Myoglobin Gene Is a Big Surprise

The first analysis of a myoglobin gene reveals some striking similarities and some unexpected differences from hemoglobin genes

Vertebrate globins are the most intensively studied family of proteins in nature, at the physical, biochemical, and molecular levels. As a result, certain blood diseases, such as the thalassemias, can be described in close molecular detail, and the evolution of the two families of globin genes, α- and β-, can be charted with revealing accuracy. It is therefore somewhat surprising that, until now, the muscle protein myoglobin, which is a close cousin of hemoglobin, has received virtually no attention at the level of its gene structure. Now that this long neglect has been swept away, the results are something of a shock.

Although the myoglobin gene is similar to those for α- and β-globin in its overall configuration of coding and noncoding regions, the introns themselves are very much bigger. As a consequence, the myoglobin gene is some ten times longer than vertebrate hemoglobin genes, and it qualifies as one of the most "stretched out" genes so far discovered: less than 5 percent of its structure codes for message.

The great disparity between the size of the introns in the hemoglobin and myoglobin genes raises the question of which represents the ancestral form and which has been modified since they diverged some 700 million years ago. Meanwhile, the long-standing puzzle of globin biology, that of the source of leghemoglobin genes in the root nodules of soybean, remains unresolved by the myoglobin data. Although the leghemoglobin gene echoes the hemoglobin gene in general structure, it contains a third intron in the middle of the central coding sequence.

The myoglobin protein has long been a classic subject of physical and biochemical study. It was, for instance, one of the first proteins to be analyzed by x-ray diffraction. The myoglobin gene, however, remained unprobed and uncharted. By contrast, the concentrated effort of molecular biology on the hemoglobin genes has produced an enormous wealth of data that gives interesting interspecies comparisons.

Several important features emerge from these comparisons. For instance, all the genes of the α- and β-globin families are made up of three coding regions interrupted by two introns. And the size of the first intron in the two families is rather similar over a wide range of species; it is typically about 120 nucleotides in length, although there are variants above and below this figure. The second intron is of a size similar to the first in the α family but is much bigger, about 600 to 800 nucleotides, in the β family.

"The very strong conservation of intron size over wide phylogenetic distances led a lot of people to believe that it was the length, not the sequence, of these noncoding regions that was important in some way in globin gene expression," says Alec Jeffreys, of the University of Leicester, England. "We were therefore completely surprised to find these very long introns in myoglobin."

Although Jeffreys and his colleagues, Alain Blanchetot, Victoria Wilson, and David Wood, are primarily interested in human myoglobin, they found it convenient to work first with the seal protein, which is found in high concentration in the muscles of these diving mammals. They initially obtained a complementary DNA copy of part of the myoglobin message (*1*), which was then used as a probe to fish for the gene. So far they have sequenced about three-quarters of the gene, which includes all the coding regions but not all the introns. The gene measures 9200 base pairs long (*2*).

> "We were therefore completely surprised to find these very long introns in myoglobin."

"Although the coding regions have diverged in sequence from those in hemoglobin genes, they are homologous in overall structure," says Jeffreys. "The coding regions are interrupted at precisely the same positions in both genes." This discovery shows the three-exon two-intron configuration of globins to be very ancient and to have preceded the divergence between myoglobin and hemoglobin genes, dated at 700 million years ago.

In contrast with hemoglobin, which functions as a tetrameric protein, myoglobin is monomeric. In this respect it is similar to leghemoglobin of soybean nodules. Because globin-like proteins have been found in no other group of plant, there has been speculation that the gene for leghemoglobin became integrated into the soybean genome after horizontal transfer from the animal world, perhaps as a "passenger" on a virus (*3*). Some people have suggested that leghemoglobin might have derived from myoglobin, but the difference in size of introns, and the possession by leghemoglobin of the third intron, now seems to rule this out.

One possible explanation of the relation between the various globins that does not involve horizontal gene transfer is as follows. Perhaps the protoglobin gene that predated the divergence between plants and animals possessed three introns. This intron trio is still present in leghemoglobin, the plant kingdom's descendant of this gene, whereas in the animal lineage the middle intron was lost. Another possibility, which nods again in the direction of horizontal gene transfer, is that the protoglobin present before the divergence of vertebrates and invertebrates had three introns. The third intron was lost in the vertebrate line while being retained in invertebrates.

"We cannot answer these kinds of questions until we have a lot more phylogenetic data," says Jeffreys. "This means analyzing globin genes in various invertebrate species." If, for example, insect globin were to contain three introns, then the possibility of horizontal gene transfer to legumes would be strengthened.

Jeffreys, meanwhile, is beginning to work on human myoglobin. It will be interesting to discover whether human myoglobin also has very long introns, and, if so, whether their structure reveals anything about their functional importance in the gene. At the same time, David Konkel, of the University of Texas Medical Center at Galveston, is just beginning to analyze chicken myoglobin, using Jeffreys's seal myoglobin clone as a probe. The phylogenetic gaps are slowly being filled in.—**ROGER LEWIN**

References

1. D. Wood, A. Blanchetot, A. J. Jeffreys, *Nucleic Acids Research* **10**, 7133 (1982).
2. A. Blanchetot, V. Wilson, D. Wood, A. J. Jeffreys, *Nature (London)* **301**, 732 (1983).
3. R. Lewin, *Science* **214**, 426 (1981).