Regular Review

Mapping haemoglobin genes

D J WEATHERALL

The work for which the Nobel prizes for medicine and physiology have been given over recent years often seems to the clinician completely devoid of any possible practical application. For example, who would have imagined that the discoveries that led to two recent prizes, the viral reverse transcriptases and DNA restriction enzymes, would within a few years provide extraordinary insights into the causes of some common genetic diseases? Yet it is these two fundamental advances which have opened up the whole field of the detailed structural analysis of human genes and which, within the last year, have started to yield some remarkable information about the molecular basis for some of the inherited diseases of haemoglobin production.

Genetic disorders of haemoglobin synthesis—Many genetic disorders result from an abnormality in the structure or rate of synthesis of a specific protein or enzyme. The molecular structure is determined by the order of nucleotide bases that constitute the DNA of the gene (or genes) for the particular protein. This genetic information is transferred from the nucleus to the cytoplasm of the cell in which the protein is synthesised by means of a molecule called messenger RNA, which is transcribed from the structural gene—and hence is an exact replica of it. The messenger RNA acts as a template for protein synthesis as amino-acids are brought to it on specific carrier molecules which "find" the right position by interacting with triplets of bases (codons). This mechanism—DNA→RNA→protein—is the central pivot of molecular biology.

Many human genetic disorders result from point mutations in the DNA of a particular gene for a protein or enzyme. A single base change in the DNA of the structural gene leads to the insertion of an incorrect amino-acid in the protein, and this may alter the structure or function of that protein and cause disease. Much more commonly, however, genetic diseases result from defects in the rate of production of a protein or enzyme. In these conditions a protein may not be made at all or may be made at a drastically reduced rate.

Probably the most common and best-studied single-gene human disorders due to a reduced rate of protein synthesis are the inherited disorders of haemoglobin synthesis, the thalassaemias. These produce major public health problems in many parts of the world and are being seen with increasing frequency in Britain. Normal adult haemoglobin consists of two α-chains and two β-chains ($\alpha_2\beta_2$), and there are two major forms of thalassaemia, α-thalassaemia and β-thalassaemia, which result from defective α- and β-chain synthesis respectively.[1] These conditions are genetically heterogeneous, and in some forms of α- or β-thalassaemia synthesis of α- or β-chains occurs at a reduced rate, while in others none are amount or absence of a gene product. While analysis of haemoglobin and its messenger RNAs has given some insight into the underlying mechanisms that cause these diseases, a complete understanding of their molecular basis can be obtained only by direct analysis of the globin genes themselves.

Analysis of the globin genes—To examine the globin genes (or any other genes for that matter) they have to be isolated. The first step is to obtain some DNA. Since every cell in the body contains an individual's entire complement of genes DNA can be obtained from peripheral blood lymphocytes, operation samples such as spleen, skin grown in tissue culture, or any other available source. Having obtained a sample of DNA, however, we are faced with another formidable difficulty. There is enough DNA in the cell nucleus to form some 10 million genes the size of a globin gene[2]—so that if the globin genes are present in only a few copies we are attempting to isolate and examine only one part of DNA in several million. Though indirect methods for doing this have been available for some years, only recently, with the advent of restriction enzyme technology, has this type of problem been really amenable to attack.

DNA restriction endonucleases are enzymes found in bacteria and capable of slicing DNA at specific base sequences.[3] Presumably bacteria synthesise these enzymes to protect them from infecting viruses; they do not chop up their own DNA with restriction enzymes because they protect the site susceptible to attack, probably by methylation of the DNA at that site. A whole series of restriction enzymes are now available, each cutting DNA at different and highly specific sequences. Suppose, therefore, an α- or β-globin gene is to be isolated and its size and structure examined. An individual globin gene might be contained in about 1600 nucleotide bases (1·6 kilobases or kb). A restriction enzyme can be selected to cut the DNA on either side of, say, the β-globin gene so that it is present in a conveniently sized piece of DNA, about 4 kb long, for example. Of course, breaking up all the DNA in one cell with this enzyme will yield many thousands of pieces of varying sizes, simply because many "restriction sites" other than those near the globin genes will be attacked by the enzyme. These pieces may be separated by electrophoresis of the mixture in a slab of supporting gel, in which smaller pieces migrate faster than the larger ones. Having done this we are left with a gel containing a smear of many fragments of DNA of different sizes. The next problem is to locate the globin genes.

It is in the isolation of mammalian genes that the viral reverse transcriptases come into the story. In 1967 Temin predicted that certain tumour viruses must produce an enzyme

to RNA→DNA. Such enzymes were discovered independently by Temin and Baltimore[4 5] in 1970, and have now become generally available. Hence they can now be used, with globin messenger RNA, to produce DNA copies. A sample of reticulocyte-rich blood is obtained and α- or β-globin messenger RNA isolated. This is then mixed together with reverse transcriptase and all the building blocks required to make DNA (including some radioactive bases or phosphate) to produce a radioactively labelled copy or complementary DNA (cDNA) which has a base composition precisely complementary to the messenger RNA from which it was copied. In other words we now have a radioactive α- or β-globin cDNA. How can this be used to find the globin genes? DNA is a double-stranded molecule, and the backbone of the strands consists of sugars and phosphates. The strands are linked by the bases adenine, guanine, thymine, and cytosine; and, because of the rules of base pairing, cytosine always pairs with guanine and adenine with thymine. In fact the bases are joined by weak hydrogen bonds, so that if DNA is heated the two strands will come apart. They come together on cooling. They will pair, however, only if there is precise matching of the bases. Thus if the total cellular DNA is heated and radioactive cDNA for an α- or β-globin gene is added to the mixture, which is then allowed to reanneal, the cDNA will bind (or hybridise) to the total cell DNA only if it finds a complementary sequence of DNA —in fact, a globin gene. In this way we can use the radioactive cDNAs to "look for" the α- or β-globin genes.

Now let us return to our smear of fragments of DNA in the gel. Clearly it is not possible to hybridise DNA in the middle of a lump of gel, but the DNA fragments can be blotted on to an absorbent filter and then hybridised with radioactive cDNA probes to locate the α- or β-globin genes. Bound radioactive cDNA can be located by placing an *x*-ray plate on top of the filter after hybridisation to obtain a radioautograph of the labelled cDNA. Surprisingly, after all these manoeuvres it is possible to obtain very precise pictures of globin genes or parts of them on fragments of DNA of different sizes, depending on which particular restriction enzymes are used to cut up the original DNA.

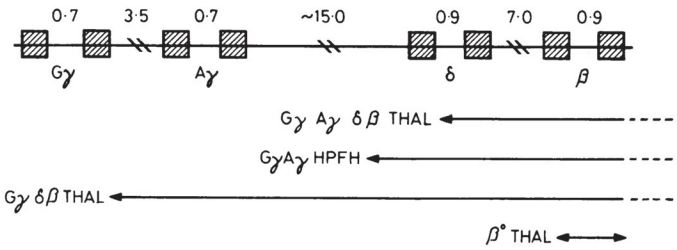
The use of restriction enzymes does not stop at producing maps of this type, however. Restricted fragments of DNA containing globin genes can be inserted into the DNA of certain plasmids or bacteriophages—organisms that replicate in certain bacteria.[6] If the insertion of the plasmid confers a selective growth advantage to the bacteria by supplying, for example, an enzyme which allows it to grow on selective media a clone of bacteria can be developed that contain replicating plasmid DNA (including the globin genes). These bacteria can be isolated for further growth by adding radioactive cDNA to the bacterial plates and finding out which clones contain the globin genes, just as was done for the restriction maps. Hence large amounts of pure globin genes can be generated in these bacterial factories. These can be sequenced to examine their structure or can be used to make extremely pure probes for locating globin genes on gels.[7 8]

Normal and abnormal human globin gene maps—In the last 18 months these complex techniques have yielded a vast amount of information about the organisation of the human globin genes. One surprising finding (which has shaken the molecular biologists) is that globin genes, apparently like most mammalian genes, are discontinuous.[9] In other words, the region of DNA that codes for the structure of a globin gene is split up into pieces which are separated by one or more inserts (or introns), lengths of DNA of up to several hundred

the whole gene is transcribed into a long piece of messenger RNA and then the inserts are excised and the coding portion of the messenger RNA is spliced before being delivered to the cytoplasm. Clearly there are many places where this process could go wrong.

A relatively complete map of the α-, β-, and related globin genes has now been produced.[10–12] Using this map, and with a whole series of restriction enzymes available to cut the globin genes either within or outside the coding sequences, their size and organisation can now be studied in a whole variety of different genetic disorders of haemoglobin production.

To make use of the gene maps obtained by restriction enzyme analysis of both normal DNA and that obtained from patients with genetic disorders of haemoglobin production, we need to consider further the heterogeneity of normal human haemoglobins.[13] In fetal life the major haemoglobin is haemoglobin F, which has two α-chains and two γ-chains ($\alpha_2\gamma_2$). The γ chains are heterogeneous and consist of two types of molecules, which differ only at position 136; in one type of haemoglobin F position 136 is occupied by glycine ($^G\gamma$) and in the other type by alanine ($^A\gamma$). The production of $^G\gamma$- and $^A\gamma$-globin chains is directed by separate gene loci. Normal adults have haemoglobin A ($\alpha_2\beta_2$) and the minor component haemoglobin A_2, which consists of α-chains and δ-chains ($\alpha_2\delta_2$). Most normal adults inherit two α-chain genes from each parent, four in all. The $^G\gamma$-, $^A\gamma$-, δ-, and β-globin genes lie in a linked cluster on chromosome 11; the anatomy of this region has been mapped and is shown in the figure.



Map of human globin genes. The distances are shown in kilobases (kb). The structural genes ($^G\gamma$, $^A\gamma$, δ and β) are shown in boxes interrupted by inserts (introns). Deletions which produce hereditary persistence of fetal haemoglobin (HPFH), δβ thalassaemia, and some forms of β^0 thalassaemia are indicated by arrows. The broken lines represent uncertainty about extent of deletions. $^G\gamma^A\gamma$ HPFH or δβ thal, or $^G\gamma$ δβ thal are forms of HPFH or δβ thalassaemia in which the haemoglobin F contains both $^G\gamma$ and $^A\gamma$ chains or $^G\gamma$ chains only.

Analysis of some of the common forms of thalassaemia by restriction endonuclease mapping has begun to disclose a variety of different molecular defects, usually involving deletions—loss of whole globin genes or parts of them. The figure also shows some of the recently discovered deletions of genetic material in the $^G\gamma$-$^A\gamma$-δ-β gene cluster, which give rise to such common clinical disorders as β^0 thalassaemia (β-thalassaemia with no β-chain production); δβ-thalassaemia, which is a disorder associated with no δ- or β-chain production; and hereditary persistence of fetal haemoglobin, another condition in which there is no δ- or β-chain production but in which γ-chain production continues into adult life and makes up for the deficit of β- and δ-chains.[14–18] Similarly, some of the α-thalassaemias have been shown to be due to deletions of one or both of the linked α-chain genes, which lie approximately 2·7 kb apart on chromosome 16.[19–21] These studies have also shown that the α- and β-thalassaemias are remarkably hetero-

α- or β-chain synthesis occurs the α- or β-globin genes appear to be intact. To elucidate these conditions we need to insert a piece of DNA containing the genes into a bacterial plasmid, isolate the particular clone containing the globin gene and grow it in large quantities, and then sequence the gene to find the precise molecular defect. Several laboratories have already succeeded in cloning human β-chain genes, from both normal and thalassaemic individuals, and answers to these problems should soon be available.

The future—The field of human gene mapping has moved incredibly fast over the last few months. Where is all this activity leading? Clearly the molecular basis for many of the genetic disorders of haemoglobin synthesis will be defined over the next year or two. Furthermore, with the increasing understanding of how the globin genes are organised we may be able to answer some of the questions about how they are regulated and switched on and off during development. As techniques become available for making appropriate radioactive cDNA probes it may be possible to apply the techniques developed for studying haemoglobin to the study of other human genetic disorders.

To the physician faced with the management of these distressing diseases the most pertinent question is whether all this sophisticated knowledge will offer any help in their management. The most immediate practical application likely to come from this work is the development of easier methods of prenatal diagnosis, at least for some of the inherited disorders of haemoglobin synthesis. Where these conditions result from major gene deletions it should be possible to identify the abnormality on relatively small samples of DNA prepared from amniotic fluid cells. This approach has already been shown to work for certain types of α- and β-thalassaemia.[22] Where the globin genes are intact, however, and no demonstrable abnormality can be found by restriction mapping we need other approaches. Almost certainly human DNA will be found to be polymorphic in areas other than those coding for the structural genes. If such a polymorphism produces a new restriction site, and it happens to be tightly associated with a common genetic disorder of haemoglobin synthesis such as β-thalassaemia or sickle-cell anaemia, then restriction mapping might offer an approach to the prenatal diagnosis of the condition. Such a polymorphism has already been found together with the sickle-cell mutation[23]; but though the association seemed to be extremely strong when first described more recent work suggests that this may not always be the case, and that the use of this particular polymorphism will probably not give foolproof prenatal diagnosis of sickle-cell anaemia. But these are early days, and possibly tighter associations will be found between one of the common haemoglobin disorders and a related polymorphism in the non-coding parts of the DNA.[23] If so, the prenatal diagnosis of these conditions will be possible on the basis of a small amniotic fluid sample rather than the current laborious methods requiring fetal blood sampling.

The ability to isolate, clone, and grow relatively large quantities of globin genes raises the possibility of using such material for specific gene replacement in patients in whom the globin genes are missing or non-functioning. Though the technical problems are formidable, this approach does not seem beyond the realms of feasibility in view of the rapid advances. But the difficulties should not be underestimated. Though globin genes can be introduced into a variety of cells, incorporated into the cell's genome, and even persuaded to synthesise some haemoglobin, there is still a long step from this remarkable technology to actual "gene therapy." For parts of the gene responsible for regulation as well as those that direct the structure of the haemoglobin chains, and somehow to ensure that the genes are active and under the normal regulatory control for protein synthesis. What kind of a cell should the globin genes be inserted into? Their placement into anything later than the stem cell would mean simply that the genes went into a terminally maturing population of cells that would eventually be lost. Yet how can the globin genes be inserted into a stem cell when we do not know what a stem cell looks like? How could the new cell line be encouraged to proliferate in preference to the abnormal cell line of the patient into whom it was inserted? Clearly these are formidable problems, but it would be a brave man who would say that they are insoluble. Progress has been so remarkable in the last few years that clinicians should probably be starting to think about some of the extremely difficult ethical problems that will be presented by the potential for such treatment.

D J WEATHERALL

Nuffield Professor of Clinical Medicine,
University of Oxford

[1] Weatherall, D J, and Clegg, J B, *The Thalassaemia Syndromes*, 3rd edn. Oxford, Blackwell Scientific, 1979.
[2] Williamson, R, Direct measurement of the number of globin genes, *British Medical Bulletin*, 1976, **32,** 246.
[3] Nathans, D, and Smith, H O, Restriction endonucleases in the analysis and restructuring of DNA molecules, *Annual Review of Biochemistry*, 1975, **44,** 273.
[4] Temin, H M, and Mizutani, S, RNA-dependent DNA polymerase in virions of rous sarcoma virus, *Nature*, 1970, **226,** 1211.
[5] Baltimore, D, RNA-dependent DNA polymerase in virions of RNA tumour viruses, *Nature*, 1970, **226,** 1209.
[6] Lawn, R M, *et al*, The isolation and characterisation of linked delta- and beta-globin genes from a cloned library of human DNA, *Cell*, 1978, **15,** 1157.
[7] Little, P, *et al*, Isolation and partial sequence of recombinant plasmids containing human α-, β-, and γ-globin cDNA fragments, *Nature*, 1978, **273,** 640.
[8] Wilson, J T, *et al*, Insertion of synthetic copies of human globin genes into bacterial plasmids, *Nucleic Acids Research*, 1978, **5,** 563.
[9] Leder, P, Discontinuous genes, *New England Journal of Medicine*, 1978, **298,** 1079.
[10] Flavell, R A, *et al*, Analysis of the beta-delta-globin gene loci in normal and Hb Lepore DNA: direct determination of gene linkage and intergene distance, *Cell*, 1978, **15,** 25.
[11] Mears, J G, *et al*, Organisation of human delta- and beta-globin genes in cellular DNA and the presence of intragenic inserts, *Cell*, 1978, **15,** 15.
[12] Little, P F R, *et al*, Structure of the human fetal globin gene locus, *Nature*, 1979, **278,** 227.
[13] Weatherall, D J, and Clegg, J B, Recent developments in the molecular genetics of human hemoglobin, *Cell*, 1979, **16,** 467.
[14] Mears, J G, *et al*, Changes in restricted human cellular DNA fragments containing globin gene sequences in thalassemias and related disorders, *Proceedings of the National Academy of Sciences of the United States of America*, 1978, **75,** 1222.
[15] Orkin, S H, *et al*, Partial deletion of β-globin gene DNA in certain patients with β^0-thalassemia, *Proceedings of the National Academy of Sciences of the United States of America*, 1978, **76,** 2400.
[16] Ottolenghi, S, *et al*, Globin gene deletion in HPFH $\delta^0\beta^0$-thalassaemia and Hb Lepore disease, *Nature*, 1979, **278,** 654.
[17] Orkin, S H, *et al*, Deletion of the $^A\gamma$ globin gene in $^G\gamma$-$\delta\beta$-thalassemia, *Journal of Clinical Investigation*, 1979, in press.
[18] Fritsch, E F, *et al*, Characterisation of deletions which affect the expression of fetal globin genes in man, *Nature*, 1979, **279,** 598.
[19] Orkin, S H, *et al*, The molecular basis of α-thalassemias: Frequent occurrence of dysfunctional α loci among non-Asians with Hb H disease, *Cell*, 1979, **17,** 33.
[20] Dozy, A M, *et al*, Alpha globin gene organisation in blacks precludes the severe form of alpha-thalassemia, *Clinical Research*, 1979, **27,** 507A.
[21] Higgs, D R, *et al*, Negro α-thalassaemia is caused by the deletion of a single α-globin gene, *Lancet*, 1979, in press.
[22] Orkin, S H, *et al*, Application of endonuclease mapping to the analysis and prenatal diagnosis of thalassemias caused by globin-gene deletion, *New England Journal of Medicine*, 1978, **299,** 166.
[23] Kan, Y W, and Dozy, A M, Polymorphism of DNA sequence adjacent to human β-globin structural gene: Relationship to sickle mutation, *Proceedings of the National Academy of Sciences of the United States of America*, 1978, **75,** 5631.