

the evolution of hemoglobin.

by Ross Hardison

A comparative study of hemoglobin was conducted to explain how an ancestral single-function molecule gave rise to descending molecules with varied functions. Hemoglobin is the molecule in red blood cells responsible for giving blood its color and for carrying oxygen throughout the body. New functions of metallo-porphyrin rings or a kind of molecular cage embedded in proteins were developed with the appearance of atmospheric oxygen. The presence of hemoglobin in both oxygen-needing and non-oxygen needing organisms suggests the same evolutionary roots.

© COPYRIGHT 1999 Sigma Xi, The Scientific Research Society

Studies of a very ancient protein suggest that changes in gene regulation are an important part of the evolutionary story

The appearance of atmospheric oxygen on earth between one and two billion years ago was a dramatic and, for the primitive single-celled creatures then living on earth, a potentially traumatic event. On the one hand, oxygen was toxic. On the other hand, oxygen presented opportunities to improve the process of metabolism, increasing the efficiency of life's energy-generating systems. Keeping oxygen under control while using it in energy production has been one of the great compromises struck in the evolution of life on earth.

The compromise was a chemical one. It appears that the apparatus that sequesters oxygen in cells, possibly to protect them, is almost identical to the one that, in different contexts, exploits oxygen for its energy-generating potential. At first this apparatus was quite primitive, probably limited to a caged metal atom capable of binding oxygen or tearing away its electrons, which are used in metabolism. But this basic chemical apparatus grew increasingly complex through time and evolution. At some point the metal atom was fixed inside a kind of flat molecular cage called a porphyrin ring, and later that porphyrin ring became embedded in larger organic compounds called proteins. These organic compounds themselves became increasingly varied through time and evolution.

The descendants of those compounds include the chlorophylls and heme. Each class of compound still contains a porphyrin ring at its center. So the basic interaction between metal atom and oxygen has not changed. What has changed are the circumstances and biochemical pathways in which these various molecules interact. Some molecules are required to distribute oxygen to the various organs and tissues of an organism, and some store oxygen in a particular tissue. Some molecules participate in making oxygen, through the process of photosynthesis; others use it up in respiration.

If similar compounds are used in all of these reactions, then how do the different functions arise? The answer lies in the specific structure of the organic component, most often a protein, that houses the porphyrin ring. The configuration of each protein determines what biochemical service the protein will perform.

Because they have such an ancient lineage, the porphyrin-containing molecules provide scientists with a rare opportunity to follow the creation of new biological compounds from existing ones. That is, how does an ancestral molecule with a single function give rise to descendant molecules with varied functions? In my laboratory, I try to answer this question through the comparative study of hemoglobin, the molecule in red blood cells that gives blood its color and that carries oxygen throughout the body.

Such studies are carried out by comparing the genes that code for the hemoglobins and their chemical relatives in a range of organisms from bacteria to people to see how the genes have changed through time. Typically research in molecular evolution has focused on portions of the gene responsible for alterations in protein structure. So it came as a great surprise that the changes in hemoglobin have not been merely structural. In fact, the three-dimensional structure of hemoglobin - its shape, with folds, pockets and surfaces - has been fairly well conserved over the protein's evolutionary history. Rather, some of the most rapid and dramatic changes in hemoglobin proteins have been in the ways these molecules are regulated - the when and how of their manufacture inside the cell. The hemoglobins and their relatives continue to evolve rapidly in subtle ways, and the changes continue to come in the genetic regulation of the proteins.

This suggests that the creation of new protein functions arises as much from changes in regulation as from changes in structure. A similar observation has recently been reported for a protein important in determining the body plans of vertebrates and invertebrates. A deeper understanding of the relationships among proteins - indeed an understanding of the evolutionary relationships among organisms - will require an analysis of changes in the entire gene, including the regulatory portions. Informed by this new understanding, the study of the hemoglobins and

the evolution of hemoglobin.

their molecular relatives affords scientists a rare glimpse into the very ancient past of protein evolution, and even possibly a look toward the future.

Portable Porphyrins

Four billion years ago, when the earth was new but before life appeared on it, there was no free oxygen in the air. Instead, earth's atmosphere contained mostly water vapor, nitrogen, methane and ammonia. The first organisms to develop, probably about 3.8 billion years ago, used these materials for food and energy. Although no one knows for sure, it seems plausible that these early metabolic reactions were facilitated, or catalyzed, by metals such as iron and magnesium.

Shortly afterwards (as geologic time goes), somewhere between 3.3 and 3.5 billion years ago, there appeared single-celled organisms called cyanobacteria (formerly called blue-green algae), which had the ability to convert energy from the sun into chemical energy through photosynthesis. Photosynthesis in cyanobacteria removes electrons from hydrogen sulfide ($[H.sub.2]S$), which was present on the young earth, to yield elemental sulfur and then shunts the electrons into a chemical relay that ultimately produces the so-called "energy molecule" adenosine triphosphate (ATP).

Then, sometime between one and two billion years ago, an amazing thing happened. Photosynthetic bacteria learned a new trick. Instead of carrying out photosynthesis with $[H.sub.2]S$, they used water, $[H.sub.2]O$. And instead of producing sulfur, this process produced molecular oxygen, $[O.sub.2]$. This remarkable event transformed the earth and all of the life on it.

The oxygen so produced was released into the atmosphere. Many of the existing organisms continued as they had before, carrying out their cellular processes as though oxygen had never existed. Indeed, many single-celled organisms remain anaerobic to this day; they do not use oxygen, and some are poisoned by it. But many new organisms appeared that acquired the chemical wherewithal to manage and make use of the oxygen. Within the chemical arsenal available to organisms engaging in oxygen-based metabolism might well have been the porphyrin ring.

The versatility of the porphyrin ring, its molecular simplicity and its ubiquity in the biosphere suggest that it was developed a very long time ago. At that time life's chemistry was less elaborate than it is now, and one molecule may have performed many functions. A porphyrin molecule is a planar group of four connected rings, each of which contains a nitrogen atom that faces

the center of the ring cluster. These four nitrogens provide an ideal environment for the insertion of a metal ion, such as iron or magnesium, which are extremely useful for a variety of oxygen-related reactions.

A particular porphyrin ring containing magnesium is the organic molecule called chlorophyll, the substance in green plants that helps harvest the electromagnetic energy of sunlight for use in photosynthesis.

A slightly different porphyrin ring containing iron is called heme. When heme is bound to globin molecules, the resulting protein is hemoglobin. But heme can bind to various other proteins to produce cytochromes, which transfer electrons in the chain of biochemical reactions of respiration; oxygenases, which catalyze the oxidation of a wide variety of compounds; and fungal ligninases, which degrade the compound lignin in decaying wood. Clearly, the specific function played by the heme depends on the properties of the protein in which it is embedded.

It is interesting to consider what the original function of the porphyrin ring may have been. One possibility is that the metallo-porphyrin ring, or a chemical relative, along with associated proteins may have participated in some kind of early sulfur-based photosynthesis in ancestral cyanobacteria. Primitive heme-complex proteins, or hemoproteins, the forerunners of contemporary cytochromes, may have served as electron-transfer agents even before the appearance of molecular oxygen.

With the advent of molecular oxygen, new roles would have been opened up for metalloproteins in general, and hemoproteins in particular. Given the capacity of oxygen to damage various cellular components, oxygen-binding hemoproteins may have functioned initially to sop up the oxygen, thus protecting cells from its toxic effects. But later, when oxygen became a useful metabolic agent for some cells, the same hemoproteins could have been recruited to participate in oxygen-based metabolism, specifically to transfer electrons eventually to oxygen during respiration. This is the function of contemporary cytochromes.

In time, the hemoproteins could have been further modified to allow them to participate in other electron-transfer reactions or to take on entirely new functions. One can imagine the need for a hemoprotein that could scavenge oxygen - still in scarce supply two billion years ago - for oxygen-based respiration. With still more time and further modifications, these hemoproteins evolved into the now familiar and nearly ubiquitous hemoglobins.

Modern Hemoglobins

the evolution of hemoglobin.

In animals, hemoglobin is the extremely abundant protein that gives red blood cells their color. It binds oxygen in the lungs and delivers it to respiring tissues in the body. Hemoglobins were first found in blood simply because they are so abundant, reaching an average concentration of 15 grams in every 100 milliliters of healthy human blood. Human - indeed all vertebrate - hemoglobins involved in oxygen transport are actually made up of four polypeptides, which are long chains of amino acids. Two of these polypeptides are of the alpha variety and are called [Alpha]-globin; the other two chains are of the beta variety and are called [Beta]-globin. Inside the red blood cell, each of these polypeptide chains forms a ball with one heme group lodged inside it. The four polypeptide chains act together as one functional hemoglobin protein molecule.

Some members of the modern hemoglobin family in vertebrates do not transport oxygen, but rather store oxygen in various tissues. Myoglobin, for example, gives red muscle its color and more important, stores oxygen there. Composed of only a single polypeptide chain, myoglobin strongly resembles both [Alpha]- and [Beta]-globins in its overall shape and design as well as in its sequence of amino acids.

Scientists have so closely associated hemoglobins with oxygen transport that they have been surprised to discover hemoglobin in organisms that have no obvious need for oxygen transport. In 1982 Kjeld Marcker and colleagues at the University of Aarhus, Denmark, and Desh Pal Verma and colleagues at McGill University first isolated genes encoding plant-associated hemoglobins found in the root nodules of legumes; these are called leghemoglobins. These nodules house nitrogen-fixing bacteria that extract nitrogen from the environment and convert it into a form the plant can use. Plants make their own oxygen, so it was difficult at first to conceive a function for plant hemoglobins. It is now believed that the leghemoglobins help supply oxygen to the bacteria, while keeping it away from the nitrogen-fixing machinery to which it is toxic. Since then, work in laboratories such as that of W. James Peacock at Commonwealth Scientific and Industrial Research Organization in Australia has identified hemoglobins in all plants studied, even those that form no symbiotic relationships with bacteria.

Doubly surprising has been the discovery of hemoglobins in organisms that are only one cell large, such as bacteria. It is commonly thought that single-celled organisms are small enough to allow oxygen to diffuse into and through them without the help of any auxiliary molecules. Biologists therefore found themselves hard pressed to explain the 1986 discovery by S. Wakabayashi and colleagues at Osaka University in Japan and D. Webster at the Illinois Institute of Technology of hemoglobin in the bacterium

Vitreoscilla. Later discoveries by other investigators uncovered hemoglobins in other bacteria as well as in single-celled fungi and protozoa. Clearly, hemoglobins in such species cannot be involved in the traditional role of transport of oxygen between cells, so they must have some other function. In fact, bacterial hemoglobins may be more like the ancient hemoproteins and perform the older function of electron transport. Findings as surprising as these raise the question of whether and how all of these hemoglobins are related.

An initial hint that these proteins are relatives comes from comparisons of the sequence of amino acids in the polypeptide, and this deduction has been strongly supported by a study of their shapes and structures. The defining characteristic of vertebrate hemoglobins is the globin fold, first identified by the seminal studies of Max Perutz, John Kendrew and their colleagues at Cambridge University when they solved the three-dimensional structures of hemoglobin and myoglobin, respectively. B. K. Vainshtein and colleagues at the Russian Academy of Sciences in Moscow showed that leghemoglobin has this same globin fold. Recently the structure for the hemoglobins found in the bacteria *Vitreoscilla* and *Alcaligenes* have been solved, respectively, by C. Tarricone and colleagues at the University of Pavia in Italy, and Ulrich Ermler and colleagues at the Max Planck Institut für Biophysik in Germany. These molecules also contain the characteristic globin fold.

Family Ties

Since it seems likely that some sort of metallo-porphyrin compounds preceded oxygen-based metabolism, it also stands to reason that some of the genes encoding the proteins to which they bind could have equally deep evolutionary roots, an idea that gains further support when one also considers how widespread the hemoglobins appear to be. Only something truly ancient would be found in life forms such as bacteria, which hark back to the earliest forms of life, as well as in more recently evolved life forms, including people.

Most scientists agree that it is quite likely that highly similar proteins from distantly related species shared a common ancestor. In the distant past, some ancestral - probably single-celled - organism had one hemoglobin gene, and therefore one kind of hemoglobin protein. But at some point, this gene was duplicated, so that each of the resulting daughter cells carried two identical copies of the ancestral hemoglobin gene. Gradually, during successive cell divisions, small variations in the sequence of nucleotides - the subunits that make up a gene - started to appear. In this way, the two genes that started out identical acquired sequence differences and later, functional

the evolution of hemoglobin.

differences. It is quite likely that additional hemoglobin genes were acquired the same way, by gene duplication followed by modifications in the nucleotide sequence.

Biologists assess relatedness between proteins from different species by comparing the nucleotide sequences of the genes encoding them, as well as the sequences of amino acids making up the proteins themselves. Since genetic alterations accrue with time, greater similarity between the gene sequences is taken to indicate a closer evolutionary relationship between the genes, whereas differences in gene sequences imply greater evolutionary distances.

Using this kind of analysis, scientists can spot the appearance of a certain sequence and determine its approximate age. They can also get a sense of the way new proteins arise from old by tracing modifications, additions and deletions to early sequences as time and evolution progress.

Applying this reasoning to the amino-acid sequences of the [Alpha]- and [Beta]-globins, and myoglobin, we find that all three proteins are very similar to each other. Evolutionary reconstructions indicate that they derive from a protein that originally appeared in an ancient vertebrate about 500 million years ago. The amino acid sequences of vertebrate globins are even similar, albeit less so, to those of hemoglobins from single-celled organisms, such as yeast, protists and bacteria. This suggests that the hemoglobins are much older still.

Comparing the amino acid sequences of hemoglobins from animals, plants, protists and the class of bacteria known as eubacteria suggests that they all share a common ancestor very early in organismal evolution, in spite of the fact that the proteins carry out different functions. This confirms that the gene for hemoglobin is truly ancient and predates the time that eukaryotic cells (the nucleus-containing cells of plants and animals) diverged from eubacterial cells - between three and four billion years ago. To date, no hemoglobins have been found in the class of bacteria known as archaeobacteria. Since archaeobacteria appear to have separated from the ancestral eukaryotes more recently than did the eubacteria, one may anticipate that hemoglobins will be found in archaeobacteria at some point. (There exists, however, the possibility that eubacteria and their hemoglobins were engulfed by, and incorporated into, the ancestral eukaryotic cells. In this case, one would expect hemoglobins to be absent from the archaeobacteria.)

The finding of a common ancestor for all of the hemoglobins has several ramifications. The first is that amino acid sequences of hemoglobins are sufficiently

informative to examine issues of ancestry as far back as the evolutionary split between the cells of eubacteria and those of plants and animals. The second is that, given a common ancestor for proteins that now differ in function, biologists can begin to examine how changes in the regulation - and not just the shape - of hemoglobin genes gave rise to these different functions.

Regulation and Function

The structure of a protein is but one of the factors determining its function. In closely related proteins, such as the family of hemoglobins, the conditions under which a protein is manufactured also determine its function. For example, two different hemoglobins may bind oxygen. But hemoglobin designed to scavenge oxygen would be expected to be manufactured, or expressed, at a particularly high rate during times when oxygen is scarce. In contrast, a protein designed to sop up excess oxygen would be expected to be expressed during times of oxygen abundance.

The broad distribution of hemoglobins throughout the phylogenetic tree provides scientists an opportunity to examine how the regulation of hemoglobin genes has changed, presumably to accommodate the different protein functions. This study, conducted in my laboratory among others, reveals that regulatory mechanisms have changed in striking ways over the course of evolution, even in cases where the changes bring no obvious benefit. In many cases, the regulatory changes and evolutionary distance are so large that no remnant of the ancestral state is left to guide inferences from sequence comparisons. Thus biologists studying this question are confined to comparisons between DNA sequences from relatively close phyla, such as comparisons among the hemoglobin genes from various bacteria or among [Beta]-globin genes of mammals.

Response to Oxygen

It seems logical, in light of the function of hemoglobins, to begin a study of regulatory mechanisms governing the expression of hemoglobin genes with an inquiry into the way those genes respond to oxygen. In most cases, organisms respond to low concentrations of oxygen - referred to as hypoxic or anaerobic conditions - by increasing the amount of hemoglobin they produce, which of course increases the amount of oxygen delivered to cells for oxygen-dependent respiration and energy production. However, the way hemoglobin production is increased can vary between organisms. An organism can directly increase the number of hemoglobin molecules it produces per cell, the mechanism observed in many bacteria, or an organism can increase its output of

the evolution of hemoglobin.

hemoglobin-producing cells, the mechanism observed in mammals and other vertebrates.

Scientists are interested in learning whether a single genetic mechanism governs this response in distantly related organisms, which would suggest that regulation is as highly conserved through evolution as is the coding portion of the genes themselves. The information available for bacteria right now is too limited to reach any firm conclusion, but it appears that there is not a high degree of conservation, even among different bacterial species. In some cases, the proteins that govern gene expression are similar in different species but are used in different ways. This, we have discovered, is a general theme of hemoglobin regulation - common proteins are used in regulation, but in different ways and in different contexts.

For example, in various bacteria, the fumarate nitrate reduction (FNR) protein activates the production of a number of proteins in response to low oxygen concentrations. Under anaerobic conditions, FNR assumes a shape that enables it to interact with and activate the hemoglobin gene in the bacterium *Vitreoscilla*, as shown by K. L. Dikshit and colleagues at the institute of Microbial Technology in India.

FNR is also required to activate a hemoglobin gene in the bacterium *Bacillus subtilis*. But in this case, activation is indirect. Michiko Nakano and colleagues at Louisiana State University showed that FNR stimulates the production of nitrite, which in turn activates the production of regulatory proteins that interact with and activate the expression of the hemoglobin gene. The same protein - FNR - is implicated in the stimulation of related hemoglobin genes in different families of bacteria.

In vertebrates, low oxygen is also a stimulus for hemoglobin synthesis, but as noted before, the increase is achieved by increasing the number of cells that produce hemoglobin, rather than increasing the amount of hemoglobin produced by each cell. The vertebrate cell that produces hemoglobin is the red blood cell, also known as the erythrocyte. Each erythrocyte is packed with hemoglobin - about 280 million molecules per cell - so increasing the number of erythrocytes effectively elevates the amount of hemoglobin available to the individual. Hypoxia is generally sensed in the liver and kidneys. The kidneys respond by manufacturing erythropoietin, a hormone that penetrates the bone marrow and stimulates red blood cells to proliferate.

Gregg Semenza and colleagues at Johns Hopkins University showed that the erythropoietin gene itself is stimulated by hypoxia induction factor 1 (HIF1), which activates gene expression when it binds to one of the

gene's regulatory regions. HIF1 belongs to a large protein family, each member of which contains a segment called the PAS domain. Proteins with a PAS domain are found in eubacteria, archaeobacteria, fungi, plants and animals and are involved in sensing a variety of stimuli, including light and oxygen levels. Although none of the PAS-containing proteins in bacteria has as yet been implicated in controlling the expression of hemoglobin genes, several appear to be oxygen sensors. It will be of considerable interest to see whether PAS-containing proteins other than HIF1 regulate the production of hemoglobin, either directly or indirectly, in nonmammalian species.

There is at least one example of a protein that does control hemoglobin in species widely separated by evolution. A complex of proteins, called HAP1, 2, 3 and 4, regulates hemoglobin production in the yeast *Saccharomyces cerevisiae* when oxygen levels are high. The HAP proteins show a high degree of sequence similarity to two subunits of CP1, a mammalian protein implicated in the expression of all of the mammalian globin genes. Thus, despite their different responses to oxygen (mammalian hemoglobin expression being increased in conditions of low oxygen, whereas yeast hemoglobin expression is stimulated by high oxygen levels) and the greater complexity of the mammalian mechanism, the proteins implicated in the regulation of homologous genes in yeast and mammals seem to be related.

Paradoxical Vertebrate Globin Evolution

One of the most powerful examples of how much regulatory regions of related genes can differ concerns the regions regulating the expression of the two polypeptides that make up vertebrate hemoglobin proteins - the [Alpha]- and [Beta]-globin polypeptide chains. Since [Alpha]- and [Beta]-globins are different peptides, they are encoded by distinct genes: An [Alpha]-globin gene codes for the [Alpha]-globin peptide, and a [Beta]-globin gene codes for the [Beta]-globin peptide. Both polypeptides, as well as heme, are manufactured when the cell needs hemoglobin, and the entire hemoglobin protein is assembled after each of the polypeptides is synthesized.

Each gene is independently regulated and controlled, but evolutionary analysis indicates that these genes arose through the duplication of a common ancestral gene about 500 million years ago. Since the two genes would have been identical after the initial duplication one would expect the portions of the gene that code for the protein to have remained similar in the course of evolution, and they have. One would also expect that the regulatory portions of the gene likewise have remained similar - and there lies a paradox. They are quite different.

the evolution of hemoglobin.

An examination of contemporary [Alpha] and [Beta]-globin genes indicates many differences with regard to their mechanism of regulation. In birds and mammals, the two genes do not even lie on the same chromosome, as they most certainly must have long ago. In mammals, this takes on additional significance in regulating these genes. It appears that the chromosomal structures differ in the regions containing [Alpha]- and [Beta]-globin genes, and this has implications for the way the genes are regulated.

Chromosomes are linear strands of DNA that may each include up to thousands of genes, arrayed end to end. Inside nonbacterial cells, such as those of fungi, protozoans, plants and animals, chromosomes are twisted up in a complex of DNA and protein that is called chromatin. Some regions of chromatin are coiled so tightly that the genes lying there are hidden from the proteins that will activate their expression. When a particular gene needs to be expressed, the chromosomal region in which the gene lies will frequently become uncoiled, exposing the gene to the expression machinery in the proper cells and at the correct stage of development. Harold Weintraub and Mark Groudine at the Fred Hutchinson Cancer Research Center in Seattle showed this to be the case for the [Beta]-globin gene, which lies on one of these supercoiled regions of its chromosome. That chromosomal region becomes uncoiled only in red blood cells that are making hemoglobin. In contrast, as demonstrated by Doug Higgs and coworkers at Oxford University, the [Alpha]-globin gene lies in a region of its chromosome that is almost always uncoiled in all cell types, not just in red blood cells and not just when they are making hemoglobin.

Promoters and Enhancers

Despite their common ancestry, the chromosomal effects on the regulation of [Alpha]- and [Beta]-globin genes have diverged considerably. To get a better handle on how changeable are the control regions of genes, it would be instructive to compare the regulatory regions of globin genes among different types of globins and in different species.

Globin gene expression is regulated both by enhancers and by the gene's "on" switch, the promoter. The promoter sits at the start of the gene and is absolutely required to initiate gene expression. The enhancer, although not essential to gene expression, greatly augments the level of expression. What is distinctive about enhancers is that, unlike the promoter, the enhancer is not found at the start of the gene, but it can lie anywhere within it. In some cases, the enhancer doesn't even have to be within the confines of the gene, but can still increase the level of globin gene expression. In the globin genes of birds and mammals, the major enhancers can be placed very far

away from the promoter. Both enhancers and promoters are DNA segments with clusters of binding sites for proteins that regulate expression.

The placement of enhancers and promoters is similar in the genes encoding the [Alpha]- and [Beta]-globins in both birds and mammals. Also, some of the proteins that bind to these promoters and enhancers are similar in these groups of genes. It has come as a surprise, then, to learn that DNA sequences of these regulatory regions are not highly conserved among these groups of genes. No statistically significant sequence matches are seen in the promoter regions or in the distant control elements, such as the enhancer, in comparisons between mammalian [Alpha] and [Beta]-globin gene clusters or between [Beta]-globin gene clusters of chickens and humans. This suggests that either the control regions evolved independently in these vertebrate lineages, or that the same control regions now function differently, or that similar function is maintained in the absence of observable similarities in DNA sequence.

Although much still remains to be discovered, it is clear that some regulatory proteins are used only in the control of [Beta]-globin genes, not [Alpha]-globin genes. In the cases of proteins that regulate both [Alpha]- and [Beta]-globin genes in both birds and mammals, the number, position and context of their binding sites in the promoters and enhancers differs. This would account for the inability to detect the common binding sites by DNA sequence comparisons, and points to informative differences in the mechanisms by which these proteins regulate globin genes.

Regulating [Beta]-Globin Genes

If the evolutionary distance between avian and mammalian genes has perhaps been too great to see much in the way of regulatory similarities, then maybe we need to narrow the evolutionary distance between proteins further. Collaborative studies involving my own laboratory have made comparisons between [Beta]-globin and related genes of various mammals, including humans, galagos (the prosimian primates known as African bush babies), rabbits, goats, cows and mice.

We found a high sequence similarity in many of the control regions of these genes. In fact comparisons between the [Beta]-globin genes from mammals as different as mice and men showed them to be similar over long stretches of noncoding regions, including many regulatory regions, such as the enhancer. These are useful pointers to functional segments. For example, within the enhancer we discovered a segment that we call the E-box, whose sequence is identical in the mammals we studied. Studies

the evolution of hemoglobin.

by Laura Elnitski in my laboratory indicate that the E-box is crucial for maximal enhancement of [Beta]-globin gene expression.

It is interesting to note that, of the multiple globin genes in each mammal, the regulatory sequences of only one matches strongly with the homologous globin gene in other mammals. For instance, the promoter of the [Beta]-globin gene expressed in adult life matches between mouse and human, as does the [Epsilon]-globin gene expressed in embryos. In contrast, comparisons between the promoters of the genes coding for adult versus fetal or embryonic forms of hemoglobin reveal very few common regulatory elements. When structurally similar proteins share no common regulatory elements, we must conclude that the regulatory elements are changing more rapidly than are the coding sequences. Hence it should come as no surprise that comparisons between genes that diverged even earlier than these in the mammalian versus avian [Beta]-globin genes, for example, or mammalian [Alpha]- versus [Beta]-globin genes - do not reveal matches in regulatory elements.

Evolutionary Logic

We have traced the evolution of hemoglobin and seen how the active core, the porphyrin ring, which is responsible for the basic chemistry of the molecule in which it sits, was eventually embedded in larger organic structures. The structure of these organic molecules imparts to the molecule its specific function, determining, for example, whether that basic chemistry was used in the service of respiratory reaction chains, oxygen transport or oxygen sequestration. Obviously structural changes in these organic molecules translate into functional changes.

We have also seen that structurally similar molecules can become further differentiated functionally by being expressed at different times in the development of an organism, as is the case for fetal and adult globins, or under different circumstances, such as the scarcity or abundance of oxygen. These distinctions are not attributable to the overall structures of the proteins themselves, since, as I already noted, they are very similar. Rather, differences in expression profiles are achieved through differences in the regulatory segments of genes. And these, as we have seen, can be vastly different even in closely related proteins from closely related organisms, which strongly suggests that regulatory regions are changing more rapidly than are the structures of proteins themselves.

This notion is relatively new in the study of molecular evolution, which has concentrated on comparisons between protein structures to determine evolutionary

relationships between species. But molecular evolutionists are starting to recognize the value of looking at relations between regulatory regions. Recently, David Stern at the Wellcome Institute in Cambridge, England, reported that changes in the pattern of expression of a member of a family of genes called Hox genes, which act to specify the body plan of vertebrates and invertebrates, account for differences in the appearance of legs among three species of flies. The amino acid sequences of the proteins are identical, but the regulation of the genes encoding them has evolved to change the body structure. Furthermore, Stern proposed that alterations in the body plans of various creatures through evolution result from changes in the regulation, rather than the structure, of the Hox gene in question.

It is important, however, to note that in our experience, changes in the regulatory and structural portions of the genes are not entirely independent. Some interplay between these two processes is expected as the encoded proteins acquire new functions. For example, the myoglobin gene, which codes for an oxygen-storing protein, should be and is - regulated differently from the [Alpha]- and [Beta]-globin genes, which of course code for the proteins that transport oxygen to different tissues.

On the other hand, if changes in regulation took place only in concert with changes in function, then genes encoding hemoglobins with the same function should be under the same type of regulation. This prediction is not observed. The survey in this article shows that even within a family of genes encoding proteins that have to function together, such as the [Alpha] and [Beta]-globins, many features of their regulation differ in unexpected ways. Furthermore, substantial changes have taken place in the [Beta]-globin family since the time that mammals and birds diverged. Some regulatory proteins are used in both families of genes in birds and mammals. Other proteins may function in the same general pathway - for example, the FNR protein, which induces the expression of hemoglobin in a variety of bacteria when they encounter anaerobic conditions - but the specific use of the protein differs in different bacteria.

In general, changes in regulatory mechanisms are considerably more rapid than are needed to accommodate the differences in functions of the hemoglobins. Perhaps there is considerable plasticity in the regulatory mechanisms, allowing multiple routes to a common end: increasing hemoglobin production. Alternatively, the truly important common features of these regulatory components may not be obvious at this level of analysis with current information.

We see in our analysis of hemoglobin genes that

the evolution of hemoglobin.

sequence comparisons between the genes of a single family of animals - mammals, for example - are more instructive than are comparisons between more distantly related species. Although the meaning of such comparisons can still be debated, we believe that comparisons of regulatory regions will continue to enrich the discussions of the evolving functions of proteins.

Acknowledgments

I am indebted to my collaborator, Dr. Webb Miller, and colleagues in my laboratory for the research that provides the foundation for this essay on broader aspects of evolution of hemoglobin genes and their regulation. I thank Dr. Max Perutz for graciously providing historical pictures of hemoglobin structures, Dr. Greg Farber for generating ribbon diagrams of protein structure and Michelle Hoffman for insightful and thorough editing. I thank the National Institutes of Health and the National Library of Medicine for continuing support of our research.

Bibliography

- Andersson, C. R., E. O. Jensen, D. J. Llewellyn, E. S. Dennis and W. J. Peacock. 1996. A new hemoglobin gene from soybean: A role for hemoglobin in all plants. *Proceedings of the National Academy of Sciences* 93:5682-5687.
- Bunn, H. F., and R. O. Poynton. 1996. Oxygen sensing and molecular adaptation to hypoxia. *Physiological Reviews* 76:839-885.
- Chapman, D., and J. W. Schopf. 1983. Biological and biochemical effects of the development of an aerobic environment. In ed., J. William Schopf, *Earth's Earliest Biosphere*. Princeton: Princeton University Press, pp. 302-320.
- Craddock, C. F., P. Vyas, J. A. Sharpe, H. Ayyub, W. G. Wood and D. R. Higgs. 1995. Contrasting effects of alpha and beta globin regulatory elements on chromatin structure may be related to their different chromosomal environments. *EMBO Journal* 14:1718-1726.
- Dickerson, R. E., and I. Geis. 1983. *Hemoglobin: Structure, Function, Evolution and Pathology*. Menlo Park, Cal.: The Benjamin/Cummings Publishing Co., Inc.
- Felsenfeld, G. 1993. Chromatin structure and the expression of globin-encoding genes. *Gene* 135:119-124.
- Fridovich, I. 1998. Oxygen toxicity: a radical explanation. *Journal of Experimental Biology* 201:1203-1209.
- Goldberg, D. E. 1995. The enigmatic oxygen-avid hemoglobin of *Ascaris*. *Bioessays* 17:177-182.
- Goodman, M., J. Czelusniak, B. Koop, D. Tagle and J. Slightom. 1987. *Globins: A case study in molecular phylogeny*. Cold Spring Harbor Symposium on Quantitative Biology 52:875-890.
- Grosveld, F., M. Antoniou, M. Berry, E. de Boer, N. Dillon, J. Ellis, P. Fraser, O. Hanscombe, J. Hurst, A. Imam, M. Lindenbaum, S. Philipsen, S. Pruzina, J. Strouboulis, S. Raguz-Bolognesi and D. Talbot. 1993. The regulation of human globin gene switching. *Philosophical Transactions of the Royal Society (London)* 339:183-191.
- Hardison, R. 1998. Hemoglobins from bacteria to man: Evolution of different patterns of gene expression. *Journal of Experimental Biology* 201:1099-1117.
- Hardison, R., J. L. Slightom, D. L. Gumucio, M. Goodman, N. Stojanovic and W. Miller. 1997. Locus control regions of mammalian [Beta]-globin gene clusters: Combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205:73-94.
- Keilin, D. 1966. *The History of Cell Respiration and Cytochrome*. Cambridge, U.K.: Cambridge University Press.
- Martin, D., S. Fiering and M. Groudine. 1996. Regulation of [Beta]-globin gene expression: straightening out the locus. *Current Opinion in Genetics and Development* 6:488-495.
- Orkin, S. 1995. Regulation of globin gene expression in erythroid cells. *European Journal of Biochemistry* 231:271-281.
- Pastore, A., and A. M. Lesk. 1990. Comparison of the structures of globins and phycocyanins: evidence for evolutionary relationship. *Proteins* 8:133-155.
- Riggs, A. F. 1991. Aspects of the origin and evolution of non-vertebrate hemoglobins. *American Zoologist* 31:535-545.
- Rouault, T. A., and R. D. Klausner. 1996. Iron-sulfur clusters as biosensors of oxidants and iron. *Trends in Biochemical Sciences* 21:174-177.
- Stamatoyannopoulos, G., and A. W. Nienhuis. 1994. Hemoglobin switching. In eds., G. Stamatoyannopoulos, A. W. Nienhuis, E. W. Majerus and H. Varmus, *The Molecular Basis of Blood Diseases*. Philadelphia: W.B. Saunders

the evolution of hemoglobin.

Co., pp. 107-155.

Zhu, H., and A. F. Riggs. 1992. Yeast flavohemoglobin is an ancient protein related to globins and a reductase family. *Proceedings of the National Academy of Sciences* 89:5015-5019.

Zhulin, I. B., B. L. Taylor and R. Dixon. 1997. PAS domain S-boxes in Archaea, bacteria and sensors for oxygen and redox. *Trends in Biochemical Sciences* 22:331-333.

Ross Hardison is a professor in the Department of Biochemistry and Molecular Biology of the Pennsylvania State University. His research combines bioinformatics and molecular genetics to study regulation of mammalian hemoglobin genes. He received his Ph. D. in biochemistry from the University of Iowa in 1977. Address: Department of Biochemistry and Molecular Biology, the Pennsylvania State University, Althouse Laboratory, University Park, PA 16802. Internet: rch8@psu.edu. Website: <http://globin.cse.psu.edu>.