The EMBO Journal vol.9 no.4 pp.1283–1288, 1990

Repetitive sequence involvement in the duplication and divergence of mouse lysozyme genes

M.Cross and R.Renkawitz

Genzentrum, Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG

Communicated by W.Zillig

Mouse M and P lysozymes are the products of separate genes, are specifically expressed in separate tissues, and are adapted to different functions. The lysozyme genes have assumed these markedly different characteristics following their generation by gene duplication 30–50 million years ago. The discovery of the lysozyme P gene only 5 kb upstream from the M gene in tandem repeat has enabled an investigation of the molecular basis of their duplication and subsequent divergence. The duplication is shown to have involved recombination between two B2 repeat sequences flanking the original gene. The resulting downstream copy has retained the myeloid specificity of expression along with just 1.7 kb of upstream sequences, while the upstream copy is inactive in macrophages and has become expressed instead in the small intestine. Although multiple gene conversion events have served to maintain a generally high homology between the genes, certain regions have been found to be specific for either one of the gene pair: two repetitive sequences peculiar to the P region may serve to protect the coding regions from gene conversion, while sequences unique to the M gene may be more directly involved in differential regulation.

Key words: B2 repeat/divergence/evolution/gene duplication/regulation

Introduction

One of the products of a gene duplication event is effectively released from the selective pressures acting on the original gene, and is hence free to adopt a new role. The evolutionary relevance of this process is supported by the occurrence of gene families with members which fulfil various, though related, functions. Nucleotide sequence comparisons between members of a gene family have indicated that duplication may occur either via RNA mediated transpositional events (Soares *et al.*, 1985), as a result of which a gene copy is inserted at a distant site, or by recombination between sequences flanking the original gene, with the consequent formation of a gene cluster (for review see Maeda and Smithies, 1986). The recruitment of one of the copies to a new function, rather than its degeneration to the status of a pseudogene, must then involve alterations in either or both of its coding and regulatory sequences. It is therefore of importance to our understanding of the processes which generate diversity in gene regulation to define not only the mechanism of gene duplication, but also the nature and possible order of those events which subsequently alter the structure and pattern of expression of the gene products.

The bacteriolytic activity of lysozyme (EC 3.2.1.17) is widely detectable throughout both the plant and animal kingdoms (for review, see Jollès and Jollès, 1984). Although the enzyme can be found in most mammalian tissues and secretions, in which it is assumed to contribute to non-specific host defence mechanisms (Fleming, 1922), both the screening of a large number of cell lines (Ralph *et al.*, 1976) and the analysis of tissue sections by immunostaining (Klockars and Osserman, 1974) or *in situ* hybridization (Ping Chung *et al.*, 1988) has localized expression of lysozyme to myeloid cells and to the Paneth cells of the small intestine.

The proposal of a non-immune role for lysozyme has been prompted by the discovery of very high levels of activity in the digestive tracts of a number of species including ruminants (Dobson *et al.*, 1984) and house mouse (Hammer *et al.*, 1987), implying recruitment of lysozyme from its immune related function to a relatively recent digestive function in these lineages. The assumption of high levels of lysozyme in the cow abomasum has been accompanied both by an increase in the number of lysozyme genes to ~10 (Irwin and Wilson, 1989), and by an apparent decrease in the levels of enzyme in extra-intestinal tissues (Prieur, 1986). In the house mouse, however, high expression in the small intestine has apparently been achieved without compromising levels of extra-intestinal enzyme, and has involved a single gene duplication (Hammer *et al.*, 1987): the predominant form of lysozyme mRNA in the mouse small intestine being encoded by the lysozyme P gene, while the lysozyme M gene accounts for the vast majority of lysozyme mRNA detectable in all other tissues, and all that found in isolated macrophages (Cross *et al.*, 1988). Linkage studies have indicated that the lysozyme M and P genes are present on the same chromosome (Hammer and Wilson, 1987), and it has been calculated from both partial protein sequence and phylogenetic data that the gene duplication giving rise to them occurred ~50 million years (Mys) ago (Hammer *et al.*, 1987).

We describe here the tandem arrangement and close proximity of the lysozyme M and P genes, and present evidence that the original gene duplication involved recombination between repetitive sequences flanking the ancestral gene. In addition, we demonstrate a number of extensive discontinuities between the genes which involve both repetitive and apparently unique sequences, and which may have influenced the maintenance of macrophage specificity for the M gene and the assumption of Paneth cell specificity for the P gene. The close proximity, high homology, and separate tissue specificities of the two genes make this an important system in the study of the molecular mechanisms of regulatory evolution.

© Oxford University Press

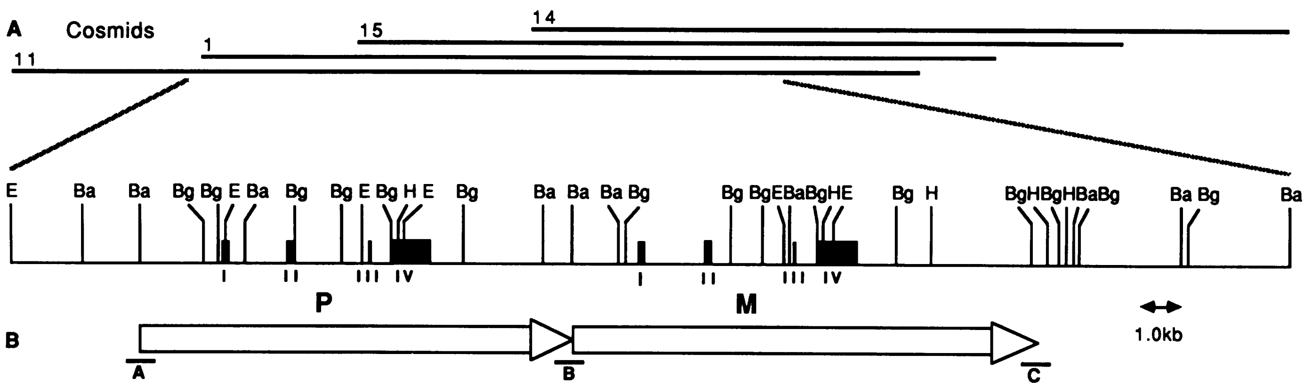

Fig. 1. The lysozyme gene pair. (**A**) A map of a 30 kb region containing the genes is presented in relation to the cosmid clones used. Exons are represented by numbered solid boxes. Restriction sites: Ba = *Bam*HI; Bg = *Bgl*II; E = *Eco*RI; H = *Hin*dIII. (**B**) Open arrows represent the extent of the duplicated region determined by hybridization experiments and by sequencing of the end points A, B and C.

Results

Mouse lysozyme gene arrangement

The lysozyme M gene structure has been described previously (Cross *et al.*, 1988). With the intention of isolating long, contiguous stretches of the lysozyme gene region, we screened a cosmid library of *Mus domesticus* with a lysozyme M cDNA probe. Four hybridizing clones (cosmlg1, 11, 14 and 15) were picked and shown by restriction mapping to be overlapping and to span a region from ~25 kb upstream to 40 kb downstream of the lysozyme M gene (Figure 1A). Hybridization of the lysozyme M cDNA probe to Southern blots of the cosmid digests not only confirmed the presence of lysozyme M gene sequences in all cosmids, but also revealed a region of very high homology to the M cDNA a short distance upstream from the M gene. Further subcloning, mapping and hybridization studies using exon-specific probes from the M gene demonstrated the existence of a complete second lysozyme gene in this position. The intron/exon distribution of the second gene (Figure 1A) was determined by hybridization studies and subsequently confirmed for exons 1, 3 and the upstream end of exon 4 by DNA sequencing (Figure 2 and data not shown). Evidence that the second lysozyme gene is that encoding lysozyme P comes from the comparison of the predicted translation of exon 1 protein coding sequences with the published N-terminal protein sequence of P lysozyme (Hammer *et al.*, 1987). Both sequences indicate the presence of the same six specific amino acid substitutions within 27 residues of the mature peptide (Figure 2). This identity is further supported at the DNA sequence level by comparison of exons 1 and 3 and partial exon 4 sequences with that of a lysozyme P cDNA (data not shown; G.Cortopassi and A.Wilson, 1990).

Hence, the lysozyme P and M genes, which have markedly different tissue specificities of expression, are arranged in close tandem repeat with their coding sequences ~5 kb apart.

Extent and nature of the gene duplication

The arrangement of the lysozyme genes is suggestive of their generation via a recombinative duplication event, following which each gene would effectively constitute an insertion mutation in the flanking regions of the other. This arrangement implies certain restrictions on the location of those regulatory sequences responsible for the different tissue

```
      *
(ATG AAG GCT CTC CTG ACT CTG GGA CTC CTC CTG CTT
(Met Lys Ala Leu Leu Thr Leu Gly Leu Leu Leu Leu

                 *                   *   *   *
TCT GTC ACT GCC CAG GCC) AAG GTC TAC AAT CGT TGT
Ser Val Thr Ala Gln Ala) Lys Val Tyr Asn Arg Cys

     *       *                               *
GAG TTG GCC AGA ATT CTG AAA AGG AAT GGA ATG GAT
Glu Leu Ala Arg Ile Leu Lys Arg Asn Gly Met Asp

     **   *     **     *
GGC TAC CGT GGT GTC AAG CTG GCT GACT
Gly Tyr Arg Gly Val Lys Leu Ala Asp
```

Fig. 2. Coding sequence and predicted translation of the lysozyme P gene exon 1. The signal peptide and corresponding nucleotide sequences are bracketed. Asterixes (above nucleotides) and underlines (amino acids) indicate mismatches to the lysozyme M sequence (Cross *et al.*, 1988).

specificities of expression. In particular, the retention by the M gene of myeloid specificity suggests that the removal of upstream sequences following duplication did not significantly compromise the original function of the gene. It was therefore of interest to determine the precise extent of the duplicated unit.

The approximate extents of homology between the intergenic region and regions upstream from P or downstream from M were determined by a series of Southern blotting experiments in which digests of plasmid subclones spanning each region were hybridized separately to probes covering the other two (data not shown). Having located the approximate end points of the duplication event in this manner (Figure 1B), the regions of interest were sequenced to define more precisely the point of recombination.

Alignment of the sequences showed a block of 180 bp to be highly conserved between all three regions (Figure 3), while sequences 5′ of this block are conserved between the intergenic and downstream regions (B and C in Figure 3), and sequences 3′ are conserved between the intergenic and upstream regions (B and A in Figure 3). This pattern is consistent with the occurrence of the original recombination event within this central element, which was found to be

Fig. 3. Nucleotide sequences of the end points of duplication. B2 repetitive elements are underlined. Chevrons mark the direct repeats flanking the B2 elements in A and C.

a member of the mouse B2 family of middle repetitive elements (Krayev *et al.*, 1982). The repeats upstream of P and downstream of M (A and C respectively in Figure 3) both show the small duplication of target sequences typical of B2 insertion, while the intergenic version (B) has no such duplication, being flanked instead by one copy each of the A and C type target sequences. The continuity of the

Fig. 4. Lysozyme gene duplication by unequal homologous crossing over between the B2 repeats A and C. One of the product chromosomes carries two copies of the gene separated by a hybrid B2 repeat (B). The other product chromosome, which would have the gene deleted, is not shown.

intergenic B2 repeat shows that the recombination was homologous. The sizes of the duplicated regions (i.e. the distance between comparable positions of the B2 repeats) are 10.5 kb (P domain) and 11.4 kb (M domain). This difference appears to be largely attributable to sequences downstream of the coding regions, the distance between exon 4 of the P gene and the intergenic B2 repeat being ~1 kb shorter than the corresponding part of the M region. These results indicate that the duplication event giving rise to the lysozyme M and P genes occurred by homologous recombination between two B2 repeats which flanked the original gene and were 10–12 kb apart (Figure 4).

Distribution of repetitive sequences throughout the lysozyme region

It is possible that recombination involving repetitive sequences may not only promote gene duplication (Shen *et al.*, 1981; Jeffreys and Harris, 1982; Barsh *et al.*, 1983; Nicholls *et al.*, 1987), but also cause rearrangements during divergence of the duplicated units (Hess *et al.*, 1983). In addition, the differential presence of repetitive sequences at particular positions would interrupt the homology of a gene pair and may hence serve to decrease the frequency of initiation and limit the extent of gene conversion events (Hess *et al.*, 1983), for which there are obvious indications in the lysozyme genes (see below). In order to determine the distribution of repetitive sequences throughout the lysozyme gene region, digests were performed such that the complete duplicated region of 22 kb was present on fragments of 100–800 bp, with flanking regions represented on fragments of up to 2.5 kb. A Southern blot of these digests was hybridized firstly to labelled mouse genomic DNA to locate general repetitive sequences, and was then stripped and hybridized to a 170 bp *Rsa*I fragment which carries the downstream B2 repeat (C in Figure 1B) but lacks the A/T rich 3′ end.

Within the 30 kb covered by the digests, the genomic probe hybridized to 19 fragments (Figure 5A), 13 of which showed homology to the B2 repeat probe (Figure 5B). Assuming that each contiguous group of hybridizing fragments corresponds to a single repetitive element, the observed pattern indicates the presence of at least 10

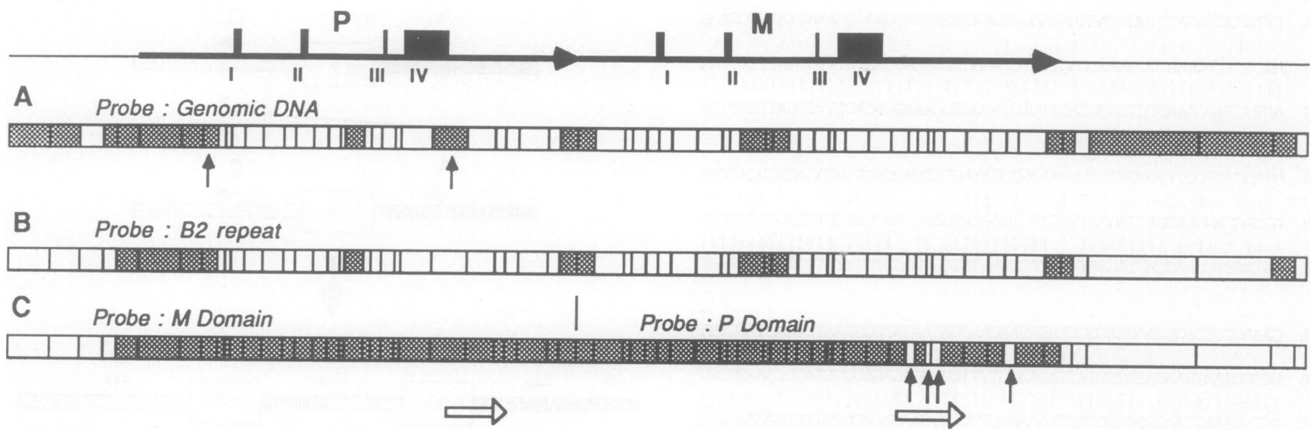

Fig. 5. Homologies of the lysozyme genes both to repetitive sequences and to each other. Restriction fragments of a 30 kb region containing the genes are shown by boxes. Shading indicates hybridization to genomic DNA (**A**), a B2 repeat probe (**B**), and to composite probes covering the opposite half of the region shown (**C**). Solid arrows mark the positions of those repetitive sequences present flanking the P but not the M genes (**A**) and the positions of four M specific fragments (**C**). The positions of sequences compared in Figure 6 are indicated by open arrows below (**C**).

repetitive sequences, six of which are B2 related. However, since it seems likely that some of the larger groups of contiguous fragments harbour more than one repetitive element, the values given above are probably underestimated. Of particular interest are two regions in which repetitive elements are obviously present in the P but absent from the M gene. Remarkably, these differences occur immediately upstream and immediately downstream of the coding sequences. The position of these discontinuities is suggestive of a possible contribution to protection against gene conversion (see below).

Comparison of the duplicated units

The repetitive sequence duplication revealed no indication of the cause of the ~1 kb difference in length between the lysozyme M and P gene downstream regions noted above. In an attempt to reveal any large deletions or insertions involving non-repetitive sequences, Southern blots similar to that described in the previous section were performed such that the M and P regions were represented on separate membranes. Radiolabelled probes were then prepared representing either the complete P or complete M gene regions with their respective flanking sequences, and each was hybridized to the blot of the other domain. The resolution of this approach is limited by the size of the digest fragments, such that small gene specific deletions or insertions are likely to remain undetected. Nonetheless, while every P region fragment hybridized to the M region probe, a total of four M region fragments was found to be specific (Figure 5C). All of the M specific fragments are in the downstream region of the gene, and the sum of their sizes corresponds approximately to the difference in length of the M and P regions.

The sequences of a 2.0 kb region containing the group of three M specific fragments nearest to the gene, and 1.6 kb of the corresponding region of the P gene were determined and are compared graphically in Figure 6. In addition to substitutions and some differences apparently due to duplication or deletion of very short sequences, the regions were found to contain one large P specific segment (X in Figure 6) of ~360 bp, and two M specific segments (Y and Z in Figure 6) of ~200 and 550 bp respectively. Specific

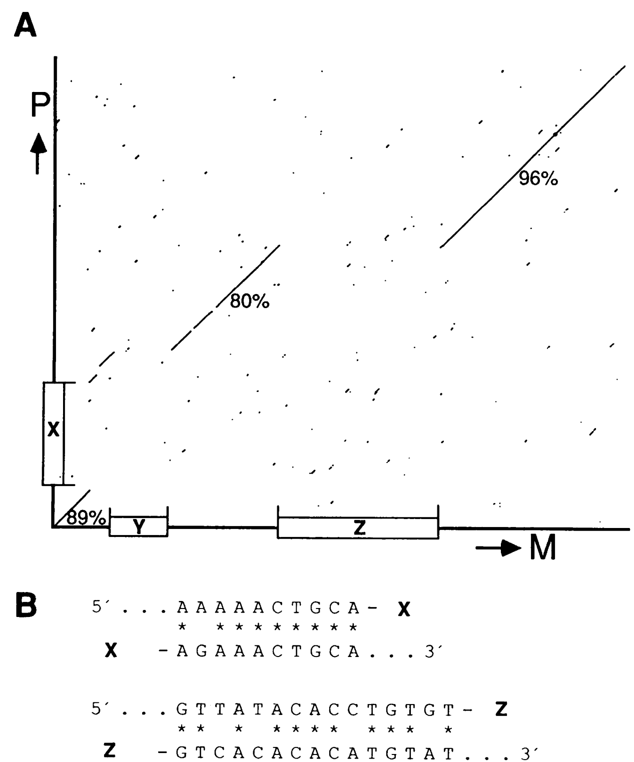

Fig. 6. (**A**) Sequence comparison of a region containing M specific fragments (horizontal axis) with the corresponding region of the P gene (vertical axis). The positions of these sequences are indicated in Figure 5C. X, Y and Z indicate the three largest gene specific segments. Extensive homologous sequences are marked with their degree of identity. (**B**) Partial direct repeats at the ends of segments X and Z.

segment X was not identified in the original hybridization experiment due to its association with sequences showing very high conservation (data not shown). While this element demonstrates a very low level of similarity to a B2 repetitive sequence, neither of the M specific elements revealed extensive similarities in data bank searches of known nucleotide sequences. In addition, a probe carrying region Y and most of region Z hybridizes strongly to a single 7.5 kb band in *Bam*HI digested mouse genomic DNA (corresponding to the expected M gene fragment), and does

not show the smear typical of repetitive sequences (not shown). Short partial repeats found at the ends of elements X (nine matches over 10 bp, Figure 6B) and Z (11 matches over 15 bp, Figure 6B) may be indicative either of staggered breaks generated during insertion, or of sites favoured for recombinative deletion of the sequences concerned, although the apparently non-repetitive nature of sequences Y and Z would be more consistent with their deletion from the P gene than with their insertion into the M gene following duplication.

Divergence of the duplicated sequences

Comparison of the protein coding nucleotide sequences of lysozyme P and M exons 1 (Figure 2) reveals a 22.6% frequency of substitution per synonymous site (calculated by the method of Wu and Li, 1985). Assuming rates of substitution per synonymous site in rodents to be between 6.5×10^{-9} and 10×10^{-9} per year (Li *et al.*, 1987), the exon 1 sequences predict a divergence time of 22–35 Mys for the M and P genes. Sequences from elsewhere within the duplicated region share very variable and often much higher degrees of homology: the sequences flanking the B2 repeats at the end points of duplication show 11.5% (upstream of repeats B and C, Figure 3) and 2.5% (downstream from repeats B and A, Figure 3) substitution; those surrounding the gene specific segments in the downstream regions show 4–20% substitution (Figure 6); and other sequences in the intron 2 regions demonstrate a 30% degree of substitution (data not shown). This latter frequency predicts divergence over 30–50 Mys, and is hence more consistent with the range 30–65 Mys previously proposed from N-terminal protein sequence and phylogenetic data (Hammer *et al.*, 1987). The relevance of these variations in sequence conservation to the molecular evolution of the lysozyme genes is discussed below.

Discussion

The occurrence of gene duplication by unequal homologous crossing over between repetitive sequences has previously been implied by work on human globin (Shen *et al.*, 1981; Jeffreys and Harris, 1982; Nicholls *et al.*, 1987) and human growth hormone (Barsh *et al.*, 1983) gene clusters. More recently, a B2 repeat within the mouse *Tla* region was proposed to be one end point of either a partial duplication or gene conversion event (Fischer *et al.*, 1989). We present here clear and complete evidence that recombination between B2 repeats was involved in the generation of the mouse lysozyme P/M gene pair, demonstrating a positive contribution of these Alu like transposable elements to the processes of molecular evolution. Although such elements have many characteristics expected of 'parasitic' DNA (Maynard Smith, 1989), the observation that their presence may indeed have beneficial effects on the evolutionary adaption of the host species is suggestive of a level of 'symbiosis'.

Estimation of the age of the gene duplication is hindered by the variability in sequence identities at different positions between the duplicated regions. Application of calculated divergence rates (Li *et al.*, 1987) suggests that while some regions have diverged over 30–50 Mys [which is in the lower range of an estimate based on partial protein sequence and phylogenetic data (Hammer *et al.*, 1987)], others appear to have diverged over as little as 2.5 Mys. The interruption of these regions by repetitive (Figure 3) and gene specific (Figure 6) sequences argues against a strong functional selection as a reason for maintenance of high homology. Rather, we propose that the highly conserved sequences have been involved in intergenic recombination events which have served to homogenize parts of the duplicated regions on several occasions since the gene duplication. In evolutionary terms, such correction events are collectively referred to as 'gene conversion' (Maeda and Smithies, 1986).

The special interest of the molecular evolution of the lysozyme genes lies in their assumption of markedly different tissue specificities of expression following duplication, despite their generally high homology and close proximity. The organization of the region reported here provides some clues to the evolutionary events involved in the generation and maintenance of differential expression. Firstly, the retention of ancestral (myeloid) specificity by the downstream copy of the gene suggests that the sequence determinants of myeloid specificity are located either within or downstream of the duplicated unit, which itself encompasses only 1.7 kb upstream from the start site of transcription. Secondly, the generally high homology of the genes is contrasted by the presence of distinct M specific non-repetitive segments in the downstream region which have probably been deleted from the P gene following duplication. It seems likely that the down regulation of the P gene in myeloid cells preceded and facilitated adaptation of the enzyme to the more stable form (Hammer *et al.*, 1987) suited to function in the small intestine, and although the location of regulatory sequences controlling M and P expression is as yet unknown, it is plausible that one of the P specific deletions has removed sequences essential for myeloid specific expression. This possibility is currently under examination.

Irrespective of any direct role of these M specific segments in gene regulation, the fact that they separate three regions of markedly different degrees of identity (Figure 6) suggests that they have interrupted the progress of branch migration in recombination intermediates involved in gene conversion events on at least two occasions. The predicted reduction in both frequency and extent of gene conversion mediated by insertion or deletion mutation would be expected to stabilize the divergence of structure required for the establishment of different functions of duplicated genes (Hess *et al.*, 1983; Shimenti and Duncan, 1984). This would be particularly relevant in the case of the lysozyme genes, where gene conversion appears to have occurred with high frequency during their relatively short history of divergence. In this respect it is interesting to note that the P lysozyme coding sequences are flanked very closely by two blocks of repetitive sequences which are absent from the M gene, and which may therefore contribute to the stability of the gene pair by limiting the initiation or progression of gene conversion from flanking into coding sequences.

Hence our results indicate that repetitive sequences have contributed substantially to the molecular evolution of a single lysozyme gene into a pair of differentially expressed genes with different functions. The initial gene duplication event involved recombination between flanking homologies provided by two copies of the Alu like B2 family of repeats, while other repetitive sequences have since interrupted

the extensive homology between the resulting gene pair in a manner which may have facilitated their divergence by limiting the homogenizing effects of frequent gene conversion.

Materials and methods

Isolation and mapping of cosmid clones

Independently amplified fractions of a mouse genomic library (male *M. domesticus* strain 129 Sv) constructed in the cosmid vector pcos2EMBL were a gift from A.-M.Frischauf (ICRF, London). Colonies corresponding to five genomic equivalents from each of four fractions were transferred to Genescreen membranes (New England Nuclear) and screened under conditions recommended by the manufacturer using a probe containing the complete coding region of the mouse lysozyme M cDNA (Cross *et al.*, 1988). One hybridizing clone from each fraction was picked for further analysis. Restriction maps were determined both by standard techniques and by end specific detection of partial digest fragments. Complete digestion of the vector *Sal*I site was followed by partial digestion with *Bam*HI, *Hind*III or *Eco*RI fragment separation by reversed field electrophoresis on 1% agarose gels (Carle *et al.*, 1986). The DNA was then transferred to nylon membranes (see below) and hybridized to one of two DNA probes each specific for one vector arm. Autoradiography revealed a ladder of bands from which the order and size of restriction fragments was calculated. The application of this technique to cosmid 11 was restricted by the presence of a *Sal*I site near one end of the insert. Insert fragments were subcloned into pBR322 or Bluescribe (Stratagene) vectors using standard procedures (Maniatis *et al.*, 1982).

Southern blotting and hybridization

Following electrophoresis through 0.6–1.8% agarose gels, DNA fragments were denatured in 0.4 M NaOH, 0.6 M NaCl for 30 min and then transferred overnight in the same solution to Hybond membranes (Amersham Buchler). Membranes were rinsed in 0.5 M Tris–Cl, pH 7.5, 0.6 M NaCl, air dried and baked for 1 h at 80°C. DNA was cross-linked to the membranes by UV irradiation (8 s duration, 20 cm from a 254 nm Herolab transilluminator).

Mouse genomic DNA was a gift from B.Möllers (Genzentrum, Martinsried). Specific plasmid DNA probe fragments were purified by two rounds of electrophoresis through low melting point agarose. All probes were radiolabelled by the random priming method (Feinberg and Vogelstein, 1983).

Prehybridization of membranes in 7% SDS, 10 mM EDTA, 0.25 M Na_2HPO_4, pH 7.2 and 1% BSA was performed at temperatures between 55 and 68°C for 15 min in a rotisserie incubator (Becker and Schütz, 1988). Hybridization was carried out for 16 h at the same temperature in prehybridization buffer containing denatured herring sperm DNA (250 μg/ml), and specific probe in at least 5-fold excess over membrane bound DNA. Membranes were washed three times in 500 ml 20 mM Na_2HPO_4, pH 7.2, 1 mM EDTA, 1% SDS and 0.1 M NaCl at temperatures between 55 and 68°C, sealed wet under plastic foil and exposed to Kodak X-Omat AR5 film. Hybridized membranes were stripped by three 15 min washes in 500 ml 0.2 M NaOH at room temperature, and neutralized in 0.1 M Na_2HPO_4, pH 7.2, 1 mM EDTA. Stripping was confirmed by autoradiography before rehybridization of membranes.

Sequencing and sequence analysis

All sequences were determined from CsCl density gradient purified plasmid DNA templates by the chain termination method (Chen and Seeburg, 1985), using the Sequenase kit (United States Biochemicals). Sequencing primers were kindly provided by Ronald Mertz and Dorit Weigand (Genzentrum, Martinsried). All sequence analysis was performed on a VAX/VMS system using software from the University of Wisconsin Genetics Computer Group (Devereux *et al.*, 1984). The 'Compare' program was executed using a window size of 21 and a stringency of 14.

Acknowledgements

We are indebted to many of our colleagues in the Genzentrum, especially Angela Wedel, Bettina Möllers, Manfred Schartl and Uwe Hinz for helpful discussion during the course of this work and for their comments on the manuscript. We would also like to thank Bettina Möllers for mouse genomic DNA, Dorit Weigand and Ronald Mertz for sequencing primers, Gino Cortopassi and Allan Wilson for making available information prior to publication, and Beate Wittbrodt and Dagmar Wolf for their invaluable technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 324).

References

Barsh,G.S., Seeburg,P.H. and Gelinas,R.E. (1983) *Nucleic Acids Res.*, **11**, 3939–3958.
Becker,P.B. and Schütz,G. (1988) *In Setlow,J.K. (ed.), Genetic Engineering Principles and Methods.* Plenum Press, New York, pp.
Carle,G.F., Frank,M. and Olson,M.V. (1986) *Science*, **232**, 65–68.
Chen,E.J. and Seeburg,P.H. (1985) *DNA*, **4**, 165–170.
Cross,M., Mangelsdorf,I., Wedel,A. and Renkawitz,R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6232–6236.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Dobson,D.E., Prager,E.M. and Wilson,A.C. (1984) *J. Biol. Chem.*, **259**, 11607–11616.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
Fischer,D.A., Pecht,M. and Hood,L. (1989) *J. Mol. Evol.*, **28**, 306–312
Fleming,A. (1922) *Proc. R. Soc. Lond., Ser. B*, **93**, 306–317.
Hammer,M.F. and Wilson,A.C. (1987) *Genetics*, **115**, 521–533.
Hammer,M.F., Schilling,J.W., Prager,E.M. and Wilson,A.C. (1987) *J Mol. Evol.*, **24**, 272–279.
Hess,J.F., Fox,M., Schmid,C. and Shen,C.-K.J. (1983) *Proc. Natl. Acad Sci. USA*, **80**, 5970–5974.
Irwin,D. and Wilson,A.C. (1989) *J. Biol. Chem.*, **264**, 11387–11393.
Jeffreys,A.J. and Harris,S. (1982) *Nature*, **296**, 9–10.
Jollès,P. and Jollès,J. (1984) *Mol. Cell. Biochem.*, **63**, 165–189.
Klockars,M. and Osserman,E.F. (1974) *J. Histochem. Cytochem.*, **22**, 139–146.
Krayev,A.S., Markusheva,T.V., Kramerov,D.A., Ryskov,A.P., Skyrabin,K.G., Bayev,A.A. and Georgiev,G.P. (1982) *Nucleic Acids Res.*, **10**, 7461–7475.
Li,W.-H., Tanimura,M. and Sharp,P.M. (1987) *J. Mol. Evol.*, **25**, 330–342.
Maeda,N. and Smithies,O. (1986) *Annu. Rev. Genet.*, **20**, 81–108.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maynard Smith,J. (1989) *Evolutionary Genetics.* Oxford University Press, Oxford.
Nicholls,R.D., Fischel-Ghodsian,N. and Higgs,D.R. (1987) *Cell*, **49**, 396–378.
Ping Chung,L., Keshav,S. and Gordon,S. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6227–6231.
Prieur,D.J. (1986) *Comp. Biochem. Physiol.*, **85B**, 349–353.
Ralph,P., Moore,M.A.S. and Nilsson,K. (1976) *J. Exp. Med.*, **143**, 1528–1533.
Shen,S., Slightom,J.L. and Smithies,O. (1981) *Cell*, **26**, 191–203.
Shimenti,J.C. and Duncan,C.H. (1984) *Nucleic Acids Res.*, **12**, 1641–1655.
Soares,M.B., Schon,E., Henderson,A., Karathanasis,S.K., Cate,R., Zeitlin,S., Chirgwin,J. and Efstratiadis,A. (1985) *Mol. Cell. Biol.*, **5**, 2090–2103.
Wilson,A. (1990) *Nucleic Acids Res.*, **18**, in press.
Wu,C.-I. and Li,W.-H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1741–1745.

Received on November 28, 1989; revised on January 11, 1990