The Structure of the Rabbit $\beta$-Globin Gene

RICHARD A. FLAVELL and CEES WAALWIJK

Section for Medical Enzymology and Molecular Biology, Laboratory of Biochemistry, University of Amsterdam, P.O. Box 60.000, 1005 GA Amsterdam, The Netherlands

and ALEC J. JEFFREYS

Department of Genetics, Adrian Building, University of Leicester, University Road, Leicester LE1 7RH, U.K.

Although we have learned much about the structure of eukaryotic genes present in multiple copies per genome, the complexity of eukaryotic genomes has so far impeded the analysis of single-copy genes. However, restriction-enzyme methodology, recombinant-DNA technology and highly specific hybridization techniques have recently made these genes accessible to study and have revealed an unexpected pattern of gene organization.

We have studied the rabbit $\beta$-globin gene using an approach devised by Southern (1975). Rabbit DNA is digested with a given restriction endonuclease; this generates about $10^6$ different DNA fragments. The digested DNA is electrophoresed on an agarose gel and transferred on to a nitrocellulose filter (Southern, 1975). The globin-DNA-containing fragments are then detected by hybridization with a $^{32}$P-labelled cloned $\beta$-globin complementary DNA plasmid (plasmid $P\beta G$1 DNA; Maniatis et al., 1976). A full description of the method is given in Jeffreys & Flavell (1977a,b).

A discontinuous $\beta$-globin gene in rabbits

Using this approach we have constructed a physical map of restriction-enzyme cleavage sites in and around the rabbit $\beta$-globin gene directly in chromosomal DNA (Fig. 1). Detailed analysis of this physical map has shown that the globin gene does not exist as a contiguous stretch of DNA on the chromosome, but is instead present in at least two segments that are interrupted by a 600-base-pair non-$\beta$-globin DNA sequence. The interruption occurs somewhere within the sequence coding for amino acids 101–120 of the 146-amino acid $\beta$-globin chain. This conclusion is based on the following.

(a) Restriction endonucleases BamHI and EcoRI cleave the $\beta$-globin complementary DNA (the DNA copy of $\beta$-globin mRNA) sequence at positions coding for amino acid residues 99–100 and 121–122 respectively to give a distance between the two sites of 67 nucleotides. On chromosomal DNA, however, the same endonucleases BamHI and EcoRI sites map about 600–700 nucleotides apart (Jeffreys & Flavell, 1977b).

(b) The $\beta$-globin complementary DNA contains two endonuclease HaeIII cleavage sites 333 base-pairs apart at positions coding for amino acid residues 27 and 138 (Efstratiadis et al., 1977). Endonuclease-HaeIII cleavage of rabbit liver DNA gives instead a $\beta$-globin DNA fragment of about 900 base-pairs containing these same $\beta$-globin-coding DNA sequences plus a 600-base-pair insert (Jeffreys & Flavell, 1977b).

(c) Endonuclease TaqYI cleaves the $\beta$-globin complementary DNA at a position coding for amino acid residues 42–43, 237 base-pairs from the intragenic endonuclease EcoRI site. In rabbit liver DNA, however, this endonuclease-TaqYI site maps 800 nucleotides from this endonuclease EcoRI site (C. Waalwijk & R. A. Flavell, unpublished work).

(d) The $\beta$-globin gene insert contains an endonuclease-HapII cleavage site about 400 nucleotides to the 3'-side of the endonuclease-BamHI site. No other endonuclease-HapII cleavage sites exist in the map shown in Fig. 1, nor are any found in $\beta$-globin complementary DNA.

A 600-base-pair intervening sequence has been found in the $\beta$-globin gene from various other rabbit tissues including sperm and erythroid cells (Jeffreys & Flavell, 1977b). The discontinuous structure of the $\beta$-globin gene seems therefore to be a
Fig. 1. Physical map of cleavage sites in and around the rabbit β-globin gene

The physical map of the β-globin gene in the rabbit genome is compared with the corresponding map of cleavage sites in the β-globin complementary DNA insert in plasmid PβG1 DNA taken from Efstratiadis et al. (1977). Cleavage sites are shown for endonucleases BamHI (B), BglII (Bg), EcoRI (E), HaeIII (H), KpnI (K), PstI (P), TaqYI (T) and HapII (Hap) [from Jeffreys & Flavell (1977a)]. [The localization of endonuclease-TaqYI and -HapII sites are from C. Waalwijk & R. A. Flavell (unpublished work).]

Incomplete cleavage of the β-globin gene insert by endonuclease HapII: preferential under-methylation of a CpG doublet?

Bird & Southern (1978) have used restriction endonucleases to study methylation of eukaryotic DNA. DNA methylation, which occurs almost exclusively at the cytidine residues of CpG doublets, prevents cleavage by certain restriction endonucleases, e.g. HpaII and HhaI. Although Bird & Southern (1978) found that the majority of endonuclease-HhaI and -HpaII sites in Xenopus ribosomal DNA were blocked by methylation, they found evidence for an average of 30–60% cleavage per ribosomal DNA repeat at a single endonuclease-HhaI site in the 28S cistron; apparently certain CpG sequences in Xenopus ribosomal DNA are preferentially under-methylated.

The cleavage of the β-globin gene insert by endonuclease HapII (C-C-G-G, an isoschizomer of endonuclease HpaII) suggests that preferential under-methylation of a CpG doublet also occurs at this site. However, only 50% of the β-globin gene inserts are cut by endonuclease HapII in rabbit liver DNA. This is not the result of partial digestion by endonuclease HapII since the pattern is unaltered when the enzyme concentration is increased to 8 times that required to give a complete digest (as monitored by digestion of an internal bacteriophage-λ DNA marker). These experiments show that this 50% digestion pattern is found with DNA from individuals of two strains of inbred rabbits (Alaska and Vienna White) and in several individual F₁ hybrids of these strains. It seems unlikely therefore that this pattern is the result of a
polymorphism at the endonuclease-\textit{Hap11} site in the \(\beta\)-globin gene insert. We favour the alternative explanation that the endonuclease-\textit{Hap11} site is only methylated in 50\% of the chromosomes in rabbit liver.

As yet we do not know whether this reflects methylation heterogeneity among various liver cell types or whether all cells methylate this endonuclease-\textit{Hap11} site selectively on only one chromosome of the relevant chromosome pair. Nevertheless, the similarity between the under-methylation in the globin and ribosomal DNA systems is intriguing.

\textit{Repeated sequences in the neighbourhood of the rabbit \(\beta\)-globin gene}

The continuing interest in repetitive DNA stems to a great extent from the idea that this DNA fulfils a regulatory role in gene expression (Britten & Davidson, 1969). Close association of repetitive DNA with structural genes has been shown in sea-urchin DNA (Davidson \textit{et al.}, 1975) and with the duck globin genes (Bishop & Freeman, 1974).

To determine the linkage of repeated DNA to the rabbit \(\beta\)-globin gene (Flavell \textit{et al.}, 1978), rabbit DNA was cleaved with a restriction endonuclease, denatured and annealed to a given \(C_{ot}\) value. The DNA was fractionated on hydroxyapatite to separate completely single-stranded DNA from DNA containing duplex regions. If the \(\beta\)-globin DNA fragment contains a sequence repeated elsewhere in the genome, then it will appear in the duplex fraction after annealing to a low \(C_{ot}\) value. The \(\beta\)-globin DNA fragments were detected by denaturation of the DNA from the respective fractions from the hydroxyapatite columns and electrophoresis in an agarose gel, followed by our standard hybridization analysis (Jeffreys & Flavell, 1977a). Comparison of reassociation rates with that of a single gene copy of a 1500-base-pair fragment of plasmid PMB9 digested with endonuclease \textit{AvaI} enabled us to determine the repetition frequency of the regions in and around the \(\beta\)-globin gene. The \(C_{ot}\) values calculated for the various DNA digests are shown in Table 1. The data suggest the presence of two types of repetitive DNA close to the rabbit \(\beta\)-globin gene. A 6000-fold repeated sequence is found in a region from 700–2500 base-pairs in the 3'-direction from the 3'-terminus of the gene. The sequences immediately bordering on the gene, however, appear to be

\begin{table}
\caption{Repetition frequency of the DNA segments around the rabbit \(\beta\)-globin gene}
\begin{tabular}{|c|c|c|c|}
\hline
10\(^{-3}\) \times \text{DNA region map position (bases) in no. of bases contained in parentheses} & \text{Endonuclease used to produce DNA fragment} & \text{Region of \(\beta\)-globin gene} & \text{No. of copies of sequence/ genome} \\
\hline
\text{Single copy marker} & \text{EcoRI (2.6)} & 5' & 1200 \text{(1)} \\
\text{EcoRI (2.6)} & 5' & 500 & 2 \\
\text{BamHI and KpnI (1.9)} & 5' & 400 & 3 \\
\text{HaeIII (0.9)} & \text{Central} & 400 & 3 \\
\text{EcoRI (0.8)} & 5' & 1100 & 1 \\
\text{PstI and KpnI (3.6)} & \text{All} & 0.16 & 7000 \\
\text{BamHI and KpnI (3.0)} & 3' & 0.23 & 5000 \\
\hline
\end{tabular}
\end{table]

1978
present with one to three copies per genome; in particular the maximum repetition frequency of DNA sequences in the 1500 base-pairs of 5'-end extragenic DNA sequences contained within the map of Fig. 1 is three copies per genome. We can find no evidence for repetitive DNA in this region, which might be expected to fulfil a regulatory role of the type proposed by Britten & Davidson (1969). The final significance of these findings, however, awaits the mapping of the primary transcript of the β-globin gene.

Transcription of the β-globin gene: RNA splicing

The fact that β-globin genes examined from erythroid tissues exhibit the same discontinuous gene structure suggests that intact β-globin mRNA has to be transcribed from this structure. RNA splicing (Berget et al., 1977; Chow et al., 1977; Klessig, 1977) seems a likely mechanism for producing β-globin mRNA from an RNA precursor containing the two parts of the mRNA plus the intervening sequences. A 15S β-globin mRNA precursor appears to be transcribed from the mouse β-globin gene (Curtis et al., 1977, and references therein), which also shows the same interrupted gene structure as the rabbit β-globin gene (see below). Recent R-loop-mapping experiments, using a cloned chromosomal mouse β-globin gene (Tilghman et al., 1978), showed that this precursor for β-globin mRNA (Curtis et al., 1977) hybridizes to a contiguous region of the chromosomal DNA that probably comprises the two discrete gene segments and the 600-base-pair intervening sequence (C. Weissmann, personal communication).

Discontinuous genes in eukaryotes: a general phenomenon

A discontinuous-gene structure of the type described above has now been demonstrated for the chicken ovalbumin genes (Breathnach et al., 1977; Doel et al., 1977; A. Sippel & G. Schütz, personal communication) and the genes coding for immunoglobulin light chains (S. Tonegawa, personal communication). It has also been known for some time that a subset of the Drosophila ribosomal cistrons contains split 28S rRNA genes (Glover & Hogness, 1977). Finally, it is noteworthy that Tilghman et al. (1978) have shown, in a simultaneous study to ours, that a cloned mouse β-globin gene has an essentially identical structure with that described above for the rabbit β-globin gene. In this case the gene is split after the codon for amino acid 104.

Not all eukaryotic genes appear to be discontinuous; however, for example the sea-urchin histone genes, the ribosomal genes and tRNA<sup>Met</sup> genes of Xenopus are 'conventional' in that the RNA- or protein-coding sequences appear to exist as a single DNA tract co-linear with the gene product. It seems likely that eukaryotic genomes are a mixture of 'conventional' genes, discontinuous genes and perhaps still other structures waiting to surprise the molecular biologist upon their discovery.

Our own work described here was supported in part by a grant to P. Borst from The Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from The Netherlands Organization for the Advancement of Pure Research (Z.W.O.).


Vol. 6 25
The Ovalbumin Structural Gene in Genome and Messenger

NORMAN H. CAREY and MICHAEL T. DOEL

Searle Research Laboratories, Lane End Road, High Wycombe, Bucks. HP12 4HL, U.K.

For more than 10 years now there has been considerable interest in the mechanism of control of expression of the eukaryotic genome. Many workers have focused attention on the observation that there is temporal control of this expression, resulting in the synthesis of a different assortment of proteins in the various tissues of an organism and at different times. The mechanism by which this is controlled has been examined in a number of cell systems and by a wide variety of approaches, but cells in which the synthesis of specific proteins is controlled by hormones have featured strongly in such studies (Tomkins et al., 1970; O'Malley & Means, 1974; Palmiter, 1975; Gorski & Gannon, 1976).

It has been presumed by some authors that, because a protein or its messenger is synthesized de novo (or from very small amounts) after hormone administration, a de-repression mechanism similar to that proposed by Jacob & Monod (1961) operates in eukaryotes as well as prokaryotes (Harris et al., 1975). The fact that hormones are found linked to their specific receptors and bound to chromatin has been a major influence in perpetuating this suggestion. However, the evidence in favour of the proposal is far from conclusive, and there are data that support alternative mechanisms, for example one in which the hormone is involved in stabilizing the messenger (Palmiter & Carey, 1974).

Attempts to examine this mechanism more closely by studying the transcription of specific genes in vitro have met with major problems. The low concentration of each gene in the genome, the uncertainty of its chemical integrity, and the complexity of the protein and RNA components that constitute chromatin have all resulted in data that are difficult to interpret. It seems unlikely that this problem will be solved without purification of those components of the genome that are concerned with the synthesis of a single protein. Of these, the structural gene and its associated control sequences must be of primary concern, since they can in principle be used in identifying and purifying other molecules (RNA and protein) that may be involved. It is now recognized that cloning these DNA structures in bacteria is the only realistic procedure that will achieve the purification required.

There is a second reason why the cloning of specific genes from the genome may be of interest. This concerns the proposal that bacteria in which eukaryotic gene sequences have been cloned may be persuaded to transcribe and translate the sequences, thus allowing the synthesis of useful polypeptides. It seemed possible that sequences derived from the genome could have some advantages, as yet unspecifiable, in this respect compared with those derived from mRNA.

For these reasons, we have been examining the structure of the ovalbumin gene in the genome. Total chicken DNA has been digested with restriction enzymes, and the fragments separated on agarose gels and then transferred to nitrocellulose filters. The ovalbumin sequences were detected by hybridization either with [32P]mRNA fragments or with [32P]RNA transcribed from a plasmid containing cloned ovalbumin sequences.