Mouse Haemoglobin Beta Chains

COMPARATIVE SEQUENCE DATA ON ADULT MAJOR AND MINOR BETA CHAINS FROM TWO SPECIES, MUS MUSCULUS AND MUS CERVICOLOR

By JOHN G. GILMAN*

Department of Genetics, University of Wisconsin, Madison, WI 53706, U.S.A.

(Received 28 April 1976)

In inbred Mus musculus several different β chains are known. In certain strains two β chains are produced in unequal amounts by the two closely linked genes of the doublet breeding unit allele Hbb*: βdmaj and βdmim. One strain has a variant doublet allele, Hbb*, which produces a variant minor β chain, βpmin (the major β chain, βpmaj, may not differ from βdmaj chain). Certain other strains have a singlet allele, Hbb*, that produces only one β chain, βs. Other species have different β-chain patterns. In M. cervicolor two variant major β chains are found, βcmaj (d-like) and βcmaj (s-like), both of which were found associated with minor β chains, M. caroli has only one, 'Lepore-like' β chain, with structural features characteristic predominantly of βdmaj chain in the N-terminal half and of βdmaj chain in the C-terminal half. The present paper presents sequence data on βs, βdmaj, βdmim, βpmin and βcmaj (d-like) chains. The data on βdmim chain cover almost the whole of that chain and show a minimum of nine differences from βdmaj chain and two from βpmin chain. It is suggested that the data on the β chains of the various species show evidence for the past occurrence of double crossovers over regions within a gene coding for only one or a few amino acids, which events can be explained by the 'hybrid DNA' models of genetic recombination. Supplementary information on the amino acid sequence of the proteins has been deposited as Supplementary Publication SUP 50067 (36 pages) at the British Library Lending Division, Boston Spa, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in Biochem. J. (1976) 153, 5.

Gene duplication followed by structural divergence is a process by means of which proteins with divergent functions may have evolved. Ingram (1963) has suggested that haemoglobin α and β chains and myoglobin evolved from a single primordial globin chain as a result of such a process involving gene duplication. Duplicate genes are also known that code for proteins with no great structural differentiation. They may represent an initial stage in the evolution of functional divergence. For example the β and δ polypeptides of adult human haemoglobin β chains differ at only ten positions in the amino acid sequence, and are produced by two closely linked genes of a doublet locus in unequal amounts (δ polypeptide is only about 2% of total chain in mature adult erythrocytes).

In mice, doublet haemoglobin β-chain loci are also known. In laboratory mice (Mus musculus musculus) the doublet breeding unit alleles Hbb and Hbb both produce two variant β chains in unequal amounts (Russell & McFarland, 1974). In Hbb/Hbb homozygotes, βdmaj chain is found as 80% of total β chain and βdmim chain as 20% and they are produced by closely linked genes (Hutton et al., 1962a,b). A singlet breeding unit allele, Hbb*, is also generally found in wild North American mouse populations and is polymorphic with Hbb allele (Selander et al., 1969). Hbb* allele is also found in several inbred mouse strains (Russell & Gerald, 1958). Tryptic-peptide composition data of Rifkin (1968) suggest that βs chain, the one β chain produced by Hbb* allele, differs from βdmaj chain in at least three positions.

Gilman (1974) reported the finding of a polymorphism in the Thai species M. cervicolor. This polymorphism has some similarity to the Hbb*, Hbb* allele polymorphism of Mus musculus with respect to differences in the major β chains. In contrast with Mus musculus, in M. cervicolor the polymorphism does not also involve a difference in gene number. Both types of M. cervicolor breeding unit alleles apparently produce a minor β chain (βcmim) as well as the major chain, and the only known sequence differences involve the variant major β chains, βcmaj (d-like) and βcmaj (s-like). Gilman (1974) also reported that the Thai species M. caroli has only one 'Lepore-like' β chain that has structural features characteristic predominantly of βdmaj chain in the N-terminal half and of βdmaj in the C-terminal half.
The present paper presents comparative sequence data on \( \beta s \), \( \beta \text{dmin} \) and \( \beta \text{cmaj} \) (d-like) chains; the Supplementary Paper (SUP 50067) presents additional data on \( \beta \text{dmin} \) and \( \beta \text{cmaj} \) (d-like) chains, as well as data on \( \beta \text{maj} \) and \( \beta \text{pmin} \) chains. The data on \( \beta s \) and \( \beta \text{dmaj} \) chains supplement those of Popp (1972, 1973) and Popp & Bailiff (1973), and suggest the absence of the small inversions in \( \beta \text{maj} \) chain relative to \( \beta s \) chain claimed by those workers; the data on \( \beta \text{dmin} \) chain cover almost the whole of that chain and show a minimum of nine differences from \( \beta \text{maj} \) chain and two differences from \( \beta \text{pmin} \) chain. The data on the \( \beta \) chains of the two species, \textit{M. musculus} and \textit{M. cervicolor}, show a striking variability at position 109, where differences in human haemoglobin has been shown to lead to functional alteration (Nute et al., 1974).

**Experimental**

**Animals**

Inbred \textit{M. musculus} strains used were from The Jackson Laboratory, Bar Harbor, ME, U.S.A.: strain C57B1/6J (for studies of \( \beta s \) chain), strain C3H/HeJ (for \( \beta \text{dmaj} \) chain \( N \)-terminal sequence and tryptic-peptide data of the Supplementary Paper), and strain Au/SsJ (for studies on \( \beta \text{pmin} \) chain given in the Supplementary Paper). Random-bred strain HA/ICR (Swiss) mice from A. R. S. Sprague-Dawley, Madison, WI, U.S.A., were used as a source of \( \beta \text{dmaj} \) chain and as the source of \( \beta \text{dmin} \) chain (except for \( N \)-terminal sequence and tryptic peptide data of the Supplementary Paper); a colony of \( \text{Hbb}^{a}/\text{Hbb}^{a} \) homozygotes was derived from the random-bred mice originally supplied, as described by Gilman (1976). The \textit{M. cervicolor} mice used in this study were generously provided by Dr. T. C. Hsu (University of Texas, M. D. Anderson Hospital and Tumor Institute, Houston, TX, U.S.A.); they had been originally trapped in Thailand and characterized by Dr. Joe T. Marshall.

**Starch-gel electrophoresis**

Starch-gel electrophoresis at pH 8.6 was as described by Gilman (1976) to identify haemoglobin components. A gel of similar composition, but containing in addition 6M-urea and 0.05M-2-mercaptoethanol, was used to identify \( \alpha \) and \( \beta \)-globin components (the bridge solution contained neither urea nor 2-mercaptoethanol, however). Gel pH was about 8.8. The samples were globin from which most of the haem had been removed by the method of Rossi-Fanelli et al. (1958); globin was dissolved in the same buffer solution used to make gels, at a concentration of 10mg/ml. Electrophoresis was at 4°C for 24h at 120V.

**Protein separations**

\( \beta \text{dmaj} \) and \( \beta \text{pmin} \) haemoglobins were isolated from haemolysates from mice with doublet and variant doublet alleles respectively by means of ionic-strength gradient chromatography with DEAE-Sephadex A-50 (Pharmacia Fine Chemicals, Piscataway, NJ, U.S.A.), as described in detail by Gilman (1976). Representative chromatograms and starch-gel electrophoretic identification of haemoglobin components (at pH 8.6) are given in the Supplementary Paper.

Haemolysate from mice with singlet alleles was used as a source of \( \beta s \) chain without any previous fractionation of haemoglobins, since \( \beta s \) chain is the only \( \beta \) chain of adult \( \text{Hbb}^{a}/\text{Hbb}^{a} \) homozygous mice. \textit{M. cervicolor} haemoglobin was not subjected to any fractionation before sequencing because only three mice were available.

Purified \( \beta \text{dmaj} \), \( \beta \text{pmin} \), \( \beta s \) and \( \beta \text{dmaj} \) chains were obtained from the respective haemoglobins by a variation of the method of Clegg et al. (1968). A column (2cm x 15cm) of carboxymethyl-cellulose (Whatman CM23, Whatman Biochemicals, Maidstone, Kent, U.K.) was equilibrated with starting buffer (0.01M-sodium acetate/8M-urea/0.05M-2-mercaptoethanol with sufficient acetic acid to lower the pH to 5.6). Globin dissolved in starting buffer was adsorbed to the column and chains were eluted from the column by means of a linear gradient made with 350ml of starting buffer and 350ml of final buffer, which has the sodium acetate concentration raised to 0.1M.

**Protein cleavage**

Tryptic digestion of purified \( \beta \text{dmaj} \) and \( \beta \text{pmin} \) chains, and purification of peptides, was as described by Gilman (1976); results are given in the Supplementary Paper. Leucine aminopeptidase ( Worthington Biochemicals Corp., Freehold, NJ, U.S.A.) digestion of peptides used the procedure described in Blackburn (1970). Tryptic digestion of maleylated \( \beta \) chain [prepared by the method of Butler et al. (1967) with minor modifications] was as follows: 10mg of maleylated \( \beta \) chain were dissolved in 1ml of 0.1M-NH$_4$HCO$_3$ to which a small amount of aq. NH$_3$ had been added to raise the pH to about 9.5; 0.2mg of 1-chloro-4-phenyl-3-L-toluene-p-sulphonamido-butane-2-one-treated trypsin ( Worthington Biochemical Corp., Freehold, NJ, U.S.A.) dissolved in 0.1ml of 1M-HCl was added and digestion proceeded for 2.5h at 37°C; another 0.2mg of trypsin was then added and digestion proceeded for another 1-2h, after which the digest was freeze-dried. With \( \beta \text{dmaj} \) chain, sequencing was done by using the freeze-dried digest without any subsequent purification procedures having been used (results are in the
Supplementary Paper). In the case of trypsin-cleaved maleylated βmin chain, the digest was partially fractionated on a column of Sephadex G-100 (2 cm x 40 cm), which was equilibrated with 0.1M-NH4HCO3 brought to pH 9.5 withaq. NH3. Three poorly resolved peaks were obtained. The first peak to be eluted was not sequenced as it contained very little material. The second peak contained primarily the largest expected peptide, with the sequence beginning at position 41 (see Supplementary Paper). The third peak contained peptides giving rise to all expected sequences (see Supplementary Paper).

CNBr cleavage was performed on globin or purified β chain as described by Gilman (1974), except that after cleavage salts were usually removed by means of Sephadex G-15 gel filtration (which did not cause any fractionation of peptides). No fractionation of peptides was performed before sequencing.

Sequencing procedures

Sequencing was carried out by using an Edman & Begg (1967) sequenator (Illiltron Division, Illinois Tool Works, Chicago, IL, U.S.A.). Sequenator methods and data-processing procedures were those of Smithies et al. (1971) with minor modifications. Each sequenator cycle consists, in essence, of the coupling of phenyl isothiocyanate to free N-terminal residues, followed by cleavage and extraction of the resultant amino acid thiazolinones. They were converted back into amino acids by HI hydrolyses or by NaOH/Na2S2O4 hydrolyses. For the N-terminal sequence of βmin chain (see Supplementary Paper) samples were split and both the HI hydrolysis and the original NaOH/Na2S2O4 hydrolysis described by Smithies et al. (1971) were performed. For the N-terminal sequence of βcma (d-like) chain given in the Supplementary Paper, HI hydrolysis was done for all 43 positions, and, in addition, a modified NaOH/Na2S2O4 hydrolysis was done for the first 30 positions. The latter type of hydrolysis was the only one used in all other experiments. It was milder than the original in that 0.15M-NaOH was used instead of 0.2M-NaOH, and hydrolysis was for 1 h instead of 3.5 h. The milder method permitted the recovery of approximately equal amounts of asparagine and aspartic acid after hydrolysis of asparagine thiazolinone; the original harsher method gave aspartic acid only.

After correction (by computer) for background amino acids, out-of-step sequencing and unequal yields of the various amino acids, the data were plotted automatically as the logarithm of the corrected recovery (in nmol) versus position number (Smithies et al., 1971). Background corrections were deliberately small so as to avoid loss of information, so that in the final Figures arbitrary cut-off lines were drawn to separate what was considered sequence information from what was considered, for the most part, background noise. In the final Figures presented here the term ‘relative recovery’ is used to label the ordinate, since, when the final Figures were drawn, the axes were arbitrarily shifted up or down (with the effect of multiplying the amount of every amino acid shown in the plot by some arbitrary constant factor).

The final plot for the N-terminal sequence of βmin chain (Fig. 3, Supplementary Paper) is that of NaOH/Na2S2O4 hydrolysis data, modified by insertion of data from HI hydrolysis; data from both types of hydrolysis were necessary in order to unambiguously distinguish alanine from either serine or cysteine (Smithies et al., 1971). The N-terminal sequence of M. cervicolor globin (Fig. 9, Supplementary Paper) is that of HI hydrolysis data, modified by insertion of NaOH/Na2S2O4 hydrolysis data for the first 30 positions. For the data on CNBr-cleaved protein, only NaOH/Na2S2O4 hydrolyses were performed, but deductions were nevertheless made with regard to distinguishing alanine from serine or cysteine. In those experiments where threonine was clearly recovered where expected (as a-aminobutyric acid), lack of recovery at a position was interpreted to imply the presence of serine or cysteine, as indicated by a ‘S’ in the interpretive sequences below the data plot (serine and cysteine are consistently destroyed by NaOH/Na2S2O4 hydrolysis). In those experiments where threonine was not consistently recovered, lack of recovery of any amino acid at a given position was interpreted to imply either threonine, serine or cysteine, as indicated by ‘t’ in the interpretive sequences; this occurred in the data on βcma (d-like) chain (Fig. 3) and in the data in the Supplementary Paper on βmaj and βmin chains. In the case of NaOH/Na2S2O4 hydrolysis data on maleylated trypsin-cleaved β chain (see Supplementary Paper), no interpretation was placed on lack of recovery at a position; the reason is that, in addition to the ambiguities just mentioned, lysine was never recovered and histidine was recovered in very low yield and sometimes not at all.

Results

Starch-gel study

Plate 1 illustrates the polymorphism for major β chains found in M. musculus in comparison with the similar major β chain polymorphism of M. cervicolor. Globin chains were separated on starch gel containing 6M-urea and 0.5M-2-mercaptoethanol at pH 8.8. Comparing two types of M. musculus haemolysates, that of an Hbb/|Hbb homozygote (slot 1) and that of an Hbb/|Hbb homozygote (slot 2), one sees that βs chain migrates more slowly towards the positive pole
than does βdmaj chain. This is expected, since βs chain has the neutral glycine at position 13 (Popp, 1972), whereas βdmaj chain has cysteine (Bonaventura & Riggs, 1967), which would have a partial negative charge at pH 8.8. The haemolysate from the Hbbb/Hbb site shows a faint band that migrated only a short distance from the origin towards the positive pole; this is presumably βdmaj chain, and it is not seen in slot 1, because Hbbb/Hbb4 homozygote produces no minor β chain.

Slots 3–5 contain haemolysates of three different mice of species M. cervicolar. Slots 3 and 4 show identical patterns. Two major bands and two minor bands are clearly seen. One major band has the same mobility as M. musculus α chain and is presumably due to α chain. The other major band has mobility slightly less than that of βdmaj chain; I have called this band βcmaj (d-like). Of the two minor bands, one is probably due to minor β chain and the other to minor α chain (Gilman, 1974), though precisely which is α chain and which is β chain is uncertain. An assignment of the names αmin and βmin has, however, been provisionally made, by using the rationale given in the legend of Plate 1.

The pattern of the M. cervicolar haemolysate of slot 5 is similar to that of slots 3 and 4, except that βcmaj (d-like) chain is not seen and instead there is a major chain with mobility similar to that of βs chain; I have called this major M. cervicolar chain βcmaj (s-like). Presumably αmin chain is also present in this haemolysate, though it is not clearly seen, because its mobility is only slightly different from that of βcmaj (s-like). Thus the only difference between the pattern of slots 3–4 and that of slot 5 is the presence of βcmaj (d-like) chain in the former and βcmaj (s-like) chain in the latter.

Sequence data: βs and βdmaj chains

The Edman & Begg (1967) sequenator may be used to obtain sequence data beginning at the N-terminus (position 1) of a protein. In order to obtain sequencer data for the region beyond position 50 or so, it is necessary to cleave the protein at one or more positions. For the data in the present paper, the method of CNBr cleavage was the one primarily used, since CNBr cleaves specifically on the C-terminal side of methionine, and very few methionine residues are present in haemoglobin chains. No purification procedures were used to isolate the various peptide fragments, since the sequenator has been shown to give interpretable data on a mixture of proteins (see Gilman, 1976).

Fig. 1 gives the sequenator data on CNBr-cleaved βs chain, for which Popp (1972) has published an almost complete sequence. Fig. 1 illustrates the results that I have obtained sequencing a mixture of peptides resulting from CNBr cleavage of β chain. At most three amino acids are seen at any position in this Figure, and the data of Popp (1972) permit a sorting-out of these data into three sequences. One of these sequences is from the original N-terminus. The other two are from the new N-termini generated by CNBr cleavage at methionine residues. One, beginning at position 56, is due to cleavage at methionine-55; the other, beginning at position 110, is due to cleavage at methionine-109. The residues from the latter sequence are circled in the plot; it is seen that recoveries from this sequence fall off rapidly, and beyond cycle 13 there is essentially no recovery of this sequence.

The data of Fig. 1 are consistent with the trypsin-peptide composition data of Rifkin et al. (1966), but are not entirely consistent with the sequence given by Popp (1972). These data show asparagine at position 72 and serine or cysteine at position 80; trypsin-peptide composition data (Rifkin et al., 1966) indicate that it must be serine rather than cysteine. Popp (1972), however, found serine at position 72 and asparagine at position 80.

An equivalent CNBr-cleavage experiment performed on βdmaj chain from random-bred HA/ICR strain mice (see Supplementary Paper) showed no differences from my data on βs chain in the region positions 72–80. However, Popp & Bailiff (1973) found inversions at positions 72–73 and positions 76–77 in βdmaj chain from Balb/c strain mice, with respect to the data of Popp (1973) on βs chain. Examining haemoglobin from Balb/c strain mice I could find no evidence that βdmaj chain from this strain was any different from βdmaj chain from HA/ICR strain mice or from βs chain, in the region positions 72–80 (Gilman, 1973).

Sequence data: βdm and βpmin chains

Fig. 2 shows the results of a sequenator experiment on CNBr-cleaved βdm chain. In contrast with the data for βs chain (Fig. 1), only two sequences are seen. Data presented in the Supplementary Paper, on the sequence for positions 1–43 and 41–58 of βdm chain and trypsin-peptide composition data on that chain, allowed deduction of two interpretive sequences consistent with the data of Fig. 2, which are shown below the data plot. One sequence begins at position 1, from the original N-terminus, and the other begins at position 56, from the peptide formed by cleavage at methionine-55. The absence of any other sequence due to CNBr cleavage at methionine suggests the absence of methionine at position 109 of βdm chain, at which methionine is found in βs and βdmaj chains.

Tryptic cleavage of maleic anhydride-treated βdm chain was another relatively specific cleavage method used to confirm directly the absence of methionine at position 109. Such maleylated β chain would have lysine residues blocked and unsusceptible
Globin from haemolysates of *M. musculus* $Hbb^d/Hbb^*$ and $Hbb^s/Hbb^s$ are shown for comparison. The haemolysates are from: slot 1, $Hbb^d/Hbb^*$ mice; slot 2, $Hbb^s/Hbb^s$ mice; slots 3, 4 and 5, three different *M. cervicolor* mice. Labels on the left are for bands from *M. musculus* samples, whereas labels on the right are for bands from *M. cervicolor* samples. The rationale for the labelling of *M. musculus* bands and the major bands of *M. cervicolor* is given in the text. For the minor bands of *M. cervicolor*, only educated guesses could be made. Starch-gel electrophoresis in the absence of urea (pH 8.6) suggests that *M. cervicolor* has a minor $\alpha$ chain ($\alpha cmin$) and a minor $\beta$ chain ($\beta cmin$) in addition to the major $\alpha$ chain ($\alpha cmaj$) and major $\beta$ chains [$\beta cmaj$ (d-like) and $\beta cmaj$ (s-like)] (Gilman, 1974). Those data show that the haemoglobin containing $\beta cmin$ and $\alpha cmaj$ chains is more positively charged at pH 8.6 than the haemoglobins containing $\alpha cmaj$ and either $\beta cmaj$ (d-like) or $\beta cmaj$ (s-like) chains; the latter haemoglobins, in turn, are more positively charged than haemoglobins containing $\alpha cmin$ and either $\beta cmaj$ (d-like) or $\beta cmaj$ (s-like) chains. Thus it is likely that $\alpha cmaj$ chain is more positively charged than $\alpha cmin$ chain, whereas $\beta cmin$ chain would be more positively charged than either $\beta cmaj$ (d-like) or $\beta cmaj$ (s-like) chains. The labelling of $\alpha cmin$ and $\beta cmin$ chains given in this Plate is consistent with that deduced from the data of Gilman (1974).

**Urea/starch-gel electrophoresis (pH 8.8) of globin from haemolysis of three mice of species *M. cervicolor***

Globin from haemolysates of *M. musculus* $Hbb^d/Hbb^*$ and $Hbb^s/Hbb^s$ are shown for comparison. The haemolysates are from: slot 1, $Hbb^d/Hbb^*$ mice; slot 2, $Hbb^s/Hbb^s$ mice; slots 3, 4 and 5, three different *M. cervicolor* mice. Labels on the left are for bands from *M. musculus* samples, whereas labels on the right are for bands from *M. cervicolor* samples. The rationale for the labelling of *M. musculus* bands and the major bands of *M. cervicolor* is given in the text. For the minor bands of *M. cervicolor*, only educated guesses could be made. Starch-gel electrophoresis in the absence of urea (pH 8.6) suggests that *M. cervicolor* has a minor $\alpha$ chain ($\alpha cmin$) and a minor $\beta$ chain ($\beta cmin$) in addition to the major $\alpha$ chain ($\alpha cmaj$) and major $\beta$ chains [$\beta cmaj$ (d-like) and $\beta cmaj$ (s-like)] (Gilman, 1974). Those data show that the haemoglobin containing $\beta cmin$ and $\alpha cmaj$ chains is more positively charged at pH 8.6 than the haemoglobins containing $\alpha cmaj$ and either $\beta cmaj$ (d-like) or $\beta cmaj$ (s-like) chains; the latter haemoglobins, in turn, are more positively charged than haemoglobins containing $\alpha cmin$ and either $\beta cmaj$ (d-like) or $\beta cmaj$ (s-like) chains. Thus it is likely that $\alpha cmaj$ chain is more positively charged than $\alpha cmin$ chain, whereas $\beta cmin$ chain would be more positively charged than either $\beta cmaj$ (d-like) or $\beta cmaj$ (s-like) chains. The labelling of $\alpha cmin$ and $\beta cmin$ chains given in this Plate is consistent with that deduced from the data of Gilman (1974).
Interpretive sequences of the various peptides, deduced as described in the text, are shown below the data plot. Lack of recovery of an amino acid at a position was taken to imply cysteine or serine at that position (see the Experimental section for the rationale). Residues interpreted as being partially or solely due to the sequence beginning at position β110 are circled in the data plot. The one-letter codes for the amino acids are: A, alanine; B, aspartic acid or asparagine; C, cysteine; D, aspartic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; Z, glutamic acid or glutamine. Three additional symbols are used: $, serine or cysteine; &, serine, cysteine or alanine; t, serine, cysteine or threonine.

These data clearly show the absence of methionine at position 109 of βdmin chain (equivalent data on βdmaj chain, also given in the Supplementary Paper, do show the presence of methionine at position 109). The sequence data on βdmin chain were interpreted as showing the presence of alanine at position 109 in that chain (see Supplementary Paper for a more detailed discussion). The ‘unexpected’ sequence beginning at position 109 in the data on trypsin-cleaved maleylated βdmin chain was interpreted as due to a cleavage (of unknown cause) between asparagine and alanine. The only Asn-Ala sequence

Vol. 159
Interpretive sequences, deduced as described in the text, are shown below the data plot. Lack of recovery of an amino acid at a position was taken to imply cysteine or serine at that position (see the Experimental section for rationale). Triangles are around lysines that were recovered in low yield from the $\beta$56 sequence. Beyond cycle 22, the yield of amino acids from the $\beta$1 sequence has declined; residues are circled that are believed to be from that sequence, but which occur below the cut-off line. See the legend to Fig. 1 for an explanation of symbols.

The Supplementary Paper presents sequence data on CNBr-cleaved $\beta$pm min chain which are similar to those of Fig. 2 on $\beta$dm min chain. Only two sequences are seen, suggesting that $\beta$pm min chain lacks methionine at position 109 and is thus similar to $\beta$dm min in this regard. The data on $\beta$pm min chain do show two sequence differences from $\beta$dm min chain: at position 22 $\beta$pm min chain has alanine and at position 23 isoleucine (compared with glutamic acid or glutamine).
MOUSE ADULT HAEMOGLOBIN BETA CHAINS

Fig. 3. Sequence of a mixture of peptides from CNBr cleavage of M. cervicolor globin containing βcmaj (d-like) chain

The globin was from the same sample as that used in slot 4 of the starch gel of Plate 1; those data show that the globin sample contained the M. cervicolor major β chain variant βcmaj (d-like) and little or no βcmaj (s-like). Interpretive sequences of the various peptides, deduced as described in the text, are shown below the data plot. In the data plot itself, underlined residues are from the α1 sequence, residues in squares are from the α33 sequence, and residues in triangles are from the β56 sequence. For the interpretive sequences below the data plot, if no amino acid was recovered at a given position, then the correct assignment was taken (where possible) from the N-terminal sequence data for 43 positions on M. cervicolor globin (Fig. 9, Supplementary Paper); where supplementary data were lacking, no recovery was taken to imply threonine, cysteine or serine at that position (see the Experimental section for the rationale). Below the cut-off line of the data plot, at cycles 18, 21 and 22, residues are circled that are believed to be from the β56 sequence of βcmaj chain (positions 73, 76 and 77). See the legend to Fig. 1 for an explanation of symbols.

Sequence data: βcmaj (d-like) chain from M. cervicolor

Only three M. cervicolor mice were available, two of which had βcmaj (d-like) chain and one of which had βcmaj (s-like) chain, as shown in Plate 1. All experiments were therefore performed on un-fractionated globin in order to obtain the maximum information from the limited material available. I here present sequence data on globin containing βcmaj (d-like) chain. Equivalent experiments were performed on globin containing βcmaj (s-like) chain (Gilman, 1973).
The globin sample was cleaved with CNBr and sequenced without any fractionation of peptides. The results are shown in Fig. 3. A separate experiment on uncleaved globin provided data on the N-terminal 43 positions (see Supplementary Paper). The N-terminal sequence data, and recourse to arguments of homology with *M. musculus* α and β chains, allowed sorting-out of the data into four sequences, as shown in the interpretive sequences below the data plot.

Of the four sequences, two are from the original free N-termini of the α and β chains; the residues deduced as being from the N-terminal α chain sequence are underlined in the data of Fig. 3, in order to emphasize that recoveries from this sequence are seen to fall off more rapidly than recoveries from the other sequences. There are also two sequences from the new free N-terminal residues generated as a result of CNBr cleavage: one begins at position α33 (residues in squares in Fig. 3), and the other begins at position β56 (residues in triangles in Fig. 3). The purpose of this experiment was to obtain data on the interior of the β chain, and it can be seen that recovery of amino acids from the sequence covering positions 56–82 of βcmaj (d-like) chain is quite good throughout all 27 sequenator cycles. It was even possible to detect minor residues in these data, deduced as being from βcmaj chain by homology with βdmin chain. At cycles 18, 21 and 22, there are minor residues of glutamic acid or glutamine, lysine and asparagine; they are enclosed in circles, and are assumed to be from positions 73, 76 and 77 of βcmaj chain.

The data of Fig. 3 and the N-terminal sequence data of the Supplementary Paper show that βcmaj (d-like) chain has serine or cysteine at position 13; starch-gel data of Gilman (1974) would favour cysteine, which is the same residue found at that position in βdmin chain. Sequence data on βcmaj (s-like) chain (Gilman, 1973) show that it has glycine at position 13, like βs chain. Additional differences among the various chains are discussed below.

**Supplementary information**

Detailed evidence for some of the amino acid sequence data reported above has been deposited with the British Library Lending Division as SUP 50067. The information comprises:

1. Details on the purification of βdmin chain. N-Terminal sequence for 43 positions of globin containing α and βdmin chains.
2. Sequence data on peptides from tryptic digestion of maleylated βdmin chain. One peptide was isolated, and data cover positions 41–58 of βdmin chain; data on a mixture of peptides are also presented as discussed above, which were interpreted to show alanine at position 109 of βdmin chain.
3. Amino acid composition of tryptic peptides of βdmin chain, with the exception of the peptide covering the region positions 105–120, which was not recovered.

4. A deduction of the partial sequence of βdmin chain, based on the data of this paper and of the Supplementary Paper; this partial sequence has already been reported (Gilman, 1976).

5. Data on CNBr-cleaved βdmaj chain (similar to that on βs chain of Fig. 1) and data on CNBr-cleaved βpmin chain (similar to that on βdmin chain of Fig. 2).

6. N-Terminal sequence data for 43 positions on unfractionated globin from *M. cervicolor* that contained βcmaj (d-like) chain (as well as acmaj, βcmin and acmin chains).

**Discussion**

**Sequence data**

In *M. musculus musculus*, three breeding unit alleles are known at the adult β-chain locus: Hbb*, Hbb* and Hbb* (Russell & McFarland, 1974). Hbb* is a singlet, producing only one β chain, βs; Hbb* and Hbb* are doublets, composed of two closely linked genes, one producing a major β chain (βdmaj and βpmin respectively) and the other a minor β chain, found as about 20% of total β chain (βdmin and βpmin respectively). Other β-chain patterns were found in mice of two other species, *M. cervicolor* and *M. caroli*, both from Thailand (Gilman, 1974). For *M. cervicolor*, the data suggest two types of doublet loci, both of which produce major and minor β chains and which differ in the type of major β chain produced [βcmaj (d-like) versus βcmaj (s-like) chains]. For *M. caroli*, only one type of β chain was found in the mice examined, whose structure was 'Lepore-like' in that it had structural landmarks characteristic predominantly of βdmin chain in the N-terminal half and of βdmaj chain in the C-terminal half.

The present paper has presented sequence data on *M. musculus* β chains, βs, βdmin, βdmaj and βpmin, and on *M. cervicolor* βcmaj (d-like) chain. A preliminary summary of the results has already been presented (Gilman, 1974); the partial sequence of βdmin chain, based on data of this paper, has also been previously reported (Gilman, 1976). Table 1 shows a comparison of these chains, and of βcmaj (s-like) chain, at positions where differences between them have been found.

In these data, summarized in Table 1, there are three important uncertainties, at position 109 in βdmin chain, and positions 13 and 76 in βcmaj (d-like) chain. For position 109 of βdmin chain, rather complex data (see Supplementary Paper) were interpreted to show the presence of alanine, but additional confirmatory data would be desirable. For position 13 of βcmaj (d-like) chain the N-terminal sequence data of the Supplementary Paper show serine or cysteine at position 13, and it is tempting to
Table 1. Comparison of M. cervicolor β chains, βcmaj (d-like) and βcmaj (s-like), with M. musculus β chains, βs, βdmaj, βdmaj and ϕβmin, at positions where differences have been found

Data on βs and βdmaj chains are primarily those of Popp (1972, 1973) and Popp & Bailiff (1973) respectively, except where their results conflicted with those of the present paper, as discussed in the text; where such conflict occurred (in the region of positions 72–80), results presented are those of the present paper, in conjunction with tryptic-peptide composition data of Rifkin (1968). Data on ϕβmin (s-like) and on position 139 of ϕβmin chain are from Gilman (1973), who also obtained data on βdmaj chain, similar to those reported in the present paper for βdmaj chain, and failed to find any differences between those two chains. Residues are given in parentheses if only tryptic-peptide composition data were available and the position in the sequence was deduced by homology with βdmaj chain. Unusual symbols are: Syr, serine or cysteine; *, serine or threonine; **, not methionine; ***, serine, cysteine or threonine.

<table>
<thead>
<tr>
<th>Position</th>
<th>9</th>
<th>13</th>
<th>16</th>
<th>20</th>
<th>22</th>
<th>23</th>
<th>58</th>
<th>73</th>
<th>76</th>
<th>77</th>
<th>80</th>
<th>109</th>
<th>139</th>
</tr>
</thead>
<tbody>
<tr>
<td>ϕβcmaj (d-like)</td>
<td>Ala</td>
<td>Syr</td>
<td>Ala</td>
<td>Syr</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Asp</td>
<td>**</td>
<td>His</td>
<td>Asn</td>
<td>*?</td>
</tr>
<tr>
<td>ϕβcmaj (s-like)</td>
<td>Ala</td>
<td>Gly</td>
<td>Gly</td>
<td>Syr</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Glx</td>
<td>**</td>
<td>His</td>
<td>Asn</td>
<td>*?</td>
</tr>
<tr>
<td>βdmaj</td>
<td>Ala</td>
<td>Cys</td>
<td>Gly</td>
<td>Ser</td>
<td>Glx</td>
<td>Val</td>
<td>Gly</td>
<td>Ala</td>
<td>Asp</td>
<td>Asn</td>
<td>His</td>
<td>*</td>
<td>Met</td>
</tr>
<tr>
<td>βs</td>
<td>Ala</td>
<td>Gly</td>
<td>Gly</td>
<td>Ala</td>
<td>Glu</td>
<td>Val</td>
<td>Gly</td>
<td>Ala</td>
<td>Asp</td>
<td>Asn</td>
<td>His</td>
<td>Ser</td>
<td>Met</td>
</tr>
<tr>
<td>βdmaj</td>
<td>Ser</td>
<td>Cys</td>
<td>Ala</td>
<td>Pro</td>
<td>Glx</td>
<td>Val</td>
<td>Gly</td>
<td>Pro</td>
<td>Glu</td>
<td>Lys</td>
<td>Asn</td>
<td>Asn</td>
<td>Ala</td>
</tr>
<tr>
<td>βϕmin</td>
<td>Ser</td>
<td>Cys</td>
<td>Ala</td>
<td>Pro</td>
<td>Ala</td>
<td>Ile</td>
<td>Gly</td>
<td>Pro</td>
<td>Glx</td>
<td>Lys</td>
<td>Asn</td>
<td>Asn</td>
<td>**</td>
</tr>
</tbody>
</table>

assume cysteine, by homology with βdmaj chain. Data from starch-gel electrophoresis of cystamine-treated haemolysates lend strength to this surmise, as they show that haemoglobin containing ϕcmaj and βcmaj (d-like) chains has the same number of cysteines available for reaction with cystamine as does the analogous M. musculus haemoglobin, which contains ϕ and βdmaj chains (Gilman, 1974). With regard to position 76 of ϕcmaj (d-like) chain, the data on CNBr-cleaved M. cervicolor globin (Fig. 3) show only alanine and aspartic acid at cycle 21, which were interpreted as due to position α53 (from the α33 sequence) and to position 21 respectively. This interpretation means that position 76 gave no recovery, which, by the reasoning given in the Experimental section, suggests either serine, threonine or cysteine at position 76 in βcmaj (d-like) chain. [The same result was also obtained for βcmaj (s-like) chain (Gilman, 1973).] This assignment for position 76 seems reasonable, as serine is at position 76 in rabbit adult β chain (Dayhoff, 1972), but since the residue at position α53 is not known for M. cervicolor, and homology with M. musculus α chain was assumed, additional data on position β76 would be desirable.

Not shown in Table 1 is the sequence information of βcmaj chain that can be deduced by the presence or absence of minor residues in the data on globin from M. cervicolor haemolysate. In Fig. 3, residues are circled that are believed to be from positions 73, 76 and 77 of βcmaj chain: glutamic acid or glutamine, lysine and asparagine respectively, which are the same residues at those positions in βdmaj chain. These CNBr-cleavage data also suggest that βcmaj chain may be like βdmaj chain at position 109, since no minor residues are seen in Fig. 3 that would indicate that a CNBr cleavage had occurred at position 109 of βcmaj chain. For those positions listed in Table 1, there are two further similarities between βcmaj and βdmaj chains. N-Terminal sequence data on globin containing βcmaj (s-like) chain (Gilman, 1973) suggest that βcmaj and βdmaj chains are alike at position 16, since those data show a minor alanine residue at that position. A similarity between βcmaj and βdmaj chains at position 13 is suggested by starch-gel data on cystamine-treated M. cervicolor haemolysate (Gilman, 1974), that demonstrate that βcmaj chain has the same number of free cysteines as does M. musculus βϕmin chain; βϕmin and βdmaj chains both have two cysteines, one of which is at position 13, suggesting that βcmaj chain may also have cysteine at position 13.

At some of the positions shown in Table 1, β cmaj chain can be shown to be clearly different from βdmaj chain. In the N-terminal data of the Supplementary Paper, minor aspartic acid or asparagine residues are found at position 12, which could be from βcmaj chain [by homology with mouse embryonic y chain (Gilman, 1976)], although the possibility cannot be excluded that the minor residue at position 12 came from βdmaj chain. At position 20, no minor proline residue was recovered in the N-terminal sequence data, nor was minor glutamic acid or glutamine residues recovered at position 22, which suggests that βcmaj chain may be different from βdmaj chain at those positions and perhaps like the major M. cervicolor β chains. Nor was minor proline residue recovered at cycle 3 of the CNBr-cleavage data of Fig. 3, implying no proline at position 58 in βcmaj chain. Thus a search for minor residues in the sequence data on M. cervicolor globin provides some information on βcmaj chain at several of the positions.
listed in Table 1. At five positions, 16, 73, 76, 77 and 109, βcmin and βdmin chains may be similar, and starch-gel data suggest that they may also be alike at position 13. βcmin and βdmin chains are probably different at positions 20, 22 and 58.

Evolutionary possibilities

Nute et al. (1974) have given evidence demonstrating a possible functional role in haemoglobin of position β109, one of the positions at which significant variation exists in mouse β chain. Such variability is not found in other mammalian species (Dayhoff, 1972), whose β chains generally have valine, except for grey kangaroo, which has isoleucine. Nute et al. (1974) showed that human haemoglobin mutant Hb San Diego has methionine at position β109 (like mouse βdmin and βs chains), and that it has higher oxygen affinity compared with normal human haemoglobin; they also suggested that Hb San Diego haemoglobin may have abnormal co-operative properties, although Gilman et al. (1975) have argued that the data of Nute et al. (1974) do not prove this, as all of their oxygen-dissociation-value data were obtained by using a mixture of normal and mutant haemoglobin. It is therefore a possibility that variation at position 109 in mouse β chain led to a variation in functional properties of the haemoglobins without reducing co-operativity, and that the presence of variability at that position may be a result of the action of natural selection.

Another sequence variability for which a role for selection may be indicated occurs at position 13. Two variant major β chains, one with cysteine at position 13 and the other with glycine, are found in each of two Mus species (musculus and cervicolor), for which DNA-hybridization data indicate a time of divergence one-third as long ago as mouse diverged from rat (Rice, 1972). If one assumes that this cysteine/glycine difference predated the species divergence, then this difference at position 13 has persisted in the two species for at least the time spanning the million years.

On the other hand, if one assumes that this difference at position β13 arose after the divergence of M. musculus and M. cervicolor, then it must have arisen and become widespread, independently in the two species, some time in the past several million years. Either hypothesis suggests that the cysteine/glycine difference at position β13 may be of selective value to mice, since it is a constant factor in the two species, although many other elements of the β-chain sequence have varied between them. Although position 13 probably does not have a functional role in haemoglobin itself, because it is external (Perutz & Lehmann, 1968), a cellular functional role can be imagined. The cysteine/glycine variation at position β13 could affect the pH-buffering properties of haemoglobin, which might have an effect on the concentration of cellular 2,3-diphosphoglycerate, or the cysteine at position 13 could help protect from oxidation other more important amino acid residues of haemoglobin or other erythrocyte proteins.

Assuming for the moment the hypothesis that the cysteine/glycine difference at position β13 preceded the M. cervicolor divergence, then cysteine at position 13 may be considered a ‘marker’ for an ancestral βdmaj chain, whereas glycine at position 13 would be a marker for ancestral βs chain; after species divergence, ancestral βdmin chain is assumed to have evolved to βdmaj chain in M. musculus and to βcmaj (d-like) chain in M. cervicolor, whereas βs chain would have evolved to βs chain in M. musculus and to βcmaj (s-like) chain in M. cervicolor. [See Gilman (1976) for one possible evolutionary scheme which makes this assumption.] Under this hypothesis, the sequence data of Table 1 force one to the conclusion that a considerable amount of convergent evolution must have occurred for βs and βdmaj in each species. For example, at position 22, the two M. musculus major β chains, βs and βdmin, have glutamic acid or glutamine, although the two M. cervicolor major β chains, βcmaj (s-like) and βcmaj (d-like), have alanine. The same type of convergence is observed at position 25, where the two M. musculus major β chains have glycine and the two major M. cervicolor β chains have alanine, and at position 109, where the two M. musculus chains have methionine, whereas neither of the two M. cervicolor chains has methionine.

One can avoid the necessity of postulating convergent evolution for βs and βdmaj chains if one assumes that the cysteine/glycine difference at position β13 arose independently in the two species after their divergence and after the major β chains of M. cervicolor had diverged from the major β chains of M. musculus at positions 22, 25 and 109. However, even with this hypothesis, a different type of convergent evolution must be postulated, in which a major β chain of M. cervicolor became like the minor β chain at a position.

The discussion under ‘Sequence data’ (above) suggested that M. musculus and M. cervicolor minor β chains βdmin and βcmaj respectively are alike at positions 16 (alanine) and 73 (glutamic acid or glutamine), as well as at a number of other positions, which implies that the ancestral minor β chain βdmin had those same residues. At position 16, M. musculus βs and βdmaj chains and M. cervicolor βcmaj (s-like) chain have glycine, which suggests that glycine was the ancestral major β-chain residue in both βs and βdmaj chains, yet M. cervicolor βcmaj (d-like) chain has alanine, like the minor chains. Similarly, at position 73, βcmin and βdmin chains have glutamic acid or glutamine residues, and βs, βdmaj and βcmaj (d-like) chains have an aspartic acid residue, yet βcmaj (s-like) chain has a glutamic acid or glutamine residue like the minor chains.

1976
Apparently, at positions 16 and 73, one or the other of the *M. cervicolor* major β chains shows convergent evolution with the minor β chain.

All such instances of convergent evolution described above can be explained by the hypothesis that the gene coding for one of the chains became like the gene coding for another chain, for a small region of the gene, as the result of a process involving genetic exchange. Hybrid-DNA models for genetic recombination (Holliday, 1964; Whitehouse, 1963, 1973) envisage the occurrence of just such a process, both with and without concurrent genetic cross-over. In fact, such hybrid-DNA models had, as one of their goals, the explaining of the phenomenon of negative interference over short genetic regions, such as had been observed in *Neurospora* (Mitchell, 1955); negative interference is the apparent excess occurrence of double cross-overs. In effect, I am postulating the occurrence of such double cross-overs in order to account for the insertion of a genetic region coding for alanine at position β16 from the gene coding for βmin chain into the gene coding for βmaj (d-like) chain; the hybrid-DNA models of genetic recombination explain how this might have occurred, and why in apparently high frequency.

The data on mouse haemoglobin β chains thus appear to show evidence for the past occurrence of unequal cross-over events, such as that which may have given rise to the 'Lepore-like' β chain of *M. caroli* (Gilman, 1974, 1976), and also for genetic exchange events not involving cross-over. Both types of exchange events would have had the effect of producing 'mosaic' β chains, which combined structural features of the previously existing β chains and which may thereby have differed selectively from the previous β chains. The different patterns of variability and polymorphism found in the various mouse species may then have come into being as a consequence of the action of selective forces on the various new β chains which were created by those exchange processes.

I thank Dr. O. Smithies, in whose laboratory this work was performed, for advice, technical assistance and freedom in the pursuit of this project. I thank Dr. E. S. Russell for encouragement and gifts of animals, Dr. T. C. Hsu for gifts of animals, and Dr. L. Rossi-Bernardi for his hospitality during the writing of this paper. This is paper no. 1849 from the Laboratory of Genetics, University of Wisconsin. The work was supported in part by the National Institutes of Health (GM20069, GM00398 and 1 FO 2 HL55564-02).

References


