The EMBO Journal vol.3 no.9 pp.2029–2032, 1984

The E_β^b gene may have acted as the donor gene in a gene conversion-like event generating the A_β^{bm12} mutant

M.Denaro, U.Hammerling, L.Rask and P.A.Peterson

Department of Cell Research, The Wallenberg Laboratory, University of Uppsala, Box 562, S-751 22 Uppsala, Sweden

Communicated by P.A.Peterson

At least two different class II histocompatibility antigens, I-A and I-E, are encoded by the murine major histocompatibility complex. Both types of class II antigens are composed of polypeptide chains called α and β. Class II antigens display extensive genetic polymorphism, the main part of which resides in the NH_2-terminal domains of the A_α, A_β and E_β chains. Recently it was shown that the mutant gene A_β^{bm12} differed from the wild-type gene A_β^b by three nucleotide substitutions, which all occur within a stretch of 14 nucleotides. Multiple substitutions of the type found in the A_β^{bm12} gene suggest that the mutant arose by a gene conversion-like event. To examine whether the E_β^b gene may have served as the donor gene in the generation of the A_β^{bm12} gene, we have isolated and sequenced a cDNA clone corresponding to the E_β^b gene. Comparisons of the E_β^b, the A_β^b and the A_β^{bm12} nucleotide sequences revealed that the E_β^b sequence is identical to that of A_β^{bm12} in the positions where the latter differs from the A_β^b sequence. This observation is consistent with the notion that the A_β^{bm12} mutant gene arose by a gene conversion-like event involving the E_β^b gene.

Key words: major histocompatibility complex/E_β^b gene/ A_β^{bm12} mutant/gene conversion event

Introduction

The major histocompatibility complex (MHC) antigens of the class I and II types belong to the same family as the immunoglobulins (Larhammar *et al.*, 1982). At least one chain of the T-cell antigen receptor is also a member of this family (Hedrick *et al.*, 1984; Yanagi *et al.*, 1984). A common feature of many members of this family of proteins is their extensive genetic polymorphism (Klein and Figuerora, 1981). However, allelic variants of some members are rare, e.g., of β_2-microglobulin (Robinson *et al.*, 1981) Thy-1 (Williams and Gagnos, 1982) and the secretory component (Mostov *et al.*, 1984). The mechanisms generating the polymorphism of the MHC antigens are not known. However, gene conversion or similar mechanisms may contribute to the polymorphism of class I MHC antigens (Weiss *et al.*, 1983; Pease *et al.*, 1983) in much the same way as has been suggested for immunoglobulin germ-line V-genes (Baltimore, 1981). Whether such gene conversion-like events may explain most of the polymorphism in the MHC or whether they just contribute marginally, remains unclear.

Class II antigens, which are cell surface-expressed molecules composed of one α and one β chain, are derived from adjacent genes in the I-region of the murine MHC (Steinmetz *et al.*, 1982). Two types of class II antigens, called I-A and I-E, having different α and β chains have been identified (Jones, 1977; Uhr *et al.*, 1979). I-A as well as I-E β genes are polymorphic (Charron and McDevitt, 1979; Cook *et al.*, 1979). Recently, McIntyre and Seidman (1984) proposed that a mutant of the I-A_β^b gene may have been generated by a conversion event involving the I-E_β^b gene as the donor gene. To establish that a mutant is generated by a conversion event it is necessary to identify the donor gene. We describe here the isolation and the characterization of a cDNA clone that codes for the murine I-E_β^b chain. Analysis of the nucleotide sequence suggests that the I-E_β^b gene may be the donor gene in the event that generated the A_β^{bm12} mutant gene.

Results and Discussion

Isolation and characterization of the cDNA clone pEBB24

A cDNA library consisting of 10 000 colonies was constructed from total poly(A)+ mRNA of spleens of C57BL/6 mice (b haplotype). The library was screened with a 600-bp restriction fragment of the $E_{\beta2}$ gene as the probe in a search for a cDNA clone corresponding to this gene. The fragment used was subcloned from the cosmid c39.1 derived from a BALB/c library and corresponds to the entire second domain exon of the $E_{\beta2}$ gene (Steinmetz *et al.*, 1982). The homology between the $E_{\beta2}$ second domain exon and that of the E_β gene is >80% (data not shown). We therefore expected that even under stringent conditions we would detect not only putative $E_{\beta2}$ cDNA clones but also clones corresponding to the E_β gene.

Five cDNA clones were isolated and subjected to partial restriction enzyme mapping. All clones contained sites characteristic for E_β gene exons. The largest E_β cDNA clone, pEBB24, was further restriction mapped and sequenced. Figure 1A depicts the restriction map of the clone together with the strategy employed to determine its nucleotide sequence.

The insert of clone pEBB24 consists of 1054 bp and, as shown in Figure 1B, its sequence starts at the last nucleotide of the codon for amino acid −4 of the signal sequence and ends with a short poly(A) tail. A termination signal is present at base pair 725 and it is followed by a 340 bp long 3′-untranslated region that includes a polyadenylation signal (see Figure 1B). The amino acid sequence predicted from the nucleotide sequence encompasses the last three amino acids of the signal sequence and the 238 amino acids that form the mature protein.

The predicted E_β^b amino acid sequence shows one glycosylation site in the first domain. The first and second domains each display two cysteine residues that probably interact to form intra-domain disulfide bonds. The E_β^b sequence contains an additional cysteine residue at position 12 in agreement with the sequence of the E_β^k allele (Mengle-Gaw and McDevitt, 1983). Thus, the E_β^b chain shows the characteristic features of other murine and human class II β chains (Larhammar *et al.*, 1982; Gustafsson *et al.*, 1984; Saito *et al.*, 1983; Larhammar *et al.*, 1983; Mengle-Gaw and McDevitt, 1983).

© IRL Press Limited, Oxford, England.

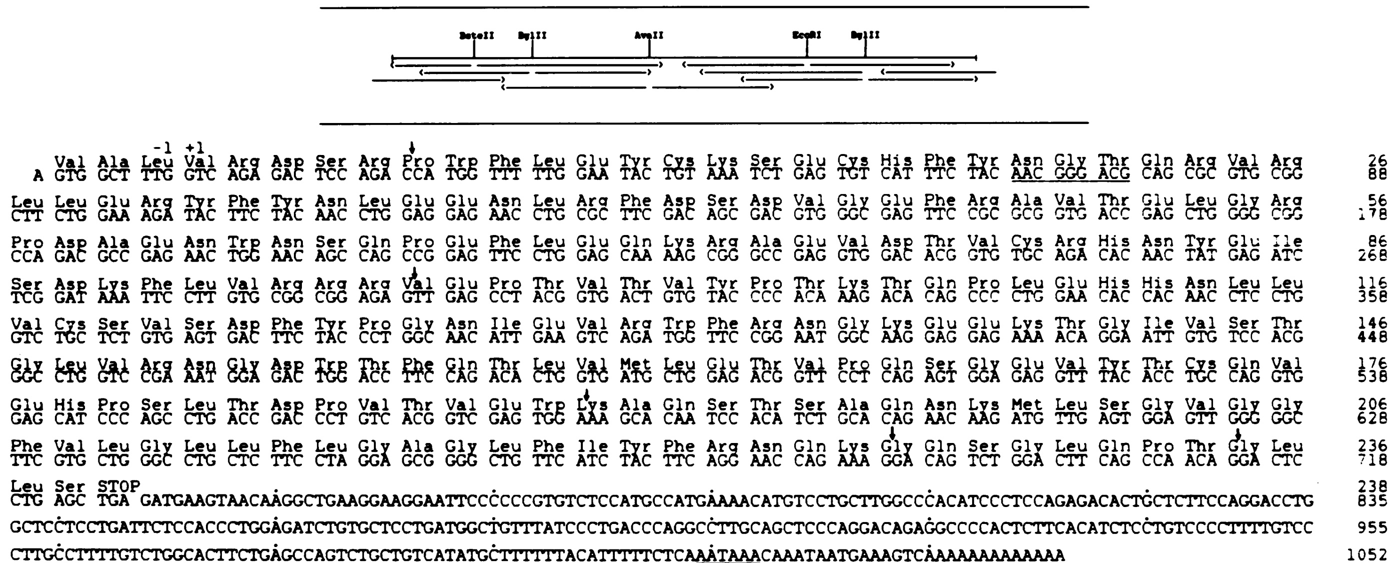

Fig. 1. (A) Restriction map of the I-E^b_β cDNA clone pEBB24. Only restriction sites used for Maxam and Gilbert chemical degradation procedure are shown. Arrows show the direciton of sequencing. **(B)** Nucleotide sequence and predicted amino acid sequence of the I-E^b_β cDNA clone pEBB24. Arrows mark exon boundaries according to the organization of the E^d_β gene (Saito *et al.*, 1983). The attachment site for the N-linked carbohydrate and the polyadenylation signal are underlined.

Comparisons of the nucleotide and amino acid sequences of three E_β alleles

The nucleotide and amino acid sequences of three E_β alleles, i.e., the b, the d (Saito *et al.*, 1983) and the k alleles (Mengle-Gaw and McDevitt, 1983), were compared (Figure 2). As for other murine and human allelic class II chains, the amino acid replacements are predominantly located in the first domain while the second domain is well conserved (Kämpe *et al.*, 1983; Benoist *et al.*, 1983; Choi *et al.*, 1983; Schenning *et al.*, 1984; Gustafsson *et al.*, 1984).

A conspicuously high number of nucleotide substitutions are present in the codons for amino acids 2, 3 and 4, notably in the k allele. No allelic variation has been found at the corresponding positions in other β genes (Choi *et al.*, 1983; Gustafsson *et al.*, 1984). Moreover, the NH_2-terminal sequences of the I-E^k and I-E^d β chains, determined at the protein level (Silver *et al.*, 1979; Cook *et al.*, 1979) are in good agreement with the amino acid sequence predicted for the I-E^b β chain. The substitutions found in the E^k_β allele and possibly also in the E^d_β allele might therefore be due to cloning artefacts or sequencing mistakes and should be investigated further by analyses of independent clones.

Disregarding the differences at codons 2, 3 and 4, only five nucleotide substitutions leading to four amino acid replacements are found in the first domain exon of the E^b_β gene compared with the E^k_β allele (Figure 2). The first domain exon of the E^d_β allele contains 22 nucleotide substitutions giving rise to 14 amino acid replacements compared with the E^b_β sequence. The nucleotide sequences encoding the second domain, the membrane, and the cytoplasmic portions are virtually identical between the E_β alleles, displaying just a few scattered substitutions. These results are consistent with those of serological and biochemical studies suggesting that the E^d_β allele is less similar to other E_β alleles than they are to each other (Kupinsky *et al.*, 1982).

Four of the five nucleotide substitutions between the E^b_β and E^k_β sequences are located within a stretch of 19 bp in codons 87 – 93. The sequence of this stretch in the E^b_β allele is completely identical to the corresponding portion of E^d_β gene. Based on these sequences it is conceivable that the E^b_β allele might have arisen by a crossing-over event that occurred between the E^k_β and E^d_β alleles at around base pairs 79 and 87. The 5′ end of the gene would then be derived from E^k_β and the 3′ end from E^d_β. However, this suggestion is only consistent with available data about the I-region provided a second cross-over event occurred between the A_α and the E_β genes, since the A^k_α gene is no more similar to A^b_β than is A^d_β. Another alternative is that the E^b_β allele has arisen by gene conversion-like events involving the E^d_β and E^k_β or related alleles. A third possibility, which by no means can be excluded, is that the three alleles have evolved independently from a common ancestral gene by accumulation of multiple point mutations. Available knowledge is not sufficient to allow a distinction to be made as to which extent, if any, the different mechanisms have contributed in generating the different E_β alleles.

Comparison of the nucleotide sequences of the E^b_β and the A^{bm12}_β

Recently, the nucleotide sequence of a mutant of the I-A^b_β gene called A^{bm12}_β, was published (McIntyre and Seidman, 1984). The A^{bm12}_β mutant displayed three nucleotide substitutions in the first domain exon compared with the A^b_β wild-type sequence. On the basis of all mutations occurring within a stretch of 14 bp the authors proposed that the substitutions might have originated from a related gene after a gene conversion-like event. The E^b_β gene was suggested to have acted as the donor gene. In order to test this hypothesis the E^b_β sequence was compared with the A^b_β and the A^{bm12}_β sequences (Figure 3). It is obvious that the nucleotide stretch of 14 bp containing the three substitutions was completely identical between the E^b_β and the A^{bm12}_β sequences. Consequently, this finding is consistent with the idea that a transfer of genetic material involving from 14 to 44 bp might have occurred from the E^b_β to the A^b_β gene to generate the A^{bm12}_β mutant.

Fig. 2. Comparison of nucleotide and predicted amino acid sequences of the E_β^b, E_β^k and E_β^d alleles. Arrows mark exon boundaries.

Conclusions

Analysis of the E_β^b nucleotide sequence showed that this gene may have acted as the donor sequence in a gene conversion-like event generating the A_β^{bm12} mutant. The size of the nucleotide stretch, 14–44 bp, implicated in the event is similar in length to the nucleotide stretch that has been suggested to have been transferred from a Qa-gene to the H-2K^b gene to create the H-2K^{bm1} mutant. This latter case is so far the only case of a possible gene conversion-like event analysed regarding class I MHC antigens (Weiss *et al.*, 1983; Pease *et al.*, 1983; Mellor *et al.*, 1984).

Gene conversion as it is known in yeast normally involves longer nucleotide stretches, 200 bp and longer. Such events tend to homogenize sequences since they lead to the loss of genetic information in the individual. To contribute to polymorphism of MHC antigens the sequences involved in genetic exchange events must be short. In fact, two cases, the A_β^{bm12} and the K^{bm1} mutants, that have been examined provide suggestive evidence that gene conversion-like events occur among the MHC genes. Although the individual loses genetic information by such genetic exchange, the genetic information of the species obviously increases. However, it should be borne in mind that the existence of molecular mechanisms allowing short nucleotide segments with several mismatches to interact specifically with homologous genes remains to be established.

It cannot either be excluded that both the A_β^{bm12} and the

```
                 50                                                    60
Aβb         CGC GCG GTG ACC GAG CTG GGG CGG CCA GAC GCC GAG TAC TGG AAC AGC CAG CCG GAG
            Arg Ala Val Thr Glu Leu Gly Arg Pro Asp Ala Glu Tyr Trp Asn Ser Gln Pro Glu
Aβbm12

Eβb                                                         A
                                                            Asn

                     70                                                      80
Aβb         ATC CTG GAG CGA ACG CGG GCC GAG CTG GAC ACG GTG TGC AGA CAC AAC TAC GAG
            Ile Leu Glu Arg Thr Arg Ala Glu Leu Asp Thr Val Cys Arg His Asn Tyr Glu
Aβbm12      [T  ]       [ A   A ]
            [Phe]       [Gln Lys]
Eβb         [T  ]       [ A   A ]         G                                T
            [Phe]       [Gln Lys]         Val
```

Fig. 3. Comparison of nucleotide and predicted amino acid sequences of the E_β^b, A_β^b and A_β^{bm12} genes from codons 49 to 85. Positions where the A_β^{bm12} sequence differs from the A_β^b sequence but is identical to that of E_β^b are within boxes.

K^{bm1} mutants might have arisen by separate point mutations accumulated over a period of time. Actually, all polymorphism of class II antigens so far examined may be accounted for by independent and random point mutations, provided there are no strong selective forces operating on the polymorphic exons (Gustafsson *et al.*, 1984). Thus, it is as yet too early to decide to which extent gene conversion-like events and point mutations, respectively, contribute to the polymorphism of class II antigens.

Materials and methods

Construction and screening of cDNA library

Total mRNA was prepared from spleens of C57BL/6 mice essentially according to the protocol of Auffray and Rougeon (1980) with minor modifications. Briefly, tissue frozen in liquid nitrogen was homogenized in 3 M LiCl, 6 M urea and 10 mM Vanadyl Ribonucleotide Complex. After 5 h on ice the homogenized material was centrifuged, re-dissolved in 10 mM Tris, pH 7.5, containing 5 mM EDTA and 1% SDS, treated with phenol and chloroform, and finally precipitated with ethanol. This RNA preparation was enriched for poly(A)$^+$ mRNA, by two passages on oligo(dT)-cellulose.

Double-stranded cDNA was synthesized according to Land *et al.* (1981). Tailing, annealing and transformation was according to Maniatis *et al.* (1982). The cDNA library was screened by colony hybridization (Maniatis *et al.*, 1982) with the nick-translated (Davies *et al.*, 1980), 600-bp *Bgl*II fragment containing the second domain exon of the $E_{\beta 2}$ gene from cosmid c39.1 as the probe (Steinmetz *et al.*, 1982).

Nucleotide sequence determination

Nucleotide sequences were determined by the chemical degradation procedure (Maxam and Gilbert, 1980).

Acknowledgements

We are grateful to Dr. M.Steinmetz for the gift of the cosmid c39.1. We also thank Ms. Elisabetta Rossi for invaluable help with the computer analyses. M.D. especially wishes to thank her for her patience and understanding. We would also like to thank Dan Larhammar and Kenth Gustafsson for advice and helpful discussions. This work was supported by grants from the Swedish Cancer Society, King Gustaf V's 80-years' Fund, and Marcus Borgström's Fund.

References

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303-314.
Baltimore,D. (1981) *Cell*, **24**, 592-594.
Benoist,C.O., Mathis,D.J., Kanter,M.R., Williams,V.E. and McDevitt,H.O. (1983) *Cell*, **34**, 169-177.
Charron,D.J. and McDevitt,H.O. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 6567-6571.
Choi,E., McIntyre,K., Germain,R.N. and Seidman,J.G. (1983) *Science (Wash.)*, **221**, 283-286.
Cook,R.G., Siegelman,M.H., Capra,J.D., Uhr,J.W. and Vitetta,E.S. (1979) *J. Immunol.*, **122**, 2232-2237.
Davies,R.W., Botstein,D. and Roth,J.R. (1980) *Advanced Bacterial Genetics*, published by Cold Spring Harbor Laboratory Press, NY.
Gustafsson,K., Wiman,K., Emmoth,E., Larhammar,D., Böhme,J., Hyldig-Nielsen,J.J., Ronne,H., Peterson,P.A. and Rask,L. (1984) *EMBO J.*, **3**, 1655-1661.
Hedrick,S.M., Nielsen,E.A., Kayalea,J., Cohen,D.I. and Davis,M.M. (1984) *Nature*, **308**, 153-158.
Jones,P. (1977) *J. Exp. Med.*, **146**, 1266-1272.
Kämpe,O., Larhammar,D., Wiman,K., Schenning,L., Claesson,L., Gustafsson,K., Pääbo,P., Hyldig-Nielsen,J.J., Rask,L. and Peterson,P.A. (1983) in Möller,E. and Möller,G. (eds.), *Genetics of the Immune Response*, Plenum Publ. Corporation, pp. 61-79.
Klein,J. and Figueroa,F. (1981) *Immunol. Rev.*, **60**, 23-57.
Kupinsky,J.M., Gamble,A.P., David,C.S. and Freed,J.F. (1982) *Immunogenetics*, **16**, 393-405.
Land,H., Grez,M., Hauser,H., Lindenmaier,W. and Schütz,G. (1981) *Nucleic Acids Res.*, **9**, 2251-2266.
Larhammar,D., Gustafsson,K., Claesson,L., Bill,P., Wiman,K., Schenning, L., Sundelin,J., Widmark,E., Peterson,P.A. and Rask,L. (1982) *Cell*, **30**, 153-161.
Larhammar,D., Hammerling,U., Denaro,M., Lund,T., Flavell,R.A., Rask, L. and Peterson,P.A. (1983) *Cell*, **34**, 179-188.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) in *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
McIntyre,K.R. and Seidman,J.G. (1984) *Nature*, **308**, 551-553.
Mellor,A.L., Weiss,E.H., Kress,M., Jay,G. and Flavell,R.A. (1984) *Cell*, **36**, 139-144.
Mengle-Gaw,L. and McDevitt,H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7621-7625.
Mostov,K., Friedlander,M. and Blobel,G. (1984) *Nature*, **308**, 37-43.
Pease,L.R., Schulze,D.H., Pfaffenbach,G.M. and Nathenson,S.G. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 242-246.
Robinson,P.J., Graf,L. and Sege,K. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1167-1170.
Saito,H., Maki,A.R., Clayton,L.K. and Tonegawa,S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5520-5524.
Schenning,L., Larhammar,D., Bill,P., Wiman,K., Jonsson,A.K., Rask,L. and Peterson,P.A. (1984) *EMBO J.*, **3**, 447-452.
Silver,J., Walker,L.E., Reisfeld,R.A., Pellegrino,M.A. and Ferrone,S. (1979) *Mol. Immunol.*, **16**, 37-42.
Steinmetz,M., Minard,K., Horvath,S., McNicholas,J., Frelinger,J., Wake, C., Long,E., Mach,B. and Hood,L. (1982) *Nature*, **300**, 35-42.
Uhr,J.W., Capra,J.D., Vitetta,E.S. and Cook,R.G. (1979) *Science (Wash.)*, **206**, 292-297.
Weiss,E.H., Mellor,A., Golden,L., Fahrner,K., Simpson,E., Hurst,J. and Flavell,R.A. (1983) *Nature*, **301**, 671-674.
Williams,A.F. and Gagnon,J. (1982) *Science (Wash.)*, **216**, 696-703.
Yanagi,Y., Yoshikai,Y., Leggett,K., Clark,S., Aleksander,I. and Mak,T. (1984) *Nature*, **308**, 145-149.

Received on 8 June 1984

Note added in proof

A sequence error involving a *Bst*NI site, whose second C is methylated in the *E. coli* strain used is present at nucleotide position 652 where it should be a G. This correction does not influence the deduced amino acid sequence and the conclusions.