Evolution of the Mouse β-Globin Genes: A Recent Gene Conversion in the Hbb^S Haplotype[1]

Mark A. Erhart, Kimberlee S. Simons, and Steven Weaver
Department of Biological Sciences, University of Illinois at Chicago

We have determined the complete nucleotide sequence of the two nonallelic adult β-globin genes of the C57BL/10 mouse. These genes, designated β^s and β', show a sequence similarity of 99.6% over the region bordered by the translational start and stop codons. Both β^s and β^t encode functional polypeptide chains that are identical. A comparison of the C57BL/10 β-globin haplotype, *Hbb^s*, with that of the BALB/c mouse, *Hbb^d*, suggests that the two haplotypes have distinct evolutionary histories. The two adult β-globin genes of the *Hbb^d* haplotype, β^{dmaj} and β^{dmin}, are 16% divergent at the nucleotide level and encode distinct polypeptides that are synthesized in differing amounts. Our analysis indicates that a gene correction mechanism has been operating on the *Hbb^s* chromosome to keep β^s and β^t evolving in concert, whereas on the *Hbb^d* chromosome, β^{dmin} has diverged considerably from β^{dmaj}. We suggest that gene conversion is responsible for the maintained similarity of the *Hbb^s* genes. Furthermore, we attribute the divergence of the *Hbb^d* genes in part to the absence of a region of simple-sequence DNA within the large intervening sequence of β^{dmin}. We propose that this region of DNA plays a role in facilitating gene conversion. The deletion of this area in β^{dmin} introduced a block of nonhomology between the β^{dmaj}-β^{dmin} gene pair and thus may have inhibited further gene correction within the *Hbb^d* haplotype.

Introduction

Polymorphism in β-globin protein production both in wild populations and among inbred strains of the mouse *Mus musculus* invites an evolutionary analysis of the β-globin gene family at the DNA level. With respect to β chain expression in the adult mouse, three allelic complexes—or haplotypes—have been defined at the β-hemoglobin (*Hbb*) locus—*Hbb^d*, *Hbb^p*, and *Hbb^s* (Hutton et al. 1962). The *Hbb^d* and *Hbb^p* haplotypes are very similar in that they each direct the synthesis of two distinct β chains that are made in unequal amounts. The predominant proteins produced by the two haplotypes βdmajor (*Hbb^d*) and βpmajor (*Hbb^p*) constitute ~80% of β chain synthesis, and the minor species, βdminor (*Hbb^d*) and βpminor (*Hbb^p*), make up the remainder (Gilman 1976). The amino acid sequences of the βdmajor and βpmajor polypeptides are identical, but the minor chains, βdminor and βpminor, differ at two residues (Gilman 1976). The βdminor and βdmajor chains differ at nine of 146 amino acids, while βpminor and βpmajor differ at these same residues as well as at two additional positions. In contrast, the third haplotype, *Hbb^s*, produces only one type of β chain in the adult. This globin chain, termed βsingle, is much like βdmajor and βpmajor, differing at only three amino acid positions (Popp 1973; Gilman 1976).

The Hbb^s haplotype of the C57BL/10 mouse includes two adult β-globin genes, designated $β^s$ and $β^t$ (for explanation of the genetic nomenclature used here, see Material and Methods), both of which have been cloned, restriction mapped, and further characterized by heteroduplex analysis (Weaver et al. 1981). When clones containing the $β^s$ and $β^t$ genes were annealed, an uninterrupted duplex of ~1,800 bp could be observed, indicating that the gross structure of these genes was similar. Analogous experiments with clones containing the $β^{dmaj}$ and $β^{dmin}$ genes from the BALB/c mouse (Tiemeier et al. 1978) had shown that large blocks of nonhomology were present between these two genes. These results suggested that the two adult genes of the Hbb^s haplotype were much more similar than their Hbb^d counterparts but left open the question of whether both $β^s$ and $β^t$ contributed to the production of βsingle polypeptide or whether one of the two genes had been silenced at some point in the recent past.

In order to understand more clearly both the relationship between $β^s$ and $β^t$ and the evolution of the mouse β-globin genes in general, knowledge of the DNA sequence of the $β^s$ and $β^t$ genes was necessary. Here we present the complete nucleotide sequence of the $β^s$ and $β^t$ genes and their flanking sequences. These nonallelic, tandemly duplicated genes exhibit a sequence similarity of 99.6% from the ATG initiation codon through the TAA termination codon. Outside this area, however, these sequences are as different as are $β^{dmaj}$ and $β^{dmin}$. We propose that the identity of $β^s$ and $β^t$ persists owing to the correction of one gene against the other since the duplication of their common ancestor. We suggest that such correction is somehow inhibited in the Hbb^d haplotype, resulting in the divergence of $β^{dmin}$ through the accumulation of both single base changes and insertions/deletions, especially in the second intron.

Material and Methods
Material

Restriction enzymes were purchased from New England Biolabs and Boehringer Mannheim or were gifts from B. Nichols. *Escherichia coli* DNA polymerase I holoenzyme and Klenow fragment were obtained from Boehringer Mannheim; T4 polynucleotide kinase and dideoxy nucleoside triphosphates were obtained from P-L Biochemicals. DNA sequencing primers were purchased from New England Biolabs. ^{32}P-labeled nucleoside triphosphates were obtained from Amersham and New England Nuclear.

Cloning and DNA Sequence Analysis

Restriction fragments containing the $β^s$ and $β^t$ genes and their flanking regions were subcloned from the λ-derived clones BA2 and BA3 (Weaver et al. 1981) into either pBR322 or the M13 vectors mp8 and mp9. DNA fragments contained in pBR322 were either labeled at the 5′ end using $γ^{32}$P-ATP and polynucleotide kinase or by filling in a recessed 3′ end with an appropriate mixture of labeled and unlabeled nucleotides. Then the sequence of the labeled fragments was determined by using the partial chemical degradation procedure of Maxam and Gilbert (1977). DNA fragments cloned into M13 vectors were sequenced via the dideoxy chain termination method described by Sanger et al. (1980).

Genetic Nomenclature

The guidelines recommended at the Mouse Globin Nomenclature meeting (Jackson Laboratory, Bar Harbor, Me., May 21–24, 1984) provide the following

formal genetic designations for the genes referred to in this paper. The nomenclature includes three components—a haplotype designation (here Hbb), a locus or structural gene designation ($b1$ or $b2$), and an allele descriptor (s or d) as a superscript. Under this system, the adult β-globin genes discussed here are as follows: $\beta^{dmaj} \equiv Hbb\text{-}b1^{d}$, $\beta^{dmin} \equiv Hbb\text{-}b2^{d}$, $\beta^{s} \equiv Hbb\text{-}b1^{s}$, and $\beta^{t} \equiv Hbb\text{-}b2^{s}$. We have retained the original gene symbols throughout this article for purposes of clarity and continuity.

Results and Discussion

The β-globin genes of the C57BL/10 mouse strain were isolated from a library of *Eco*RI partial digestion products cloned into the λ-derived vector Charon 4A (Weaver et al. 1981). In figure 1, the physical map of this Hbb^{s} gene cluster is aligned with the map of the Hbb^{d} haplotype of the BALB/c mouse strain (Jahn et al. 1980; Leder et al. 1980). In general, homology between the two haplotypes is extensive, and allelic correspondence between different clones is unambiguous from blot hybridization and heteroduplex analyses (Weaver et al. 1981 and unpublished results).

Restriction fragments containing portions of the β^{s} and β^{t} genes and their immediately flanking sequences were subcloned and sequenced according to the strategy outlined in figure 2. DNA sequence was determined using both the chemical (Maxam and Gilbert 1977) and dideoxy chain termination (Sanger et al. 1980; Messing et al. 1981) procedures. The DNA sequence presented here covers a region
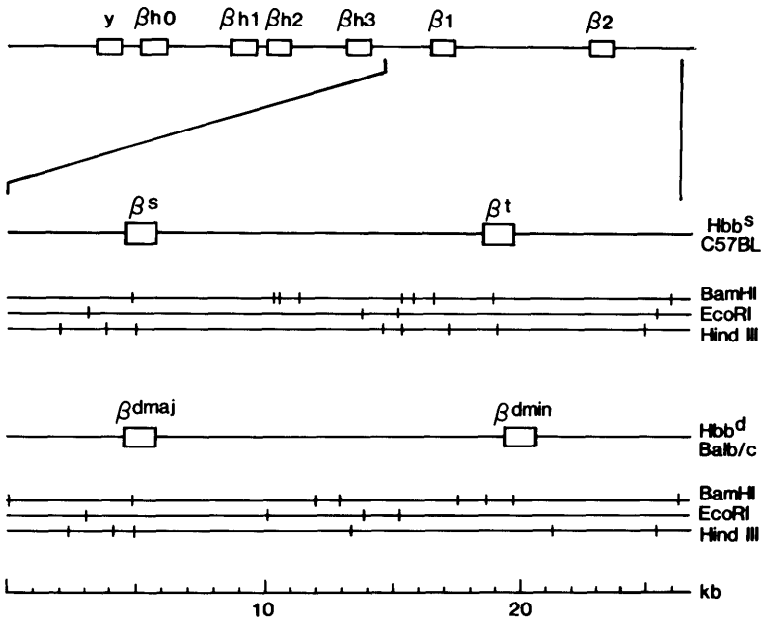
FIG. 1.—The linkage arrangement of the mouse β-like globin gene cluster. At the top, the organization of β-like globin genes in the mouse is shown; it includes the embryonic globin gene *y*; β*h0*, which may encode a minor early embryonic β-like polypeptide (Hill et al. 1984); β*h1*, which encodes the embryonic z protein (Hill et al. 1984); and β*h2* and β*h3*, both of which have been classified as pseudogenes (Jahn et al. 1980; Hutchison et al. 1984; Phillips et al. 1984). The adult β-globin genes, β*1* and β*2*, lie at the right end of the cluster. At the bottom, the restriction maps of the adult regions in both C57BL/10 (Hbb^{s}) and BALB/c (Hbb^{d}) are shown. β^{s} and β^{dmaj} are alleles at the β*1* locus, and β^{t} and β^{dmin} are alleles at the β*2* locus. The location of gene sequences is indicated by the open boxes. The 5′ to 3′ direction of transcription is from left to right. The scale is in kbp.
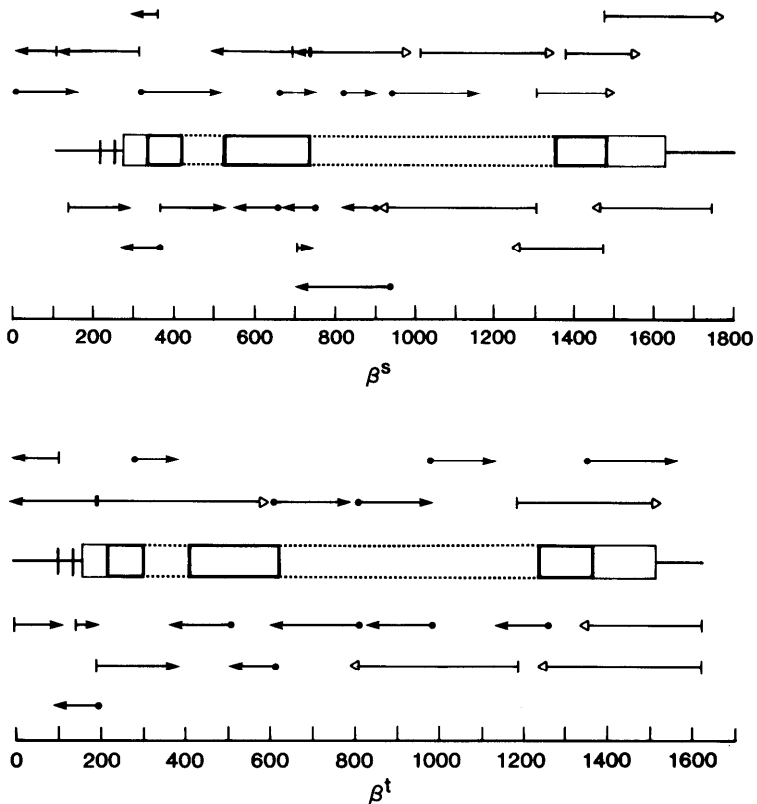
FIG. 2.—DNA sequencing strategy for the β^s and β^t genes. The β^s (top) and β^t (bottom) genes are shown divided into sequenced flanking regions (solid line), nontranslated regions (thin boxes), coding regions (thick boxes), and intervening sequences (dashed boxes). Vertical bars indicate the positions of the CAAT and TATA boxes. Horizontal arrows denote the lengths of the regions of DNA that were sequenced. Sequences were read on both the coding strand (top) and noncoding strand (bottom). Open arrowheads: sequences obtained by the dideoxy chain termination method; solid arrowheads: sequences obtained by the chemical degradation method. The 3'-labeled ends are indicated by solid circles. The scale is in bp.

extending from 200 bp upstream of the mRNA capping site to 150 bp downstream from the poly(A) addition site. Ninety-eight percent of the sequence of both genes was confirmed on both strands, and all restriction sites that separate adjacent fragments were sequenced across.

The DNA sequences of β^s and β^t are aligned with one another and with those of the β^{dmaj} and β^{dmin} genes of the Hbb^d type mouse, BALB/c (Konkel et al. 1979), in figure 3. Both the β^s and β^t genes code for the same polypeptide; this is consistent with the fact that mice homozygous for the Hbb^s haplotype express only one type of β chain as adults. In the following discussion, the relationship between allelic sequences will be considered first, followed by an analysis of the relationship between the β^s and β^t genes.

The β^s and β^{dmaj} Alleles Exhibit Extensive Homology

The DNA sequence of the β^s and β^t genes from the C57BL/10 mouse confirms the amino acid sequence of the βsingle protein of the C57BL/6 mouse reported by Gilman (1976). However, Popp (1973) has reported a slightly different amino acid

```
                                                              -179  cctgagacgtcctaagccagtgagtggcacagc  β^dmaj

-241  aggatgctttagagacagtgttctctgcacagataaggacaaacattattcagagggagtacccagagctgagac-tcctaagccagtcagtggcacagc  β^s

                                    -189  ctcagagtgag-acccagaactcaaatgtcttaagcc------tgggccata  β^t

                      -221  cgaattagctgcgaggataagaacagacactactcagagtgag-acccagaactcaaatgtcttaagcc------tgtgccata  β^dmin

                                                                         CAAT
-146  atgtccagggagaaatatcgcttcgtcctcaccgaagcctgattccgtagagccaca-ccctggtaagggccaatctgctcacacaggatagagagggca  β^dmaj
-142  at--ccagggagaaatat-gcttc-gtcatcaccgaagcctgattccgtagagccaca-ccctggtaagggccaatctgctcacacaggatagagagggca  β^s
-144  gccaccgtgtgtagatat-ggtt-gtcatctctgaagcctcaccctgcagggtaaca-ccctggcattggccaatctgctcagagaggacagagtgggca  β^t
-145  gccaccgtgtagatat-ggtt-gtcatctctgaagcctcaccctgcaaggtaacacccctggcattggccaatctgctcagagaggacagagtgggca  β^dmin

            TATAA                          cAp
-47  ggagccagggcagagcatataaagg-tgaggtaggatcagttgctcctcACATTTGCTTCTGACATAGTTGTGTTGACTCACAACCCCAGAAACAGACATC  β^dmaj
-47  ggagccagggcagagcatataaagg-tgaggtaggatcagttgctcctcACATTTGCTTCTGACATAGTTGTGTTGACTCACAACCCCAGAAACAGACATC  β^s
-47  ggagccagcattgggtatataaagctgag-cagggtcagttgcttcttACGTTTGCTTCTGATTCTGTTGTGTTGACTTGCAACCTCAGAAACAGACATC  β^t
-47  ggagccagcattgggtatataaagctgag-cagggtcagttgcttcttACGTTTGCTTCTGAGTCTGTTGTGTTGACTTGCAACCTCAGAAACAGACATC  β^dmin

     INI              10        Cys           Ser              30
53  ATGGTGCACCTGACTGATGCTGAGAAGGCTGCTGTCTCTTGCCTGTGGGAAAGGTGAACTCCGATGAAGTTGGTGGTGAGGCCCTGGGCAGgttggtat  β^dmaj
53  ATGGTGCACCTGACTGATGCTGAGAAGGCTGCTGTCTCTGGCCTGTGGGAAAGGTGAACGCCGATGAAGTTGGTGGTGAGGCCCTGGGCAGgttggtat  β^s
     ValHisLeuThrAspAlaGluLysAlaAlaValSerGlyLeuTrpGlyLysValAsnAlaAspGluValGlyGlyGluAlaLeuGlyAr
53  ATGGTGCACCTGACTGATGCTGAGAAGTCTGCTGTCTCTTGCCTGTGGGAAAGGTGAACCCCGATGAAGTTGGTGGTGAGGCCCTGGGCAGgttggtat  β^t
53  ATGGTGCACCTGACTGATGCTGAGAAGTCTGCTGTCTCTTGCCTGTGGGCAAAGGTGAACCCCGATGAAGTTGGTGGTGAGGCCCTGGGCAGgttggtat  β^dmin
                                Ser          Cys        Ala         Pro

153  ccaggttacaaggcagctcacaagaagaagtggggtgcttggagacagaggtctgctttccagcagacactaacttcagtgtccctgtctatgtttcc  β^dmaj
153  ccaggttacaaggcagctcacaag---aagttggggtgcttggagacagaggtctgctttccagcaggcactaactttgagtgtccctgtctatgtttcc  β^s
153  ccaggttacaaggcagctcacaagtagaagtggggtgcttggagacagaggtctgctttccagcaggcactaacttttcagtgtccctgtctatgtttcc  β^t
153  ccaggttacaaggcagctcacaagtagaagtggggtgcttggagacagaggtctgctttccagcaggcactaacttttcagtgtccctgtctatgtttcc  β^dmin

       31              40                 50                  60
253  cttttagGCTGCTGGTTGTCTACCCTTGGACCGAGCGGTACTTTGATAGCTTTGGAGACCTATCCTCTGCCTCTGCTATCATGGGTAATGCCAAAGTG  β^dmaj
250  cttttagGCTGCTGGTTGTCTACCCTTGGACCCAGCGGTACTTTGATAGCTTTGGAGACCTATCCTCTGCCTCTGCTATCATGGGTAATGCCAAAGTG  β^s
       gLeuLeuValValTyrProTrpThrGlnArgTyrPheAspSerPheGlyAspLeuSerSerAlaSerAlaIleMetGlyAsnAlaLysVal
253  cttttagGCTGCTGGTTGTCTACCCTTGGACCCAGCGGTACTTTGATAGCTTTGGAGACCTATCCTCTGCCTCTGCTATCATGGGTAATGCCAAAGTG  β^t
253  cttttagGCTGCTGGTTGTCTACCCTTGGACCCAGCGGTACTTTGATAGCTTTGGAGACCTATCCTCTGCCTCTGCTATCATGGGTAATCCCAAGGTG  β^dmin
                                                                                                     Pro

        70               80                 90
352  AAGGCCCATGGCAAGAAGGTGATAACTGCCTTTAACGATGGCCTGAATCACTTGGACAGCCTCAAGGGCACCTTTGCCAGCCTCAGTGAGCTCCACTGT  β^dmaj
349  AAGGCCCATGGCAAGAAAGTGATAACTGCCTTTAACGATGGCCTGAATCACTTGGACAGCCTCAAGGGCACCTTTGCCAGCCTCAGTGAGCTCCACTGT  β^s
       LysAlaHisGlyLysLysValIleThrAlaPheAsnAspGlyLeuAsnHisLeuAspSerLeuLysGlyThrPheAlaSerLeuSerGluLeuHisCys
352  AAGGCCCATGGCAAGAAGGTGATAACTGCCTTTAACGATGGCCTGAATCACTTGGACAGCCTCAAGGGCACCTTTGCCAGCCTCAGTGAGCTCCACTGT  β^t
352  AAGGCCCATGGCAAAAAGGTGATAACTGCCTTTAACGAGGGCCTGAAAACTTGGACAACCTCAAGGGCACCTTTGCCAGCCTCAGTGAGCTCCACTGT  β^dmin
                                 Glu            LysAsn         Asn

             100      104
451  GACAAGCTGCATGTGGATCCTGAGAACTTCAGGgtgagtctgatgggcacctcctgggtttccttccccctggctattctgctcaaccttcctatcagaaaa  β^dmaj
448  GACAAGCTGCATGTGGATCCTGAGAACTTCAGGgtgagtctgatgggcacctcctgggtttccttccccctggctattctgctcaaccttcctatcagaagg  β^s
       AspLysLeuHisValAspProGluAsnPheArg
451  GACAAGCTGCATGTGGATCCTGAGAACTTCAGGgtgagtctgatgggcacctcctgggtttccttccccctggctattctgctcaaccttcctatcagaagg  β^t
451  GACAAGCTGCATGTGGATCCTGAGAACTTCAGGgtgagtctgatgggcacctcctgggtttccttccccctggctattctgctcaaccttcctatcagagg  β^dmin

552  aaaggggaagcgattctagggagcagtctccatgactgtgtgtg-----------gagtgttgacaagagttcggatatttattctctactcagaattg  β^dmaj
549  aaaggggaagcgattctagggagcagtctccatgactgtgtgtg-----------gagtgttgacaagagtttggatatttttattctctactcagaatcg  β^s
552  aaaggggaagcgattctagggagcagtctccatgactgtgtgtg-----------gagtgttgacaagagtttggatatttttattctctactcagaatcg  β^t
552  aaaggggaagcgattctagggagcagtttcgatcatagtgtgtgggatgtgaattgtgagtgtgactagagtttggatatttttattctctactcagaattg  β^dmin

641  ctgctcccctcactctgttctgt-------gttgtcattttcctctttctttggt-aagcttttaattttccagttgca-ttttactaaattaattaagc  β^dmaj
638  ctgctcccctcactctgttctgt-------gttgtcattttcctctttctttggt-aagcttttaattttccagttgca-ttttactaaattaattaagc  β^s
641  ctgctcccctcactctgttctgt-------gttgtcattttcctctttctttggt-aagcttttaattttccagttgca-ttttactaaattaattaagc  β^t
653  ctgctccctctcactctgttctgttcttctgtgtgtcattttcctctttctt-ggttaag-ttttaattttcagttgcagtttta---------------  β^dmin
```

FIG. 3.—DNA sequence comparison of the adult β-globin genes in C57BL/10 and BALB/c mice. The four sequences have been aligned by inspection to maximize the homology between them. Uppercase letters represent positions that are part of the mature mRNA, whereas lowercase letters denote flanking and intervening sequences. Single nucleotide differences between adjacent sequences are marked by dots; dashed lines are gaps introduced to maximize homology. Each sequence is numbered beginning at the cap site; there is no nucleotide 0. The amino acid sequence of the βsingle protein is listed in its entirety.

```
731   tctggttatttacttcccatcctgatatcagcttccctcctcctttcctcccagtccttctctctctctcctctctctttctctaatcctttcctttccct  β^dmaj
728   t--ggttatttacttcccatcctgatatcagcttccctcctcctttcctcccagtccttctctctctctcctctctctttctctaatcctttcctttccct  β^s
731   t--ggttatttacttcccatcctgatatcagcttccctcctcctttcctcccagtccttctctctctctctctctttctctaatcctttcctttccct    β^t
(736) ----------------------------------------------------------------------------------------------------   β^dmin

831   cagttcatt-ct-ct-c---------tt-gatctacgtt-tgtttgtctctttaaatattgccttgtaacttgctcagaggacaaggaagatatgtcc--  β^dmaj
826   cagttcatttcttcttc---------tttgatctacttt-tgtttgtctcttttaaatattgccttgtaacttactcagaggacaaggaagatatgtcc--  β^s
829   cagttcatttcttcttc---------tttgatctacttt-tgtttgtctctttaaatattgccttgtaacttactcagaggacaaggaagatatgtcc--  β^t
(736) -----------------aagtgcatctttttatctactttctgtttt-tcttttcaaatatatcctggtatctactctgaggacaa-aaagatataatgatt  β^dmin

915   ctgtttcttctcatagctctc-----aagaatagtagcataattggctttttatgccagggtgacagggggaagaatatatttttacatataaatttctgtttg  β^dmaj
914   ctgtttcttctcatagctctc-----aagaatagtagcataattggctttttatgccagggtgacagggggaagaatatatttttacatataaatttctgtttg  β^s
917   ctgtttcttctcatagctctc-----aagaatagtagcataattggctttttatgccagggtgacagggggaagaatattttttacatataaattctgtttg  β^t
817   ctctgatccttccacagtctaatgaatataggggcgataattggcttttta-ggataggacaataaggaagagtatattttgtggctaaatttttgtgtg  β^dmin

1010  acatagg--------------------------------------------attcttataataatttgtcagtag-----------------------t  β^dmaj
1009  acatagg--------------------------------------------attcttataataatttgtcagcag-----------------------t  β^s
1012  acatagg--------------------------------------------attcttataataatttgtcagcag-----------------------t  β^t
916   tcatagaataattcttttttattttttatttattatttttttcatagaataattcttatcaaaattgaccagtagaaccagctgccattttacctatct  β^dmin

1042  tt-aaggttgcaaac--aaatgtc-tttgtaaataagcctgcaggtatctggtattttgctctacagttatgttgatggttcttccatattcccacag   β^dmaj
1041  tt-aaggttgcaaac--aaatgtc-tttataaataagcctgcag-tatctggtattttgctctacagttatgttgatgattcttccatcttcccacag   β^s
1044  tt-aaggttgcaaac--aaatgtc-tttataaataagcctgcag-tatctggtattttgctctacagttatgttgatggttcttccatcttcccacag   β^t
1016  ttgaagattataactgtaaattccatttg-aaatcggccttcag-catcttgtattgttgctctactc-atattgatgcgtcttctctcttcccacag   β^dmin

                   105                    110            115            120            125            130
1137  CTCCTGGGCAATATGATCGTGATTGTGCTGGGCCACCACCTTGGCAAGGATTTCACCCCCGCTGCACAGGCTGCCTTCCAGAAGGTGGTGGCTCGAGTC  β^dmaj
1135  CTCCTGGGCAATATGATCGTGATTGTGCTGGGCCACCACCTGGGCAAGGATTTCACCCCCGCTGCACAGGCTGCCTTCCAGAAGGTGGTGGCTCGAGTC  β^s
            LeuLeuGlyAsnMetIleValIleValLeuGlyHisHisLeuGlyLysAspPheThrProAlaAlaGlnAlaAlaPheGlnLysValValAlaGlyVal
1138  CTCCTGGGCAATATGATCGTGATTGTGCTGGGCCACCACCTGGGCAAGGATTTCACCCCCGCTGCACAGGCTGCCTTCCAGAAGGTGGTGGCTCGAGTC  β^t
1112  CTCCTAGGCAATGCGATCGTGATTGTGCTGGGCCACCACCTGGGCAAGGATTTCACCCCCGCTGCACAGGCTGCCTTCCAGAAGGTGGTGGCTCGAGTG  β^dmin
                      Ala

                   Thr140                          TER
1236  GCCACTGCCTTGGCTCACAAGTACCACTAAACCCCCTTTCCTGCTCTTGCCTGTGAACAATGGTTAATTGTTCCCAAGAGAGCATCTGTCAGTTGTTGGC  β^dmaj
1234  GCCGCTGCCCTGGCTCACAAGTACCACTAAACCCCCTTTCCTGCTCTTGCCTGTGAACAATGGTTAATTGTTCCCAAGAGAGCATCTGTCAGTTGTTGGC  β^s
            AlaAlaAlaLeuAlaHisLysTyrHis
1237  GCTGCTGCCCTGGCTCACAAGTACCACTAAGCCCCTTTT-CTGCTATTGTCTATGCACAAAGGTTATATGTCCCCTAGAGAAAAACTGTCAATTGTGGGG  β^t
1211  GCCACTGCCCTGGCTCACAAGTACCACTAAGCCCCTTTT-CTGCTATTGTCTATGCACAAAGGTTATATGTCCCCTAGAGAAAAACTGTCAAGTGTGGGG  β^dmin
          Thr

                                              pA box           +pA site
1336  AAAAATGAT--AGACATTTGAAAATCTGTCTTCTGACAAATAAAAAGCCATTTATGTTCACTGCaatgatgtgttttaaattatttgtctgtgtcatagaaggg  β^dmaj
1334  AAAAATGATAAAGACATTTGAAAATCTGTCTTCTGACAAATAAAAAGCATTTAT-TTCACTGCaatgatgtgtttaaattatttgtctgtgtcatagaaggg  β^s
1336  AAA-TGATGAAGACCCTTTGGGCATCTAGCTTTTATCTAATAAATGATATTTACTGTCATCTCaattctgtgtttgattacttttgttttcttgcaaggat  β^t
1310  AAA-TGATGAAGACCCTTTGGGCATCTAGCTTTTATCTAATAAATGATATTTACTGTCATCTCaattctgtgtttgattacttgtgtttcttacaaggat  β^dmin

1434  tttatgctaagttttcaagatacaaagaagtgaggggttcagtctgacctttgggganataaatgaattacacttcaaattgtgttgt-cagctaagcagc  β^dmaj
1433  tttatgctaagttttcaagatacaaagaagtgaggggttcagtctgacctttcgggaaataaatgaattacacttcaaa-tgtataggacagc-aagcagt  β^s
1435  taatgtgaaatatttattatataaagcagttggggcatactgagggaagaaagtgaggggtaaaagtcatgccattctatttcagttaaaaccattcaag  β^t
1409  taatgtgaaatatttattatataaagcagttggggcatactgagggatggaagtgaggggtaaaagtcatgtcattctatttcagttaaaaccattcaag  β^dmin

1533  agtagccacagatc  β^dmaj
(1531)--aagccacagatc  β^s
1535  aataaaataccaaagaggaagtgagagttcagaaaacaaggtattcactttgacttcctggtaattgacttctgcag  β^t
1509  aataaaataccaaagaagaagtgagagttcagaaaacaaggtattcactttgacttcctggtaattgacttctgcag  β^dmin
```

Polymorphic amino acids in β^dmaj and β^dmin are also shown. The CAAT box, TATA box, cap site, poly(A) box, and poly(A) addition site are all indicated. The sequences of β^dmaj and β^dmin are taken from Konkel et al. (1979).

sequence for the βsingle protein from a different C57BL line. This protein was found to contain serine at position 72 and asparagine at position 80. This inconsistency may be due to the fact that the analyses were performed on βsingle

proteins isolated from two different C57BL strains that may harbor authentic variants of the "single" haplotype.

Our DNA sequence of β^s and β^t corroborates Gilman's (1976) finding of three amino acid differences between βsingle and βdmajor. Residue 13 is a glycine in βsingle, while in βdmajor it is a cysteine. Residue 20 of βsingle is an alanine, but in βdmajor it is a serine. Each of these replacements reflects a first position G ↔ T transversion. In coding block three, a G ↔ A transition at the first nucleotide of codon 139 results in an alanine ↔ threonine replacement. In addition, silent substitutions, resulting in no amino acid replacements, occur within codons 66 (A ↔ G), 118 (G ↔ T), and 141 (C ↔ T). It is surprising that the number of silent differences observed here is equal to the number of replacement differences, since silent differences have been shown to occur at a higher rate than those that result in amino acid replacements when one compares two functionally related sequences that are thought to share a common evolutionary origin (Perler et al. 1980). A more extensive discussion of this aspect of allelic comparisons follows in the next section.

Both intervening sequences of β^s and β^{dmaj} show a >90% similarity. In the 116 bp of IVS 1, only two 1-bp changes and one 3-bp gap are present. We consider the gap to be a deletion in β^s because the three nucleotides are present in β^t and β^{dmin} as well as in β^{dmaj}. The deletion removes one triplet of an $(AAG)_3$ directly repeated sequence. It might have occurred as a result of slipped mispairing during DNA replication (see Moore 1983). Within IVS 2, β^s and β^{dmaj} are also strikingly similar. There are only 16 differences over its 650-bp length. Eleven of these differences are 1-bp substitutions, four are 1-bp deletions, and one is a 2-bp gap.

The sequence similarity exhibited by the β^s and β^{dmaj} alleles extends well into the areas that flank the genes. The two alleles are identical in the 52 nucleotides that make up the 5' nontranslated regions of the genes. The identity between β^s and β^{dmaj} persists from the mRNA capping site to a point 42 nucleotides upstream from the putative CAAT box (at −79), where the first nucleotide difference is located. The sequences remain >90% similar, however, for as far upstream as our analysis extends. Thus, the mRNA capping site, TATA homology, and CAAT box, as well as the distances between them, have been conserved.

The 3' noncoding regions of these two genes are highly homologous. Two small deletions—a 2-bp deletion in β^{dmaj} and a 1-bp deletion in β^s—are the only differences that occur in the 132 nucleotides between the TAA stop codon and the poly(A) addition site. Beyond this point, a similarity of 98% is observed for another 100 nucleotides into the 3' flanking region, which is as far as our sequence data extends. The substantial similarity between β^s and β^{dmaj} extends over at least a 7-kb area as evidenced by the ability of cloned fragments containing β^s and β^{dmaj} to form a heteroduplex (Weaver et al. 1981).

The β^t and β^{dmin} Alleles Are More Divergent than Are β^s and β^{dmaj}

The βsingle and βdminor proteins differ at 11 of 146 amino acid positions. Nucleotide differences in codons 9, 13, 16, and 20 in exon one; in codons 58, 73, 76, 77, and 80 in exon two; and in codons 109 and 139 in exon three all correspond to amino acid replacements. As is the case with β^s and β^{dmaj}, silent substitutions are unexpectedly infrequent in relation to replacement substitutions: only six silent differences occur in the coding sequences of the β^t and β^{dmin} alleles, compared to the 11 replacement differences.

    In a previous analysis, Perler et al. (1980) showed that, in interspecific comparisons of homologous genes, silent substitutions exceed replacement differences by a factor of seven. In our examination of the mouse β-globin alleles, we observe an equal number of replacement and silent differences in the coding regions of β^s and β^{dmaj} and an almost twofold predominance of replacement over silent differences in the β^t and β^{dmin} coding regions. A similar situation exists in the case of the two rabbit β-globin alleles (Efstratiadis et al. 1977; Hardison et al. 1979; van Ooyen et al. 1979). There are four replacement differences but no silent differences in the coding sequences of these genes. Furthermore, only two 1-bp substitutions are evident in the intervening sequences. Thus, in both the mouse and rabbit β-globin alleles, either the rate of replacement substitutions has been accelerated (due perhaps to the relaxation of functional constraints at some time) or the occurrence of these mutations was so recent that silent differences have had little chance to accumulate. These explanations seem more likely than one that invokes positive selection of advantageous mutations because none of the polymorphic residues in the mouse and rabbit β-globin genes appears to occupy functionally important positions within the hemoglobin tetramer (Dickerson and Geis 1983).

    With respect to their intervening sequences, the relationship between the alleles β^t and β^{dmin} is quite different from that between the alleles β^s and β^{dmaj}. Whereas the β^s and β^{dmaj} genes exhibit a high degree of homology in both intervening sequences, β^{dmin} has an extremely divergent IVS 2. On the other hand, β^t resembles β^s and β^{dmaj} in both introns. In order to maintain the alignment of the DNA sequence of the IVS 2 of β^{dmin} with that of the other three genes (fig. 3), one must provide one 132-bp deletion and six insertions of 12, 8, 9, 5, 44, and 24 bp. In addition, there are a number of smaller insertions and deletions in this region, as well as numerous single-nucleotide substitutions. In contrast, IVS 1 has been highly conserved between β^t and β^{dmin}, with only two 1-bp differences over 116 nucleotides.

    Most of the aforementioned insertions in the second intron of β^{dmin} occur in regions of DNA consisting of short, simple-sequence direct repeats. These regions would tend to promote slipped mispairing during DNA replication which leads to insertions and deletions (Moore 1983). For example, the 9-bp insertion at position +753 in β^{dmin} lies in a simple-sequence environment made up of short runs of pyrimidines. Also, the 44-bp insertion at +923 contains short repeats of A+T-rich segments and is bordered by A+T-rich regions. Interestingly, the 132-bp deletion at +721 in β^{dmin} corresponds to a simple-sequence region in the other three genes. Ninety-seven of the last 104 nucleotides of this deleted region are (C_nT_n), where n = 1-4. The significance of the absence of this C_nT_n simple-sequence region in β^{dmin} will be discussed further below.

    Despite their comparatively extensive differences between coding blocks and between introns, β^t and β^{dmin} are much more homologous in their nontranslated and flanking sequences. In other words, β^t and β^{dmin} are more similar outside the interval bounded by the ATG and TAA codons than within it. The 5' nontranslated regions of β^t and β^{dmin} show a >95% similarity, with only two single-nucleotide differences. Our sequence data for β^t extends to 200 bp upstream from the mRNA capping site. In this region, only three differences exist between these alleles; a 1-bp substitution and deletions of 1- and 6-bp in the β^t sequence. All three differences occur well upstream of the CAAT box (at −78). The 3' nontranslated portions of the β^t and β^{dmin} alleles show a >99% similarity, with only one difference, a G ↔ T

transversion, in 130 bp. Further downstream, a similarity of 98% is evident over the 200 nucleotides of 3′-flanking DNA that can be compared.

Panels *A* and *B* of figure 4 depict the differences seen in the relationship between β^s and β^{dmaj} (occupying the $\beta1$ locus), on the one hand, and between β^t and β^{dmin} (at the $\beta2$ locus), on the other. There is a striking contrast between the two relationships. The β^s and β^{dmaj} alleles show a >95% similarity in the exons, introns, and flanking sequences, and their protein products differ at only three of 146 amino acid positions. On the other hand, while the β^t and β^{dmin} alleles exhibit comparable similarity over flanking regions and IVS 1, they are distinctly less similar in the coding blocks and are highly divergent in IVS 2. In addition, the proteins encoded by the β^t and β^{dmin} alleles differ at 11 of 146 amino acid positions.

A Comparison of β^s and β^t: Anatomy of the *Hbb*^{s} Haplotype

The relationship between the two adjacent β-globin genes of the *Hbb*^{s} haplotype, β^s and β^t, is displayed graphically in panel *C* of figure 4. In the region bounded by the translational start and stop codons, the similarity between β^s and β^t is dramatically greater than that present in either of the interallelic comparisons (fig. 4, panels *A* and *B*). Both genes encode the same amino acid sequence, that of the βsingle polypeptide. Within this extremely similar region, only three differences are observed: a 3-bp deletion in IVS 1 of β^s, an A $\leftrightarrow$ G transition near the 3′ end of IVS 2, and a silent C $\leftrightarrow$ T transition within codon 138. Clearly, the nonallelic β^s and β^t genes are more similar to one another than either one is to its allelic counterpart in the *Hbb*^{d} haplotype.

Outside the region of identity, the noncoding and flanking regions of β^s and β^t become abruptly more divergent—as divergent as the corresponding regions in the β^{dmaj}-β^{dmin} gene pair. This is not surprising, because it follows from the patterns of allelic similarities discussed above. From the initiation codon to the mRNA capping site, there are seven single-nucleotide differences. The similarity in the region 5′ to the cap site declines to 70% in the interval between the TATA box (-33) and the CAAT box (-79). Beginning at a point 100 bp upstream from the cap site, the β^s and β^t sequences show an average of $\sim60\%$ similarity over the remaining 100 nucleotides that can be compared.

To the 3′ side of the termination codon, the decrease in similarity between β^s and β^t is both more abrupt and more striking. The 3′ noncoding regions of β^s and β^t are only 70% similar (41/134 single-nucleotide differences), which makes this the only region in which the corresponding mRNAs may be readily distinguished. Pronounced divergence of the 3′ nontranslated regions of genes that are otherwise highly homologous has been noted in other systems, for instance in the tandemly duplicated human adult α-globin genes (Michelson and Orkin 1980). Further downstream, the β^s and β^t sequences become as dissimilar as randomly generated sequences just 11 nucleotides beyond the poly(A) addition site. In summary, the β^s and β^t genes show a >99% similarity between the start and stop codons, but they are as dissimilar as are the β^{dmaj} and β^{dmin} genes outside this area.

The Concerted Evolution of β^s and β^t Is Due to Gene Conversion

From the number of differences between the β^{dmaj} and β^{dmin} genes, Perler et al. (1980) calculated that the β-globin genes on the *Hbb*^{d} chromosome have been diverging for about 30 Myr. Therefore, one would expect that the other contemporary
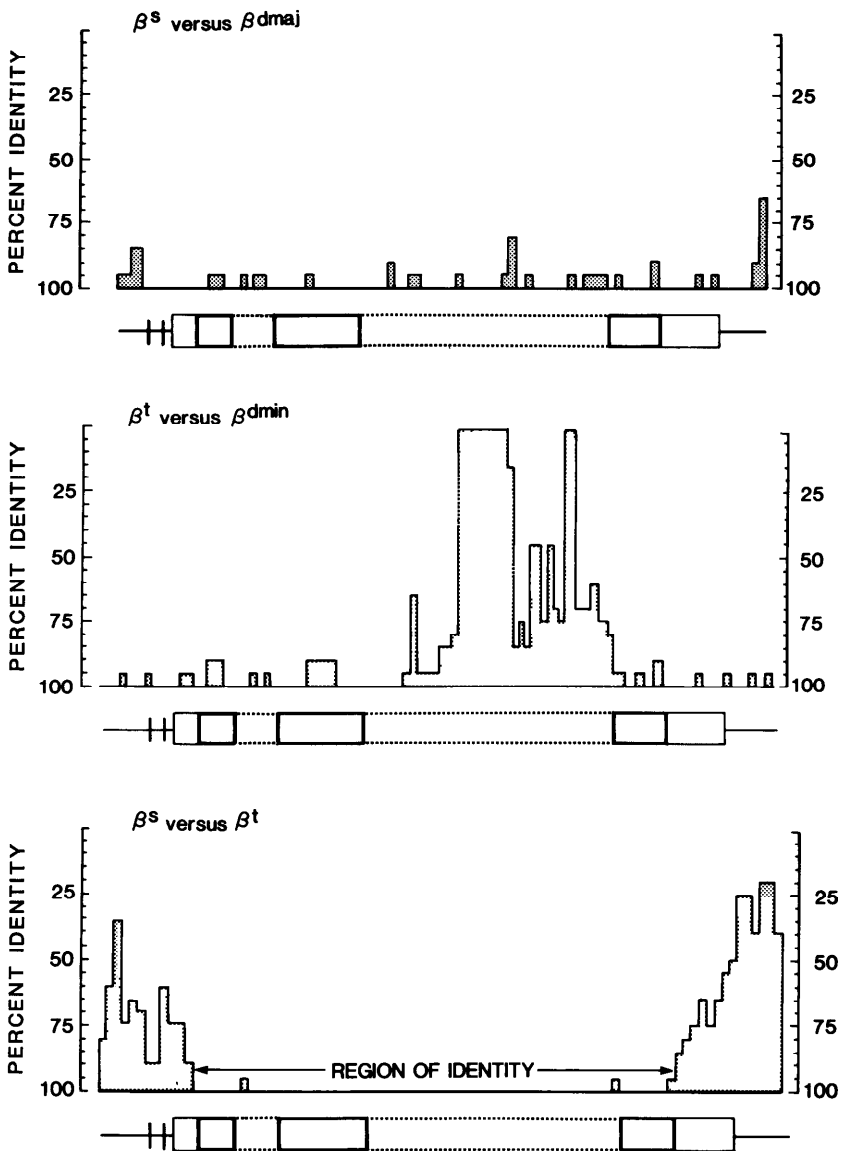
FIG. 4.—Homology plots for pairs of adult mouse β-globin genes. For each comparison, nucleotide differences (due to both point mutations and deletions/insertions) were counted along the entire length of the genes. The sequences were compared at intervals of 20 nucleotides, and homology values were plotted along the vertical axis. A diagramatic representation of the mouse β-globin gene is shown along the horizontal axis of each panel, with nontranslated regions (thin boxes), coding blocks (thick boxes), introns (dashed boxes), and flanking sequences (solid lines) indicated. The vertical bars depict the position of the CAAT and TATA boxes.

mouse β-globin haplotypes would harbor genes that have diverged to the same degree, so long as these haplotypes descended from the same duplication event as Hbb^d. However, the degree of divergence present within each of the nonallelic gene pairs described here is markedly different and suggests that the genes on the Hbb^s

chromosome are not evolving independently, since they are more alike than would be expected. That is, β^s and β^t have diverged <1%, while β^{dmaj} and β^{dmin} have diverged 16%. It is clear that the evolutionary histories of the Hbb^s and Hbb^d haplotypes are quite distinct. The identity between β^s and β^t has been maintained by a gene correction mechanism—resulting in 99.6% similarity over an interval extending from a point 14 nucleotides upstream of the initiation codon to the TAA termination codon—while at the same time divergence of the β^{dmin} gene of the Hbb^d haplotype has taken place.

In theory, identity between tandemly duplicated genes can be maintained by either of at least two nonexclusive mechanisms: multiple rounds of unequal crossing-over between tandemly duplicated sequences and gene conversion (see Hood et al. 1975). The former process involves intermediates with differing numbers of repeating units and results in the rectification of chromosome segments whose length is a multiple of the fundamental length of the tandem duplication. Gene conversion, on the other hand, involves no change in gene number, and there is no obvious rule about the length of the interval that can be rectified (see Shen et al. 1981). Most instances of concerted evolution between tandemly duplicated globin genes have been attributed to gene conversion events. The two functional human α-globin genes are believed to have evolved in concert via a gene conversion mechanism (Lauer et al. 1980; Zimmer et al. 1980), as have the two human ζ-globin genes (Proudfoot et al. 1982). In the latter case, however, a nonsense mutation in one of the two genes, ψζ, has rendered it nonfunctional. The goat α-globin genes (Schon et al. 1982) show a similarity of 99%, their protein products differing at three amino acid positions owing to single-nucleotide changes. Finally, the two human fetal globin genes have undergone sequence correction via partial gene conversion (Slightom et al. 1980; Shen et al. 1981). The most recent conversion event has restored DNA sequence identity over a 1.5-kb region that includes the 5' two-thirds of these genes.

We propose that the concerted evolution of β^s and β^t is due to gene conversion. The most recent conversion event corrected the sequences from a point 14 nucleotides upstream from the ATG start codon through the TAA termination codon. We rule out the possibility that the correction of these two sequences is due to unequal crossing-over. Such an event would have rendered the resultant genes identical in one or both flanking regions. Instead, the sequence comparison of β^s and β^t shows significantly reduced similarity in the 5' flanking regions and <25% similarity in the 3' flanking sequences.

Features of the Hbb^s Gene Conversion Unit

Gene conversion units have been well-defined for a number of concomitantly evolving globin gene pairs (Liebhaber et al. 1980; Shen et al. 1981; Proudfoot et al. 1982; Schon et al. 1982), and both the size and scope of these converted intervals vary significantly from one case to the next. The Hbb^s conversion unit described here covers ~1.2 kb and extends from just upstream of the translational start codon through the termination codon. Thus, the gene conversion unit in which β^s and β^t lie is different from previously described globin gene conversion units with respect to one important feature: the contemporary converted interval does not include the putative transcriptional regulatory signals that are found upstream of the gene. Although it is possible that this region (defined here as extending from the cap site

to the CAAT box) was at one time included within the conversion unit, it now harbors 17/79 nucleotide differences between β^s and β^t. Schon et al. (1982) point out that the goat α-globin conversion unit contains all nearby regions that are thought to be important for transcription (i.e., CAAT box, TATA box, cap site, poly[A] box, and poly[A] addition site), and these authors suggest that it is important, with regard to gene expression, for the transcription unit and the conversion unit to be congruent. Indeed, all previously described globin gene conversion units do encompass these upstream transcriptional regulatory signals. However, both the β^s and β^t genes are expressed (S. Lewis, personal communication), even though the Hbb^s conversion unit extends only 14 nucleotides upstream of the coding region. Similarly, the 3' border of this unit lies immediately outside of the coding region. As is evident in panel C of figure 4, the similarity between β^s and β^t declines from 99% to 70% in the 3' nontranslated region.

In some instances, the borders of globin gene conversion units are defined by unusual sequence elements. For example, the human fetal globin gene conversion is demarcated at the 3' end by a region of simple-sequence DNA consisting primarily of the nucleotides T and G (Shen et al. 1981). This stretch of DNA has the potential for forming a Z-DNA structure owing to its alternating purine-pyrimidine nature (Arnott et al. 1980). Shen et al. (1981) propose that this sequence acts as a "hotspot" for DNA exchange between genes on adjacent chromatids. Although the boundaries of the Hbb^s conversion unit exhibit no unusual repeats or potential Z-DNA-forming sequences, an extended region of simple-sequence DNA (consisting of the nucleotides T and C) is present in IVS 2 of both β^s and β^t, as well as in β^{dmaj}. Radding (1978) has pointed out that generalized recombination is significantly enhanced by the presence of short tandem repeats of simple-sequence DNA. It is conceivable, then, that the sequence present in IVS 2 of β^s and β^t could help to promote the kind of recombination event that is prerequisite for a gene conversion. The fact that this sequence has been deleted in the divergent β^{dmin} gene on the Hbb^d chromosome reinforces the idea that the simple-sequence region in IVS 2 is important in maintaining identity between β^s and β^t and that its absence in β^{dmin} may have inhibited correction within the Hbb^d haplotype.

Evolutionary History of the Hbb^s and Hbb^d Haplotypes

We propose (cf. fig. 5) a sequence of events leading to the present arrangements of the two major Hbb haplotypes in *Mus musculus*. Following the duplication of an ancestral β-globin sequence, $\beta0$, two identical tandem segments were present on the chromosome—namely, two nonallelic β-globin loci, $\beta1$ and $\beta2$, and their flanking sequences. Subsequent to this duplication event, the two genes would be free to independently accumulate nucleotide changes unless some mechanism of maintaining sequence identity between these duplication units were involved. It is clear from the DNA sequences analyzed here that two types of chromosomal arrangements of the mouse adult β-globin genes have been derived from the ancestral chromosome—the Hbb^s haplotype, in which identity has been maintained between two genes, and the Hbb^d haplotype, which consists of two divergent genes.

The third line of figure 5 illustrates the establishment of the ancestral Hbb^s and Hbb^d haplotypes as distinct entities. Although the details of these two haplotypes at the time of their appearance in the population is unknown, we propose that at this point in their history the two haplotypes differed in one salient regard: the β^s
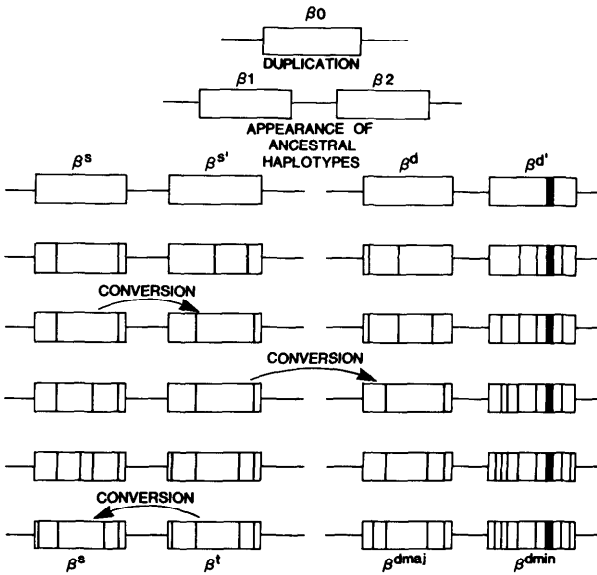
FIG. 5.—The evolutionary history of the Hbb^s and Hbb^d haplotypes. This diagram depicts in a generalized way a model for the origin of the two haplotypes that is consistent with the available DNA sequence information. We propose that the duplication of a single ancestral β-globin gene (β0) gave rise to two nonallelic genes (β1 and β2). This was followed by the appearance of two distinct ancestral haplotypes ($β^s$ and $β^{s'}$, $β^d$, and $β^{d'}$) in the population; the vertical box in $β^{d'}$ gene indicates a deletion of simple-sequence DNA in IVS 2 that may have played a significant role. The overall pattern of similarity between $β^s$, $β^t$, and $β^{dmaj}$ suggests the liklihood of a long history of gene conversion events among their evolutionary forerunners ($β^s$, $β^{s'}$, and $β^d$). This figure illustrates the independent accumulation of mutations (vertical bars) in each of the four genes, with repeated gene conversion (arrows) maintaining the identity between $β^s$, $β^{s'}$, and $β^d$. The $β^{d'}$ gene does not participate in these exchanges, owing to the deletion in IVS2. The latest conversion event, which occurred 0.5 Myr ago (see text), gave rise to $β^s$ and $β^t$ (bottom left); the contemporary $β^{dmaj}$ and $β^{dmin}$ genes are pictured at the right.

and $β^{s'}$ genes were able to undergo sequence correction, whereas the $β^d$ and $β^{d'}$ genes had suddenly become so nonhomologous, owing to a large deletion of simple-sequence DNA in the evolutionary forerunner of $β^{dmin}$ (illustrated as solid box in $β^{d'}$ in fig. 5), that they could no longer maintain identity through gene conversion. Notice, however, that the ancestor of $β^{dmaj}$ was able to take part in gene conversion events with the $β^s$ and $β^{s'}$ sequences. Events of this nature could presumably occur between homologous chromosomes in an Hbb^s/Hbb^d heterozygote. It is worth noting that the direction of the conversion events as indicated by the arrows in figure 5 are purely for illustrative purposes. One cannot deduce from the DNA sequence data, for example, in which direction the latest conversion event occurred.

We have calculated estimates of the evolutionary times that have elapsed since each of the events outlined in figure 5 took place. To estimate the age of the duplication event that gave rise to the nonallelic mouse β-globin genes, we determined the percentage of divergence between both the $β^s$-$β^t$ and $β^{dmaj}$-$β^{dmin}$ gene pairs. We compared the same region in both cases, namely, the area 5′ to the converted interval in the Hbb^s haplotype. This region falls within the putative duplication unit and thus reflects the extent to which these nonallelic genes have diverged since the

ancestral duplication event. Using Zimmer's (1981) value of 1% sequence divergence/ 2.2 Myr for substantial lengths of noncoding DNA in human and ape globin genes, we obtained estimates of 57 Myr (β^s-β^t) and 68 Myr (β^{dmaj}-β^{dmin}) for the duplication of the mouse β-globin genes. The somewhat lower value for the β^s-β^t gene pair suggests that the Hbb^s conversion interval once extended further upstream from its present 5' border. When these calculations are performed using only the area upstream of the CAAT box, slightly higher values of 68 Myr (β^s-β^t) and 88 Myr (β^{dmaj}-β^{dmin}) are obtained. These numbers may more accurately predict the time of the duplication since selection presumably acts to prevent base changes within the region containing the TATA box and CAAT box, resulting in a lower estimate when this region is included in the analysis. At any rate, each of these values places the duplication at or before the time of the mammalian radiation (65–80 Myr ago), in agreement with the data of Hardies et al. (1984).

In order to analyze quantitatively the evolutionary histories of the Hbb^s and Hbb^d haplotypes, one needs to incorporate several estimates of evolutionary distances among the four genes. On the basis of the number of replacement differences that occur between these genes, Perler et al. (1980) have calculated that the β^{dmaj} and β^{dmin} genes have been diverging for 30 Myr. However, it is probable that the two haplotypes were still able to interact at that point in time; that is, they have not evolved as totally separate entities in the last 30 Myr. First, the fact that wild populations of *Mus musculus* maintain a balanced polymorphism with respect to the two haplotypes (Berry 1978; Berry et al. 1981; Sage 1981) suggests that there may be selective advantage associated with heterozygosity at the *Hbb* locus. Although the possibility that these haplotypes have long maintained a neutral coexistence cannot be ruled out, it has been shown (Berry 1978; Berry et al. 1981) that distinct wild populations of *Mus musculus* maintain both the Hbb^s and Hbb^d alleles at similar frequencies; thus, these haplotypes have probably not been isolated from one another at the populational level during the past 30 Myr. Second, on the basis of our calculations performed according to Perler et al. (1980), we calculate that the β^s and β^{dmaj} genes have only been diverging for 3–8 Myr, depending on whether the accumulation of nucleotide differences in these alleles is due to random drift or selection. This estimate agrees closely with that obtained if one assumes that β^s and β^{dmaj} are neutral alleles (Tajima 1983; M. Nei, personal communication). These data suggest that β^{dmaj} was probably able to interact with the β^s and β^t genes via gene conversion by virtue of the high homology shared by these sequences—notably in IVS 2, where the stretch of simple-sequence DNA lies. On the other hand, as was previously pointed out (Weaver et al. 1981), β^{dmin} could not take part in this type of correction owing to its abberant IVS 2, resulting in its divergence from the other β-globin genes. Hence, in the time since β^{dmin} began to diverge from β^{dmaj}, repeated gene conversion has kept the β^s, β^t, and β^{dmaj} genes highly homologous, with the most recent conversion event having occurred between β^s and β^t. We have estimated the time since the β^s and β^t genes were last corrected using both the method of Perler et al. (1980) and that employed by Zimmer (1981). Either of these two methods places the time of the latest conversion event at <500,000 years ago.

As we have suggested here, gene correction within the Hbb^d haplotype has not occurred in the last 30 Myr, owing in part to a deletion of simple-sequence DNA in IVS 2, which we believe has a role in promoting gene conversion in the Hbb^s haplotype. The loss of this sequence in β^{dmin} not only removed this putative

recombination-enhancing region but also introduced a large block of nonhomology in the Hbb^d genes. Thus, even if a conversion event could be initiated between β^{dmaj} and β^{dmin}, the exchange would most likely be terminated once it reached the region of nonhomology. An analysis of the human α-globin genes shows that such blocks of nonhomology can play a major role in the whittling down and even in the disruption of conversion units (Hess et al. 1983; Michelson and Orkin 1983).

Conclusions

Both the divergence and the concerted evolution of tandem gene pairs have been elucidated through the DNA sequence analysis of globin genes from many species. The globins have helped to illustrate the entire spectrum of the molecular evolution of tandemly duplicated gene pairs—concerted evolution through sequence correction, sequence divergence to the extent that two genes encode functionally different polypeptides, and the silencing of one member of a duplicated gene pair, whereupon it diverges to become a pseudogene. We have presented an evolutionary analysis of the β-globin genes of the mouse (*Mus musculus*), which are unique among other mammalian β-globin genes in that two distinct haplotypes are maintained in wild populations (Berry 1978; Berry et al. 1981; Sage 1981). These two haplotypes represent the different degrees to which a duplicated gene pair can evolve: in one case, the genes remain identical via gene conversion, while in the other case the haplotype illustrates how the same gene pair has diverged owing in part to a large deletion in the intron of one of the two genes. To this end, inbred mouse strains are invaluable in that they allow one to examine and compare variant haplotypes, thereby leading to a clearer understanding of the mechanisms that play a role in the evolution of gene families.

LITERATURE CITED

ARNOTT, S., R. CHANDRASEKARAN, D. L. BIRDSALL, A. G. W. LESLIE, and R. L. RATLIFF. 1980. Left-handed DNA helices. Nature 283:743–745.

BERRY, R. J. 1978. Genetic variation in wild house mice: where natural selection and history meet. Am. Sci. 66:52–60.

BERRY, R. J., R. D. SAGE, W. Z. LIDICKER, JR., and W. B. JACKSON. 1981. Genetical variation in three pacific house mouse populations. J. Zool. 193:391–406.

DICKERSON, R. E., and I. GEIS. 1983. Hemoglobin: structure, function, evolution, and pathology. Benjamin/Cummings, Menlo Park, Calif.

EFSTRATIADIS, A., F. C. KAFATOS, and T. MANIATIS. 1977. The primary structure of rabbit β-globin mRNA as determined from cloned DNA. Cell 10:571–585.

GILMAN, J. G. 1976. Mouse hemoglobin beta chains: comparative sequence data on adult major and minor beta chain from two species. Biochem. J. 159:43–53.

HARDIES, S. C., M. H. EDGELL, and C. A. HUTCHISON III. 1984. Evolution of the mammalian β-globin gene cluster. J. Biol. Chem. 259:3748–3756.

HARDISON, R. C., E. T. BUTLER III, E. LACY, T. MANIATIS, N. ROSENTHAL, and A. EFSTRATIADIS. 1979. The structure and transcription of four linked rabbit β-like globin genes. Cell 18:1285–1297.

HESS, J. F., M. FOX, C. SCHMID, and C.-K. J. SHEN. 1983. Molecular evolution of the human adult α-globin-like gene region: insertion and deletion of Alu family repeats and non-Alu DNA sequences. Proc. Natl. Acad. Sci. USA 80:5970–5974.

HILL, A., S. C. HARDIES, S. J. PHILLIPS, M. G. DAVIS, C. A. HUTCHISON III, and M. H. EDGELL. 1984. Two mouse early embryonic β-globin gene sequences. J. Biol. Chem. 259:3739–3747.

HOOD, L., J. H. CAMPBELL, and S. C. R. ELGIN. 1975. The organization, expression, and evolution of antibody genes and other multigene families. Annu. Rev. Genet. 9:305–353.

HUTCHISON, C. A., III, S. C. HARDIES, R. W. PADGETT, S. WEAVER, and M. H. EDGELL. 1984. The mouse globin pseudogene βh3 is descended from a premammalian δ-globin gene. J. Biol. Chem. 259:12881–12889.

HUTTON, J. J., J. BISHOP, R. SCHWEET, and E. S. RUSSELL. 1962. Hemoglobin inheritance in inbred mouse strains. I. Structural differences. Proc. Natl. Acad. Sci. USA 48:1505–1513.

JAHN, C. L., C. A. HUTCHINSON III, S. J. PHILLIPS, S. WEAVER, N. L. HAIGWOOD, C. F. VOLIVA, and M. H. EDGELL. 1980. DNA sequence organization of the β-globin complex in the BALB/c mouse. Cell 21:159–168.

KONKEL, D. A., J. V. MAIZEL, JR., and P. LEDER. 1979. The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell 18:865–873.

LAUER, J., C.-K. J. SHEN, and T. MANIATIS. 1980. The chromosomal arrangement of human α-like globin genes: sequence homology and α-globin gene deletions. Cell 20:119–130.

LEDER, P., J. N. HANSEN, D. KONKEL, A. LEDER, Y. NISHIOKA, and C. TALKINGTON. 1980. Mouse globin system: a functional and evolutionary analysis. Science 209:1336–1342.

LIEBHABER, S. A., M. GOOSENS, and Y. W. KAN. 1980. Homology and concerted evolution at the α1 and α2 loci of human α-globin. Nature 290:26–29.

MAXAM, A. M., and W. GILBERT. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74:560–574.

MESSING, J., R. CREA, and P. H. SEEBURG. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9:309–321.

MICHELSON, A. M., and S. H. ORKIN. 1980. The 3' untranslated regions of the duplicated human α-globin genes are unexpectedly divergent. Cell 22:371–377.

———. 1983. Boundaries of gene conversion within the duplicated human α-globin genes. J. Biol. Chem. 258:15245–15254.

MOORE, G. P. 1983. Slipped-mispairing and the evolution of introns. Trends Biochem. Sci. 8:411–414.

PERLER, F., A. EFSTRATIADIS, P. LOMEDICO, W. GILBERT, R. KOLODNER, and J. DODGSON. 1980. The evolution of genes: the chicken preproinsulin gene. Cell 20:555–566.

PHILLIPS, S. J., S. C. HARDIES, C. L. JAHN, M. H. EDGELL, and C. A. HUTCHISON III. 1984. The complete nucleotide sequence of a β-globin-like structure, βh2, from the Hbb^d mouse BALB/c. J. Biol. Chem. 259:7947–7954.

POPP, R. A. 1973. Sequence of amino acids in the β chain of single hemoglobins from C57BL, SWR, and NB mice. Biochim. Biophys. Acta 303:52–60.

PROUDFOOT, N. J., A. GIL, and T. MANIATIS. 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 31:553–563.

RADDING, C. M. 1978. Genetic recombination: strand transfer and mismatch repair. Annu. Rev. Biochem. 47:847–880.

SAGE, R. D. 1981. Wild mice. Pp. 39–90 in H. L. FOSTER, J. D. SMALL, and J. G. FOX, eds. The mouse in biomedical research. Vol. 1. Academic Press, New York.

SANGER, F., A. R. COULSON, B. G. BARELL, A. J. H. SMITH, and B. A. ROE. 1980. Cloning in single stranded bacteriophage as an aid to rapid sequencing. J. Mol. Biol. **143**:161–178.

SCHON, E. A., S. M. WERNKE, and J. B. LINGREL. 1982. Gene conversion of two functional goat α-globin genes preserves only minimal flanking sequences. J. Biol. Chem. **257**:6825–6835.

SHEN, S.-H., J. L. SLIGHTOM, and O. SMITHIES. 1981. A history of the human fetal globin gene duplication. Cell **26**:191–203.

SLIGHTOM, J. L., A. E. BLECHL, and O. SMITHIES. 1980. Human fetal $^G\gamma$-globin and $^A\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.

TAJIMA, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437–460.

TIEMEIER, D. C., S. M. TILGHMAN, F. I. POLSKY, J. G. SEIDMAN, A. LEDER, M. H. EDGELL, and P. LEDER. 1978. A comparison of two cloned mouse β-globin genes and their surrounding and intervening sequences. Cell **14**:237–245.

VAN OOYEN, A., J. VAN DEN BERG, N. MANTEI, and C. WEISSMAN. 1979. Comparison of total sequence of a cloned rabbit β-globin gene and its flanking regions with a homologous mouse sequence. Science **206**:337–344.

WEAVER, S., M. B. COMER, C. L. JAHN, C. A. HUTCHINSON III, and M. H. EDGELL. 1981. The adult β-globin genes of the "single" type mouse C57BL. Cell **24**:403–411.

ZIMMER, E. A. 1981. Evolution of primate globin genes. Ph.D. thesis, University of California, Berkeley.

ZIMMER, E. A., S. L. MARTIN, S. M. BEVERLY, Y. W. KAN, and A. C. WILSON. 1980. Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc. Nat. Acad. Sci. USA **77**:2158–2162.