of 5′ and 3′ foldback structures. Hairpin loops are shown as circles. Duplex stem structures are drawn to the same scale as horizontal lines. The proximal ends of the bacteriophage arms are shown as arrowheads. (A) BS102 and BS109 (5′ hairpin loop). The bubble structure in the stem is drawn to scale in the horizontal direction. The duplexed 5′ ends of the two transcription units are shown as an arrowhead across the line which represents the stem. The nonoverlapping mouse DNA at one end of the cloned sequence is shown as a diagonal line. (B) BS201 (3′ hairpin loop). The nonduplexed single strands adjacent to the bacteriophage arms are shown as horizontal lines to the same scale as the duplex stem. Restriction site symbols are defined in Fig. 3.

two 5′ flanking regions which did not contain sequences with inverted symmetry.

The structures formed by the two clones were very similar. Allowing for the amount of overlap at the 5′ end of the transcription unit, the length of the 5′ flanking region of homology is 3.50 kb. The duplex in this region was continuous in 10 molecules. In 16 molecules it was interrupted by a two-stranded bubble upstream of the cap site. Bubbles also occurred infrequently at two other positions. All of these presumably represent regions of lower homology or high A+T content or both.

Some of the 45-kb units are arranged in tandem and are separated by about 4 kb of intervening DNA. This means that the 3′ ends of two genes are opposed to each other and are separated by about 26 kb. The length of DNA between two 3′ ends is too great to be cloned in bacteriophage lambda. However, we isolated clones that contained the 3′ ends of two 3′ flanking regions together with an intervening region without inverted symmetry. One of these clones was used to prepare foldback structures (Fig. 5 and 6). These showed a 4-kb loop and a 1.1-kb stem. Beyond the stem nonduplexed mouse DNA extended for at least the length of the shorter of the two flanking regions of DNA.

Few common repetitive elements within the 45-kb unit. The clones shown in Fig. 1 and additional clones (Figure 7 in reference 7) (BS220, BS203, BS201, BS209, BS222, and BS224, all containing regions 3′ to the MUP genes) were probed with plasmid DNA probes containing the abundant repetitive sequences B1 (17), B2 (18), R (5, 14), and MIF (4, 19, 20). According to the reported frequencies of these elements, B1, B2, and R would be expected, on a random basis, to be present in about five, two, and two copies, respectively, in a 45-kb region of DNA. One copy of MIF could be expected per 100 kb. In fact, only two repetitive elements were consistently observed. A B1 element was present 5′ to the group 1 genes in three of three cases. Similarly, an R element was found near the 3′ junction sequence in seven of seven cases. A B1 element was found 3′ to one of three group 1 genes, a B2 element was adjacent to R in two of seven cases, and an MIF element was adjacent to R in three of seven cases. These results permit two conclusions: (i) the 45-kb unit shows a low incidence of occurrence of the more common repetitive elements, compared with some other murine gene arrays, and (ii) repetitive elements contribute to the variation between members of the array.

The R element is located roughly midway between the 3′ ends of the genes. Fig. 7A shows the result of probing homologous 2-kb *Eco*RI fragments of six clones with a cloned R element. The element was found both in clones which are known or believed to be parts of the 45-kb units (Fig. 7, lanes 1 to 4), as well as at a homologous site in clones (9) with a distinctly different organization (Fig. 7, lanes 5 and 6). The same fragments were also probed with BS3-B3, a 1.7-kb *Hin*dIII fragment of BS3 which overlaps the 2-kb *Eco*RI fragment by 0.7 kb (9). Because this fragment is nonrepetitive (9), it evidently does not overlap the R element.

Lueders and Paterson (19) have demonstrated heterogeneity among R elements. Figure 7A shows that when the R elements associated with the MUP genes are probed with the cloned R element, a wide variation in signal strength is observed. Restriction site mapping (9) shows that BS220 and BS203 (Fig. 7A, lanes 1 and 2) are closely related, as are BS201 and BS209. The results of hybridization of the nonrepetitive BS3-B3 probe is consistent with this relationship because the two fragments of each pair gave signals of similar strength. The R element probe, however, gave a very strong signal with the *Eco*RI fragment of BS201, but it gave a weak signal with the homologous fragment of BS209, the other member of the second pair. It seems unlikely that different R elements were separately inserted at the same

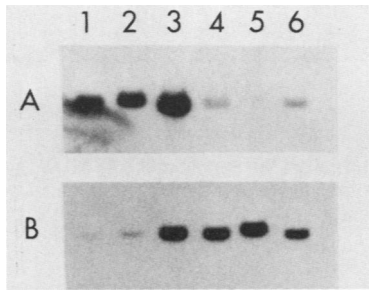

FIG. 7. Hybridization of a cloned R element with the flanking regions of cloned MUP genes. Six Charon 4A clones were digested with *Eco*RI, run in parallel on an agarose gel, and transferred to a nitrocellulose filter. The filter was probed first with the R element probe pM290 (1), exposed, washed to remove the probe, probed with BS3-B3 (7), and again exposed. The regions of the autoradiograms corresponding to a 2-kb *Eco*RI fragment are shown. (A) pue290; (B) BS3-B3. Lane 1, BS220; lane 2, BS203; lane 3, BS201; lane 4, BS209; lane 5, BS222; lane 6, BS224.

place in different units. It is more likely that the R elements were amplified along with the MUP genes in the array. If so, they might be expected to evolve along with the adjoining unit or units in the manner of the sequence detected with the BS3-B3 probe. Instead, there is a very large difference between BS201 and BS209. This might be due either to the deletion of part of the R element from BS209 or to the conversion of either element by a dissimilar R element. The size conservation of the 2-kb *Eco*RI fragment argues in favor of the latter alternative.

DISCUSSION

We estimate that between 26 and 30 of the 35 MUP genes are contained in 13 to 15 45-kb units. We previously defined the 45-kb units by restriction site mapping and Southern blotting with subcloned probes (9). These methods have limitations. Restriction sites can be absent in those cases in which overall similarity exists, or they can be displaced by deletions or insertions. Probing methods are limited by the availability of probes and suitable restriction sites in the target DNA. Thus, it is often difficult or impossible to define precisely regions of homology by these methods. In this study we examined the same units by measuring duplex and single-stranded regions with an electron microscope. Heteroduplexes may break down to single strands in regions which preserve a limited similarity. Consequently, regions of homology may go undetected. However, in the experiments described here there was no conflict with the previous results.

Regions of symmetry and asymmetry within the 45-kb palindrome. Data obtained with an electron microscope show two pairs of regions of inverted symmetry within the 45-kb unit interspersed with nonsymmetrical regions (Fig. 8). The results are consistent with the mapping data described previously, but they give a more precise definition of the boundaries between the different regions. In Fig. 8 is shown a 45-kb unit containing BS25 and its flanking sequences (group 2) and BS6 and its flanking sequences (group 1) and elements of BS102, BS201, and BS225. This combination gives the configuration of a typical 45-kb unit, although it is unlikely that any unit takes this precise form.

The main symmetrical components are the group 1 and group 2 genes and their immediate flanking regions, which together make up a pair of inverted symmetrical sequences 9.5 kb long. A second pair of symmetrical sequences is found 3′ to the transcription units. These regions (C1 and C2) are defined by the electron microscope studies of heteroduplexes and of the hairpin stem of the BS201 foldback structure (Fig. 6 and 7), as well as by results of previous mapping experiments (probes c and i of reference 7). The 3′ hairpin stem structure formed by BS201 (Fig. 5 and 6) was only 1.1 kb long. We interpret this as being due to an insertion of at least 2.2 kb in C1. This interpretation is supported by restriction site and probe homologies with other clones (9).

The nonsymmetrical 5′ and 3′ regions (Fig. 8, 5L and 3L) may never have possessed inverted symmetry and instead may have arisen from nonhomologous sequences distantly flanking one of the two genes that were first brought together by inversion. On the other hand, it seems likely that the nonsymmetrical regions A1 and A2 (Fig. 8) arose at some time within regions that were previously continuously homologous.

There are two main ways in which these regions may have diverged: by the accumulation of point mutations and small deletions and insertions to such an extent that an ancestral homology can no longer be detected, or by the insertion or deletion of relatively long DNA sequences or both. Although the former possibility cannot be excluded at this time, it seems to be less likely because the divergence between

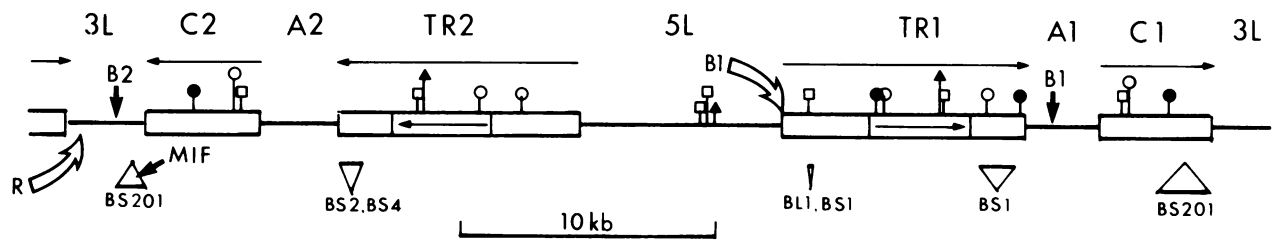

FIG. 8. Sequence homologies within and between 45-kb gene pair units. The 45-kb unit shown is assembled from BS25 (group 2), BS6 (group 1), BS102, and BS225. Regions of inverted symmetry are shown as boxes marked by arrows. TR1 and TR2, Group 1 and group 2 transcription units, respectively, and their 5′ and 3′ flanking regions deduced from heteroduplexes and 5′ foldback structures (Fig. 3 and 6); C, regions containing the conserved sequence that hybridizes with probe BS6-3, further defined by heteroduplexes (Fig. 3). Continuous horizontal lines show regions without inverted symmetry as defined by electron microscopy. These regions are frequently similar or identical in different repeats of the 45-kb unit. A, Regions that interrupt the arms of the palindrome; 5L, the 5′ foldback loop; 3L, the 3′ foldback loop. A complete 3′ loop and the end of an adjacent palindrome are shown at the left end of the diagram. Known deletions or insertions are shown by triangles. The approximate positions of four repetitive elements are marked by arrows. Open arrows mark the two elements that are present in every case examined. Closed arrows mark elements that are present only in some cases. Only selected restriction sites are shown. Restriction site symbols are defined in Fig. 3.

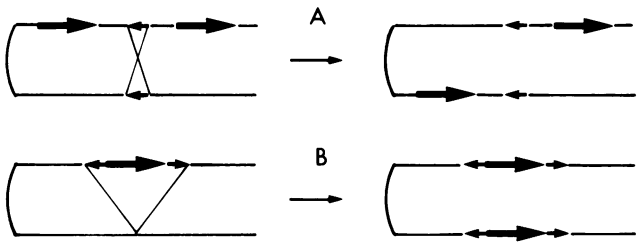

FIG. 9. Mechanisms of inversion. Large arrows represent ancestral MUP genes and flanking sequences. Small arrows represent repetitive (A) or transposable (B) elements. (A) Inversion by reciprocal exchange; (B) inversion by transposon-mediated replicative transposition.

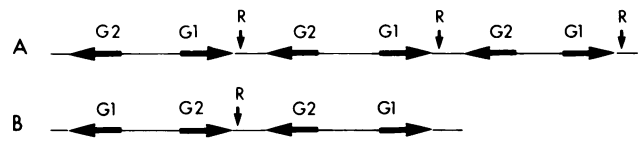

FIG. 10. Alternative arrangements of MUP genes. G1, G2, Group 1 and group 2 genes and flanking sequences, respectively; R, repetitive element. (A) Direct tandem repetition; (B) pair of inverted 45-kb units.

group 1 and group 2 genes is only about 10% in introns, exons, and flanking sequences (Clark et al., unpublished data). The more likely explanation that the differences are due to large deletions or insertions or both (Fig. 4) is supported circumstantially by the occurrence of known deletions or insertions between different 45-kb units (Fig. 8; see below).

Among the inverted symmetrical regions, only the group 1 genes are known to be functional. Because they are functional, they may be presumed to have been conserved by selection against deleterious changes. Although many and perhaps all of the group 2 genes are pseudogenes, all descended from an ancestral pseudogene (Ghazal et al., in press), it is possible that the pseudogene mutation occurred fairly recently and that before the mutation occurred they were conserved in the same way. Again, it is possible that they are not in fact pseudogenes. Instead, they may be transcribed differently from the group 1 genes, or their transcripts may be processed differently. However, this possibility is not supported by the fact that the putative promoter regions and splice junctions of the group 1 and group 2 genes are very similar (Clark et al., unpublished data). On the other hand, the symmetrical homology between the group 1 and group 2 genes and their flanking regions may be a chance phenomenon, due to the chance that inactive group 2 genes have not been deleted or otherwise altered whereas adjacent regions have. The conservation of region C is open to the same alternative explanations.

Differences between the 45-kb units. The exonic nucleotide sequences of four group 1 genes differ from each other by about 1% (Clark et al., unpublished data). The exon 1 sequences of five of six group 1 genes were identical, and the sixth differed in two positions. The exon 1 sequences of four group 2 genes differed by an average of 5%. In addition, the signal peptide regions of the group 2 genes were all different in length (Ghazal et al., in press). Because the difference between a group 1 and a group 2 gene is about 10% in both exons and introns, it is clear that the two genes in the arms of a 45-kb palindrome differ from each other more than do the corresponding gene regions of different palindromes. The comparison of different 45-kb palindromes by mapping and probing (9) and by electron microscopy shows that some differ from others by a limited number of substantial deletions or insertions (Fig. 8).

The 2-kb deletion or insertion event between BS1 and BS6 is probably a deletion. BS6 is the more common type of group 1 MUP gene and shares the sequence in question with BL14 and other group 1 genes. The 2-kb sequence is found in group 2 genes, as shown by the BL14 × BS2 heteroduplex

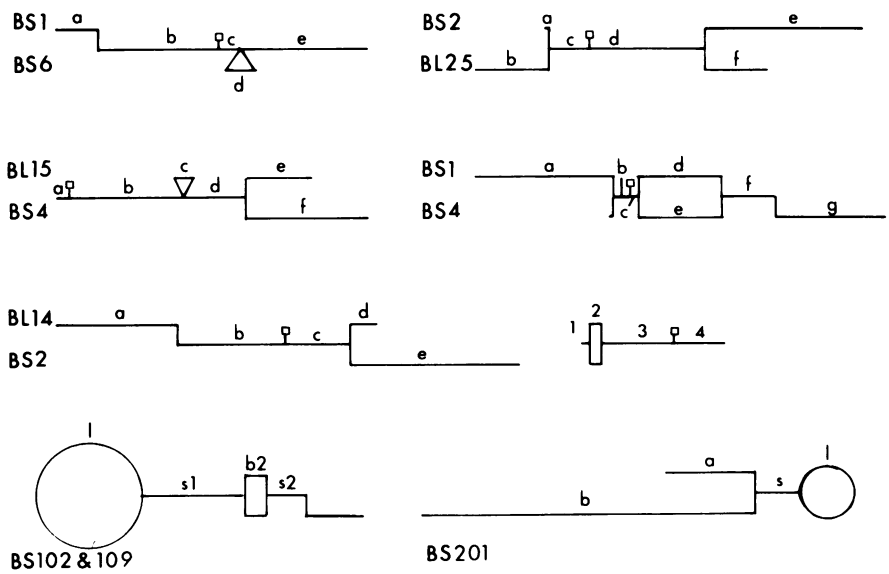

FIG. A1. Heteroduplex and foldback structures. In the diagrams of heteroduplexes, horizontal lines show the lengths of single- or double-stranded regions of the cloned mouse DNA. Boundaries between the two are linked by vertical and diagonal lines. All diagrams show the ubiquitous *Hin*dIII site (□), where present. The estimated lengths of regions a through g and their derivation are tabulated. The central (duplex) region of the BL14 × BS2 heteroduplex has been redrawn to show the dimensions of a bubble that was observed in 5 of 19 structures. The same conventions are followed in the diagrams of foldback structures, except that the single-stranded loops are shown as circles.

(Fig. 3). Thus, it is likely that the sequence was present in the common ancestor of the group 1 and group 2 genes and was deleted from BS1 subsequent to the divergence of the two groups (and presumably subsequent to the inversion that created the 45-kb unit). It is not clear whether the deletion or insertion between BL15 and BS4 is a deletion in BS4 or an insertion into BL15. The group 2 genes differ from each other more than do the group 1 genes (3; Ghazal et al., in press), and there is no restriction site consensus in this region (Fig. 3, BL15 × BS4 and BS2 × BL25). The presence or absence of the sequence adjacent to the R element, which is marked as a deletion in BS2 (Fig. 8), correlates with the presence or absence of the MIF element and may relate to the transposition or excision of MIF.

Origin and arrangement of 45-kb palindromes. The inversion event that brought the 45-kb palindrome into existence could have occurred in at least two ways, both of which involve repeated elements. First, recombination between inverted repeat elements produces an inversion of the intervening DNA. This explanation requires that there be a preexisting array of at least two genes (and flanking regions) in direct orientation and requires the presence of two copies of a repeat element in inverted orientation. To explain the long nonsymmetrical region between the genes (5L in Fig. 8), at least one of the elements would have to be some distance from one of the tandemly repeated genes (Fig. 9A). Second, the replicative transposition of bacteriophage Mu and other transposable elements to a second point on the same chromosome can produce an inversion of the transposon (13, 21). This might have occurred if a MUP gene were bracketed by inverted repeat sequences with insertion sequence-like properties. In such an event only a single gene would have to be present before the inversion event. The two copies of the transposable element would be duplicated along with the gene and its flanking sequences (Fig. 9B).

We failed to detect pairs of repetitive elements at appropriate positions within the 45-kb palindrome. Either of the two sequences that are consistently present (R and B1) might be the remaining element of such a pair, with the other having been lost in the meantime, an element that we did not investigate may have been involved, or the mechanism may not have been one of those proposed.

Previously (9), we have pointed out that because the 45-kb units have polarity, they may be arranged either as direct or as inverted repeats. The location of the R elements in different 3′ nonsymmetrical regions (3L in Fig. 8) suggests that both types of arrangements exist. If the units were all arranged tandemly, we would expect region 3L always to be oriented in the same way with respect to the neighboring group 2 arm (Fig. 10A). The 3L elements of BS220 and BS203 are orientated in this way. However, the 3L regions of BS201 and BS209 are oppositely oriented with respect to a probable group 2 neighbor. This suggests that the inverted arrangement of 45-kb units shown in Fig. 10B also exists.

The uniformity of the 45-kb units argues for their concerted evolution by interchromosomal or intrachromosomal recombination or gene conversion (9; Ghazal et al., in press). The two types of gene arrangement (Fig. 10) would have similar consequences in interchromosomal exchanges. In intrachromosomal exchanges, they have different consequences that have already been well documented. If we assume that exchanges occur only between 45-kb units in the same chromosomal orientation, then interchromosomal unequal crossing-over leads to the expansion and contraction of the array more or less irrespective of the arrangement of the units (23). Intrachromosomal exchange between directly oriented units leads to the deletion of the intervening DNA. Between oppositely oriented units it leads to inversion of the intervening DNA (21). Pure gene conversion events need produce none of these chromosomal rearrangements. Any or all of these mechanisms may contribute to the homogenization of the array of 45-kb palindromes.

TABLE A1. Derivation of the interpretation of heteroduplex data with restriction site mapping data[a]

Heteroduplex (*n*)	Segment	EM	Mapping distance (kb)
BS1 × BS6 (11)	a	2.1 ± 0.1	2.1
	b		5.8
	b + c	6.8 ± 0.03	
	d[b]	2 ± 0.1	1.6[c]
	e		6.1
BS2 × BL25 (1)	a	0.2	
	b	3.5	
	a + c		5.2
	b + c		8.7[c]
	c + d	7.6	
	e	7.5	
	f	3	
BL15 × BS4 (11)	a		1.5[c]
	c[d]	0.9 ± 0.17	
	d	3.1 ± 0.27	
	e[d]	3.7 ± 0.24	
	f[d]	5.8 ± 0.32	
	All BL15		14.0
	All BS4		14.8
BS1 × BS4 (7)	a	6.6 ± 0.53	6.3
	b		1.6
	b + c	2 ± 0.23	
	d[e]	4.1 ± 0.31	
	e[e]	4.4 ± 0.27	
	f	2.3 ± 0.12	
	g	5.4 ± 0.42	
BL14 × BS2 (19)	a	5.9 ± 0.1	6[c]
	b		5.2
	b + c	8.3 ± 0.2	
	d	1.2 ± 0.16	
	e	7.9 ± 0.2	
	1	0.1	
	2	0.6[f]	
	3	4.5	
	4	3.1	

[a] From the electron micrographs shown in Fig. 2 and 5.
[b] Assigned to BS6 by mapping.
[c] Mapping distance calculated as the difference between two measurements (e.g., vector arm and end fragment minus vector arm).
[d] c was assigned to BL15 by using mapped distances between homologous sites. e and f were then assigned by using the lengths of the cloned sequences.
[e] Arbitrary assignment.
[f] $n = 5$.

ACKNOWLEDGMENTS

Much of the electron microscope analysis was done by R.D.P.S. (1982), L.B. (1983), and J.H. (1984) in fulfillment of the B.Sc. degree. We are grateful to J. Jacob, P. Highton, and Pam Beattie for help with these analyses.

J.O.B. and A.J.C. were supported by the Cancer Research Campaign and the Medical Research Council.

APPENDIX

Heteroduplexes. Measurements (± standard error) are related to the position of a ubiquitous *Hin*dIII site which lies 3 kb from the cap site and 1 kb from the poly(A) site (Fig. A1). Table A1 shows the

TABLE A2. Derivation of the interpretation of foldback structure data

Clone	Segment	EM (kb ± SE)
BS102	l	7.4 ± 0.21 (14)[a]
	s1	2.6 ± 0.22 (6)
	b2	0.4 ± 0.18 (6)
	s2	1.1 ± 0.2 (6)
	s1 + b2 + S2	4.2 ± 0.43 (14)
BS109	l	8.0 ± 0.24 (17)
	s1	2.3 ± 0.11 (10)
	b2	0.6 ± 0.08 (10)
	s2	0.8 ± 0.08 (10)
	s1 + b2 + S2	3.7 ± 0.16 (10)
BS201	l	4.0 ± 0.09 (10)
	s	1.1 ± 0.10 (10)
	a[b]	2.2 ± 0.31 (6)
	b[b]	8.0 ± 0.51 (6)

[a] Numbers in parentheses are the number of measurements.
[b] Estimated by subtracting the lengths of the vector arms from the measured lengths of the tails of the hairpin loop structures.

derivation of the interpretation of heteroduplex data with restriction site mapping data.

Foldback structures. Table A2 shows the derivation of the interpretation of foldback structure data with restriction site mapping data.

LITERATURE CITED

1. **Bennett, K. L., R. E. Hill, D. F. Pietras, M. Woodworth-Gutai, C. Kane-Haas, J. M. Houston, J. K. Heath, and N. D. Hastie.** 1984. Most highly repeated dispersed DNA families in the mouse genome. Mol. Cell. Biol. **4**:1561–1571.
2. **Bennett, K. L., P. A. Lalley, R. K. Barth, and N. D. Hastie.** 1982. Mapping the structural genes coding for the major urinary proteins in the mouse: combined use of recombinant inbred strains and somatic cell hybrids. Proc. Natl. Acad. Sci. U.S.A. **79**:1220–1224.
3. **Bishop, J. O., A. J. Clark, P. M. Clissold, S. Hainey, and U. Francke.** 1982. Two main groups of mouse major urinary protein genes, both largely located on chromosome 4. EMBO J. **1**:615–620.
4. **Brown, S. D. M., and G. A. Dover.** 1981. The organisation and evolutionary progress of a dispersed repetitive family in widely separated rodent families. J. Mol. Biol. **150**:441–446.
5. **Cheng, S. M., and C. L. Schildkraut.** 1980. A family of moderately repetitive sequences in mouse DNA. Nucleic Acids Res. **8**:4075–4090.
6. **Chow, L. T., J. M. Roberts, J. B. Lewis, and T. R. Broker.** 1977. A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. Cell **11**:819–836.
7. **Clark, A. J., P. M. Clissold, R. Al-Shawi, P. Beattie, and J. O. Bishop.** 1984. Structure of the mouse major urinary protein genes: differential splicing configurations in the 3'-non-coding region. EMBO J. **3**:1045–1052.
8. **Clark, A. J., P. M. Clissold, and J. O. Bishop.** 1982. Variation between mouse major urinary protein genes isolated from a single inbred line. Gene **18**:221–230.
9. **Clark, A. J., J. Hickman, and J. O. Bishop.** 1984. A 45-kb domain with two divergently orientated genes is the unit of organisation of the murine major urinary protein genes. EMBO J. **3**:2055–2064.
10. **Clissold, P. M., and J. O. Bishop.** 1982. Variation in mouse MUP genes and the MUP gene products within and between inbred strains. Gene **18**:211–220.
11. **Clissold, P. M., S. Hainey, and J. O. Bishop.** 1984. Messenger RNAs coding for mouse major urinary proteins are differentially induced by testosterone. Biochem. Genet. **22**:379–387.
12. **Finlayson, J. S., M. Potter, C. S. Shinnick, and O. Smithies.** 1974. Components of the major urinary protein complex of inbred mice: determination of amino-terminal sequences and comparisons with homologous components from wild mice. Biochem. Genet. **11**:325–335.
13. **Harshay, R. M., and A. L. Bukhari.** 1981. A mechanism of DNA transposition. Proc. Natl. Acad. Sci. U.S.A. **78**:1090–1094.
14. **Heller, R., and N. Arnheim.** 1980. Structure and organization of the highly repeated and interspersed 1.3kb EcoRI-BglII sequence family in mice. Nucleic Acids Res. **8**:5031–5042.
15. **Knopf, J. L., J. F. Gallagher, and W. A. Held.** 1983. Differential, multihormonal regulation of the mouse major urinary protein gene family in the liver. Mol. Cell. Biol. **3**:2232–2240.
16. **Krauter, K., L. Leinwald, P. D'Eustachio, F. Ruddle, and J. Darnell.** 1982. Structural genes of the mouse major urinary protein are on chromosome 4. J. Cell. Biol. **94**:414–417.
17. **Krayev, A. S., D. A. Kramerov, K. G. Skryabin, A. P. Ryskov, A. A. Bayev, and G. P. Giorgiev.** 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res. **8**:1202–1215.
18. **Krayev, A. S., T. V. Markusheva, D. A. Kramerov, A. P. Ryskov, K. G. Skryabin, A. A. Bayev, and G. P. Giorgiev.** 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res. **10**:7461–7475.
19. **Lueders, K. K., and B. M. Paterson.** 1982. A short interspersed repetitive element found near some mouse structural genes. Nucleic Acids Res. **10**:7715–7729.
20. **Meunier-Rotival, M. P., P. Soriano, G. Cuny, F. Strauss, and G. Bernardi.** 1982. Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. Proc. Natl. Acad. Sci. U.S.A. **79**:355–359.
21. **Roeder, G. S., and G. R. Fink.** 1980. Genetic rearrangements associated with a transposable element in yeast. Cell **21**:239–249.
22. **Shaw, P. H., W. Held, and N. D. Hastie.** 1983. The gene family for major urinary proteins: expression in several secretory tissues of the mouse. Cell **32**:755–761.
23. **Smith, G. P.** 1976. Evolution of repeated DNA sequences by unequal crossover. Science **191**:528–535.