Volume 14 Number 5 1986 Nucleic Acids Research

Allotypic differences in murine μ genes

P.H.Schreier[1,2], S.Quester[2] and A.Bothwell[3]

[1]Max-Planck-Institut für Züchtungsforschung, Egelspfad, D5000 Köln 30, [2]Institut für Genetik der Universität zu Köln, Weyertal 121, D5000 Köln 41, FRG and [3]Departments of Pathology and Biology, Yale University School of Medicine, and the Howard Hughes Medical Institute, 310 Cedar Street, New Haven, CT 06510, USA

Received 21 October 1985; Accepted 31 January 1986

ABSTRACT

We report the complete DNA sequence of a c-DNA clone of the heavy chain μ^b allele of the C57BL/6 mouse. Comparisons have been made with the nucleotide sequences of the germ line BALB/c μ^a and the plasmacytoma TEPC-183 μ^a alleles reported elsewhere over the entire length of the coding and the 3' untranslated region. In contrast to the extensive differences between the γ2a a and b alleles we have reported earlier we see a very high degree of homology between the μ alleles. Only one of the nucleotide differences between C57BL/6 μ^b and BALB/c μ^a leads to an amino acid substitution. This single amino acid exchange must form the allotypic determinant of the μ^b allele. A comparison of four different DNA sequences indicates that they are all distinct IgM alleles.

INTRODUCTION

The heavy chain immunoglobulin loci are inherited as a cluster of closely linked allotype markers. Polymorphic differences that alter the amino acid sequence of immunoglobulins can create altered antigenic determinants detectable by allospecific antisera. Seven of the eight heavy chain genes (γ2a, α, γ2b, γ1, δ, μ and ε) have allotypic loci defined as Igh-1 to Igh-7, respectively (1,2). The locus encoding the μ protein was defined as Igh-6 (3). The fine structure of this heavy chain region in BALB/c has been defined by molecular cloning but only a limited number of polymorphic differences were apparent from an analysis of restriction endonuclease cleavage sites (4). A significant improvement of our definition of allotypic differences was obtained by DNA sequence analysis (5,6) and amino acid analysis (7) of γ2 allotypes. Our study of the γ2a allele in C57BL/6 (5) and its relationship to the BALB/c allele revealed 2 major findings. The sequence differences were extensive and thus the relative ease in generating anti-allotype sera and monoclonal antibodies with unique specificities did reside in multiple amino acid sequence differences. In addition, the segmental homologies apparent in our DNA sequence and their relationship to existing γ2 alleles suggested that the

process of gene conversion was the origin of this diversity (5). To further characterize these phenomena at heavy chain loci and define primary structural differences that are detected as allotypic differences we have determined the DNA sequence encoding the exons of the μ^b allele from the C57BL/6 mouse strain. Comparison has therefore been made with three other μ^a DNA sequences from BALB/c mice which permit us to define the allotypic determinant generated by a single amino acid exchange. The characterization of the μ gene is also valuable because of its use in expression systems in E. coli (8), yeast (9) and in transgenic mice (10).

MATERIALS AND METHODS

Bacterial strains used were reported earlier (5). Subcloning into M13 was done with M13 mp8 (11). The plasmids pABμ-1 and pABμ-11 were derived from the B1-8 hybridoma. The isolation and characterization of the variable region sequence has been published (12). The B1-8 hybridoma was derived from a C57BL/6 mouse after a primary immunization (13). Methods of subcloning into M13 phage, DNA sequence determination and computer analysis have been reported earlier (5).

RESULTS

Nucleotide Sequence of the C57BL/6 μ^b Heavy Chain Constant Region:

The clones used to determine the μ DNA sequence of the Igh b haplotype were derived from cDNA of the B1-8 hybridoma. The spleen cell partner in this hybridoma was from a C57BL/6 mouse and makes an anti-NP antibody having a μ heavy chain and a lambda light chain (13). The nucleotide sequence of the μ^b constant region gene was determined by using plasmids pABμ-1 and pABμ-11 (Fig. 1) which contain inserts of about 1000bp and 1900bp, respectively (12). The sequence of the region coding for the variable (V_H) region, D_H and J_H regions and about 150 bp of the CH1 domain and the sequence marked in Fig. 1 was determined using the chemical sequencing procedures (14). The sequence of the rest of the μ constant region and the 3' untranslated region (3' ut-region) were determined mainly by the chain termination procedure (15) using subcloned restriction fragments in derivatives of the single stranded DNA phage M13 (11). The strategy for determining the constant region DNA sequence is presented in Fig. 1.

Comparison of μ^b (C57BL/6) and μ^a (BALB/c) Sequences

The DNA sequence of the μ^b constant region from the C57BL/6 cDNA clone is presented in Fig. 2. It contains all four domains, the c-terminal

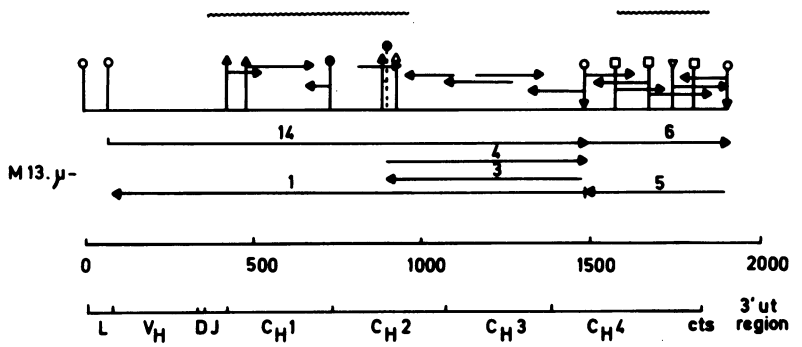

Figure 1. DNA sequencing strategy for the μ^b cDNA clone. The regions where DNA sequence determination was accomplished by the chemical sequencingprocedures are indicated by wavy lines. Regions determined by the chain termination procédure are indicated with horizontal arrows. Vertical arrows indicate restriction endonuclease cleavage sites relevant for primer isolation. The symbols used to designate restriction endonuclease sites are the following: open circle (Pst I), open circle crossed located in the middle of C_H2 (synthetic site at terminus of pABμ-1 clone), solid triangle (Hinf I), solid circle (Pvu II), open square (Hae III), open triangle pointing up (Bam HI) and open triangle pointing down (Sau 3A).

secreted region (CTS) and the 3' ut-region. There are three other complete μ DNA sequences available for comparison and all are derived from BALB/c and are thus μ^a. The first two were determined using genomic DNA clones derived from a newborn mouse (16) referred to here as BALB/c-1 (designated BALB/c in Fig. 2) and from the Bailey subline designated here BALB/c-2(17). The BALB/c-2 sequence is not shown in Fig.2 (see below). The third was from a cDNA clone from the plasmacytoma TEPC-183 (18).

A comparison of the BALB/c-1 and C57BL/6 DNA sequences reveals 11 nucleotide differences out of the approximately 1500 nucleotides or 0.67% sequence differences. The silent nucleotide substitutions appear randomly located in the sequence with the possible clustering in two regions (codons 466-467 and 536-546). These differences are identified in Fig. 3. Only one of these differences at codon 222 results in an altered amino acid. This is also true of the BALB/c-2 sequence. This sequence is not shown in Figure 2 because it is nearly identical to our C57BL/6 sequence. In addition to the common difference at codon 222 (BALB/c-2 has AGA) only four silent changes exist at codons 154, 211, 475 and in the 3' ut region which are identified in Figure 3. Therefore the residue at codon 222 must be the source of the allotypic difference between these two inbred mouse strains. The μ^b allele has a lysine residues while the μ^a allele has an arginine codon.

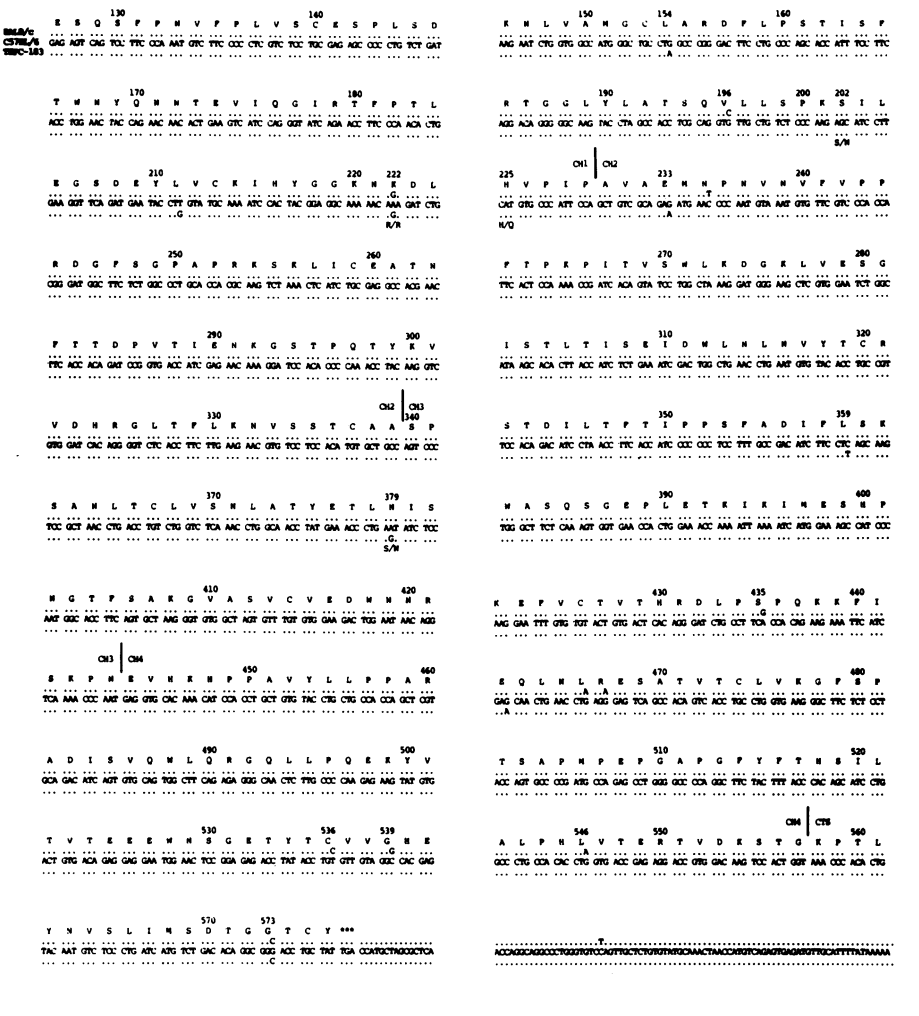

Figure 2. Comparison of nucleotide sequences of the μ^b and μ^a constant regions. The numbers above the amino acid sequence refer to the OU index (25). Letters below the TEPC-183 DNA sequence indicate first, the amino acid predicted by the TEPC-183 codon at that position and second, the amino acid actually found in the MOPC-104E protein (20). All three differences between the amino acid sequence of MOPC-104E and the predicted amino acid sequence from the BALB/c germ line are shown. The C57BL/6 DNA sequence is the only DNA sequence that is fully printed. The letters above the DNA sequences indicate the amino acid sequence predicted from the C57BL/6 DNA sequence. We assume that codon 340 is serine (AGT) in the BALB/c sequence. The codon indicated in Kawakami et al. (16) is GGT which can only encode glycine. Since serine is the amino acid deduced from their DNA sequence and the serine codon (AGT) is found in all other DNA sequences

and in the MOPC-104E protein sequence (20) we assume it is accurate. Codon 211 in the TEPC-183 sequence was recently corrected by Auffray and Rougeon in Kabat et al. (26) from CTT to CTG.

In addition, two regions encoding residues 344-352 and 485-497 have been corrected in the TEPC-183 sequence by Auffray and Rougeon (personal communication) and now are identical to the C57BL/6 and BALB/c-1 sequences. These corrections involved about 20 out of 66 bp in these regions.

This exchange does not affect any of the known glycosylation sites. The μ protein of the B1-8 cell line is indeed recognized with allospecific antisera directed against the μ^b allotype (Herzenberg, L. personal communication).

The comparison becomes more complex when made with the TEPC-183 DNA sequence. This plasmacytoma is of BALB/c origin and should be very similar to the germ line BALB/c DNA sequence. There are 15 nucleotide differences between the BALB/c-1 and the TEPC-183 DNA sequences (Fig. 2) The differences indicated in Fig. 3 in codons 154, 196, 211, 233, 235, 359, 379, 435, 462, 466, 467, 536, 539, 546 and 3' ut are located randomly throughout the sequence with the possible exception of the same two regions (codons

Codon	154	196	202	211	222	225	233	235	359	379
C57BL/6	CTG	CTG	AGC	CTT	AAA	CAT	GAG	AAC	CTC	AAT
BALB/c-1	...	..C	...	...	.G.	...	...	..T	...	...
BALB/c-2	..A	...	...	..G	.G.	...	...	...	...	...
TEPC-183	..A	...	...	..G	.G.	...	..A	...	..T	.G.
MOPC-104E	...	...	.A.	...	.G.	..R	...	...	...	...

Codon	435	462	466	467	475	489	536	539	546	573	3' ut-1	3'ut-2
C57BL/6	TCA	GAG	CTG	AGG	CTG	CTT	TGT	GGC	CTG	GGG	C	C
BALB/c-1	..G	...	..A	..A	...	...	..C	..G	..A	..C	.	T
BALB/c-2	...	...	...	...	T..	...	...	...	...	...	T	.
TEPC-183	...	..A	...	...	...	...	...	...	...	..C	.	.
MOPC-104E	...	...	...	...	...	AAR	...	...	...	...	.	.

Figure 3. Comparison of codon differences between μ sequences. BALB/c-1 refers to the sequence determined by Kawakami et al. (16) and BALB/c-2 refers to the sequence of Goldberg et al. (17). The position 3' ut-2 is indicated as a difference in Fig. 2. The position 3' ut-1 is 9 nucleotides 5' to 3' ut-2. Only limited DNA sequence studies on the μ cDNA from MOPC-104E has been performed (20). The codon indicated shows the minimum number of differences that must have occurred to obtain the amino acid actually detected from the BALB/c germ line sequence. The letter R is used to indicate a purine. The DNA sequence of codon 379 corresponded to the germ line sequence which encodes N. Amino acid sequence analyses suggested T occurred at that position.

Codon	202	222	225	347	379	489
C57BL/6	S	K	H	T	N	L
BALB/c	-	R	-	-	-	-
TEPC-183	-	R	-	-	S	-
MOPC-104E	N	R	Q	-(N)	-(T)	K

Figure 4. Amino acid differences between μ sequences. All of the amino acid differences between the predicted sequence from C57BL/6, BALB/c and TEPC-183 and the sequence determined from MOPC-104E are shown. References to codon numbers are from Fig.2. Since there were no differences between the amino acid sequences predicted from the DNA sequences determined by Kawakami et al., (16) (BALB/c-1) and Goldberg et al, (17) (BALB/c-2) only BALB/c is used to denote the sequence. From the limited DNA sequence studies on a MOPC-104E cDNA clone codons 347 and 379 were identical to the germ line of BALB/c. However, the amino acid residues detected at those positions were N and T, respectively (20).

466-467 and 536-546) that were different between C57BL/6 and BALB/c-1. The BALB/c-2 sequence appears closer to the TEPC-183 sequence than does BALB/c-1 in that there are only seven differences (codons 233, 359, 379, 462, 475, and 573 and in the 3' ut region). The C57BL/6 sequence is also close to the TEPC-183 sequence--only 7 bp are different plus the allotypic difference.

The partial DNA sequence of one other μ plasmacytoma, HPC-76, has been determined between codons 127-318 (19). Since the available sequence of HPC-76 is relatively small, we have listed the differences here and not included them in Figures 3 and 4. Five silent differences from the C57BL/6 sequence in Fig. 2 exist and all are in the third codon position. The HPC-76 sequences at codons 163, 185, 194, 195 and 270 are ATC, AGA, TCA, CAA and TCT, respectively.

The protein sequences differences are summarized in Fig. 4. It is clear that the BALB/c and C57BL/6 sequences are identical except for residue 222. The two plasmacytoma sequences have sequence differences which are either the result of somatic mutations in the μ constant region that occurred during passage of the tumor, polymorphic differences in the BALB/c strains, or are errors in sequencing. For MOPC-104E, three amino acids differ from the common BALB/c sequence: codons 202, 225 and 489. The DNA sequence of codons 202 and 489 has not been determined. The TEPC-183 protein may have only one amino acid which is different from the consensus BALB/c sequence at codon 379.

DISCUSSION

The differences observed between C57BL/6 and BALB/c germ line sequences indicate a single amino acid at codon 222 generates the difference detected by allospecific antisera. The Arg-Lys variation at codon 222 is a very conservative amino acid replacement. Both are hydrophilic residues and are therefore likely to be on the surface of the μ protein. We have detected differences in all μ sequences examined which include the sequences of C57BL/6, two BALB/c germ line isolates and the TEPC-183 plasmacytoma. Since the sequence of TEPC-183 is derived from a plasmacytoma it may have some somatic mutations but it is also possible that it represents a germ line sequence. This argument can also be applied to MOPC-104E. Although only amino acid sequence data are available, the MOPC-104E sequence may be a germ line encoded sequence from another BALB/c substrain. The BALB/c-2 sequence is remarkably similar to our C57BL/6 DNA sequence. Except for the allotypic difference at codon 222 only 4 silent differences exist. There were 12 silent differences in coding regions and 2 differences in the 3' ut region between the BALB/c-1 and BALB/c-2 sequences (see Figure 3). this may reflect an accumulation of silent mutations in the BALB/c strain maintained in Japan, (16).

A knowledge of the sequence differences between the μ genes of BALB/c and C57BL/6 mice should permit one to use DNA probes to easily determine the haplotype of the gene expressed in F1 animals or transgenic mice (10). A synthetic oligonucleotide complementary to μ mRNA having its 3' terminus as the T complementary to the A in the third position of codon 222 could be used as a primer for reverse transcription. If radiolabeled T could be incorporated by reverse transcriptase then the μ RNA in the cell would be of C57BL/6 origin. If radiolabeled dC was incorporated then the RNA would be from BALB/c.

The four murine μ genes may represent a functional subfamily of the eight heavy chain constant region genes. A strong selection on the μ alleles may be due in part because it is a unique gene. A multigene family may reveal more frequent mutational alterations because there is significantly less selective pressure on any given member. The γ2b genes in those two mouse strains contained only seven nucleotide sequence differences (21). Four of these differences resulted in replacement changes which affect 3 allotypic determinants. The γ2a genes showed a high degree of polymorphism which we suggested might have been a consequence of a process similar to gene conversion.

The predominantly expressed lambda light chain genes are the lambda I genes and both the lambda I V and C genes showed no DNA sequence differences (22). This situation is contrasted dramatically when the V_H genes are examined. The level of polymorphism is extremely high (12,23,24). The data would suggest that at least in the large NP^b V_H gene family selection on a given member might be weak.

The constant regions of μ sequences do seem to be able to deviate from the BALB/c or C57BL/6 μ sequence. Indeed we have suggested that TEPC-183 and MOPC-104E have accumulated differences, either a result of somatic mutation or divergence of two alleles. Certainly the human μ protein has many differences from the BALB/c protein including the allotypic difference at codon 222 (25). These results suggest some plasmacytoma μ antibodies may contain somatic mutations in their constant regions.

ACKNOWLEDGEMENTS

This work was supported in part by the Deutsche Forschungsgemeinschaft (SFB74) to Benno Müller-Hill. P. Schreier is an employee of BAYER A.G. Leverkusen. A. Bothwell is an Associate Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Green, M.C. (1979) Immunogenetics 8, 89-97.
2. Herzenberg, L.A., Huang, C., Oi, V.T., and Parsons, M. (1981) In: Immunoglobulin Idiotypes and their Expression, ICN-UCLA Symposia on Molecular and Cellular Biology, vol. XX. C. Janeway, E.E. Sercarz, H. Wigzell and C.F. Fox (eds.) Academic Press, New York, pp. 199-208.
3. Warner, N.L., Goding, J.W., Gutman, G.A. Warr, G.W., Herzenberg, L.A. Osborne, B.A., van der Loo, W., Black, S.J. and Loken, M.R. (1977) Nature 265, 447-449.
4. Shimizu, A., Takahashi, N., Yoaita, Y., and Honjo, T. (1982) Cell 28, 499-506.
5. Schreier, P.H., Bothwell, A.L.M., Muller-Hill, B., and Baltimore, D. (1981) Proc. Natl. Acad. Sci. USA 78, 4495-4499.
6. Ollo, R., and Rougeon, F. (1983) Cell 32, 515-523.
7. Dognin, M.J., Lauwereys, M., and Strosberg, A.D. (1981) Proc. Natl. Acad. Sci. USA 78, 4031-4035.
8. Boss, M.A., Kenten, J.H., Wood, C.R. and Emtage, J.S. (1984) Nucl. Acids Res. 12, 3791-3806.
9. Wood, C.R., Boss, M.A., Kenten, J.H. Calvert, J.E., Roberts, N.A. and Emtage,J.S. (1985) Nature 314, 446-449.
10. Grosschedl, R., Weaver, D., Baltimore, D. and Costantini, F. (1984) Cell 38, 647-658.
11. Messing, J., and Vieira, J. (1982) Gene 19, 269-276.
12. Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari,T., Rajewsky, K., and Baltimore, D. (1981) Cell 24, 625-637.
13. Reth, M., Hammerling, G.J. and Rajewsky, K. (1978) Eur. J. Immunol. 8, 393-400.

14. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Kawakami, T., Takahashi, N., and Honjo, T. (1980) Nucl. Acids Res. 8, 3933-3945.
17. Goldberg, G.I., Vanin, E.F., Zrolka, A.M. and Blattner, F.R. (1981) Gene 15, 33-42.
18. Auffray, C. and Rougeon, F. (1980) Gene 12, 77-86.
19. Bernard, O. and Gough, N.M. (1980) Proc. Natl. Acad. Sci. USA 77, 3630-3634.
20. Kehry, M.R., Fuhrman, J.S., Schilling, J.W., Rogers, J., Sibley, C.H. and Hood, L.E. (1982) Biochemistry 21, 5415-5424.
21. Ollo, R. and Rougeon, F. (1982) Nature 296, 761-763.
22. Bothwell, A.L.M., Paskind, M., Reth,M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1982) Nature 298, 380-382.
23. Loh, D.Y., Bothwell, A.L.M., White-Scharf, M.E., Imanishi-Kari, T. and Baltimore, D. (1983) Cell 33,85-93.
24. Bothwell, A.L.M. (1984) In: The Biology of Idiotypes, M.I. Green and A. Nisonoff, eds., (New York, Plenum Publishing Corp.), pp. 19-34.
25. Putnam, F.W., Florent, G., Paul, C. Shinoda, T., and Shimizu, A. (1973) Science 182, 287-291.
26. Kabat, E.A., Wu, T.T., Bilofsky, H., Reid-Miller, M. and Perry, H. (1983) Sequences of Proteins of Immunological Interests. Government Printing Office Publication NIH, Washington, DC.