The Molecular Genetics of Hemophilia

Hemophiliacs bleed because a defective gene deprives them of a key blood-clotting protein. The protein has now been made artificially by isolating the normal gene and then inserting it into cultured cells

by Richard M. Lawn and Gordon A. Vehar

A small defect in a single human gene, and the resulting absence or deficiency of the protein it encodes, can lead to debilitating disease. Such a disease is hemophilia. Hemophiliacs lack a crucial blood protein, one that takes part in the cascade of enzymatic reactions that causes blood to clot at the site of a wound. If a severely ill hemophiliac is not treated, he may suffer internal hemorrhaging after a minor bump; he will probably die at an early age from the effects of a bleeding crisis.

Fortunately hemophiliacs can be treated with regular transfusions of a concentrate of the missing protein. Since the early 1960's, when this form of treatment first became available, the lives of hemophiliacs in developed countries have improved dramatically, and their life expectancy, once only about 20 years, is now nearly normal. The protein concentrate, however, has to be prepared from the pooled blood of a large number of donors, and so it is expensive. In the U.S. the amount required by a typical hemophiliac in a year costs between $6,000 and $10,000; in poor countries the concentrate is often not available at all. Moreover, because it is made from pooled blood, it may spread viral diseases. Most hemophiliacs are chronically infected with hepatitis viruses, and they risk contracting the acquired immune deficiency syndrome (AIDS).

Hence there has been a strong interest in finding a way to make the antihemophilic protein by means of genetic engineering. In most cases hemophilia is caused by a defect in the stretch of DNA that encodes a clotting protein called factor VIII. Research groups at two biotechnology companies, including our own group at Genentech, Inc., in South San Francisco, have recently succeeded in isolating the factor VIII gene from the cells of healthy people and recombining it with the DNA of cells cultured in the laboratory. The recombinant cells replicate, and in so doing they make many clones, or copies, of the factor VIII gene. Each recombinant cell expresses the gene's instructions; together the cells synthesize a significant quantity of factor VIII.

The bioengineered protein works. In laboratory tests it causes blood drawn from hemophiliacs to clot, and it has proved effective in hemophilic dogs. Both Genentech and the Genetics Institute in Cambridge, the other firm involved in this research, are now developing methods for synthesizing factor VIII on a commercial scale. Although the protein must still undergo further tests in animals and then in human patients, it seems likely that within a few years abundant supplies of pure, virus-free factor VIII will be on the market.

The availability of the cloned gene is already transforming the study of hemophilia, which had been hampered by the fact that factor VIII is extremely difficult to purify from blood. In addition to being scarce, factor VIII is an unusually large and unstable protein. When we began our work, its structure was not known, nor where in the body it is synthesized. Now that it can be synthesized in the laboratory much more of it is available for study. Furthermore, the fundamental structure of the protein has been read from its genetic blueprint. In a few cases we and other workers have even been able to pinpoint the genetic mutations that give rise to hemophilia and are passed on from generation to generation.

The knowledge that hemophilia is inherited goes back at least to the writers of the Talmud: they decreed that boys whose older brothers or cousins had bled to death after circumcision need not undergo the procedure. The distinctive inheritance pattern of the disease—generally only males are afflicted, but females may be carriers—was first described accurately early in the 19th century. Perhaps the most celebrated carrier was Queen Victoria. One of her sons was hemophilic and at least two of her daughters were carriers. Through the marriages of her daughters Victoria's mutant gene spread to the royal families of Germany, Russia and Spain.

It is now known that hemophilia is sex-linked because the gene for factor VIII happens to lie on the X chromosome. Female cells contain two X chromosomes; male cells contain one X and one Y chromosome. Since a male has just one factor VIII gene, inherited from his mother, he will be hemophilic if the gene is defective. A female, in contrast, has two factor VIII genes, one inherited from each parent. She can therefore carry a defective gene without suffering from hemophilia, because the normal gene on her other X chromosome protects her. Only rarely will both genes be defective, and so there are only a handful of female hemophiliacs. Carrier females will, on the average, pass their mutant gene to half of their daughters, who will be carriers, and to half of their sons, who will be hemophilic.

The process of blood clotting that goes awry in a hemophiliac is understood only in outline. It is initiated by platelets that adhere to the site of a wound. The platelets would be easily dislodged, however, were they not bound in place by strands of fibrin, an insoluble polymer. The formation of a network of fibrin from its soluble precursor, fibrinogen, is the key event in clotting; it is the end result of a complex cascade of protein interactions that is somehow set in motion by an

injury to a blood vessel. At each step in the cascade a protein precursor is cleaved to form an active enzyme called a protease. The protease thereupon cleaves another protein, converting it into a protease. Most of the cleavage steps involve cofactors, which in some cases are themselves proteins that exist in both active and inactive forms. Factor VIII, in spite of its name, is such a cofactor. It helps the protease factor IX to activate factor X in the middle of the cascade.

The clotting cascade incorporates positive-feedback loops to accelerate the clotting response and negative-feedback loops to help stop clotting. For example, thrombin, the protease that converts fibrinogen into fibrin, also activates factor VIII. At the same time, though, it activates a protease called protein *C*, which deactivates factor VIII. Since the concentration of factor VIII in normal blood is extremely small (for every molecule of factor VIII there are about a million molecules of albumin, the major blood protein), it may well be a limiting factor. In other words, the ready activation and deactivation of factor VIII may in part account for the delicate balance in healthy people between clot formation and the free flow of blood.

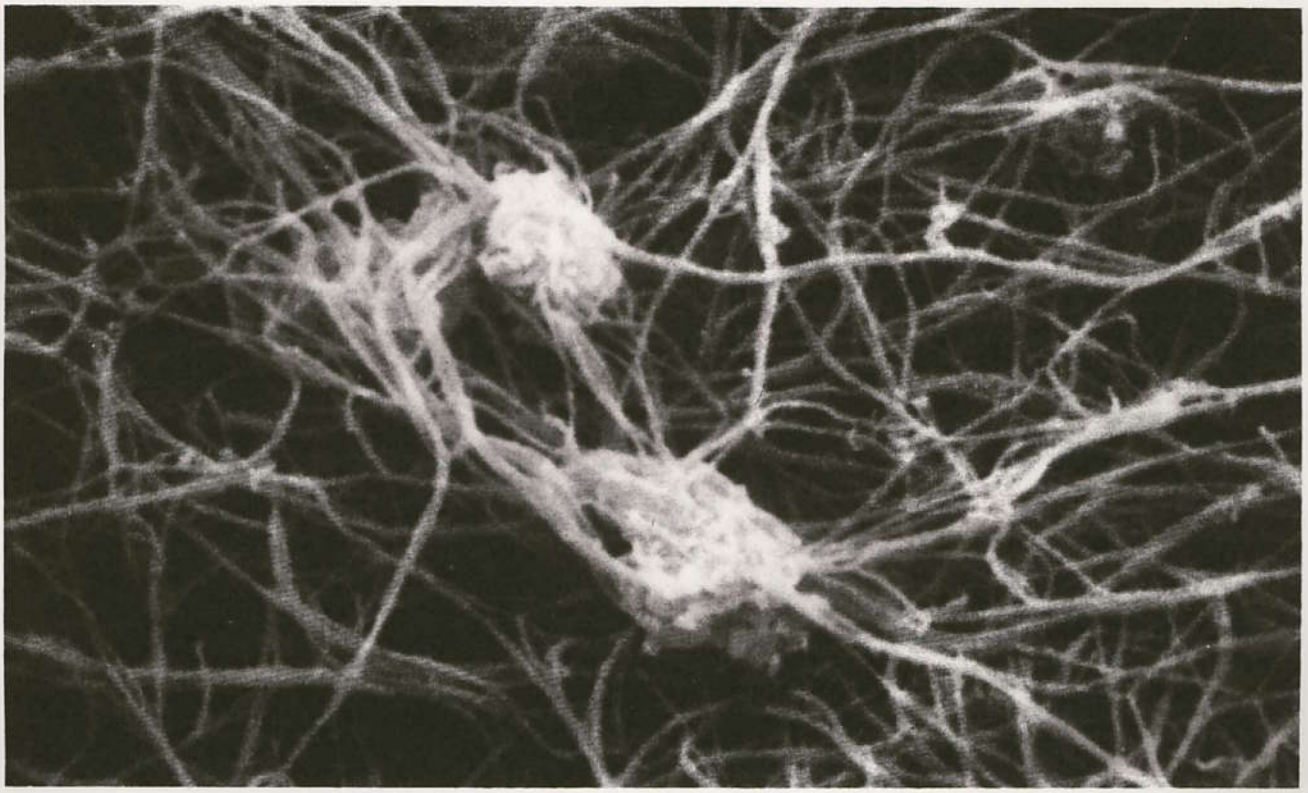
In hemophiliacs the balance is disrupted. About 85 percent of them, or roughly one male in 10,000, suffer from classic hemophilia (hemophilia *A*), in which the absence of functional factor VIII halts the clotting cascade before fibrin can form. Nearly all the rest suffer from hemophilia *B*, which is caused by a factor IX deficiency. The gene for factor IX has been cloned, and several biotechnology companies are trying to develop a bioengineered factor IX product. Bioengineered factor VIII is the greater prize, however, because hemophilia *A* afflicts more people than hemophilia *B*.

Manufacturing a protein as large and as scarce as factor VIII offered unprecedented technical challenges. The difficulties forced us to modify the standard method by which genes are cloned and manipulated to direct the synthesis of proteins.

A protein such as factor VIII is a chain of amino acids; its gene is a stretch of DNA, that is, a chain of nucleotides. The sequence of amino acids is determined by the sequence of nucleotides. Each nucleotide carries one of four bases: adenine (*A*), thymine (*T*), guanine (*G*) or cytosine (*C*). A set of three bases, called a codon, specifies one amino acid. The structure of the bases is such that they form complementary pairs: adenine forms hydrogen bonds with thymine, whereas guanine binds to cytosine. Base pairing holds the two strands of the DNA double helix together. It also governs the transcription of a gene into messenger RNA (mRNA) and the subsequent translation of mRNA into protein.

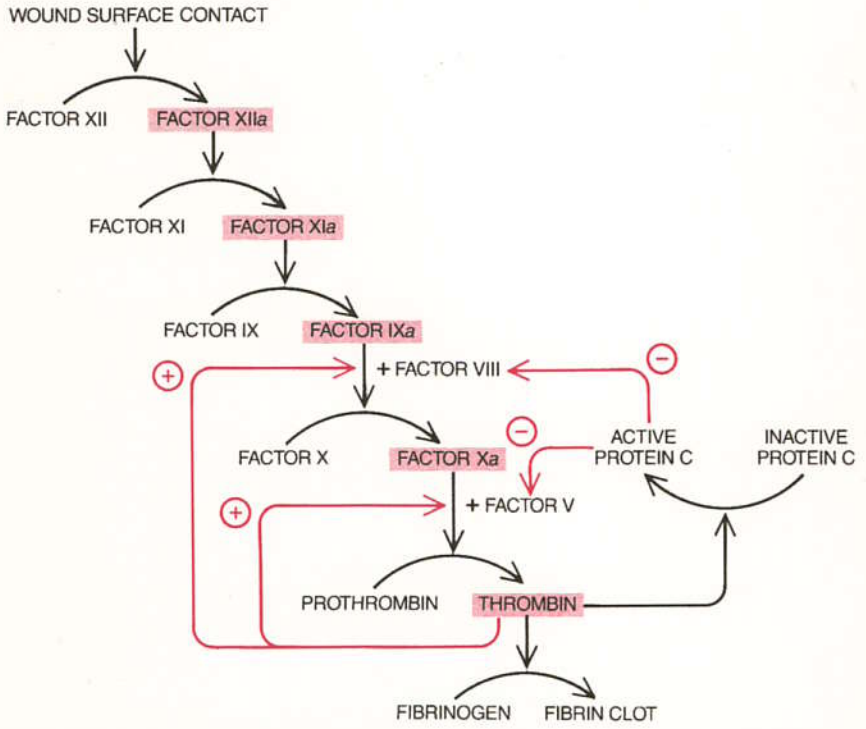
In manufacturing a protein in the laboratory the key problem is to find the right gene among the thousands in a cell. Base pairing provides the solution. A small piece of DNA or RNA whose base sequence is complementary to part of the desired gene serves as a probe for the gene. A DNA probe can be made, for example, by reverse-translating part of the desired protein's amino acid sequence according to the genetic code. The synthetic probe is labeled with a radioactive nucleotide. When the probe is mixed with the DNA in a gene "library," it "hybridizes" only with the desired gene, which is thereby labeled too.

The smaller the gene library is, the easier it is to select a specific gene. The commonest cloning method, called cDNA cloning, reduces the size of the library by taking advantage of the fact that not all genes are active in every cell. In a given cell only some genes are transcribed into mRNA and then translated into protein. If it is known

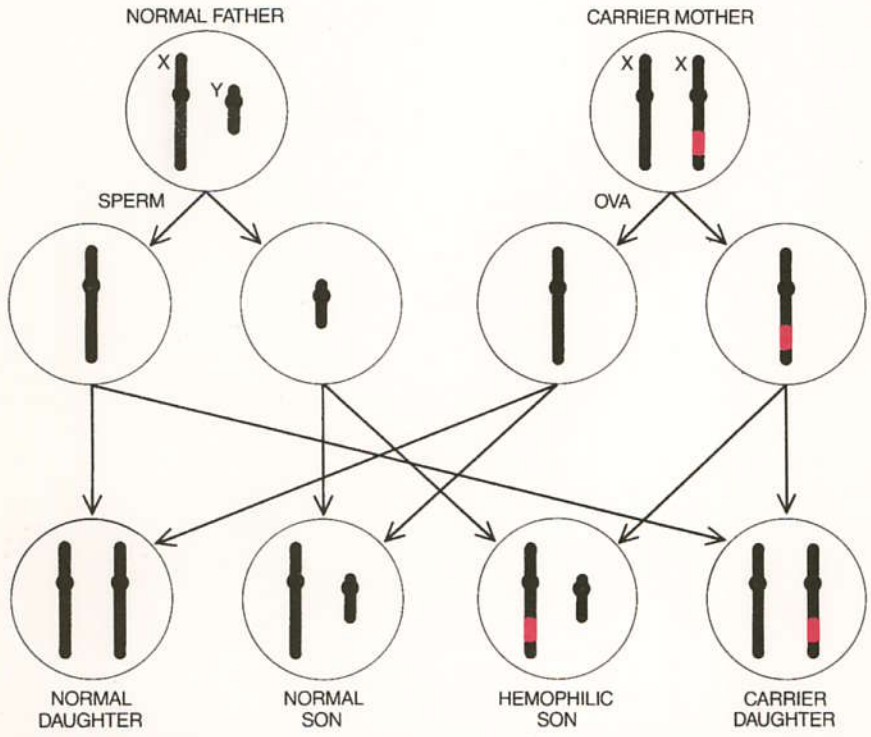


FIBRIN STRANDS stabilize a blood clot at the site of a wound by trapping the platelets that form the bulk of the clot. The electron micrograph, which was made by Jon C. Lewis of Wake Forest University, shows a clot formed in a suspension of platelets and fibrin.

A clot in the bloodstream is the result of a complex cascade of enzymatic reactions culminating in the conversion of fibrinogen, a soluble protein, into insoluble fibrin strands. In hemophiliacs a crucial protein in the blood-clotting cascade is either missing or defective.



CLOTTING CASCADE begins when cell damage at a wound somehow activates the enzyme factor XII; it ends with the conversion of fibrinogen into fibrin by thrombin. At each step an inactive protein is converted into a protease, or protein-cutting enzyme (*color*), which activates the next protein. Some steps require cofactors such as factors VIII and V. The cascade includes positive- and negative-feedback loops (*colored arrows*). Thrombin activates factors VIII and V; it also deactivates them (by activating protein *C*), which helps to halt clotting. Some 85 percent of hemophiliacs lack factor VIII. The rest lack factor IX.



SEX-LINKED INHERITANCE of hemophilia results from the location of the factor VIII gene on the X chromosome. A male carrying a mutant factor VIII gene lacks normal factor VIII and is hemophilic. A female carrier is protected by the normal gene on her second X chromosome, but half of her daughters will be carriers and half of her sons will be hemophilic. In the case of a hemophilic father (not shown), his sons will not be hemophilic, because they receive his Y (not his X) chromosome, but his daughters will be carriers.

which cells make the desired protein, one has only to screen the mRNA molecules from those cells. Among them there must be some transcripts of the desired gene.

To find the gene one first copies all the mRNA back into DNA with the help of an enzyme called reverse transcriptase. Individual pieces of copy DNA, or cDNA, are then enzymatically linked to the genetic material of a vector, which is often the bacterial virus phage lambda. The phages are introduced into bacteria in such a way that each phage multiplies in a separate region of a petri dish, producing a distinct plaque of phages and dead bacteria. Together the plaques constitute a cDNA library. At least one of them contains the desired cDNA fragment; that plaque is identified by hybridization with a probe.

The cDNA-cloning strategy works only if one knows what cells in the body produce the desired protein. Moreover, it is most likely to be successful if the protein is made in abundance; in that case the cells will contain many copies of the mRNA, and many plaques in the cDNA library will contain copies of the gene. Neither of these conditions applied to factor VIII. Factor VIII is scarce, and when we began our work, no one knew what organs produce it. Had we tried to construct a cDNA library, we might well have chosen the wrong cell type and ended up with a library that did not include the factor VIII gene.

We therefore decided to look for the factor VIII gene where we could be sure of finding it: in a recombinant library derived from the genome, or complete set of genes, of a cell. A genomic library is constructed by extracting the chromosomes from cells, cleaving the DNA into fragments with enzymes and joining the fragments to phage-lambda DNA. Because a genomic library contains hundreds of times more DNA than a cDNA library, it is more difficult to screen with a probe.

Before we could make a probe we first needed to know part of factor VIII's amino acid sequence. Determining even a small part of the sequence was no mean feat. The protein had not even been purified until 1980, when one of us (Vehar), working in the laboratory of Earl W. Davie of the University of Washington School of Medicine, laboriously extracted several milligrams of pure factor VIII from 25,000 liters of cows' blood. Subsequently Edward Tuddenham and his colleagues at the Royal Free Hospital in London obtained enough human factor VIII to enable workers at Genentech to sequence a short stretch of

the protein. A group at the Genetics Institute achieved similar results with porcine factor VIII purified by David N. Fass at the Mayo Clinic.

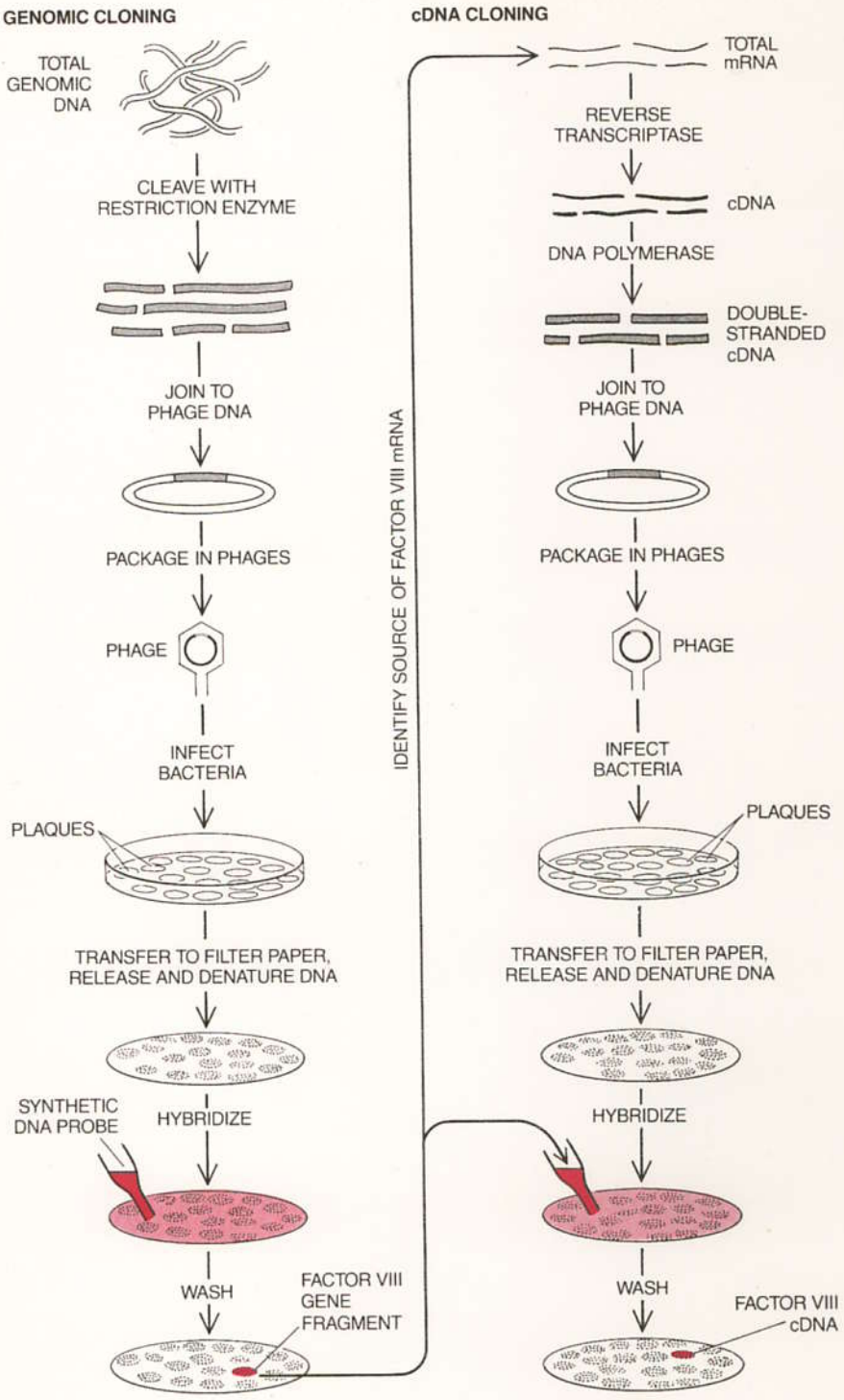
The next step was to reverse translate the protein sequence into DNA. In doing this one encounters a problem: the genetic code is redundant. An amino acid can be encoded by as many as six different codons. (There are 64 possible three-base codons but only 20 different amino acids.) One solution is to synthesize a pool of short (about 17 bases long) DNA probes that covers all the possibilities. The shorter the probe, however, the more likely it is to match randomly with a stretch of DNA other than the desired gene. When many short probes are used to screen a large genomic library, the false-positive problem becomes acute.

To avoid the problem we relied on a single, relatively long (36 bases) probe derived from a 12-amino-acid stretch of factor VIII. We had to choose among 147,456 ways of encoding this particular sequence. Fortunately we were able to make highly educated guesses, because we knew that some codons are more prevalent than others in mammalian genes. As it turned out, we got 30 of the 36 bases right, which was a close enough match. When we screened the genomic library, the synthetic probe hybridized with overlapping segments of factor VIII DNA, thereby identifying plaques containing parts of the gene.

The entire gene is too large to fit into a single phage. To find the rest of it, William I. Wood, Jane Gitschier and other workers in our laboratory rescreened the library, this time using fragments of the identified gene segments as probes. By repeating this procedure, known as chromosome walking, they eventually obtained a series of overlapping segments constituting the complete gene sequence.

The gene is 186,000 bases long. The information for factor VIII, however, is spread among 26 exons, or coding sequences, that together account for less than a twentieth of the total length of the gene. The reason is that the exons are separated by 25 introns, or noncoding intervening sequences. After the entire gene is transcribed into RNA in a living cell, the introns are cut out of the transcript. The exons are then spliced to form the mRNA that directs protein synthesis. To make factor VIII in cultured cells, we too needed a gene without introns. In other words, we had to find factor VIII mRNA and convert it into cDNA.

With pieces of the real gene available as probes it became a straightforward task to find out what cells make factor VIII and its mRNA. If



GENE CLONING involves finding a specific gene among thousands in a human cell. The standard method, if one knows which cells make the desired protein, is to screen a copy DNA (cDNA) library derived by reverse transcription from the messenger RNA (mRNA) of those cells (*right*). In looking for the factor VIII gene, however, the authors did not know where the protein is produced. Hence they screened the entire human genome (*left*). Chromosomal DNA fragments were joined to the DNA of the bacterial virus phage lambda. Each phage contained one human DNA fragment; each phage multiplied and formed a plaque in a distinct region of a bacterial culture. To identify the plaque containing the factor VIII gene, the phages were blotted onto filter paper and broken open to release their DNA. The DNA was exposed to a radioactive probe: a small piece of synthetic DNA encoding part of factor VIII. The probe hybridized with part of the factor VIII gene, thereby labeling it. To produce factor VIII in cultured cells, it was still necessary to make factor VIII cDNA, which lacks the introns (noncoding sequences) that complicate the full gene. Now fragments of the cloned gene could serve as reliable probes, first for identifying cells that make factor VIII mRNA and then for finding factor VIII cDNA in the cDNA library.

the mRNA from a particular cell type failed to hybridize with the probe, one could now be sure it was not because the probe was faulty; with the synthetic probe there would always have been uncertainty. A group led by Daniel Capon at Genentech found tiny amounts of hybridizing mRNA in the mRNA of a cultured cell line, while John J. Toole and his colleagues at the Genetics Institute identified liver cells as a source. Since then factor VIII mRNA has also been found in other tissues, including kidney, spleen and lymph cells, but the liver seems to be the primary source. Most of the factor VIII in healthy people is probably synthesized in liver cells and secreted into the bloodstream.

Once a source of factor VIII mRNA had been found, workers at both companies constructed cDNA libraries. Again pieces of the cloned gene served as probes, this time to pick the rare factor VIII cDNA out of a library of thousands of clones. Actually the factor VIII cDNA had to be stitched together from several overlapping segments; the mRNA is about 9,000 bases long, and with current techniques it is not possible to copy so large a molecule in one stretch. Finally, control sequences had to be attached to the cDNA. Control sequences direct enzymes in the recombinant cell to start and stop transcribing a gene.

Recombinant bacteria, usually *Escherichia coli,* were suitable for manufacturing the first bioengineered proteins, such as insulin and interferon, because these molecules are relatively small. Bacteria are generally not equipped, however, to produce a large and complex protein such as factor VIII. Nor do they always have the right enzymes for modifying or folding a large protein after it has been synthesized.

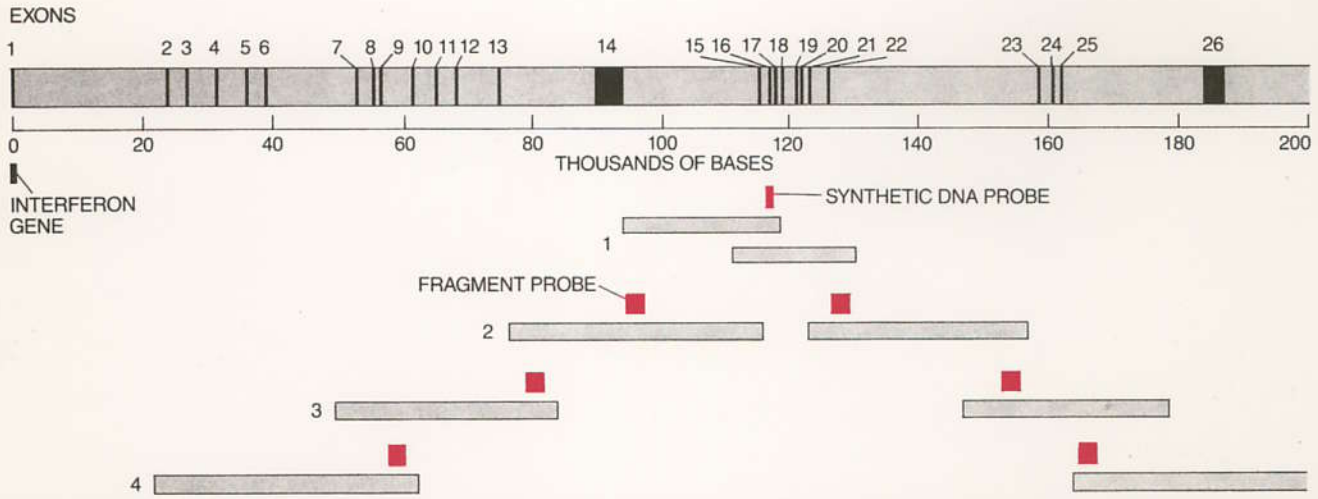
For these reasons we chose to insert the cloned gene for factor VIII into hamster cells, which are easily grown in the laboratory. The recombinant hamster cells thereupon secreted human factor VIII into the culture broth. That is, we hoped it was factor VIII; we could not be sure the protein was fully functional until it had been shown to make blood from hemophiliacs clot normally. It was possible, for instance, that our bioengineered protein was just one part of a protein complex that is missing in hemophiliacs. It was even conceivable, given how difficult it had been to purify factor VIII, that we had cloned a gene encoding some impurity in the preparation. Such unsettling possibilities had to be considered, but neither proved to be true. Bioengineered factor VIII does clot hemophilic blood. Indeed, it has been found to be equivalent in every way to the blood-derived protein.

What is the structure of factor VIII? The advent of gene cloning has produced the novel situation in which some protein sequences are determined indirectly, from the DNA sequence of the gene. Such was the case with factor VIII, which is much too large and scarce to be sequenced directly in its entirety. Before its gene was cloned workers did not even agree on its approximate size; estimates differed by a factor of nearly 100.

Now the question can be answered. The 9,000-nucleotide cDNA for factor VIII encodes a protein 2,351 amino acids long. (Nearly 2,000 bases at the ends of the gene are transcribed into mRNA but are not translated into protein.) The first 19 amino acids form a hydrophobic sequence typical of secreted proteins. This "signal peptide" is generally cut off the protein as it is secreted, and so a mature factor VIII molecule must consist of the remaining 2,332 amino acids. Its molecular weight must be about 330,000 daltons. (One dalton is 1.66×10^{-24} grams.) In comparison, a molecule of interferon is only 166 amino acids long and weighs about 19,000 daltons. The factor VIII gene is by far the largest ever cloned and expressed in foreign cells.

An analysis of the amino acid sequence of factor VIII reveals that the protein is constructed of repeated similar segments. Three of these are designated *A*. Of the roughly 350 amino acids that make up the sequence of each *A* segment, approximately a third are common to all three segments. A comparable level of homology also exists between the two 150-amino-acid segments designated *C,* which are not homologous to the *A* segments. Since there are 20 different amino acids, a one-third homology among sequences is almost certainly not a random occurrence; the homologous segments must be related.

A surprising clue to the evolutionary history of factor VIII came from a computer-aided comparison of its sequence with that of other proteins. The three *A* segments turn out to be nearly



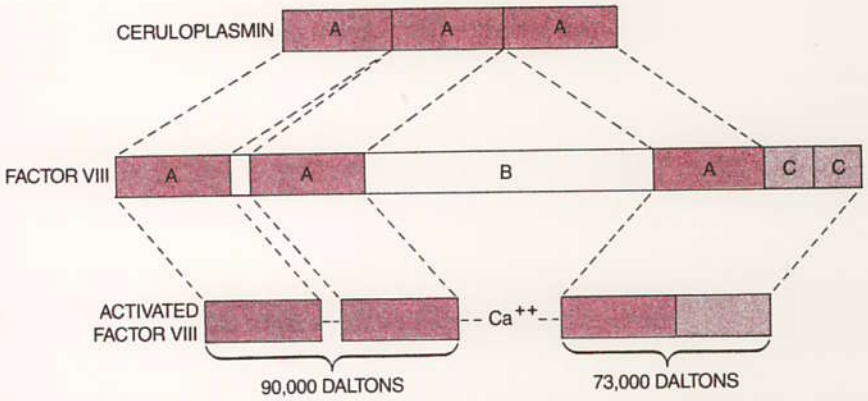
TREMENDOUS SIZE of the factor VIII gene, the largest gene cloned to date, forced workers to apply a cloning technique called chromosome walking. The factor VIII gene is 186,000 bases long. In contrast the interferon gene, which was cloned in 1980, incorporates only about 600 bases. Because the factor VIII gene is too large to fit into a single phage, segments of it were found in different plaques in the genomic library. When the library was screened with a synthetic DNA probe, the probe hybridized with overlapping segments (*1*). Pieces of the segments then served as probes to rescreen the library and identify further segments (*2*). By repeating this procedure nearly all of the gene was identified (*3, 4*). (Its beginning was found once factor VIII cDNA was available as a probe.) Less than one-twentieth of the gene consists of exons, or coding sequences (*black bands*); the 26 exons are separated by 25 introns.

as similar to the three domains of ceruloplasmin, a protein that carries copper in the bloodstream, as they are to one another. Previously there had been no reason to suspect a connection between factor VIII and ceruloplasmin, but the homology of their domains suggests their genes evolved from a common ancestor. The ancestral protein may have consisted of three identical domains, precursors of the modern *A* domains. In ceruloplasmin the three *A* domains are contiguous and make up the entire molecule, whereas in factor VIII the second and third *A* domains are separated by nearly 1,000 amino acids. In the factor VIII gene this intermediate region is encoded by a single huge exon, which may have been inserted into the ancestral gene. The regions encoding the *C* segments of factor VIII were also added to the end of the gene.

Factor VIII purified from donated blood is almost never identical with the full, 330,000-dalton molecule encoded by the gene. It is now thought that factor VIII undergoes a series of cleavages, including the removal of the region between the second and third *A* domains (the *B* region), when it is activated in the bloodstream. Evidence for this view was found by sequencing small parts of the active protein directly and comparing the results with the sequence of the cloned gene. Active factor VIII seems to consist of a 90,000-dalton subunit joined to a 73,000-dalton subunit. The first subunit consists of the first two *A* domains; the peptide bond between the two is cut but they remain linked. The second subunit consists of the third *A* domain and the two *C* domains. Just as the molecule is activated by cleavages, so too is it readily deactivated by further cleavages within the subunits. It is probably through these reactions that the clotting cascade is brought to a timely halt.

The details of how factor VIII functions in the cascade are far from clear. It is bound to a carrier protein called von Willebrand factor, which keeps factor VIII circulating in the blood and may help to position it on the surface of a platelet at the site of a wound. Once on the platelet, factor VIII probably separates from von Willebrand factor and forms a complex with factor IX and factor X. The binding sites on these proteins have not been found. All that is really known is that without factor VIII the activation of factor X by factor IX does not take place.

A standard method of learning more about how a protein works is to study the abnormal forms that result from genetic mutations. We can now begin to apply this approach to factor VIII and at the same time identify the mutations that cause hemophilia. The ultimate result will be an improved understanding of the disease and improved treatment for its victims.



AMINO ACID SEQUENCE of factor VIII resembles that of the blood protein ceruloplasmin. The three *A* domains of factor VIII have about a third of their amino acids in common and are also homologous to the three domains of ceruloplasmin. The two proteins must have evolved from a common ancestor. Factor VIII probably diverged from ceruloplasmin when a large exon (number 14) encoding the intermediate *B* segment was inserted into the ancestral gene; exons encoding the *C* segments were added to the end. When factor VIII is activated, the *B* segment is excised by proteases, and the two resulting subunits are held together by a calcium ion. Another cleavage takes place between the first two *A* domains.

As expected, we and other workers have found there is no single mutation underlying hemophilia. Fifty years ago the British geneticist J. B. S. Haldane pointed out that serious diseases linked to the X chromosome must constantly arise anew through random, spontaneous mutations; otherwise the diseases would eventually die out. Indeed, roughly a third of the cases of hemophilia observed today occur in families with no history of the disease. A hemophilic gene is of course transmitted to offspring, but before the advent of effective treatment a particular mutation soon became extinct, simply because there were fewer surviving children in hemophilic families. In contrast, a recessive mutation on an autosomal chromosome, of which a cell has two copies, can spread through the population, because it affects only those rare individuals who inherit two defective genes.

In principle it is possible to identify the mutation that gives rise to hemophilia in an individual by isolating and sequencing his factor VIII gene. Since it would take several months to sequence each 186,000-base gene, however, the method is not practical. Fortunately there is a quicker procedure, albeit one that is applicable to only a small set of cases. The procedure is based on a hybridization technique called Southern blotting.

The first step is to extract the DNA from the blood cells of a hemophiliac. The DNA is cleaved into a million or so fragments with a restriction endonuclease, an enzyme that cuts DNA wherever it recognizes a specific four-to-six-base sequence. The fragments are then separated according to size by electrophoresis: the smaller the fragment is, the farther it migrates through an agarose gel when an electric current is applied to the gel. Next the DNA is unraveled into single strands and blotted onto special filter paper. In the process the fragments retain the relative positions they occupied in the gel. Finally the filter is bathed in a solution of radioactive factor VIII cDNA. The cDNA probe hybridizes with fragments of the factor VIII gene. The sizes of the fragments can be deduced from their positions on the paper, which form a distinctive pattern.

To find factor VIII mutations, we and our colleagues compared the hybridization patterns of normal and hemophilic DNA's. Two types of gene alteration can be detected with this method. The easiest to recognize is a gross deletion of part of the factor VIII gene; some of the hybridizing fragments are missing or are altered in size. Occasionally, though, one can also detect a change of a single DNA base, provided the change happens to occur within the recognition sequence of a restriction enzyme. Such a mutation prevents the enzyme from cleaving the gene. Two hybridizing fragments in the normal pattern are therefore replaced by a single larger fragment in the hemophilic pattern. (Conversely, a mutation can also change a blot pattern by creating a new restriction site.)

One example of a single-base mutation will suffice to illustrate the procedure. DNA from a severe hemophiliac, from his parents and from his three

siblings was cut with the restriction enzyme *Taq*I, which recognizes the base sequence *TCGA*. The Southern blots of the five unaffected relatives all showed two hybridizing fragments, one consisting of 1,400 nucleotides and the other of 2,800 nucleotides. On the hemophiliac's Southern blot the two fragments were replaced by a single fragment 4,200 nucleotides long. Because we knew the sequence of the normal factor VIII gene, we could determine the position of the altered *Taq*I site. By cloning and sequencing that part of the hemophiliac's gene we found the DNA sequence of the *Taq*I site had been changed from *TCGA* to *TTGA*. The mutation prevents *Taq*I from cleaving the gene at that point.

More important—and coincidentally—the mutation changes a codon for the amino acid arginine (*CGA*) to a "stop" codon (*TGA*), which brings the synthesis of factor VIII to a premature halt. The truncated protein is probably either inactive or too unstable to survive in the bloodstream. Since the hemophiliac's parents lack the mutation, he did not inherit it. It must be a new mutation that occurred in the egg from which his cells developed.

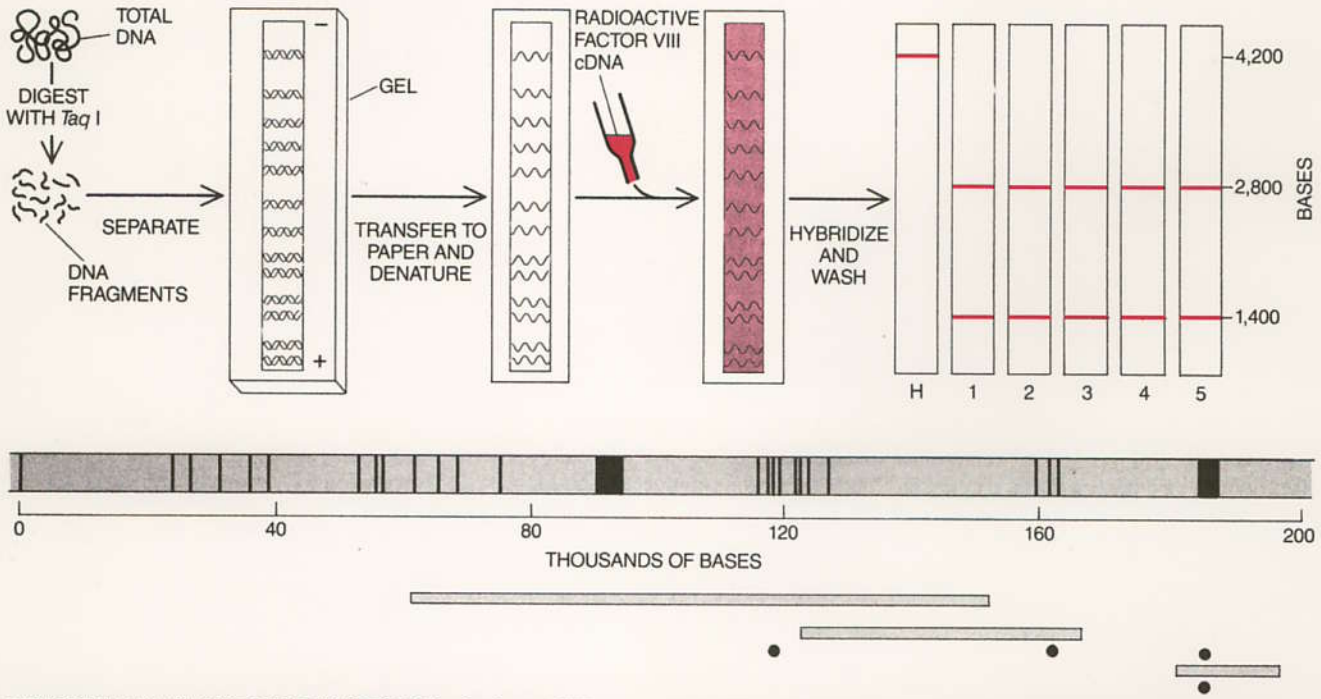
So far a total of 200 hemophilic factor VIII genes have been examined in our laboratory and by workers at the Genetics Institute and at Johns Hopkins University. Seven different mutations have been pinpointed, and none has been observed in more than one family. Four of the mutations are single-base changes, of which three lead to a truncated factor VIII and severe hemophilia; the fourth causes the substitution of an incorrect amino acid and results in a relatively mild form of the disease. The other three mutations are deletions of several thousand nucleotides from the gene. All three deletions cause severe hemophilia.

In the future investigators may have access to more efficient techniques for locating single-base mutations. By analyzing a large number of mutations one may be able to correlate types of mutation with the level of clinical severity of the disease. It would be of particular importance, for example, to understand why some 10 percent of hemophiliacs suffer immune reactions to exogenous factor VIII; these people are the hardest to treat.

In principle it would be possible to cure hemophiliacs by introducing functional factor VIII genes into their cells. Yet gene therapy for any disease is probably years away. One of the chief obstacles is the problem of controlling the productivity of the inserted genes: too much factor VIII, for example, may be as dangerous as too little.

The cloned factor VIII gene is already serving as the basis for more reliable methods of diagnosing female carriers and of detecting hemophilia in fetuses. Essentially these methods involve blot hybridization tests using the cloned gene as a probe to track the inheritance of a defective gene. The prenatal-diagnosis technique is being practiced at some 70 medical centers around the world. Although it is not yet applicable in all cases, it is more reliable than the old method of measuring the concentration of factor VIII in fetal blood, and it does not require an incision. In addition, whereas a fetal blood test cannot be done before the 20th week of pregnancy, DNA-based diagnosis is feasible in the eighth week. If the parents choose an abortion, it is less risky for the mother at that stage.

Clearly the most significant immediate medical implication of the cloning of the factor VIII gene is the prospect of a safe and abundant supply of factor VIII. The bioengineered protein is scheduled to begin several years of clinical trials within a year or two. When it becomes commercially available, hemophiliacs will be liberated from the menace of transfused infection. Many of those who live in underdeveloped countries, and who today still die young, will receive effective treatment for the first time.



HEMOPHILIA-CAUSING MUTATIONS in the factor VIII gene can be detected by Southern blotting (*top*) if they happen to change the way the gene is fragmented by a restriction enzyme. DNA from blood cells is cut into millions of fragments, in this case with the enzyme *Taq*I. The fragments are separated according to size by electrophoresis, unraveled into single strands and blotted onto filter paper. The filter is bathed in a solution of radioactive factor VIII cDNA, which hybridizes only with fragments of the factor VIII gene. The size of the hybridizing fragments is revealed by exposing X-ray film to the filter. In the example shown here a point mutation in the factor VIII gene of a hemophiliac (*H*) has eliminated a *Taq*I cleavage site. The 2,800- and 1,400-base fragments on the blot patterns of his relatives (*1–5*) are replaced by a single, uncut 4,200-base fragment. So far seven different mutations have been located on hemophilic factor VIII genes (*bottom*). Four are point mutations, or changes of a single base (*dots*); three are extensive deletions (*bars*).