Hemoglobin Beta Chain Structural Variation in Mice: Evolutionary and Functional Implications

Abstract. *The* Hbb^d *allele at the hemoglobin beta chain locus of* Mus musculus *is composed of two linked genes, coding for structurally different beta chains, βdmin and βdmaj.* Mus caroli *has only one beta chain, which combines structural features of both βdmin and βdmaj and thus may be a Lepore type of beta chain. Sequence data suggest that selection may have been important in the evolution of the mouse beta chains.*

Three hemoglobin beta chain alleles have been indentified at the *Hbb* locus of *Mus musculus* (*1, 2*). Two of them, Hbb^d and Hbb^p, give rise to multiple-component phenotypes, as shown after separation by starch gel electrophoresis, while Hbb^s produces a single homogeneous band (*2*). Fresh hemolyzates from Hbb^d/Hbb^d and Hbb^p/Hbb^p mice both show three bands, a major band and two minor bands, and aged solutions show additional minor bands. To explain the multiple band pattern of Hbb^d/Hbb^d mice, Hutton *et al.* (*3*) provided evidence that this homozygote has two hemoglobins that differ in their beta chains. They are found in unequal amounts, the minor one as 20 percent and the major one as 80 percent of the total, and their beta chains are produced by closely linked loci. Riggs (*4*) and Morton (*5*) showed that some of the minor bands of Hbb^d/Hbb^d and Hbb^p/Hbb^p mice are polymers of hemoglobin molecules, but none of the studies verified the report of Hutton *et al.* (*3*).

Partial amino acid sequences of the beta chains from several strains of mice are presented in this report (*6*). These studies establish that there are two beta chains, βdmaj and βdmin, in Hbb^d/Hbb^d homozygous mice, and that these chains differ extensively from each other and from the sole beta chain of Hbb^s/Hbb^s mice. Mice of genotype Hbb^p/Hbb^p also have two beta chains—βpmaj, which is like βdmaj, and βpmin, which is a variant of βdmin. While searching for additional variants, I discovered a new beta chain in a Thai mouse species, *Mus caroli*. This beta chain closely resembles βdmin in its NH_2-terminal region, but it resembles βdmaj in its COOH-terminal half. This new mouse beta chain, βcaroli, might therefore be considered a Lepore type of beta chain (*7*).

Minor hemoglobin was isolated from the multicomponent hemolyzates by means of diethylaminoethyl Sephadex chromatography with an ionic strength gradient of ammonium acetate at alkaline *p*H. Most of the tryptic peptides of the beta chain were obtained by established procedures (*8, 9*). To supplement peptide data, I obtained NH_2-terminal sequences up to 50 residues long using an Edman-Begg sequenator (*10*) and procedures described elsewhere (*11*).

These methods provided peptide composition or sequence data on all of βdmin, 70 percent of βpmin, 60 percent of βcaroli, 50 percent of βdmaj, and 30 percent of βpmaj. A comparison of these beta chains for positions at which they differ is shown in Fig. 1 (βpmaj is identical to βdmaj to the extent of this analysis). The βcaroli chain is most like βdmin up to position 58 and is most like βdmaj from position 73 on. The fact that βcaroli combines NH_2-terminal features of βdmin and COOH-terminal features of βdmaj and is the sole beta chain of the species (*12*) is consistent with the hypothesis that it arose in the same way proposed for the human Lepore hemoglobins (*7*). This hypothesis asserts that the structural gene coding for βcaroli is a recombinant gene, the product of an unequal crossover between the gene of the Hbb^d complex that codes for βdmin and, on the other chromosome, the gene of Hbb^d that codes for βdmaj.

Other simple evolutionary schemes do not require that the structural gene coding for βcaroli be a recombinant gene. For example, a hypothetical ancestral mouse population might have two beta chain alleles, $Hbb^{\hat{s}}$ and $Hbb^{\hat{c}}$, which produce beta chains with the sequences in Fig. 1. A breakage and rejoining event putting the two alleles in tandem on the same chromosome could produce the ancestor of the duplicated Hbb^d allele, while the unduplicated $Hbb^{\hat{s}}$ allele might evolve to the present unduplicated Hbb^s allele, and the $Hbb^{\hat{c}}$ allele could evolve to the present structural gene for βcaroli. This scheme requires the least number of events to account for the evolution from an unduplicated ancestral beta chain locus to the present complex situation. However, the existence of several types of human Lepore hemoglobins (*7*) suggests that unequal crossing-over between similar, closely linked beta chain loci is not an unusual event. The constraints imposed by the present data are not rigorous enough to allow a definitive choice between a Lepore hypothesis and some others.

Any theory explaining the evolution of the mouse beta chains will have to account for the existence and possible fixation of caroli hemoglobin in that species (*12*) and for the Hbb^d, Hbb^s, Hbb^p polymorphism of musculus. Two hypotheses can do this. A selection hypothesis might state that the caroli hemoglobin was superior to the other hemoglobins in the population and so became its only hemoglobin, and a selection hypothesis could also suggest that the polymorphism of musculus is maintained because on the average the heterozygotes are superior to the homozygotes. A random drift or non-Darwinian hypothesis (*13*) might suggest

	9	14	16	20	22	23	58	73	76	77	80	109	111	121
βdmaj	Ala	Leu	Gly	Ser	Glx	Val	Ala	Asp	Asn	His	*	Met	Val	Asp
βdmin	Ser	Leu	Ala	Pro	Glx	Val	Pro	Glu	Lys	Asn	Asn	**	Val	(Asx)
βpmin	Syr	Leu	Ala	Pro	Ala	Ile	(Pro	Glx	Lys	Asn	Asn)	?	?	(Asx)
βcaroli	Ala	Met	Ala	Pro	Glx	Val	Pro	Asp	Asn	His	Asn	Met	Leu	Glx
Hypothetical Ancestral Sequences:														
β$\hat{s}$	Ala	Leu	Gly	Pro	Glu	Val	Ala	Asp	Asn	His	Asn	Met	Val	Asp
β$\hat{c}$	Ala	Leu	Ala	Pro	Glu	Val	Pro	Asp	Asn	His	Asn	Met	Val	Asp

Fig. 1. A comparison of βdmaj, βdmin, βpmin, and βcaroli at the positions where differences were found. No differences were found between βpmaj and βdmaj; β$\hat{s}$ and β$\hat{c}$ are hypothetical sequences required by the non-Lepore hypothesis presented; differences between them are underlined. Some of the data on βdmin have been reported (*18*). The sequence of βs has been reported by Popp (*20*), and sequence data on beta chains produced by the Hbb^d allele were reported by Bonaventura and Riggs (*21*) and Rifkin (*9*). Amino acid symbols are given in (*22*).

that only by chance have these phenomena occurred and that no hemoglobin need be functionally different from any other.

Population studies indicate that selection may be involved in maintaining the musculus polymorphism. Every adequately sampled population of wild musculus in the United States and Denmark has both Hbb^s and Hbb^d alleles, and in some of the populations Selander *et al.* (*14*) found an excess of Hbb^s/Hbb^d heterozygotes. No population or functional data on caroli hemoglobin have yet been obtained, so it is not known whether selection has influenced the structural evolution of its beta chain. However, the Perutz model of hemoglobin functioning (*15*) suggests to me that functional differences may exist between caroli hemoglobin and the other mouse hemoglobins. Two residues in caroli (at positions 14 and 111), with side chains directed internally, differ from all other mouse beta chains. The methionine residue at position 14 is found in no other beta chain that has been sequenced (*16*).

It is possible to posit a specific functional role for this methionine residue. In the three-dimensional model of myoglobin, the methionine side chain at position 14 would be located between the A and E helices in contact with the tryptophan at position 15. This tryptophan serves as a "spacer" (*15*) between the A and E helices; therefore, it and any residues contacting it could be important in regulating the spacing between the A and E helices.

Experiments of Arnone (*17*) indicate that the packing in the region between the A and E helices could be functionally important. He showed that the binding of 2,3-diphosphoglycerate (DPG) to hemoglobin causes a change in the tertiary structure of the beta chain in this region. In deoxyhemoglobin with bound DPG, the A helix is shifted closer to the E helix than it is in deoxyhemoglobin without DPG. As Arnone points out, this change in tertiary structure caused by DPG may directly decrease the oxygen affinity of deoxyhemoglobin. One of the residues of the E helix (E11) must shift away from the heme before oxygenation of the beta chain can occur (*15*), and this shift may be resisted by tighter packing between the A and E helices caused by the binding of DPG. An evolutionary change at a position such as 14, which is in the path of the movement of A toward E, may have been selected because of the effect of that residue in mediating the interaction between DPG and hemoglobin. One might predict, therefore, that caroli hemoglobin to which DPG is bound may differ functionally from the other mouse hemoglobins to which DPG is bound (*18*). Caroli hemoglobin might also differ from the other mouse hemoglobins in its DPG binding constant, since the shift of A toward E is apparently necessary for proper DPG binding (*17*), and different packing arrangements between the A and E helices may make the shift easier or harder.

Functional studies will be necessary to determine whether the fixation of caroli hemoglobin and the maintenance of the Hbb^s, Hbb^d polymorphism in natural populations are due to selection or random drift. If the predicted functional difference between caroli hemoglobin and the various musculus hemoglobins is found, then it is likely that selection has influenced the structural evolution of the mouse hemoglobins.

JOHN G. GILMAN

Laboratory of Genetics,
University of Wisconsin,
Madison 53706

References and Notes

1. H. M. Ranney and S. Gluecksohn-Waelsch, *Ann. Hum. Genet.* **19**, 269 (1955); R. A. Popp, *J. Hered.* **53**, 73 (1962).
2. H. M. Ranney, G. M. Smith, S. Gluecksohn-Waelsch, *Nature* **188**, 212 (1960); J. R. Morton, *ibid.* **194**, 383 (1962).
3. J. J. Hutton, J. Bishop, R. Schweet, E. S. Russell, *Proc. Nat. Acad. Sci. U.S.A.* **48**, 1505 (1962); *ibid.*, p. 1718.
4. A. Riggs, *Science* **147**, 621 (1965).
5. J. R. Morton, *Genet. Res.* **7**, 76 (1966).
6. Studies on βdmin were done with three strains of mice, C3H/HeJ and AKR/J (from Jackson Laboratory, Bar Harbor, Maine) and Swiss mice (from ARS Sprague-Dawley, Madison, Wis.). Swiss mice and Balb/CAu mice (a gift of R. Auerbach, University of Wisconsin, Madison) were used to study βdmaj. Chains βpmin and βpmaj were isolated from hemolyzates of Au/SsJ mice, a gift of E. S. Russell, Jackson Laboratory; and hemoglobin from caroli wild mice was a gift of M. Potter, National Cancer Institute, Bethesda, Md. Subsequently, T. C. Hsu, University of Texas, Houston, sent some caroli mice.
7. C. Baglioni, *Proc. Nat. Acad. Sci. U.S.A.* **48**, 1880 (1962); J. Barnabas and C. J. Muller, *Nature* **194**, 931 (1962); W. Ostertag and E. W. Smith, *Eur. J. Biochem.* **10**, 371 (1969); Y. Ohta, K. Yamaoka, I. Sumida, T. Yanase, *Nature New Biol.* **234**, 218 (1971).
8. Alpha and beta chains were separated by the method of Clegg *et al.* [J. B. Clegg, M. A. Naughton, D. J. Weatherall, *Nature* **207**, 945 (1965)]. Purified βcaroli was impossible to obtain using this technique; consequently, all results on βcaroli were obtained with unfractionated globin. In the case of βdmin, the protein I analyzed most thoroughly, all tryptic peptides except peptide 12 were obtained by methods similar to those of Rifkin (*9*).
9. D. Rifkin, thesis, Rockefeller University (1968).
10. P. Edman and G. Begg, *Eur. J. Biochem.* **1**, 80 (1967). The instrument used was manufactured by Illitron Division, Illinois Tool Works, Chicago.
11. O. Smithies, D. Gibson, E. M. Fanning, R. M. Goodfliesh, J. G. Gilman, D. L. Ballantyne, *Biochemistry* **10**, 4912 (1971); J. G. Gilman and O. Smithies, in preparation. The data on βcaroli were obtained by cyanogen bromide treatment [E. Gross, *Methods Enzymol.* **11**, 238 (1967)] of unfractionated globin, followed by partial fractionation of peptides on Sephadex G-50 and sequencing of mixtures of peptides with the sequenator. Sequence data covering the region of peptide 12 of βdmin were obtained after digestion of beta chains treated with maleic anhydride [P. J. G. Butler, J. I. Harris, B. S. Hartley, R. Leberman, *Biochem. J.* **103**, 78p (1967)] and partial fractionation of peptides on Sephadex G-100.
12. Hemoglobin from nine mice, trapped and identified by J. T. Marshall in two regions of Thailand, was examined on starch gel, and only one hemoglobin type was found. This is consistent with but does not prove fixation.
13. M. Kimura, *Nature* **217**, 624 (1968); ——— and T. Ohta, *ibid.* **229**, 467 (1971).
14. R. K. Selander, S. Y. Yang, W. G. Hunt, *Studies in Genetics V* (Publ. 6918, Univ. of Texas Press, Austin, 1969).
15. M. F. Perutz, H. Muirhead, J. M. Cox, L. C. G. Goaman, *Nature* **219**, 131 (1968); M. F. Perutz, *ibid.* **228**, 726 (1970).
16. M. O. Dayhoff, Ed., *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Md., 1969), vol. 4, p. D219.
17. A. Arnone, *Nature* **237**, 146 (1972).
18. Experiments T. R. Lindstrom and C. Ho [*Proc. Nat. Acad. Sci. U.S.A.* **69**, 1707 (1972)] show how this hypothesis can be tested. The beta chain of deoxyhemoglobin to which DPG is bound has a lower oxygen affinity than the alpha chain. This difference between alpha and beta chains is not found in deoxyhemoglobin without DPG. Thus, my hypothesis predicts that the beta chain of caroli hemoglobin bound to DPG will have a different oxygen affinity than the beta chains of the other mouse hemoglobins bound to DPG.
19. J. G. Gilman, paper presented at the annual meeting, American Society of Human Genetics, 1970 [*Amer. J. Hum. Genet.* **22**, 33a (1970)].
20. R. A. Popp, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Silver Spring, Md., in press), vol. 5.
21. J. Bonaventura and A. Riggs, *Science* **158**, 800 (1967).
22. Uncommon symbols are Syr, serine or cysteine; ?, an unknown residue; *, not asparagine, but probably serine, as judged by Rifkin's peptide data for beta chain of AKR mice (*9*); **, not methionine, but either serine, threonine, or alanine. Standard abbreviations are Ala, alanine; Leu, leucine; Gly, glycine; Ser, serine; Glx, glutamic acid or glutamine; Val, valine; Asp, aspartic acid; Asn, asparagine; His, histidine; Met, methionine; Pro, proline; Glu, glutamic acid; Lys, lysine; Asx, aspartic acid or asparagine; Ile, isoleucine. Parentheses around a residue or residues mean that tryptic peptide composition data are all that is available, and that the position in the sequence has been inferred by homology. Cyanogen bromide fragments were used to determine the sequence of residues beyond position 40, and the positions of methionines in the beta chains discussed here were assumed to be the same as in βs, as determined by Popp (*20*).
23. I thank my major professor, O. Smithies, University of Wisconsin, for his help and guidance; R. A. Popp, Oak Ridge National Laboratory, Oak Ridge, Tennessee, for discussions and for making available sequence data; R. Bogardt, Indiana University, Bloomington, for advice on the three-dimensional structure of myoglobin; W. Fitch, University of Wisconsin, for discussions of evolutionary schemes; J. F. Crow and C. Denniston, University of Wisconsin, for discussions of population genetics; and E. S. Russell, M. Potter, T. C. Hsu, R. Auerbach for supplying mice. This is paper 1556 from the Laboratory of Genetics, University of Wisconsin, Madison. This work was supported in part by grant GB4362 from the National Science Foundation and by grants GM00398 and GM15422 from the National Institutes of Health.

19 July 1972