Identification of Mutations in the COL4A5 Collagen Gene in Alport Syndrome

DAVID F. BARKER,* SIRKKA LIISA HOSTIKKA, JING ZHOU, LOUISE T. CHOW, ARNOLD R. OLIPHANT, STEVEN C. GERKEN, MARTIN C. GREGORY, MARK H. SKOLNICK, CURTIS L. ATKIN, KARL TRYGGVASON

X-linked Alport syndrome is a hereditary glomerulonephritis in which progressive loss of kidney function is often accompanied by progressive loss of hearing. Ultrastructural defects in glomerular basement membranes (GBM) of Alport syndrome patients implicate an altered structural protein as the cause of nephritis. The product of COL4A5, the α5(IV) collagen chain, is a specific component of GBM within the kidney, and the gene maps to the same X chromosomal region as does Alport syndrome. Three structural aberrations were found in COL4A5, an intragenic deletion, a Pst I site variant, and an uncharacterized abnormality, which appear to cause nephritis and deafness, with allele-specific severity, in three Alport syndrome kindreds in Utah.

ALPORT SYNDROME (AS) IS A HETerogeneous group of hereditary glomerulonephritides, the majority of which appear to be X-linked. The estimated frequency of the gene in the population is 1 in 5000 (*1*). Phenotypic expression of the disease in different kindreds varies with respect to apparently allele-specific age ("juvenile" versus "adult") of terminal renal failure of males; the severity of associated deafness, which ranges from subtle to profound; and the presence of other nonrenal features including lenticonus, retinal abnormalities, megathrombocytic thrombocytopenia, and Fechtner inclusions in granulocytes (*1*). Although reports of male-to-male transmission in a few AS families suggest the possibility of an autosomal form, several studies have shown close linkage of the AS locus to restriction fragment length polymorphism (RFLP) markers near Xq22 with little evidence for genetic heterogeneity (*2*). Progressive nonimmune ultrastructural abnormalities, such as patchy thickening and thinning of GBM and splitting of the lamina densa, led to the suggestion of a defect in an important structural element, possibly type IV collagen, which is the major GBM structural component (*3*). Studies with antibodies directed against the noncollagenous (NC) domain of GBM type IV collagen indicated alteration or absence of a type IV collagen α chain (*4, 5*). Since the genes for the α1(IV) and α2(IV) collagen chains are located on chromosome 13 (*6*), intensive efforts have been directed toward finding X-linked genes for type IV collagen α chains. Complementary DNA clones have recently been identified for a gene, COL4A5, that codes for the distinct α5(IV) chain and is located at Xq22 (*7*). Furthermore, it was demonstrated with chain-specific antibodies that α5(IV) in the kidney is primarily located in the GBM (*7*). The locations of COL4A5 and its product are thus both consistent with a role in AS, prompting this survey of cases for instances of mutation at the gene.

Overlapping genomic clones covering about 35 kb at the 3′ end of the human COL4A5 gene have been isolated. The 14 terminal exons have been sequenced and mapped by electron microscopic heteroduplex analysis (*8*). A map (Fig. 1) of the Eco RI sites relative to the exons has been determined. DNAs from one or two affected male members from each of 18 independently ascertained AS kindreds were examined by conventional Southern (DNA) blot analysis (*9*) for structural alterations at COL4A5. Using a pool of insert fragments from cDNA clones MD-6, PL-35, and PL-31 (Fig. 1), we found no fragment length variation in normal controls for Eco RI (63 chromosomes), Taq I (91 chromosomes), or Pst I (102 chromosomes). The observation of any fragment pattern variation in AS kindreds is therefore a strong indication of a disease-causing mutation. In 3 of the 18 kindreds tested, variant fragment patterns (Fig. 2) (*10*), which appear to reflect specific and distinct mutations in the COL4A5 gene, were detected.

With the pooled cDNA probe, DNA from affected male 197701 (lanes 4 in Fig. 2) from kindred EP in Utah (*11*) showed the absence of 7.5-, 2.4-, and 1.1-kb Eco RI fragments. He also was missing 16-, 10-, and 5.5-kb Pst I fragments and 3.5-, 3.2-,

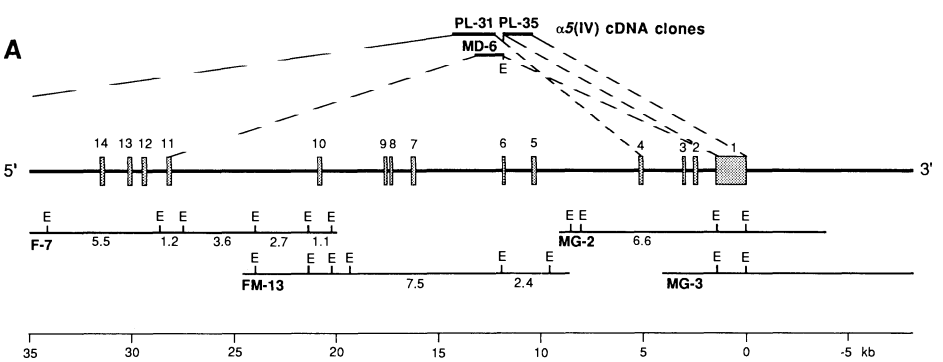

Fig. 1. Mapping of restriction sites in genomic clones with respect to exons at the 3′ end of COL4A5. (**A**) The stippled boxes indicate exon regions of COL4A5, numbered from the 3′ end. Bars labeled PL-31, PL-35, and MD-6 represent previously described cDNA clones (*7*) that span the indicated exons. The position of the Eco RI site in exon 1 is indicated for clones MD-6 and PL-35. Bars labeled F-7, FM-13, MG-3, and MG-2 represent genomic phage vector clones (*8*) with the positions of Eco RI sites marked "E." The lengths, in kilobases, of Eco RI fragments mentioned in the text are shown. The polyadenylation site at the end of the 3′ untranslated region was chosen as the origin for the indicated length scale (in kilobases) for the entire region. Exons have been sequenced and intron sizes were also confirmed with heteroduplex analysis (*8*). (**B**) Map of Pst I and Eco RI sites for the region that includes the variable Pst I fragments observed in AS kindred P. The 2.2-kb Pst I fragment indicated is lost and a novel 1.9-kb Pst I fragment is found on mutation-bearing chromosomes in kindred P. A potential novel Pst I site would result in shortening from the 5′ end of this fragment.

D. F. Barker, A. R. Oliphant, S. C. Gerken, M. H. Skolnick, Department of Medical Informatics, University of Utah School of Medicine, Salt Lake City, UT 84132.
S. L. Hostikka, J. Zhou, K. Tryggvason, Biocenter and Department of Biochemistry, University of Oulu, SF-90570 Oulu, Finland.
L. T. Chow, Department of Biochemistry, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.
M. C. Gregory, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT 84132.
C. L. Atkin, Departments of Biochemistry and Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT 84132.

*To whom correspondence should be addressed.

and 2.7-kb Taq I fragments. The concordant absence of fragments produced by all three restriction enzymes indicates the deletion of a portion of the 3′ end of the gene. That the deletion is internal to the COL4A5 gene was shown by hybridization of the genomic probes F-7, FM-13, and MG-2 (Fig. 1) to the patient DNA (*10*). The MG-2 probe, which spans the 3′ end of the gene, hybridized with a 6.6-kb Eco RI fragment that contains a 5′ portion of exon 1 and the complete exons 2, 3, and 4 as counted from the 3′ end (Fig. 1A). In contrast, the FM-13 probe, which contains exons 5 through 10, hybridized to several fragments in the control samples that were not present in 197701. The fragments affected include the 2.4-kb fragment containing exons 5 and 6, the 7.5-kb fragment containing exons 7, 8, and 9, and a 2.7-kb genomic fragment that is 5′ to exon 10. The presence of the 5.5-kb Eco RI fragment containing exons 12, 13, and 14, and the 3.6-kb fragment that is 3′ to exon 11, was confirmed by hybridization of the patient sample with the genomic probe F-7. Thus, exons 1 to 4 and exons 11 and upstream are intact in 197701 and the deletion affects the region containing exons 5 through 10.

The deletion in kindred EP is a mutation that would result in an α5(IV) chain with severely altered structure. If the deletion includes all of exons 5 through 10, the reading frame would not be changed. However, the variant chain would lack 193 residues from the carboxyl-terminal end of the collagenous domain and 47 residues from the NC domain. The NC domain is essential for triple helix formation and cross-linking of type IV collagen molecules. Consequently, the abnormally short variant chain would not be properly incorporated into homo- or heterotrimers to form a normal type IV collagen network. If a deletion end point occurs within exon 5 or 10, then it is likely that the effects on RNA splicing or translation frame would result in more extreme changes in the gene product.

The extensively studied kindred P from Utah (*1, 2, 4*) was represented by affected males 500081 and 500242. Their DNA Pst I fragment patterns (lanes 1 and 2 in Fig. 2) show the absence of a 2.2-kb fragment and the appearance of a 1.9-kb fragment not found in normal or affected individuals from other kindreds. Both the 2.2-kb and 1.9-kb fragments hybridized with the cDNA inserts of clones PL-35 and MD-6 (*10*). PL-35 and MD-6 contain coding sequences extending in opposite directions from the Eco RI site in exon 1 (Fig. 1B) and, therefore, the affected fragment must be the 2.2-kb Pst I fragment that contains the 5′ end of exon 1 as well as exons 2 and 3. (Fig. 1, A and B).

The association of the Pst I variation with the AS phenotype was confirmed in 122 individual members of kindred P by probing Southern blots of Pst I–digested DNA samples with the pooled cDNA probe. Fragment patterns, exemplified in Fig. 3, showed full concordance of clinically diagnosed AS with the 1.9-kb Pst I variant fragment. In particular, all 23 gene-carrier mothers were found to carry one copy of each of the 1.9-kb and 2.2-kb fragments, which segregated to a total of 45 sons in complete linkage with the presence or absence, respectively, of the AS phenotype. Within kindred P, all 30 affected males carried only the 1.9-kb fragment, whereas all 32 unaffected males had only the 2.2-kb fragment.

The molecular basis for the Pst I site variation in kindred P is not as yet completely understood. However, it is more likely to be a point mutation or a change of a few nucleotides rather than a 300-bp deletion, since both Eco RI and Taq I digests show apparently normal lengths of fragments that include the same part of the gene. A point mutation leading to a new Pst I site must reside in the 5′ end of the Pst I fragment containing the 5′ end of exon 1 and exons 2 and 3 (Fig. 1B), since the region at the 3′ end of the fragment is untranslated. In that case, the new Pst I site is likely to be located close to or within exon 3.

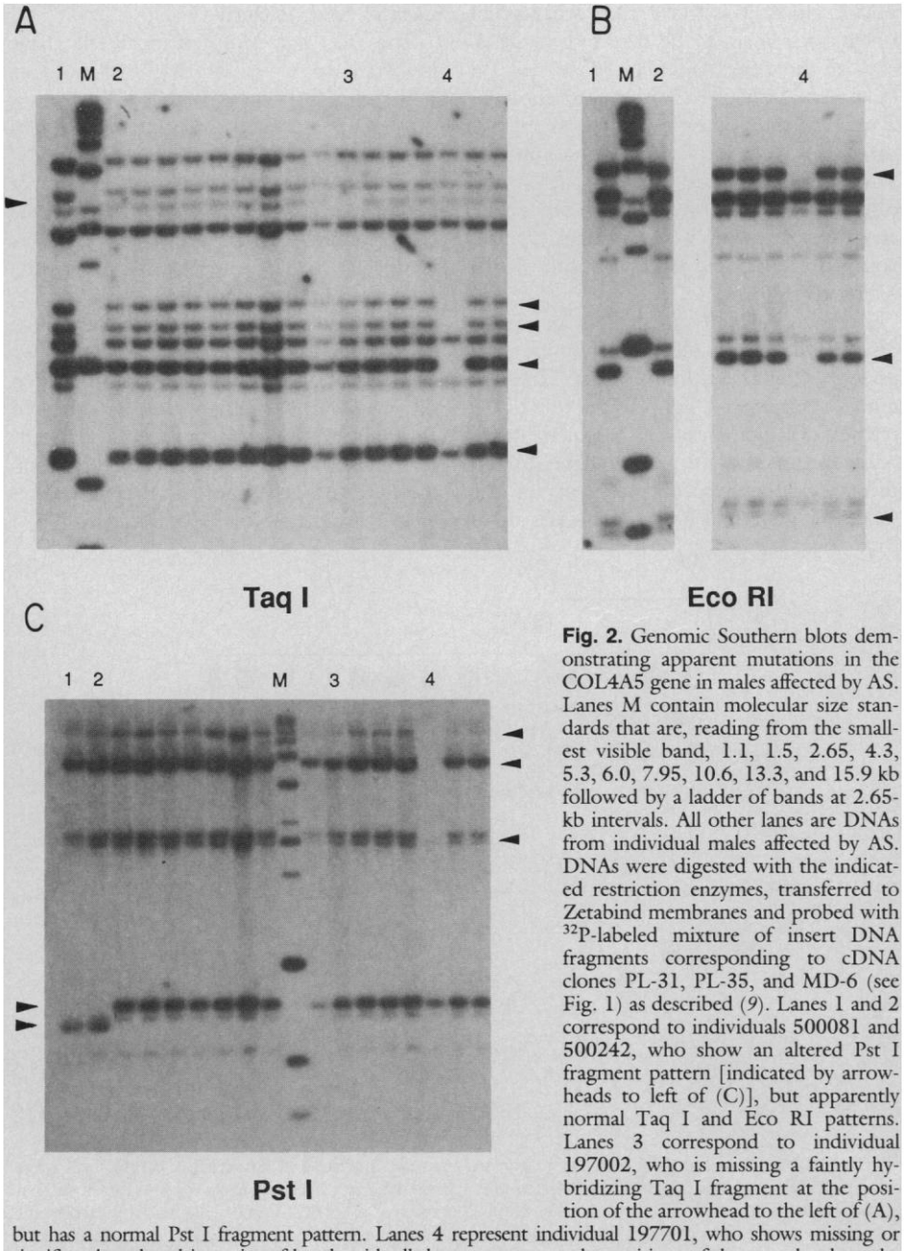

Fig. 2. Genomic Southern blots demonstrating apparent mutations in the COL4A5 gene in males affected by AS. Lanes M contain molecular size standards that are, reading from the smallest visible band, 1.1, 1.5, 2.65, 4.3, 5.3, 6.0, 7.95, 10.6, 13.3, and 15.9 kb followed by a ladder of bands at 2.65-kb intervals. All other lanes are DNAs from individual males affected by AS. DNAs were digested with the indicated restriction enzymes, transferred to Zetabind membranes and probed with ^{32}P-labeled mixture of insert DNA fragments corresponding to cDNA clones PL-31, PL-35, and MD-6 (see Fig. 1) as described (*9*). Lanes 1 and 2 correspond to individuals 500081 and 500242, who show an altered Pst I fragment pattern [indicated by arrowheads to left of (C)], but apparently normal Taq I and Eco RI patterns. Lanes 3 correspond to individual 197002, who is missing a faintly hybridizing Taq I fragment at the position of the arrowhead to the left of (A), but has a normal Pst I fragment pattern. Lanes 4 represent individual 197701, who shows missing or significantly reduced intensity of bands with all three enzymes, at the positions of the arrowheads to the right of each panel.

Affected male 197002 from kindred 1970 in Utah shows the absence of a faintly hybridizing 6.1-kb Taq I fragment (lanes 3 in Fig. 2). The weak hybridization signal has made further characterization of this variant difficult to pursue with the currently available probes.

The present results demonstrate three different mutations in the COL4A5 gene in three kindreds with AS, supporting the conclusion that mutations in this gene account for at least a portion and possibly all X-linked AS. The 5′ portion of the α5(IV) gene has not yet been characterized, and it is possible that mutations in that region could be responsible for the disease in kindreds that did not show an abnormal Southern pattern with the 3′ cDNA probes used here. Furthermore, it is likely that other point mutations remain to be detected in the 3′ end of the gene, since we have not yet examined it exhaustively with restriction enzymes or other detection methods. Finally the α1(IV) and α2(IV) genes on chromosome 13 are located head to head at a distance of only 42 bp from each other (*12*), and similarly adjacent but distinct type IV collagen genes could together represent the X-linked AS locus.

Thus, we have identified a human genetic disease affecting a basement membrane component. Characterization of the different mutations may shed light on the role of the type IV collagen chains in basement membrane structure. The present data suggest that a wide variety of mutations of the COL4A5 gene will be found in AS, as in osteogenesis imperfecta, where mutations of varying phenotypic severity occur in COL1A1 and COL1A2 (*13*). Basement membranes in different tissues must have differences in structural composition reflecting biological function: monoclonal antibodies to NC fragments of GBM type IV collagen indeed showed organ-specific distribution in basement membranes (*5*). Although it has been shown that the α5(IV) chain is expressed in many other tissues, such as in lung and spleen (*7*), AS patients have not been reported to have complications in these organs. The basis for the organ specificity and a molecular understanding of how a structurally defective α5(IV) chain results in the loss of GBM integrity and hematuria in AS patients remain to be established.

The fact that AS patients usually have moderate to severe hearing loss suggests that the α5(IV) chain is also an important component in the basement membrane of a sensory portion of the inner ear. The finding of mutations that lie entirely within the boundaries of COL4A5 in two kindreds cosegregating for nephritis and deafness demonstrates that mutations in a single gene can cause both phenotypes. Furthermore, the deafness of the individual with the large deletion is among the most profound we have seen in AS kindreds (*15*), implying that the severity of deafness is directly correlated with the severity of the defect in COL4A5.

It is likely that further detailed studies on the COL4A5 gene in other affected families will lead to the molecular characterization of specific disease-causing mutations, allowing prenatal diagnosis and detection of clinically normal gene-carrier females in families at risk. To determine the total genetic burden associated with mutations of COL4A5, it will be important to examine this gene in sporadic cases of nephritis and nephritis associated with deafness as well as cases of familial or sporadic deafness in which nephritis is not apparent. Eventually, understanding of the relation between the nature of the molecular defects of COL4A5 and phenotypic expression at the clinical and ultrastructural levels may lead to the design of appropriate therapeutic strategies.

Note added in proof: By PCR amplification and DNA sequencing, the Pst I variant in kindred P has been shown to be due to a single base pair change that converts a highly conserved, NC domain cysteine codon to a codon for serine (*17*).

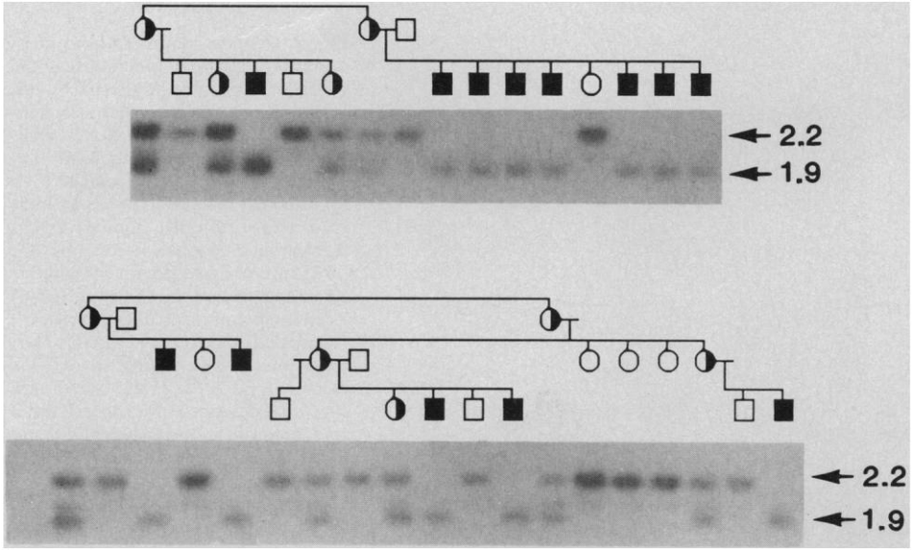

Fig. 3. Inheritance of the COL4A5 Pst I site variant in two portions of kindred P. Southern transfers were prepared and probed as described in the legend to Fig. 2. Only the region of the pattern containing the relevant 2.2- and 1.9-kb bands (arrows) is shown. A faintly hybridizing constant fragment is present at a position corresponding to approximately 1.95 kb. Obligate carrier mothers or females with diagnosed renal phenotype are shown as half-darkened circles. Unaffected females are open circles. Unaffected males are open boxes, and affected males are darkened boxes.

REFERENCES AND NOTES

1. C. L. Atkin, M. C. Gregory, W. A. Border, in *Diseases of the Kidney*, R. W. Schrier and C. W. Gottschalk, Eds. (Little, Brown, Boston, ed. 4, 1988), pp. 617–641.
2. C. L. Atkin *et al.*, *Am. J. Hum. Genet.* **42**, 249 (1988); H. Brunner *et al.*, *Kidney Int.* **34**, 507 (1988); S. Szpiro-Tapia *et al.*, *Hum. Genet.* **81**, 85 (1988); F. A. Flinter, S. Abbs, M. Bobrow, *Genomics* **4**, 335 (1989); D. F. Barker *et al.*, in preparation.
3. G. S. Spear, *Clin. Nephrol.* **1**, 336 (1973).
4. C. E. Kashtan, C. L. Atkin, M. C. Gregory, A. F. Michael, *Kidney Int.* **36**, 669 (1989); C. O. S. Savage *et al.*, *Lab. Invest.* **60**, 613 (1989).
5. M. M. Kleppel, C. Kashtan, P. A. Santi, J. Wieslander, A. F. Michael, *Lab. Invest.* **61**, 278 (1989).
6. C. D. Boyd *et al.*, *Hum. Genet.* **74**, 121 (1986); C. Griffin, B. S. Emanuel, J. R. Hansen, W. K. Cavenee, J. C. Myers, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 512 (1987).
7. S. L. Hostikka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1606 (1990).
8. J. Zhou, S. L. Hostikka, L. T. Chow, K. Tryggvason, in preparation.
9. D. Barker *et al.*, *Am. J. Hum. Genet.* **36**, 1159 (1984); L.-C. Tsui *et al.*, *Science* **230**, 1054 (1985).
10. D. F. Barker *et al.*, unpublished observations.
11. S. J. Hasstedt, C. L. Atkin, A. C. San Juan, Jr., *Am. J. Hum. Genet.* **38**, 940, 1986.
12. S. Soininen, M. Huotari, S. L. Hostikka, D. J. Prockop, K. Tryggvason, *J. Biol. Chem.* **263**, 17217 (1988).
13. D. J. Prockop and K. I. Kivirikko, *N. Engl. J. Med.* **311**, 376 (1984); B. Sykes *et al.*, *Am. J. Hum. Genet.* **46**, 293 (1990).
14. J. Saus, J. Wieslander, J. P. M. Langveld, S. Quinones, B. G. Hudson, *J. Biol. Chem.* **263**, 11374 (1988); R. J. Butkowski, J. Wieslander, M. Kleppel, A. F. Michael, A. J. Fish, *Kidney Int.* **35**, 1195 (1989).
15. D. Wester, thesis, University of Utah, Salt Lake City (1990); unpublished data.
16. R. Soininen, M. Huotari, A. Ganguly, D. J. Prockop, K. Tryggvason, *J. Biol. Chem.* **264**, 13565 (1989); S. L. Hostikka and K. Tryggvason, *FEBS Lett.* **224**, 297 (1987).
17. J. Zhou, D. F. Barker, S. L. Hostikka, S. C. Gerken, C. L. Atkin, K. Tryggvason, in preparation.

18. We thank the families of patients with Alport syndrome for their cooperation and P. R. Fain for a critical reading of the manuscript. M. Risman organized the collection of blood specimens and maintained clinical records. Technical assistance was provided by L. P. Wilder, W. L. Andersen, and L. Arffman. Supported by NIH grants DK 39497 (C.L.A.) and DK 36200 (L.T.C.) from the National

Institute of Diabetes and Digestive and Kidney Diseases, Department of Energy grant DE-FG02-88ER60689 (D.F.B.), and grants from the Hereditary Nephritis Foundation (C.L.A.), the Academy of Finland (K.T. and S.L.H.), and the Juselius Foundation (K.T.). C.L.A. acknowledges further support from PHS grant M01 RR 00064 from the Division of Research Resources.

19 March 1990; accepted 2 May 1990

Normal Development of Mice Deficient in β_2M, MHC Class I Proteins, and CD8$^+$ T Cells

BEVERLY H. KOLLER, PHILIPPA MARRACK, JOHN W. KAPPLER, OLIVER SMITHIES*

Major histocompatibility class I proteins display viral and self antigens to potentially responsive cells and are important for the maturation of T cells; β_2-microglobulin (β_2M) is required for their normal expression. Mouse chimeras derived from embryonic stem cells with a disrupted β_2M gene transmitted the inactivated gene to their progeny. Animals homozygous for the mutated β_2M gene were obtained at expected frequencies after further breeding. The homozygotes appeared normal, although no class I antigens could be detected on their cells and the animals are grossly deficient in CD4$^-$CD8$^+$ T cells, which normally mediate cytotoxic T cell function.

AT LEAST 26 CLASS I HEAVY CHAIN genes are encoded in the classic major histocompatibility complex (MHC) and the adjacent Qa/Tla region of the mouse genome (*1*). On cell surfaces these heavy chains are noncovalently associated with a 99–amino acid light chain, β_2M (*2*). There have been reports suggesting that β_2M has additional functions that are independent of its association with the class I heavy chains (*3*); most notably, β_2M is identical to thymotaxin and may be involved in migration of hemopoetic stem cells from the bone marrow (*4*). In cell lines that fail to express β_2M, class I heavy chains accumulate in the endoplasmic reticulum and are not further processed to give cell surface proteins (*5*). Beginning at the midsomite stage of embryogenesis (*6*), the classical class I antigens (H-2K and H-2D/L in the mouse) are expressed on the cell surfaces of almost all somatic tissues where, upon binding fragments of self and foreign antigens, they become potential ligands for αβ T cell receptors (*7*). During development, T cells possessing receptors that bind strongly to complexes of MHC proteins and self antigens are eliminated by the thymus (*8*). T cells possessing receptors with low affinity for self MHC are allowed to mature by a process known as positive selection, and subsequently they migrate to the periphery. Positive selection is also thought to determine the phenotypes of the mature T cells. Interaction of the αβ receptors on thymocytes with class I MHC leads to a mature T cell bearing CD8; interaction of the thymocyte αβ receptor with class II MHC molecules leads to a mature T cell bearing CD4 (*9*).

A great deal has been learned about the selection of class I restricted T cells, that is, T cells that recognize antigens associated with specific class I proteins. Less is known about the involvement of class I restricted T cells in the immune response to various pathogens, in immune surveillance, and in the regulation of the immune system. Furthermore, although the structure and expression of the nonpolymorphic class I related proteins (Qa/Tla) has been well characterized, their function has remained elusive. Some of these nonpolymorphic class I genes are expressed during early stages of embryogenesis (*10*), and the suggestion has been made that they may be involved in cell-cell interactions during development. Their early expression and their restriction to specific lymphoid populations in the mature animal has led to their designation as differentiation antigens. In addition, there have been reports that the interaction of class I molecules with a number of cell surface proteins may be of biological significance (*11*) and that class I molecules secreted in urine may provide olfactory signals that influence mating strategies (*12*).

The various functions that have been ascribed to class I proteins, and vicariously to β_2M, suggested to us the value of studying the effects in embryos and mature animals of an absence of β_2M and its dependent proteins. We therefore disrupted the β_2M gene in mouse embryonic stem (ES) cells by homologous recombination (*13*), as did Zijlstra *et al.* (*14*). ES cells are pluripotent and, when introduced into mouse blastocysts, can contribute to all tissues of the resulting chimeric animals, including the gonads (*15*). Alterations made in the ES cell genome can therefore be passed on to subsequent generations (*14, 16*). Two independent ES cell lines, E1439B and E1422A, carrying a disrupted β_2M gene were isolated (*13*), but only chimeras generated with cell line E1422A showed transmission (Table 1). The five male and two female (*17*) transmitting chimeras were mated to

Table 1. Generation of chimeric offspring. Chimeras were scored by the presence of agouti, chinchilla, or cream coat color contributed by the E14TG2a-derived ES cells (129/Ola, A^w, c^{ch}, p) on the black background of C57BL/6 mice.

Cell line	Blastocysts transferred	Progeny born (% of blastocysts transferred)	Chimeras (% of progeny born)	Female	Male	Transmitting chimeras
E1439A	144	51(35)	39(76)	17	22	0
E1422A	108	38(35)	30(79)	17	13	7

Table 2. Breeding data from transmitting chimeras. The percent transmission of the ES genome was scored by the presence of agouti offspring after breeding with either C57BL/6 or B6/D2 mice.

Mouse number	Sex	Number of litters	ES cell-derived progeny/total progeny (% transmission)
28.2	M	6	3/28 (11)
29.2	M	6	13/50 (26)
30.1	F	2	3/6 (50)
31.2	M	2	8/8 (100)
31.2	F	1	3/3 (100)
33.1	M	4	10/23 (43)
35.2	M	4	10/21 (48)

B. H. Koller and O. Smithies, Department of Pathology, University of North Carolina, Chapel Hill, NC 27599–7525.
P. Marrack and J. W. Kappler, Howard Hughes Medical Institute at Denver, Department of Medicine, National Jewish Center for Immunology and Respiratory Medicine, Denver, CO 80206.