The genetic control of protein synthesis: The haemoglobin model

D. J. WEATHERALL

From the Department of Haematology, University of Liverpool

More is known about the structure and function of haemoglobin than practically any other protein. Furthermore the rabbit and human reticulocyte systems have been used extensively for studies of haemoglobin synthesis and most of our present understanding of protein production in man has come from these sources. Thus a general outline of human haemoglobin synthesis offers a useful introduction, not only to the first section of this symposium which deals with haemoglobin specifically, but also to the discussions of other disorders resulting from abnormalities of structure or function of proteins, which follow. This brief review will highlight only a few areas of current interest in the field. For more extensive cover the reader is referred to several recent articles (Weatherall and Clegg, 1972; Huehns, 1974; Clegg, 1974; Neinhuis and Anderson, 1974).

The Human Haemoglobins

The normal human haemoglobins (Hbs) are summarized in table I. Each of them has a similar basic structure which consists of two distinct pairs of globin chains, each chain being associated with a haem group (see Huehns, 1974). In adult life there is a major haemoglobin, Hb A ($\alpha_2\beta_2$), and a minor fraction which makes up about 2.5% of the total, Hb A₂ ($\alpha_2\delta_2$). In intrauterine life the major haemoglobin is Hb F ($\alpha_2\gamma_2$). It has been shown recently that the Hb F is a mixture of molecular species with the formulae $\alpha_2\gamma_2^{136\text{gly}}$ and $\alpha_2\gamma_2^{136\text{ala}}$ (Schroeder Huisman, Shelton, Kleihauer, Dozy, and Robberson,

1968). The γ chains containing glycine at position 136 are designated G γ chains and those which contain alanine at this position are called A γ chains. The two forms of Hb F can only be separated by detailed chemical analysis of the γ chains, and by such criteria as alkali denaturation and electrophoresis which are used in clinical laboratories to demonstrate Hb F, it behaves as a single entity. At birth the ratio of molecules containing G γ chains to those containing A γ chains is about 3:1 and this ratio changes to about 2:3 in the trace amounts of Hb F present in some normal adults (Schroeder and Huisman, 1970).

Before the eighth week of intrauterine life there are two embryonic haemoglobins called Hbs Gower 1 and 2. Hb Gower 2 has the formula $\alpha_2\epsilon_2$ but the formula of Hb Gower 1 is uncertain. It has been shown recently that there is another intrauterine haemoglobin present in small amounts throughout fetal life which is called Hb Portland 1 and has the formula $\zeta_2\gamma_2$. It is not certain whether the Hb Gower 1 has the molecular formula ϵ_4 or $\zeta_2\epsilon_2$.

During fetal development there is an orderly switch from ϵ to γ chain production followed by β and δ chain production in adult life; α chain synthesis occurs throughout all stages of development but it is still not known whether α or ϵ chains are the first products of the primitive mesenchyme. The haemoglobin changes during development have been reviewed recently (Weatherall, Pembrey, and Pritchard, 1974).

The Genetic Control of Human Haemoglobin Production

The genetic control of human haemoglobin is summarized in figure 1. There is good genetic evidence that the structure of each of the human adult globin chains and the γ chains of Hb F is directed by one or more structural genes. The β and δ chains of Hbs A and A₂ are controlled by single structural genes whereas the α -chain structural gene is duplicated, at least in some populations. Similarly there is clear genetic evidence that the A γ and G γ chains are the

Haemoglobin A	$\alpha_2\beta_2$
Haemoglobin A ₂	$\alpha_2\delta_2$
Haemoglobin F	$\alpha_2\gamma_2$
Haemoglobin Gower 2	$\alpha_2\epsilon_2$
Haemoglobin Gower 1	? ϵ_4
Haemoglobin Portland	$\zeta_2\gamma_2$

Table I The normal human haemoglobins¹

¹Haemoglobin F is a mixture of two species, $\alpha_2\gamma_2^{136\text{gly}}$ and $\alpha_2\gamma_2^{136\text{ala}}$. Both adult and fetal haemoglobins also contain additional minor components with 'blocked' N-terminal ends of the β and γ chains respectively.

Fig 1 The genetic control of haemoglobin production.

products of separate gene loci. Presumably the ϵ and ζ chains are also the products of separate loci. The α chain loci remain active throughout development and α chains combine sequentially with ϵ , ζ , $G\gamma$, $A\gamma$, β , and δ chains (for more complete reviews see Weatherall and Clegg, 1972; Huehns, 1974).

The General Pattern of Protein and Nucleic Acid Synthesis in the Red Cell Precursors

The general pattern of protein synthesis in the red cell precursors was summarized well by Borsook in 1964. In this review Borsook attempted to correlate morphological maturation with the pattern of DNA, RNA, and protein synthesis during development. Although protein synthesis is active throughout the erythroid maturation, mainly non-haemoglobin protein is produced up to the polychromatophilic normoblast stage. Haemoglobin synthesis is then fully activated and continues to the reticulocyte stage of development, after which there is no further protein synthesis in the red cell. During maturation many of the non-haemoglobin proteins appear to turn over rapidly but there is no turnover of haemoglobin during the life span of the red cell. RNA synthesis continues until the orthochromatic stage of development after which nucleic-acid synthesis ceases. Interestingly, in the fetal rabbit liver, RNA synthesis appears to stop at the basophilic erythroblast stage, ie, at approximately one stage earlier than in adult animal marrow. DNA synthesis is active in pronormoblasts and basophilic and early polychromatophilic erythroblasts; the late polychromatic and orthochromatic erythroblasts do not synthesize DNA. This phenomenon has been related to the necessity of DNA synthesis for mitosis and fits in nicely with the observation that there is no further

cell division after the early polychromatic stage of development.

The information which determines the structure of a globin chain is contained in the sequences of the base pairs which make up the DNA of the particular gene for that chain. The genetic code for proteins is a triplet code with three bases coding for each particular amino acid. It is degenerate in that there may be several triplets for any given amino acid. The information which is contained in the globin-chain gene has to be transferred into the cytoplasm to be translated into a peptide chain. This is achieved by the synthesis of a form of RNA called messenger RNA (mRNA) on the DNA template which forms the gene for that particular globin chain. Messenger RNA diffuses into the cell cytoplasm and acts as a template for protein synthesis. Since no RNA synthesis occurs after the orthochromatic normoblast stage all haemoglobin production in later normoblasts and reticulocytes must depend on the stability of preformed mRNA.

Since amino acids cannot interact directly with RNA they must be brought to the mRNA template on a series of carrier molecules. These are small-molecular-weight RNA molecules called transfer RNAs (tRNAs). Each amino acid has one or more specific tRNAs which contain a triplet of bases, the anticodon, which can interact with complementary bases of the mRNA codons. The growing peptide chains are held in place on the mRNA template by structures called ribosomes, each of which consists of two subunits.

Haemoglobin synthesis is carried out on groups of ribosomes strung along mRNA rather like a string of beads, called polysomes. The number of ribosomes per polysome varies but on average there are about five. Ribosomes have two binding sites for

tRNA, the peptide site which binds the tRNA-peptide complex, and the amino-acid site which binds the incoming tRNA residue.

Protein synthesis results from the movement of ribosomes along the mRNA strand so exposing successive codons to tRNA anticodons. As a ribosome moves from one codon to the next the aminoacyl tRNA in the amino site moves to the peptide site and the peptide chain is transferred to it from the tRNA previously in that site and the chain is thus lengthened by one amino-acid residue. The first step in this process appears to be the binding of methionyl tRNA to the peptide site at a unique initiating codon (AUG or GUG) in the mRNA. Chain synthesis then continues until a specific terminating codon is reached. In mammalian cells this is probably UAG or UAA. At this point the completed peptide chain drops off the mRNA-ribosomal complex as do the ribosomes, and the subunits of the latter are transferred to the subunit pool for further protein synthesis.

The Chromosomal Arrangement of the Human Haemoglobin Genes

There is still no definite evidence as to which chromosomes carry the human haemoglobin genes although they have been designated tentatively to chromosomes 2 and group B by DNA/RNA hybridization studies (Price, Conover, and Hirschhorn, 1972).

Both formal genetic studies and the structural

analysis of the $\delta\beta$ chain of haemoglobin Lepore has provided strong evidence that the δ and β loci are closely linked, with the δ locus proximal to the β locus. Recent studies of the structure of the variant Hb Kenya (Huisman, Wrightstone, Wilson, Schroeder, and Kendall, 1972; Smith, Clegg, Weatherall, and Gilles, 1973) indicate that its non- α chains consist of the N-terminal residues of the γ chain fused to the C-terminal residues of the β chain. A tentative model for the production of this abnormal haemoglobin variant is shown in figure 2. The fact that all affected individuals produce some Hb F which is entirely of the G γ type suggests that the Hb Kenya has arisen by unequal crossing over in which the N-terminal end of the A γ locus has become fused to the C-terminal end of the β locus with the exclusion of parts of the A γ and δ loci and the whole of the δ locus. Thus the order of genes on the chromosome appears to be G γ -A γ - δ - β . The significance of this finding in terms of gene regulation will be considered in a later section.

Human Messenger RNA

Messenger RNA (mRNA) was first identified in bacterial systems in 1961 and four years later a fraction with the properties required of mRNA was isolated from rabbit reticulocyte polysomes (Burny and Marbaix, 1965). This had a sedimentation coefficient of about 9 to 10S, the range expected for a molecule of approximately 450 nucleotides. Subsequent studies have shown that a 9S fraction from

Fig 2 The arrangement of the haemoglobin genes: evidence obtained from the structure of haemoglobin Kenya.

A Normal arrangement and alignment of haemoglobin genes.

B Misalignment during crossing over.

C The gene products due to abnormal crossing over. The chromosome containing the non- α -chain gene for haemoglobin Kenya also has a remaining G γ locus.

human reticulocytes will stimulate globin synthesis in a variety of cell-free systems (Nienhuis and Anderson, 1971; Benz and Forget, 1971). More recently it has been possible to obtain milligram quantities of globin mRNA from rabbit reticulocytes and to obtain accurate molecular-weight estimations. The molecular weight of rabbit globin mRNA is estimated at about $220\,000 \pm 10\,000$, and this exceeds by some 30% the size required to code for 450 nucleotides, ie, to act as a template for a chain of 150 amino-acid residues long. A considerable part of this additional RNA consists of 50 to 75 bases of polyadenylic-acid(poly-A) residues near the 3' end, and indeed nearly all eukaryotic mRNAs seem to have such poly-A-rich regions.

A further clue to the possible identity of the additional mRNA material has come from the studies of the human α -chain mutant Hb Constant Spring (Clegg, Weatherall, and Milner, 1971). This variant contains 31 additional residues at the C-terminal end of the α -chain and it has been suggested that it has arisen from a mutation of the chain-terminating codon UAA to UGA which codes for glutamine. Thus instead of stopping at residue 141 the message is read for a further 90 bases until another termination codon is reached, thus producing a chain with 31 extra amino-acid residues at the end. Further evidence that there are areas of mRNA at the 3' end of the α -chain mRNA which are not normally translated has come from studies of the frame-shift mutant Hb Wayne (Seid-Akhaven, Winter, and Rucknagel, 1972); these new data are summarized in figure 3. Furthermore direct base sequence analysis of some

fractions of human α -chain mRNA have shown sequences which are compatible with the additional residues attached to the C-terminal end of the Hb Constant Spring α chain (Marotta, Forget, Weissman, Verma, McCaffrey and Baltimore, 1974). From data of this type it is possible to construct a tentative model of the human α -chain mRNA as shown in figure 4. The additional genetic material can be accounted for by the extra bases at the 3' end which are not usually translated and by the poly-A-rich zones distal to this area. This leaves a small number of bases still to be accounted for and these will almost certainly be found at the 5' end; the function of these non-translated base sequences is unknown.

Fig 4 A model of human α -chain mRNA based on the data shown in figure 3.

There is now considerable evidence that mRNA is synthesized from a much larger precursor molecule which is present in all mammalian nuclei called heterogeneous nuclear RNA (Hn RNA) (Scherrer, Latham, and Darrell, 1963). Thus it seems likely that genetic information is transferred first into the large nuclear precursor mRNA which contains the mRNA sequences at its 3' end. Most of this giant nuclear RNA is not transferred to the cytoplasm and presumably the smaller mRNA fraction is cleaved and diffuses into the cytoplasm to form a template for protein synthesis. The poly-A-rich sequences appear to be added after transcription from DNA.

So far little is known about the relative stabilities

Normal sequence	Ala	-	Lys	-	Tyr	-	Arg	
Code words	GCU		AAA		UAC		CGU	
Loss of A	GCU		AAU		ACC		GU?	
New sequence	Ala	-	Asn	-	Thr	-	?	

Fig 3A Frame-shift mutation

Residue number	138	139	140	141	142	143	144	145	146	147	148		172
normal α chain	Ser	Lys	Tyr	Arg	Term								
α^{CS} chain	Ser	Lys	Tyr	Arg	Gln	Ala	Gly	Ala	Ser	Val	Ala		Glu
Normal RNA codons		AA(A)G	UAC	CGU	UAA	GCU	GGA	GCC	UCG	GUA	GC	[U,C]	
new α^{Wayne} codons		AA	UAC	CGU	UAA	GCU	GGA	GCC	UCG	GUA	G		
α^{Wayne} chain	Ser	Asn	Thr	Val	Lys	Leu	Glu	Pro	Arg	Term			

Fig 3B

Fig 3 A schematic representation of a frame-shift mutation.

B The structure of the α chains of haemoglobins Constant Spring and Wayne. This diagram shows how haemoglobin Constant Spring has arisen by a mutation of a chain-terminating codon UAA whereas haemoglobin Wayne has arisen by a frame-shift mutation in which the third base in the codon for lysine 139 is lost. If this is replaced, the sequence for the normal α -chain N-terminus and the extra haemoglobin-Constant Spring sequence is obtained.

The genetic control of protein synthesis: The haemoglobin model

of human mRNAs. However there is good evidence that human δ -chain synthesis declines more rapidly than β -chain synthesis during erythroid maturation. This suggests that δ -chain mRNA is less stable than β -chain mRNA. Interestingly the $\delta\beta$ and $\beta\delta$ chains of Hbs Lepore and Miyada show similar synthetic properties to the δ chain. This indicates that regions outside the 5' and 3' ends of the β -chain structural mRNA may be important in mRNA stability and that these have been lost in the production of the δ , $\delta\beta$, and $\beta\delta$ chains respectively (Roberts, Clegg, Weatherall, and Ohta, 1973).

The isolation of human mRNA has important applications to the study of a wide spectrum of pathological processes. For example, it has been possible to show that the thalassaemic defect can be reproduced in rabbit or other cell-free systems using 9S mRNA from β -thalassaemic reticulocytes. These findings suggest that the defect in haemoglobin production in thalassaemia is due to a reduced amount of mRNA for the particular chain involved (Neinhuis and Anderson, 1971; Benz and Forget, 1971). It has been possible to obtain direct confirmation of this prediction by making highly radioactive DNA from rabbit or human α and β mRNA templates using a viral reverse transcriptase. The DNA produced in this way (cDNA) can then be used as a probe for hybridization experiments to measure the amounts of α and β mRNA in thalassaemia samples. This approach has shown a deficiency of mRNA for the α or β chain in α or β thalassaemia respectively (Kacian, Gambino, Dow, Grossbard, Natta, Ramirez, Spiegelman, Marks and Bank, 1973; Housman, Forget, Skoultchi, and Benz, 1973). Furthermore it should now be possible to use labelled cDNA produced on human mRNA templates to measure the number of gene copies for the α or β -chain loci, and this has important applications both in the determination of the number of copies normally present and for examining the possibility that some forms of thalassaemia or similar disorders result from specific gene deletions.¹

Protein Synthesis in the Red Cell

The three main steps in protein synthesis, namely, initiation, elongation, and termination, can each be discussed separately.

CHAIN INITIATION

Globin-chain initiation requires the basic ingredients of ribosomal subunits, mRNA, a specific met-tRNA, and several initiation factors. The initiation codon

has been shown to be AUG and of the two major met-tRNA species in the red cell only one, met-tRNA_i, recognizes this codon; this is thus the initiating tRNA. In bacterial systems the methionine of this tRNA carries a formyl group but no such modified methionine residue has been demonstrated in eukaryotes. However it is customary to designate the initiator tRNA in mammals as tRNA_i to distinguish it from the other met-tRNA, met-tRNA_m, which inserts methionine into the peptide chain but is not involved in chain initiation. It has been suggested that the sequence of events in initiation is as follows:

There are at least four protein initiation factors designated M₁, M_{2a}, M_{2b}, and M₃ (this nomenclature is confusing since it differs from that used by authors for bacterial initiation systems). The precise interaction of these factors in the initiation complex is not fully understood, nor is the action of haem which also seems to be involved in chain initiation. The initiation factors are required, together with GTP, for the binding of met-tRNA_i to the 40S ribosomal subunit, and at least one factor, possibly M₃, is required for the binding of mRNA to this complex. They are also required for the continued cyclization of ribosomes during protein synthesis (for a review see Crystal, Nienhuis, Pritchard, Picciano, Elson, Merrick, Graf, Shafritz, Laycock, Last, and Anderson, 1972). Haem is required in the initiation process since it prevents the action of an inhibitor which interferes with the formation of the initiating met-tRNA_i-40S complex (Legon, Jackson, and Hurst, 1973).

Experiments using 9S mRNA from different sources in heterologous cell-free systems have suggested that many of the initiation factors are common to several species. However there is evidence that there are at least some message-specific initiation factors. Furthermore, it appears that in some systems for mRNA, α and β chains are initiated with different degrees of efficiency, an important observation when such systems are used for the study of the defect in thalassaemia.

CHAIN ELONGATION

After the completion of the initiation complex the system is able to proceed with stepwise addition of the amino acids from the N-terminal end to form a growing peptide chain. This occurs in several distinct stages. The anticodon of the next amino acyl tRNA binds to the ribosomal site next to that occu-

¹ This is indeed the case. Infants with the haemoglobin Barts hydrops syndrome have been shown to have no γ -chain genes by this technique (Ottolenghi, Lanyon, Paul, Weatherall, Clegg, Pritchard, Pootrakul and Wong, 1974, submitted for publication).

pied by the initiator or peptidyl tRNA. Peptide transfer then takes place between the newly bound amino-acyl tRNA and the met-tRNA. Finally the newly made peptidyl tRNA and mRNA are both shifted from the acceptor to the donor site on the ribosomes. In this way the acceptor site is freed and the process can be repeated by stepwise addition of amino acids, so resulting in elongation of the growing chain. At least two protein elongation factors have been implicated in this process, as well as the peptide-bond-forming enzyme peptidyl transferase.

Experiments in both rabbit and human reticulocytes have indicated that both α and β -chain synthesis proceeds in a linear fashion at a uniform rate from N to C-terminal ends. The time for β -chain translation in human cells is approximately 90 seconds at 37°C (Clegg, Weatherall, Na-Nakorn, and Wasi, 1968).

CHAIN TERMINATION

At least two chain-terminating codons, UAA and UAG, have been identified in prokaryotes. Studies determining the nucleotide sequence at the end of the coat-protein cistron of the R17 bacteriophage have identified two termination codons next to each other in the order of 5' UAA UAG. This suggests that natural messenger RNAs may have a kind of 'fail-safe' mechanism in case termination by one codon is not fully effective. Recent studies on the human α -chain mutant, Hb Constant Spring, indicate that the human α -mRNA has only one termination codon, UAA (Clegg *et al*, 1971b). Furthermore the findings in this variant and in the Hb Wayne frame-shift mutant (fig 3), indicate that both UAA and UAG can act as terminators for human globin synthesis. In addition, in prokaryotes at least, termination requires the presence of at least two protein release factors, RI and RII, and polypeptidyl transferase. Little is known about the role of factors of this type in chain termination in higher organisms although at least one release factor appears to be necessary as well as GTP. The most likely sequence of events is that a polypeptidyl transferase catalyses the transfer of the peptidyl group of peptidyl tRNA to water rather than aminoacyl tRNA, in the presence of R factor. The peptide chain is thus released from the ribosomes and the tRNA which is left on

the ribosome is then released by the action of another factor. The ribosomal subunits are available for further protein synthesis.

Haem Synthesis and the Assembly of Haemoglobin

The biosynthesis of haem takes place in a series of enzyme-controlled steps (see Granick and Levere, 1964) beginning with the condensation of glycine and succinyl CoA to form α -amino- β -keto-adipic acid which is then rapidly decarboxylated to form δ -amino laevulinic acid (δ ALA). This reaction, which is catalyzed by the enzyme ALA synthetase, and which requires pyridoxal phosphate and ferrous iron as cofactors, is the only step in the pathway which requires energy. All the rest are essentially irreversible. Two molecules of ALA combine to form porphobilinogen under the action of ALA dehydrase and then four of the porphobilinogen molecules combine to form the basic tetrapyrrole ring, uroporphyrinogen. After various modifications this is converted to protoporphyrinogen which is oxidized to protoporphyrin. Iron is then inserted by the enzyme haem synthetase to form haem. The first and last steps on the pathway are mitochondrial while the rest occur in the cell cytoplasm.

The precise details of the assembly of the finished haemoglobin tetramer have yet to be worked out (see Huehns, 1974). Evidence from both rabbit and human reticulocyte experiments suggests that there is a slight degree of chain imbalance in normal cells with the formation of a small pool of α chains. It seems likely that the $\alpha\beta$ dimer is an intermediate in haemoglobin synthesis and the likely sequence of events is that $\alpha\beta$ globin combines with haem and then immediately combines with a like subunit to form the stable tetramer. It is not certain whether haem combines with globin while still on the ribosomes although this seems likely since it is probable that the chain assumes its complicated folded configuration as it is assembled and that haem fits into the haem pocket during this process.

The Regulation of Haemoglobin Synthesis

Although there is now general agreement about the main steps involved in the synthesis of haemoglobin very little is known about how this process is controlled. The erythrocyte is a highly specialized cell

Step No.	Synthesis Procedure
1	Induction of structural genes for haem and globin in committed precursor
2	'Switching on and off' of globin genes during development
3	Repression of genes other than those for Hb synthesis during erythroid maturation
4	Synchronization of α and β -chain production and haem and globin synthesis
5	Adjustment of δ -chain production at 1/40 of β -chain production
6	'Turning off' of haemoglobin synthesis
7	Control of non-haem protein, RNA and DNA synthesis, and synchronization to cell maturity

Table II Control steps required in haemoglobin synthesis

whose differentiation is directed towards the production of large amounts of a single protein. Control is required at all stages of differentiation (table II). DNA and RNA synthesis must be geared to the various stages of erythroid-precursor maturation. Control mechanisms must be involved in the turning off of these genes and in activating the haemoglobin genes half way through maturation. During the later steps of maturation protein synthesis is almost totally directed towards the production of haemoglobin and all other genes must be repressed. During this time, in adult life, haemoglobins A and A₂ are being produced at markedly different rates and clearly some form of control must maintain this ratio at a constant level. All the other areas of DNA in the red cell precursors, which are not involved in the production of red-cell proteins, must be suppressed throughout maturation.

Some form of control mechanism must be involved in the orderly changeover of globin chains which occurs during differentiation, ie, the switch between embryonic, fetal, and adult globin chains. Throughout maturation there is fairly good coordination between α and non- α -chain production, and, since the genes controlling the α and β chains are at some distance apart, some form of coordinating system must be invoked to maintain these gene products at roughly equal levels. Similarly, since there is no evidence for a large excess of haem or globin at any stage of maturation, mechanisms must exist which coordinate these completely different metabolic pathways to produce similar amounts of end product. Finally some form of relatively crude control is required to stop haemoglobin synthesis when the cellular haemoglobin content reaches a critical level.

Quite clearly control of protein synthesis in the red cell can occur either at gene level, ie, such processes as gene duplication or amplification; at transcriptional level, ie, the rate of RNA production; or during the posttranscriptional processes of RNA processing, transport, or stability and the rate of chain initiation, elongation, and termination (table III). In addition some form of coarse control might occur at the later stages of protein synthesis such as

subunit interaction, haem binding, and the stability of the finished molecule.

REGULATION AT GENE LEVEL

As already mentioned, it seems likely that the human haemoglobin chains are coded for by as many as 12 genes (4 α , 2 β , 2 δ , and 4 γ) in addition to those responsible for the embryonic haemoglobins. Both the α and γ -chain genes appear to have duplicated quite recently in evolutionary time. It is possible that these duplications have some selective value since their presence reduces the impact of serious mutations. Since the red cell is highly specialized for making large amounts of one type of protein, it seemed possible that, to achieve this end, there might be the specific amplification of appropriate pieces of genetic information such as occurs in the amphibian ribosomal-RNA genes. Attempts to demonstrate such amplification in nucleated red cells or in other sources of human DNA have been unsuccessful. Thus DNA/cDNA hybridization experiments using cDNA made from globin mRNA have suggested that only a few copies of the globin genes are present in mouse or human DNA and have provided no evidence or amplification during protein synthesis.

The subject of transcriptional regulation (fig 5) is one of intense interest at the moment (Paul, 1972; Neinhuis and Anderson, 1974). The bacterial chromosome exists in the cytoplasm as metabolically active DNA with only small amounts of basic protein, but in eukaryotes the DNA is associated with

Fig 5 A model of the synthesis of the large molecular weight nuclear RNA (Hn RNA) and its cleavage to 9S mRNA.

The Hn RNA is synthesized by an enzyme RNA polymerase and the repeating adenine bases (A) are added at the 3' end. Before transportation to the cytoplasm a large portion at the 5' end of the Hn RNA is cleaved and the functional mRNA is transported by an unknown mechanism into the cytoplasm, where it acts as a template for protein synthesis.

1	Gene level
	Duplication
	Amplification
2	Transcriptional
	mRNA synthesis
	Specificity
	Timing
	Quantity etc
3	Posttranscriptional
	mRNA processing, transport, stability; chain initiation, elongation, release; subunit affinity; haem interaction; stability

Table III Regulation of protein synthesis

large amounts of protein and is restricted in its ability to direct RNA synthesis. Two classes of protein have been found in chromatin, the histones and non-histone acidic proteins. The observation that there has been conservation of amino-acid composition of histones during evolution, and the fact that there are limited numbers present in various tissues, has suggested that these molecules are an integral part of chromatin, and impose non-specific restriction on gene expression. It seems unlikely that they have the unique properties necessary for controlling expression in differentiated cells. There are certain exceptions to this observation, however, and a unique histone has been found in immature chick red-cell precursors, which increases in amount during maturation, suggesting that it has a role in the condensation of the cell nucleus. The acidic proteins on the other hand are a highly heterogeneous group which includes many of the enzymes involved in DNA and RNA metabolism. They turn over rapidly and it is possible that some of them may represent specific repressor proteins. It has been shown recently that globin-mRNA sequences can be transcribed from erythroid chromatin using RNA polymerase from *E. coli*. Gilmour (1973) has been able to transfer the ability to direct the transcription of the globin mRNA sequence to brain chromatin using the acidic protein fraction from mouse embryonic liver erythroblasts. These experiments suggest that these molecules may play a key role in the control of structural gene transcription.

It has been shown that only a small fraction of eukaryotic DNA is required for directing the structure of protein and there is a mass of DNA which has no obvious function. Much of this consists of repetitive sequences and it has been suggested that the repetitious DNA plays a regulatory role in the control of protein synthesis in higher organisms. Based on these observations Crick, Britten, Davidson, and others have evolved models involving the interaction of various batteries of control genes which are designed to explain such complex activities as differentiation in higher organisms (Britten and Davidson, 1969; Crick, 1971). The most fully worked out model is that of Britten and Davidson (1969). In brief their model requires the interaction of various sensor and activator RNAs. A sensor DNA locus is able to recognize specific inducer molecules (hormones for example) and initiates transcription on an adjacent integrator gene after receipt of an appropriate stimulus. An activator RNA is produced at the integrator-gene loci and complexes with DNA at a distant receptor gene allowing transcription of a series of producer genes coding for several proteins with interrelated metabolic functions. This system would explain, for example, how hormones may

produce a variety of gene products and would form a basis for the complex processes of multitissue differentiation. It is relatively easy to apply this model to systems such as red-cell differentiation where there are many genes involved in the production of haemoglobins, enzymes, membrane proteins, etc, at different stages of development. One of the requirements of this model is the existence of a class of nuclear RNA located in chromatin and bound to DNA in a sequence-specific manner. Mayfield and Bonner (1972) have provided evidence that such a class exists. Thus after partial hepatectomy they found that the first response was the synthesis of a rapidly labelled high-molecular-weight RNA with novel sequences. This event was followed by increased amounts of chromosomal or low molecular weight RNA, again with sequences not usually produced by liver. Subsequently mRNA synthesis for liver protein increased before initiation of DNA synthesis occurred. This sequence of events is exactly as predicted by Britten and Davidson (1969). Thus Mayfield and Bonner suggested that the chromosomal RNAs may be gene activators.

It has been tempting to apply also some of the models of control of protein synthesis derived from bacteria to the human haemoglobin genes. The operon model of Jacob and Monod (1961) was applied by several workers once it was known that the human γ , β , and δ -chain genes were closely linked. The operator model requires that there are a series of linked structural loci controlled by a common operator locus, and that repressor substances are coded for by an adjacent regulator gene. The repressor acts by binding to the operator locus, therefore preventing transcription of the operon-specific polygenic RNA. For each repressor/operator system there exists an effector molecule which interacts with the repressor affecting its affinity for the operator. Thus the repressors regulate the transcription of all the cistrons in the operon. There is also evidence to suggest that some microbial systems are controlled in a positive manner, and more recently it has been shown that the initiation specificity of RNA polymerase can be influenced by complexing with various protein factors (sigma factor for example). Unfortunately virtually nothing is known about comparable processes which occur during transcription of DNA in eukaryotes and there is certainly no evidence for a repressor mechanism of gene regulation in higher organisms. These problems have been discussed at length by Stent (1964) and Elston and Glassman (1967) and they feel that current evidence for a eukaryote system suggests a lack of negative control.

Despite these observations, however, there is increasing evidence that there are areas of the chromosomes carrying the γ , β , and δ -chain genes which are

involved in the neonatal suppression or activation of the γ and β -chain genes. Thus where areas of the $A\gamma$, δ , and β genes have been deleted in the production of Hb Kenya, there is persistent activity of the remaining $G\gamma$ locus (Huisman *et al*, 1972; Smith *et al*, 1973). This suggests that areas of DNA have been removed in the deletion which are involved normally in the neonatal suppression of this locus. A similar argument can be used to explain the findings in many forms of hereditary persistence of haemoglobin F in man (Huisman, Schroeder, Efremov, Duma, Mladenovski, Hyman, Rachmilewitz, Bouver, Miller, Brodie, Shelton, Shelton, and Apell, 1974).

POSTTRANSCRIPTIONAL CONTROL

As already mentioned, the immediate product of DNA transcription in eukaryotes is a large precursor mRNA molecule, Hn RNA. There is no doubt that the processing of this molecule with the ultimate production of a small-molecular-weight mRNA must require a series of enzymes and this process could certainly be rate limiting for protein production. Furthermore varying stability of mRNA during erythroid maturation will play some part in determining the final amount of any globin chain in the mature red cell. A clue to this mechanism has come from the study of human haemoglobins A and A₂ as mentioned earlier. Thus Hb-A₂ synthesis is completed by the late normoblast stage of maturation while the Hb-A synthesis continues to the reticulocyte stage. This observation suggests that δ -chain message is less stable than β -chain message. Furthermore it has been shown that the synthesis of the $\delta\beta$ and $\beta\delta$ fusion genes of haemoglobins Lepore and Miyada have a similar instability, as compared with the β chain of haemoglobin A. It is tempting, therefore, to speculate that some feature of the structure of the 5' and 3' ends of the α and β mRNA enhances their stability, and that this is lacking in δ mRNA.

TRANSLATIONAL CONTROL

It is obvious that some form of control could exist at the steps of initiation, translation, and termination of protein synthesis. Similarly the rate of folding and subunit association may be rate limiting and there is evidence that the availability of haem may be more rate-limiting for one globin chain than for the other. The numbers of α and β chains made during maturation are almost similar although there is a slight excess of α chains. Exhaustive studies on the processes of chain initiation and elongation have led to the conclusion that very little control occurs at the cytoplasmic level and that the ratio of α and β chains produced in rabbit reticulocytes is determined largely by the level of mRNA synthesis in early precursors (Hunter and Jackson, 1972).

The relative number of ribosomes making α and β chains are approximately equal while the average size of the α -chain polysomes is somewhat smaller than those making β chains. This would indicate that there is slightly more α -chain mRNA being utilized than β mRNA. This situation occurs in both rabbit and human reticulocytes (Hunt, Hunter, and Munro, 1968; Clegg, Weatherall, and Eunson, 1971) and further studies by Lodish and Jacobson (1972) have shown that the difference in polysome sizes is due entirely to the faster rate of initiation of β chains. All these experiments suggest that there is a slight excess of α mRNA which leads to an overproduction of α chains in the order of 10 to 20%; presumably most of these excess chains are destroyed by proteolysis.

Studies of the relative rates of α and β chain synthesis in different cell-free systems suggests that there may be some form of positive control of the initiation rates for α and β chains since there seems to be a marked difference in the relative rates of initiation between heterologous systems. Furthermore haem seems to be involved in the control of initiation, probably by interacting with a specific inhibitor of protein synthesis. There is some evidence that haem deficiency affects α chain synthesis more than that of β chains.

As mentioned earlier, there is a slight excess of α chains in human red cell precursors. After association of α and β chains to produce an $\alpha\beta$ dimer the stable tetramer forms rapidly. It is possible that some control occurs at this level due to unequal affinity of various haemoglobin subunits for each other. Thus, for example, when there is a deficiency of α chains as in α thalassaemia, Hb A seems to be produced at the expense of mutant β chain variants such as Hbs S or E. How much differential subunit affinity can account for the levels of different haemoglobins during development is uncertain, but this seems unlikely to be an important control mechanism in normal haemoglobin production.

The Control of Haemoglobin Production during Maturation

If knowledge about the control mechanisms of normal adult haemoglobin production is scanty, there is virtually no information whatever as to how the normal processes of switching from embryonic to fetal and fetal to adult haemoglobin production are controlled during maturation. This problem has been the subject of several recent reviews (Lorkin, 1972; Weatherall *et al*, 1974).

It is now well established that the fetal-to-adult-haemoglobin switch is related to the degree of maturation of the fetus and is not related to the process of birth itself. It has been shown that Hb F is the main

intrauterine haemoglobin from early in embryonic life but recent studies have provided clear evidence that β chain synthesis is activated by about eight weeks, and haemoglobin A makes up about 10% of the total haemoglobin up to about 36 weeks, after which its synthesis suddenly increases as Hb F synthesis declines.

Recent studies by Wood and Weatherall (1973) have shown that the switch from Hb F to Hb A is completely synchronous throughout all sites of fetal erythropoiesis. The type of control required to achieve this remarkable degree of synchronization could be based on the models suggested by Britten and Davidson as suggested in an earlier section. Since the level of Hb F rises in some normal women during pregnancy (Pembrey, Weatherall, and Clegg, 1973), it is tempting to invoke some form of hormonal control for the switch mechanism but the nature of this is completely obscure. As already mentioned there is strong evidence that areas of DNA of the human haemoglobin-gene complex exist which are involved specifically in the repression of γ chain production in adult life; when these are deleted the normal neonatal suppression of γ chain synthesis fails. Thus we have a series of tantalizing clues to the general nature of the Hb F to Hb A switch but no idea whatever of the precise mechanisms involved.

One of the major difficulties in understanding the switch mechanisms is the observation that in most conditions in which Hb F production continues into adult life it is heterogeneously distributed within the red cells. Indeed it is only in some forms of hereditary persistence of fetal haemoglobin that the Hb F is uniformly distributed. The heterogeneity of Hb F synthesis is difficult to ascribe to a primary genetic mechanism and obviously there must be local tissue factors involved in fixing the final level of Hb F and Hb A in specific precursor cell lines. The cause of this phenomenon is completely obscure.

Summary

The major steps in haemoglobin synthesis in red cell precursors are now worked out and only details of specific mechanisms remain to be filled in. Thus the major steps in the production of a globin chain are the transcription of large-molecular-weight precursor mRNA (Hn RNA) from the appropriate gene, the cleavage of Hn RNA to produce definitive template mRNA which diffuses into the cell cytoplasm, the processes of chain initiation, translation, and termination, and finally the association of subunits to form a stable tetramer. From what little information there is it appears that this complex series of events is controlled at several levels but that the major control mechanisms are mediated in the processes of

transcription rather than translation. There is increasing evidence that the chromatin proteins, particularly the acidic proteins, have specific functions in maintaining areas of DNA repressed and that activation of the haemoglobin loci results from complex interactions with external inducers, the nature of which is not yet known. Virtually nothing is known about the factors involved in the switch from the intrauterine to adult haemoglobins although this appears to be a coordinated event throughout the erythropoietic tissue of the fetus.

The isolation and characterization of human mRNA has made possible both the study of the function of thalassaemic mRNA in heterologous systems and, more recently, its quantitation in abnormal red-cell precursors. Furthermore the ability to make cDNA using viral reverse transcriptase has made possible the estimation of the number of haemoglobin genes, both in normal human subjects and in those with different forms of thalassaemia and other genetic disorders of haemoglobin production.

References

- Benz, E. J., Jr., and Forget, B. G. (1971). Defect in messenger RNA for human hemoglobin synthesis in beta thalassemia. *J. clin. Invest.*, **50**, 2755-2760.
- Borsook, H. (1964). DNA, RNA and protein synthesis after acute severe blood loss: a picture of erythropoiesis at the combined morphological and molecular levels. *Ann. N.Y. Acad. Sci.*, **119**, 523-539.
- Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science*, **165**, 349-357.
- Burny, A., and Marbaix, G. (1965). Isolement de RNA messenger des réticulocytes de lapin. *Biochem. biophys. Acta*, **103**, 409-417.
- Clegg, J. B. (1974). Haemoglobin synthesis. *Clin. Haemat.*, **3**, in press.
- Clegg, J. B., Weatherall, D. J., and Eunsou, C. E. (1971a). The distribution of nascent globin chains on human reticulocyte polysomes. *Biochem. biophys. Acta. (Amst.)*, **247**, 109-112.
- Clegg, J. B., Weatherall, D. J., and Milner, P. F. (1971b). Haemoglobin Constant Spring—a chain termination mutant? *Nature (Lond.)*, **234**, 337-340.
- Clegg, J. B., Weatherall, D. J., Na Narkon, S., and Wasi, P. (1968). Haemoglobin synthesis in β -thalassaemic. *Nature (Lond.)*, **220**, 664.
- Crick, F. H. C. (1971). General model for the chromosomes of higher organisms. *Nature (Lond.)*, **234**, 25-27.
- Crystal, R. G., Nienhuis, A. W., Pritchard, P. M., Picciano, D., Elson, N. A., Merrick, W. C., Graf, H., Shafritz, D. A., Laycock, D. G., Last, J. A., and Anderson, W. F. (1972). Initiation of globin synthesis. *FEBS Letters*, **24**, 310-314.
- Elston, R. C., and Glassman, E. (1967). An approach to the problem of whether clustering of functionally related genes occur in higher organisms. *Genet. Res.*, **9**, 141-147.
- Gilmour, S. (1973). Transcription of the globin gene. (Abstr.) In *International Colloquium on Normal and Pathological Protein Synthesis in Higher Organisms*.
- Granick, S., and Levere, R. D. (1964). Heme synthesis in erythroid cells. *Progr. Hemat.*, **4**, 1-47.
- Housman, D., Forget, B. G., Skoultschi, A., and Benz, E. J. (1973). Quantitative deficiency of chain-specific globin messenger ribonucleic acids in the thalassemia syndromes. *Proc. nat. Acad. Sci. (Wash.)*, **70**, 1809-1813.
- Huehns, E. R. (1974). Haemoglobin: the haemoglobinopathies. In *Blood and Its Disorders*, edited by R. M. Hardisty and D. J. Weatherall. Blackwell, Oxford. In press.
- Huisman, T. H. J., Schroeder, W. A., Efremov, G. D., Duma, H., Mladenovski, B., Hyman, C. B., Rachmilewitz, E. A., Bouver, N., Miller, A., Brodie, A., Shelton, J. R., Shelton, J. B., and Apell, G. (1974). The present status of the heterogeneity of

- fetal hemoglobin in β -thalassemia: an attempt to unify some observations in thalassemia and related conditions. *Ann. N.Y. Acad. Sci.*, **232**, 107-124.
- Huisman, T. H. J., Wrightstone, R. N., Wilson, J. B., Schroeder, W. A., and Kendall, A. G. (1972). Haemoglobin Kenya, the product of fusion of γ and β polypeptide chains. *Arch. biochem.*, **153**, 850-853.
- Hunt, R. T., Hunter, A. R., and Munro, A. J. (1968). Control of haemoglobin synthesis: a difference in the size of the polysomes making α and β chains. *Nature (Lond.)*, **220**, 481-483.
- Hunter, A. R., and Jackson, R. J. (1972). Control of haemoglobin synthesis: coordination of α and β -chain synthesis. In *Synthesis, Structure and Function of Hemoglobin* (Hämatologie und Bluttransfusion, Bd. 10), edited by H. Martin and L. Nowicki, p. 95. Lehmann, Munich.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. molec. Biol.*, **3**, 318-356.
- Kacian, D. L., Gambino, R., Dow, L. W., Grossbard, E., Natta, C., Ramirez, F., Spiegelman, S., Marks, P. A., and Bank, A. (1973). Decreased globin messenger RNA in thalassemia detected by molecular hybridisation. *Proc. nat. Acad. Sci. (Wash.)*, **70**, 1886-1890.
- Legon, S., Jackson, R. T., and Hunt, T. (1973). Control of protein synthesis in reticulocyte lysates by haemin. *Nature [new Biol.]*, **241**, 150-152.
- Lodish, H. F., and Jacobsen, M. (1972). Regulation of hemoglobin synthesis; equal rates of translation and termination of α and β -globin chains. *J. biol. Chem.*, **247**, 3622-3629.
- Lorkin, P. A. (1973). Fetal and embryonic haemoglobins. *J. med. Genet.*, **10**, 50-64.
- Marotta, C. A., Forget, B. G., Weissman, S. M., Verma, I. M., McCaffrey, R. P., and Baltimore, D. (1974). Nucleotide sequences of human globin messenger RNA. *Proc. nat. Acad. Sci. (Wash.)*, in press.
- Mayfield, J. E., and Bonner, J. (1972). A partial sequence of nuclear events in regenerating rat liver (gene regulation chromosomal RNA). *Proc. nat. Acad. Sci. (Wash.)*, **69**, 7-10.
- Nienhuis, A. W., and Anderson, W. F. (1971). Isolation and translation of hemoglobin messenger RNA from thalassemia, sickle cell anemia and normal human reticulocytes. *J. clin. Invest.*, **50**, 2458-2460.
- Nienhuis, A. W., and Anderson, W. F. (1974). The molecular defect in thalassaemia. *Clin. Haemat.*, **3**, in press.
- Paul, J. (1972). General theory of chromosome structure and gene activation in eukaryotes. *Nature (Lond.)*, **238**, 444-446.
- Pembrey, M. E., Weatherall, D. J., and Clegg, J. B. (1973). Maternal synthesis of haemoglobin F in pregnancy. *Lancet*, **1**, 1350-1355.
- Price, P. M., Conover, J. H., and Hirschhorn, K. (1972). Chromosomal localization of human haemoglobin structural genes. *Nature (Lond.)*, **237**, 340-342.
- Roberts, A. V., Clegg, J. B., Weatherall, D. J., and Ohta, Y. (1973). Synthesis *in vitro* of Anti-Lepore haemoglobin. *Nature [new Biol.]*, **245**, 23-24.
- Scherrer, K., Latham, H., and Darnell, J. E. (1963). Demonstration of an unstable RNA and of a precursor to ribosomal RNA in Hela cells. *Proc. nat. Acad. Sci. (Wash.)*, **49**, 240-248.
- Schroeder, W. A., and Huisman, T. H. J. (1970). Nonallelic structural genes and hemoglobin synthesis. In *Proceedings of the XII International Congress on Haematology. Plenary Sessions*, p. 26. Lehmann, Munich.
- Schroeder, W. A., Huisman, T. H. J., Shelton, J. R., Shelton, J. B., Kleihauer, E. F., Dozy, A. M., and Robberson, B. (1968). Evidence for multiple structural genes for the γ -chain of human fetal hemoglobin. *Proc. nat. Acad. Sci. (Wash.)*, **60**, 537-544.
- Seid-Akhaven, M., Winter, W. P., and Rucknagel, D. L. (1972). Hemoglobin Wayne: a frameshift variant occurring in two distinct forms. (Abstr.) *Blood*, **40**, 927.
- Smith, D., Clegg, J. B., Weatherall, D. J., and Gilles, H. M. (1973). Hereditary persistence of fetal haemoglobin associated with a $\gamma\beta$ -fusion variant haemoglobin Kenya. *Nature [new Biol.]*, **246**, 184-186.
- Stent, G. S. (1964). The operon: on its third anniversary: modulation of transfer RNA species can provide a workable model of an operator-less operon. *Science*, **144**, 816-820.
- Weatherall, D. J., and Clegg, J. B. (1972). *The Thalassaemia Syndromes* 2nd ed. Blackwell, Oxford.
- Weatherall, D. J., Pembrey, M. E., and Pritchard, J. (1974). Fetal haemoglobin. *Clin. Haematol.*, **3**, in press.
- Wood, W. G., and Weatherall, D. J. (1973). Haemoglobin synthesis during human foetal development. *Nature (Lond.)*, **244**, 162-165.