BLOOD LINE

Genes that describe globins—important blood proteins—are revealing new principles for development, gene organization and evolution

BY JULIE ANN MILLER

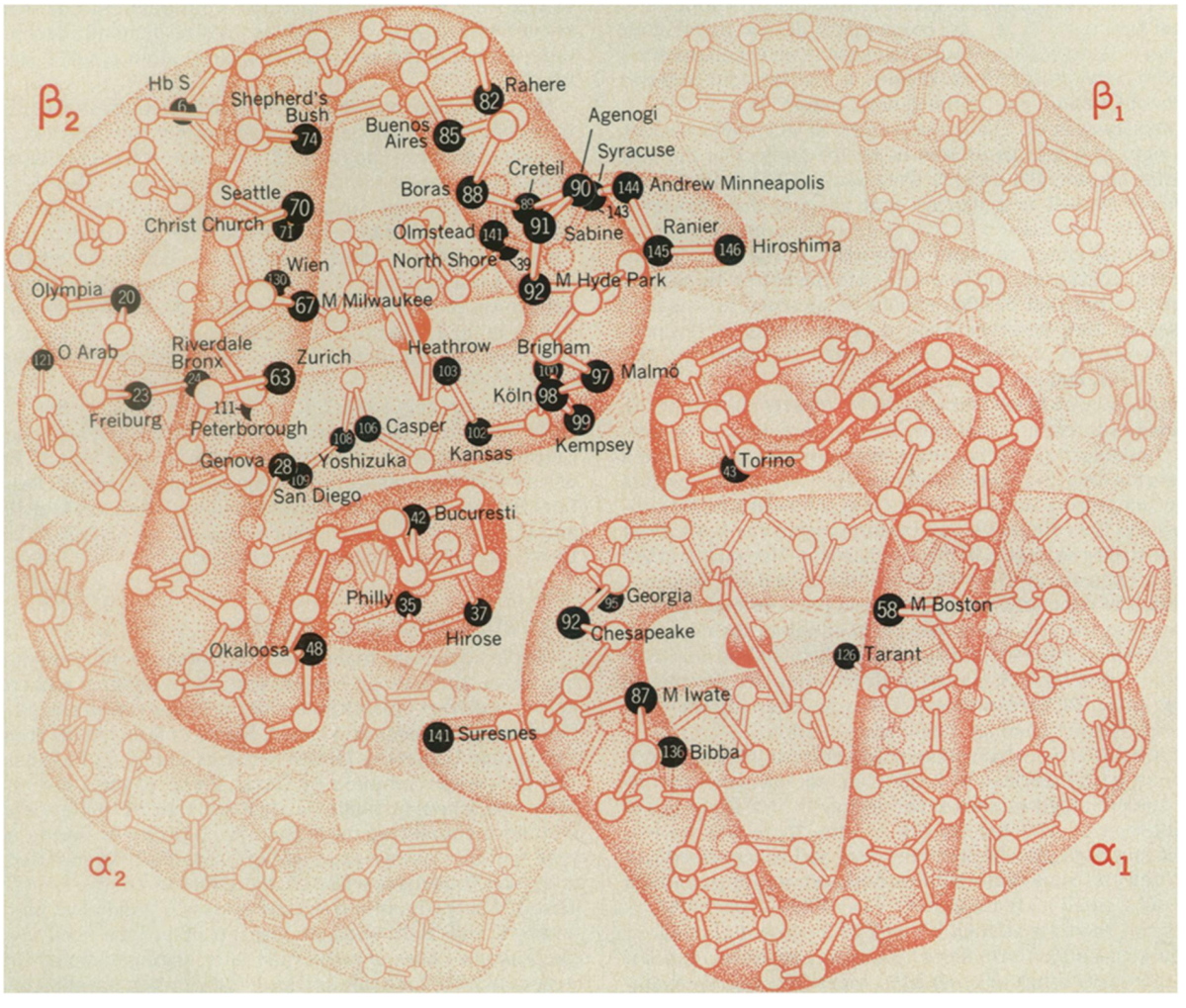
Blood is the symbolic bond between father and son, sister and brother. Blood represents what is shared by all a person's descendents. Modern science may recognize genes as the biological basis of family ties, but in common speech, it is blood, not the gene, that is thicker than water.

Blood, in the literal sense, is now taking on as much importance in the biological study of heredity as it has had in rhetoric. The powers of recombinant DNA techniques, for example, focused early on red blood cells, specifically on the genes that code for the protein portions, or globins, of hemoglobin, the oxygen-carrying pigment of red blood cells (SN: 12/13/80, p. 378). From the analysis of globin genes of humans, sheep, goats, rabbits, rats and mice, scientists are now discovering principles of gene organization that are expected to apply to many other genes and to other animals and plants.

The focus on globin genes for basic genetic studies stems from circumstances both historical and pragmatic. Because abnormalities of these genes underlie a very common group of diseases, including the thalassemias and sickle cell anemia, intense study over a period of decades by the more classical methods of human genetics has created a background of information and an array of questions at last accessible to investigation. Moreover, because different hemoglobins are produced characteristically during different stages of human life, the globin genes also offer the opportunity to examine the different mechanisms of gene control seen during development.

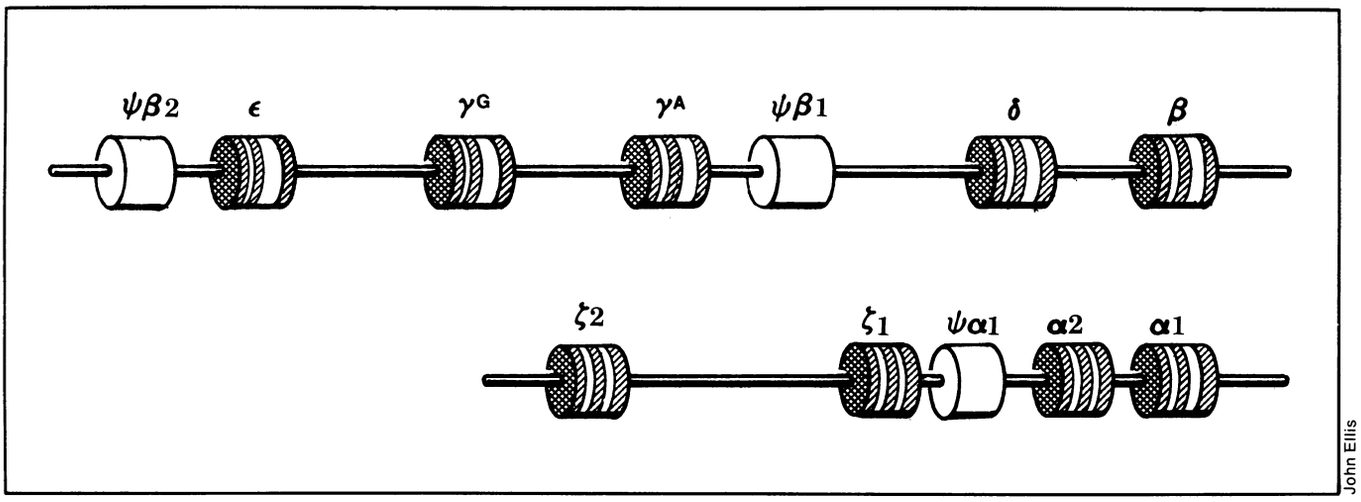
Certain characteristics of red blood cells make the globin genes especially good candidates for detailed analysis. Production of hemoglobin, for example, overwhelmingly predominates the immature red blood cell's activities. At one stage, 95 percent of the messenger RNA, the instructions for protein synthesis, conveys the globin message. That messenger RNA conveniently provides scientists with a readily available template for producing copies of globin genes needed for their investigations.

Among the principles of gene organization established by globin research is the concept of the gene family. A hemoglobin molecule contains, in addition to its oxygen-carrying heme group, two pairs of amino acid chains. One pair is called



Irving Geis

An altered amino acid in each of the positions shown causes a genetically determined human blood disorder.



The two families of globin genes form clusters on different chromosomes. Each family includes functioning genes with intervening noncoding sequences (white stripes) and pseudogenes (white), which do not code for any protein.

alpha, the other *beta*. During normal development, different genes for *alpha* chains and different genes for *beta* chains come into play in an orderly sequence. Some chains are characteristic of embryos, some of the later fetus and some of the adult animal. In all animals so far examined, the genes for the *alpha*-type chains are grouped together — in a "family" — on one chromosome and the genes for *beta*-type chains are located in a cluster on another. Recent detailed mapping shows that in each chromosome the genes of a family are lined up in developmental order — from the genes expressed in the early embryo to those expressed in the adult. Such an arrangement may be a clue to how the genes are turned on and off, or it may be a consequence of the way in which the gene families evolved.

This all seems very orderly, but unexpected skeletons in the chromosomal closet have turned up in the recent probes of the gene globin families. These unsuspected relations show a clear family likeness, but they make no protein and so had not previously been detected. Such inactive members, which the scientists call "pseudogenes," have been found in the *alpha* family of human and mouse globin and in the *beta* family of human, mouse, rabbit and goat globin genes. Although each pseudogene shares approximately 75 percent of its nucleotide subunit sequence with the respective active gene, pseudogenes cannot code for functional proteins. One reason is that "frameshift" mutations — small changes in the DNA of the pseudogenes—have shifted the grouping of the nucleotide "letters" into three-nucleotide "words" so that the protein-synthesizing machinery prematurely comes across stop signals. In addition, some of the pseudogenes lack the sequences that scientists believe are necessary for normal production and processing of messenger RNA. A role for the pseudogenes is a matter of open speculation. These "unaccountable sequences" could be genes caught in the midst of evolving from duplicates of functional globin genes to genes with new functions. It is also possible that pseudogenes play some role in development. In all the globin gene families analyzed, for instance, there is a pseudogene located between the embryonic or fetal genes and the adult genes.

The globin family regions of chromosomes have provided fertile ground for the application of recombinant DNA and DNA sequencing techniques. The exact nucleotide sequence of entire regions of the chromosome has allowed scientists to begin answering questions about how genes work and how they evolve. The sequences of the human globin gene regions have been determined primarily by Tom Maniatis at the California Institute of Technology in Pasadena, Sherman M. Weissman of Yale University in New Haven, Conn., Oliver Smithies of the University of Wisconsin in Madison, Nicholas J. Proudfoot of the Medical Research Council in Cambridge, England, and their many collaborators.

The sequencing of the globin genes contributed to the realization that many genes of animals and plants are interrupted by stretches of DNA called intervening sequences, or introns, that do not code for any protein (SN: 7/7/79, p. 12). The regions just beyond the ends of the globin genes also are being explored for clues as to how genes are controlled. Because certain nucleotide sequences crop up repeatedly among different globin genes and among different species, it is likely that they are essential to the activity of the genes.

For example, in the DNA stretch just "upstream" of each globin gene, scientists have found the same two short sequences. These "boxes" have also been identified near many other animal genes, including ones from sea urchins, fruit flies, frogs and chickens. The sequences are thought to play some role in starting the process in which a gene acts as a template for production of messenger RNA.

Rules of genetic evolution are another product of the globin studies. The experimental results emphasize the importance of such large-scale changes as the duplication of stretches of chromosome. The results also provide data for estimating the timing of the evolution of a gene.

One consequence of gene duplication, which produces identical adjoining lengths of DNA, may be that many of the globin genes occur in pairs, arranged side-by-side on the chromosome. Both members of the pair are expressed during the appropriate developmental stage of an animal's life, although one may be expressed more extensively than the other. Among the pairs are human *alpha* genes: *alpha*-1 and *alpha*-2 in the adult and *zeta*-1 and *zeta*-2 in the embryo. The *beta* family gene pairs include the human fetal *gamma*-A and *gamma*-G, whose amino acid chains differ at only one position. The human genes *delta* and *beta*, whose chains differ in 10 of 146 amino acids, also can be considered a pair. In addition, several pairs have been found among the globin genes of other animals.

Paired genes may provide the means by which genes for new functions arise. One can speculate that one gene of a pair can change over time to lose its original function with little detriment to the animal if the other gene maintains the original, essential function. It is likely, for example, that pseudogenes once were members of pairs, or threesomes, but that they accumulated debilitating nucleotide changes.

Duplications, and thus paired genes,

	Alpha Family	Beta Family	Hemoglobin Name
Embryo:	ζ_2	ϵ_2	Gower I
	ζ_2	$\gamma_2^{G,\text{ or }A}$	Portland
	α_2	ϵ_2	Gower 2
Fetus:	α_2	$\gamma_2^{G\text{ or }A}$	F
Adult:	α_2	δ_2	A_2
	α_2	β_2	A

Normal Human Hemoglobins

probably arise from errors in alignment of the chromosomes just before cell division. Homologous DNA strands frequently break and the strands "cross over" to join the opposite chromosome. Usually the crossover exchanges DNA from equivalent positions, but if the chromosomes pair out of alignment, the result is that one chromosome has extra DNA and the other is missing a stretch of DNA (SN: 7/7/79, p. 14). The longer chromosome could thus include an extra copy of a gene and its flanking sequences.

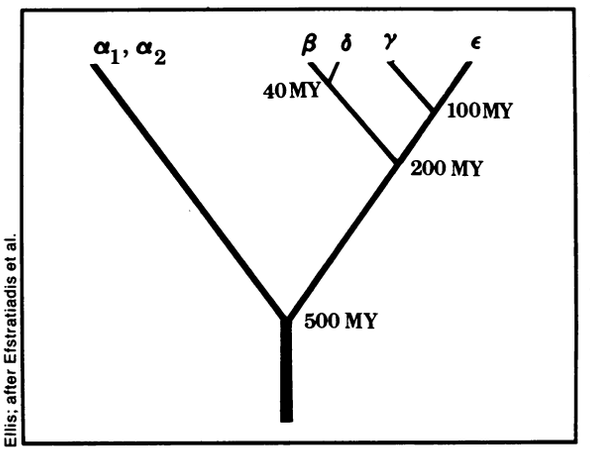
The frequent occurrence in some human populations of chromosomes with one or three adult *alpha*-globin genes is evidence that unequal crossing over is an important evolutionary event. A single *alpha*-globin gene, for example, instead of a pair, characterizes one genetic disease called *alpha*-thalassemia-2. Recently, scientists discovered human chromosomes containing three *alpha*-globin genes. Unequal crossovers would be expected to generate both singlets and triplets in the place of paired genes.

A specific mechanism seems to exist to prevent some paired genes from mutating to become different from each other. In many cases, paired genes within a species have more in common with each other than with the corresponding genes in another species. In addition, genes of a pair often are far more similar to one another than are adjacent regions that were duplicated at the same time. In the case of the human *gamma*-globin genes, Jerry L. Slightom, Ann E. Blechl and Smithies suggest that the genes were kept matched by a crossover process similar to the one that generates duplicate genes. They hypothesize that the process is used by plants and animals to enforce co-evolution of certain paired genes without requiring co-evolution of the DNA between or surrounding these genes.

Some regions of DNA change much more rapidly during evolution than scientists had expected. Before the development of the recombinant DNA methods, biologists could only examine those stretches of DNA that encode detectable products. Because those products are beneficial to the organism, drastic modification of their genes is not likely to survive natural selection. Scientists, therefore, saw little evolution of those genes.

Now, direct sequencing of DNA that encodes no product has revealed that much of the chromosome is in very rapid flux. "Big pieces of DNA are deleted and added, dramatically altering the structure of the chromosome," says Philip Leder of the National Institutes of Health. "DNA is being continually ripped up and laid down."

While this "non-essential" DNA changes too rapidly to be useful in calibrating evolutionary progress, counting the differences in the more stable nucleotide sequences has provided a new evolutionary tree. In the regions of genes that code for various proteins, single base changes are



Ellis; after Efstratiadis et al.

An evolutionary tree for human globin genes was constructed by comparing their nucleotide sequences. Branch points show how long ago (in millions of years) the genes began to differ from each other.

the most common evolutionary step. An evolutionary clock for divergent genes was calibrated using cases in which the fossil record revealed the time when species diverged. Then comparison of two genes, such as *delta* and *beta*, can tell how long they have been drifting apart. In the October CELL, evolution of the *beta*-globin gene family is considered by Argiris Efstratiadis of Harvard Medical School and other scientists. The tree they have drawn for the *beta*-globin genes is the first description of evolutionary relationships within a gene family based entirely on direct nucleotide sequence comparisons. The researchers estimate that the *beta* and *delta* genes, which appear to be present in all higher primates, diverged from each other approximately 40 million years ago. That is just the time of divergence of the New World monkeys from the Old World monkeys and great apes. The ancestral gene of *beta* and *delta*-globin diverged from the ancestor of the *gamma* and *epsilon* genes approximately 200 million years ago — during the evolution of the reptiles that gave rise to mammals.

The wide variety of hemoglobin genes in the human population demonstrates that the families are still actively evolving. More than 300 variants of human hemoglobin have been identified, including gene fusions and abhormal recombinations, insertions and deletions. Strategies are being examined for converting laboratory success in analyzing genes to therapeutic methods for the many people suffering from deficient or insufficient hemoglobin.

One strategy is to insert a gene for the missing or defective globin chain into the appropriate cells. Preliminary work along this line is already underway (SN: 4/19/80, p. 244), and two human patients have received gene transplants in a therapeutic attempt that many scientists consider premature (SN: 10/18/80, p. 245). Globin genes introduced into animal bone marrow cells have not yet been turned on to produce a high level of protein. The problem of regulating the genes' expression remains unresolved.

The array of globin genes that comprise the *alpha* and *beta* families permits another therapeutic approach. This approach takes advantage of the fact that patients with defects in genes for the adult globin chain may be able to produce perfectly good globins of the fetal type. For instance, the fetus of a person with sickle cell anemia (HbS), a defect in the *beta*-gene, develops normally before birth because the *epsilon* and *gamma* genes are intact — as is shown by the normal fetal hemoglobin that persists in small amounts in the adult. If scientists can determine the mechanism by which the fetal gene is turned down at birth and the adult gene is turned on, perhaps physicians will be able to reactivate fetal hemoglobin production in patients with the sickle cell disease.

This possibility seems especially feasible because it mimics a rare natural condition, Arthur W. Nienhuis of the NIH points out. In the genetic abnormality called hereditary persistence of fetal hemoglobin (HPFH) many cells in the adult continue to make hemoglobin of the fetal form. Occasionally, a thalassemia patient with insufficient *beta* gene globin expression also has HPFH. These patients have a uniquely mild form of thalassemia, Nienhuis says. He is currently using blood cells from sheep, which like humans have embryonic, fetal and adult hemoglobin, to develop methods for boosting fetal hemoglobin production in patients with severe thalassemia.

The portrait of the globin genes that has been rapidly sketched with recombinant DNA techniques shows dynamic gene families undergoing rapid evolutionary change. The unexpectedly large families include genes specialized to function at different developmental stages, genes that seem able no longer to function at all and genes that are currently being deleted and duplicated in human and animal populations. The scientists now need further details to extend the family sketch into a home movie showing how the member genes interact and how they respond to the animal's development. Using the methods worked out for the globin genes, scientists can turn to other inherited characteristics and determine how representative the globin family portraits are in a gallery of human genes. □