Genetic variants of human red-cell membrane sialoglycoprotein β

Study of the alterations occurring in the sialoglycoprotein- β gene

Michael J. A. TANNER,* Stephen HIGH,* Peter G. MARTIN,* David J. ANSTEE,† Philip A. JUDSON† and Trevor J. JONES‡

*Department of Biochemistry, University of Bristol, Bristol BS8 1TD, †South West Regional Blood Transfusion Service, Bristol BS10 5ND, and ‡U.K. Transplant Service, Bristol BS10 5ND, U.K.

We have studied the DNA of individuals who express an altered sialoglycoprotein β on their red cells by using Southern blotting with sialoglycoprotein- β cDNA probes. Individuals of the Leach phenotype do not express any β (sialoglycoprotein β) or γ (sialoglycoprotein γ) on their red cells, and we show that about 7 kb of DNA, including the 3' end of the β gene, is deleted in this DNA. Any protein product of this gene is likely to lack the membrane-associating domain of β . We have also examined the DNA of two types of other individuals (Yus-type and Gerbich-type) who have red cells that lack β and γ , but contain abnormal sialoglycoproteins related to β . These two types of DNA contain different internal deletions of about 6 kb in the β gene. We suggest that these deletions result from the presence of two different sets of internal homology in the β gene, and on this basis we propose structures for the abnormal Yus-type and Gerbich-type sialoglycoproteins which are consistent with the other evidence that is available. We provide evidence that β and γ are products of the same gene and suggest a possible mechanism for the origin of γ based on leaky initiation of translation of β mRNA.

INTRODUCTION

The human red-cell membrane sialoglycoproteins β (β ; also known as glycophorin C) and γ appear to be involved in maintaining the shape of the normal red cell and are associated with the red-cell skeleton (for a recent review, see Anstee & Tanner, 1986). There is evidence that the cytoplasmic portion of β provides one of the membrane attachment sites for the skeleton by interacting with protein band 4.1 (Mueller & Morrison, 1981).

Six individuals from four unrelated kindred have been found with red cells that lack β and γ , and these all have cells that are elliptocytic (Anstee *et al.*, 1984a; Daniels *et al.*, 1986; Reid *et al.*, 1986). The red cells of these individuals are described as being of the Leach phenotype and they also lack the very common Gerbich blood-group antigen (Anstee *et al.*, 1984b). Other individuals are known with red cells that lack the Gerbich antigen, but these differ from the Leach phenotype. Two types of these Gerbich-negative red cells have been studied (Gerbich-type and Yus-type), and the red cells of both types lack β and γ but contain an unusual sialoglycoprotein which is structurally related to β , but is not found in normal red cells (Anstee *et al.*, 1984b; Dahr *et al.*, 1985a; Reid *et al.*, 1987). The abnormal sialoglycoproteins found in these two cell types differ in their mobilities on SDS/polyacrylamide-gel electrophoresis (see Anstee *et al.*, 1984b).

The structure of β has been studied in some detail, and the complete amino acid sequence of this protein is known (Dahr *et al.*, 1982; Colin *et al.*, 1986). The protein contains a single membrane-traversing segment and is oriented with the N-terminus at the cell surface and C-terminus in the cytoplasm. In contrast, there is very little structural information on γ . Immunochemical experi-

ments have shown that the cytoplasmic portions of β and γ are structurally related to each other (Reid *et al.*, 1987). This region of these proteins is also structurally related to the cytoplasmic portions of the abnormal sialoglycoproteins found in the Gerbich-negative red cells of the Gerbich type and the Yus type (Reid *et al.*, 1987). Studies with monoclonal antibodies suggest that β and the abnormal sialoglycoproteins also share a common extracellular epitope (Anstee *et al.*, 1984b).

cDNA clones for β have recently been isolated (Colin *et al.*, 1986; High & Tanner, 1987). The availability of these cDNA probes makes it possible to study the differences between the normal β gene and β gene in individuals with the different types of variant red cells that lack the Gerbich antigen. In the present paper we show that the genome of a Leach-phenotype individual retains the 5' portion of the β gene, but lacks the entire 3' portion. If any polypeptide products result from this gene, they probably would lack the membrane-associating domain of normal β and thus would not be inserted into the red-cell membrane. The genomes of the Gerbich-negative individuals of the Gerbich type and Yus type have different internal deletions entirely within the region of the β gene containing the protein-coding exons. Our results are also consistent with β and γ being products of the same gene, and we suggest a possible mechanism for the origin of γ based on the leaky initiation of translation of the mRNA for β .

MATERIALS AND METHODS

Materials

Restriction enzymes and DNA-modifying enzymes were from Amersham International, Amersham, Bucks.,

Abbreviations used: α , β , γ and δ , sialoglycoproteins α , β , γ and δ ; nt, nucleotide; poly(A)⁺, polyadenylated; kb, kilobase(s); bp, base-pair(s); RFLP, restriction-fragment-length polymorphism.

U.K., Pharmacia Biochemicals, Milton Keynes, Bucks., U.K., and Gibco-BRL, Paisley, Renfrewshire, Scotland. Blood samples from the Leach phenotype donor P. L. and the Gerbich-negative donor of the Gerbich type, G. I., were obtained at the South West Regional Blood Transfusion Service, Bristol, U.K. Blood from the Yust-type donor Cho was kindly given by Dr. B. Cedergren, Umeå, Sweden, and that from the Gerbich-type donor Gow by Dr. M. Mohammed, American Red Cross, San José, CA, U.S.A.

Preparation of lymphoblastoid cell lines from normal and Gerbich-negative individuals and isolation of DNA

Anticoagulant-containing blood samples (25 ml) were collected in plastic containers and maintained and transported at ambient temperature. The mononuclear cells were recovered by using the Ficoll-Hypaque method and transformed as soon as possible after collection. The transformation of these peripheral-blood lymphocytes with Epstein-Barr virus and subsequent culture of the cell lines was as described by Doyle *et al.* (1985). The virus preparation (culture supernatant from a B95.8D marmoset cell line) was incubated with lymphocytes for 1 h at 37 °C.

Chromosomal DNA was isolated from the cultured transformed cells by a modification of the method described by Bidwell & Jarrold (1986). Washed cells $[(3-7) \times 10^7]$ were resuspended in 3.3 ml of 10 mM-Tris/HCl/1 mM-EDTA, pH 7.4, and 6.7 ml of lysis buffer (Bidwell & Jarrold, 1986) containing 20 µg of heat-treated bovine pancreatic ribonuclease. The mixture was incubated for 3 h at 55 °C, extracted with phenol and chloroform, and the DNA was precipitated with ethanol.

Preparation of DNA probes

The cDNA clone for β , BET2 (High & Tanner, 1987) in pUC9 was excised from the vector with restriction endonuclease *Eco*R1 (the gift of Dr. S. Halford) and purified by agarose-gel electrophoresis. The probes N and C (see below) were prepared by digestion of BET2 with *Pst*I (Amersham International) and purified by agarose-gel electrophoresis. The probes were labelled with $[\alpha^{32}\text{P}]\text{dATP}$ by using a nick-translation kit (Amersham International).

The synthetic oligonucleotide probe P has the sequence:

GATCCGGGGATGGCCTCTGCCTCCACCACAATGCATACTACCACCATTGC

which corresponds to the 50 nts from nt 55 to nt 104 of the coding sequence of β mRNA. Probe P was constructed by the ligation of the two synthetic oligonucleotides C1 and C2 in the presence of the overlapping complementary oligonucleotide NC2 (High & Tanner, 1987). C1 corresponds to the 27 nucleotides from nt 55 to nt 81 and C2 is the 23-nucleotide sequence from nt 82 to nt 104 of the mRNA. NC2 is the 15-nucleotide sequence

TGCATTGTGGTGGAG

which is complementary to the 7 nts at the 3' end of C1 and the 8 nts at the 5' end of C2. To prepare the labelled probe 150 ng of C1 was phosphorylated using $[\gamma^{32}\text{P}]\text{ATP}$ and T_4 polynucleotide kinase. The labelled C1 was ligated to 240 ng of C2 in the presence of 240 ng of NC2 using T_4 DNA ligase at 12 °C for 4 h. The labelled probe was then purified by polyacrylamide-gel electrophoresis in the presence of 8 M-urea.

Southern blotting of DNA samples

DNA samples were digested with restriction endonucleases as recommended by the suppliers of the enzymes. After extraction with phenol and ethanol precipitation, 10 µg samples of the cleaved DNA were separated by electrophoresis in a gel containing 0.5% agarose and ethidium bromide (0.5 µg/ml). The DNA was transferred to Genescreen Plus (New England Nuclear) and hybridized with cDNA probes by following the procedure recommended by the manufacturer. Hybridization with the synthetic oligonucleotide probe P was carried out as follows. The blot was prehybridized at 37 °C for 4 h in a solution containing 50% (w/v) formamide, 10% (w/v) dextran sulphate, 1% (w/v) SDS and 1 M-NaCl. Hybridization was done in the same solution containing the labelled probe at 37 °C for 16 h. The blot was washed twice at room temperature with $2 \times \text{SSC}$ (SSC is 0.15 M-NaCl containing 15 mM-sodium citrate, pH 7.0). The blot was then washed twice at 55 °C in $2 \times \text{SSC}$ containing 1% SDS and finally washed twice at 27 °C in $0.1 \times \text{SSC}$.

Primer extension of mRNA

This was based on the procedure of Maniatis *et al.* (1982) for first-strand cDNA synthesis. A 0.5 ng portion of the oligonucleotide NC1 was hybridized with 2 µg of K562 poly(A)⁺ RNA (prepared as described by High & Tanner, 1987) in 50 mM-NaCl/9 mM-MgCl₂/50 mM-Tris/HCl, pH 8.3. The mixture was heated at 100 °C for 1 min and then at 65 °C for 10 min. It was then allowed to cool from 65 °C to below 37 °C over a period of 3 h. Primer extension was done in 100 mM-Tris/HCl (pH 8.3)/120 mM-KCl/9.6 mM-MgCl₂/14 mM-2-mercaptoethanol/actinomycin D (20 µg/ml)/12 µM non-radioactive dATP/1 mM other dNTPs/ $[\alpha^{32}\text{P}]\text{dATP}$ (50 µCi). A 12.5-unit portion of avian-myeloblastosis-virus reverse transcriptase was added and the mixture incubated at 37 °C. After 15 min the mixture was made 1.25 M in dATP and, after a further 15 min, the reaction was stopped by the addition of EDTA to a final concentration of 20 mM. After ethanol precipitation the products were separated on a DNA sequencing gel. The size of the extension product was measured by parallel electrophoresis of the DNA sequencing reactions of a DNA

template of known sequence (*Sma*I fragment of BET2 cDNA clone).

RESULTS AND DISCUSSION

Size of β mRNA

Since the cDNA clones available for β extended to the 3' end of the mRNA, but did not extend to the 5' end (High & Tanner, 1987), it was important to determine the size of the β mRNA to ensure the cDNA clones were capable of detecting the bulk of the β gene. Primer extension was used to locate the 5' end of the mRNA using poly(A)⁺ RNA from the K562 erythroleukaemic cell line. The 24-mer oligonucleotide NC1 was used to prime synthesis from nt 1 of the coding region of the mRNA using reverse transcriptase. Primer extension

Fig. 1. DNA probes used to examine Southern blots of normal and variant DNA

The diagram shows the locations of the probe sequences. Details of the probes are given in the text. The open area in BET2 shows the location of protein-coding region, and the hatched area indicates the location of the region coding for the membrane-spanning segment of the protein.

resulted in a band of 180 nts that included the 24-nt primer. Thus the 5' end of the mRNA is located 156 nts from the start of the protein-coding region, and the mRNA for β contains 1072 nts excluding the poly(A) tail. The largest cDNA clone so far isolated, BET2, lacks only 36 nts of the 5' end of the mRNA sequence and contains 120 nts of the 5' non-coding region and all but 15 nts of the 3' non-coding region, together with the entire protein-coding region of the β mRNA (High & Tanner, 1987). BET2 should detect the major part of the β gene.

Southern blotting of normal and variant DNA

Epstein-Barr-virus-transformed lymphoblastoid cell lines were prepared from the peripheral-blood lymphocytes of individuals with Leach-phenotype red cells and individuals with red cells of the other Gerbich-negative types (Gerbich-type and Yus-type), and the DNA from these cell lines was used for study. Transformed cell lines were used because of the difficulty of obtaining regular samples of fresh peripheral blood from these rare individuals. The strategy used to define the alterations occurring in the variant DNA was to probe Southern blots of restriction-enzyme digests of the normal and variant DNA samples in turn with the complete β probe BET2, a probe (N) containing the 5' portion of BET2 and most of the protein coding sequence, a probe (C) containing the 3' non-coding region of the mRNA and a synthetic oligonucleotide (P) containing a portion of the protein-coding sequence near the N-terminus of the protein. Probe N was the 449-bp *Pst*I fragment of BET2 extending from the 5' end to the *Pst*I site. This fragment contains 329 nts of the protein-coding sequence. Probe C was the 576-bp fragment of BET2 extending from the *Pst*I site to the 3' end. C contains all but 15 nts of the 3' non-coding region of the β mRNA as well as 55 nts of the end of the coding sequence. Probe P was a 50-nt synthetic oligonucleotide prepared by the ligation of two smaller oligonucleotides and represents nts 55–104 of the protein-coding sequence (amino acids 19–35). Fig. 1 shows the location of the probe sequences in β mRNA.

We initially studied DNA from a normal Gerbich-positive individual, a donor with the Leach phenotype (P.L.; Anstee *et al.*, 1984a) and donors of Yus-type (Cho; Anstee *et al.*, 1984b) and Gerbich-type (G.I) and assigned the restriction fragments binding the probes described above. It became apparent from these experiments that the Gerbich-type donor (G.I) was in fact heterozygous for the Gerbich and Leach phenotypes. These red cells, like those of the homozygous Gerbich-type, contain only the abnormal β and are difficult to distinguish from the

latter. Subsequent experiments were carried out using DNA from another Gerbich-type individual (Gow; Reid *et al.*, 1987), but the Southern blots were probed only with the intact BET2 probe and probe P.

*Hind*III digests of DNA

The results of probing *Hind*III digests of the DNA samples with the intact β cDNA probe BET2 are shown in Fig. 2. Normal DNA gave bands of 11 and 2.7 kb, but each variant DNA gave a pattern different from this.

Table 1 shows the results of hybridization of the *Hind*III digests with the probes N, C and P. The 11-kb band of normal DNA represents the region extending from the 5' end of the cDNA clone to a point beyond nt 104 of the protein-coding sequence, which is also 5' to the *Pst*I site. The 2.7-kb normal band contains the remainder of the gene to the 3' end. The 3' region of the gene contained in this 2.7-kb band is absent from the Leach-phenotype DNA. However, both N and P detect the 6.7-kb band in the Leach-phenotype DNA, and this suggests that a portion of the 5' region of the gene which includes part of the protein-coding region (amino acids 19–35) is still present in the Leach genome. In contrast, the other Gerbich-negative types (Gerbich-type and Yus-type) retain the 2.7-kb band. In these instances alterations have occurred in the 5' portion of the gene. The Yus-type variant DNA gives a 4.8-kb band which binds probe N, which is found instead of the normal 11-kb band. None of the Yus-type bands bind the oligonucleotide probe P, suggesting that an alteration has occurred in the region of the gene which encompasses the exon(s) detected by this probe.

The situation in the Gerbich-type examples is more complicated. The 2.7-kb normal band is present, but is noticeably weaker in the Gerbich-type DNA from donor G.I than in normal DNA or the Gerbich-type DNA of donor Gow. Probes N and P both bind to the two new bands of 6.7 and 4.8 kb in donor-G.I. DNA, and the normal 11-kb band is absent. The 6.7-kb band is of the same size as that found in Leach DNA, which also binds N and P. This suggested that the donor G.I, who apparently has red cells of the Gerbich-type phenotype, might in fact be heterozygous for both the Gerbich-type and the Leach phenotype. Confirmation that this was indeed the case came from study of the other Gerbich-type donor (Gow), which showed strong binding of BET2 to the 4.8-kb and 2.7-kb fragments, but no 6.7-kb fragment (Fig. 2). In other experiments using DNA from donor Gow, only the 4.8-kb band was found to bind probe P. Thus, unlike the Yus-type, the Gerbich-type DNA retains most of the region coding for amino acids 19–35 of the normal protein.

Fig. 2. Southern blots of normal and variant DNA probed with BET2

(a) *Hind*III digests of DNA: N(6), normal DNA; C, DNA from donor Cho (Yus-type); G, DNA from donor Gow (Gerbich-type); I, DNA from donor G.I. (heterozygous for Gerbich-type and Leach phenotype); L, DNA from donor P.L. (Leach phenotype). (b) *Bam*HI digests of DNA: N(6), normal DNA; C, DNA from donor Cho (Yus-type); G, DNA from donor Gow (Gerbich-type); I, DNA from donor G.I. (heterozygous for Gerbich-type and Leach phenotype); L, DNA from donor P.L. (Leach phenotype). (c) *Pst*I digests of DNA: N19 and N6, DNA from two different normal donors; Ch, DNA from donor Cho (Yus-type); Go, DNA from donor Gow (Gerbich-type); I, DNA from donor G.I. (heterozygous for Gerbich-type and Leach phenotype); L, DNA from donor P.L. (Leach phenotype).

Table 1. Bands in restriction-enzyme digests of normal and variant DNA which hybridize with β probes

Digest	Probe	DNA source ...	Fragment size (kb)			
			Normal	Leach (P.L)	Gerbich-type (G.I.)	Yus-type (Cho)
<i>Hind</i> III	N		11, 2.7	6.7	6.7, 4.8, 2.7	4.8, 2.7
	C		2.7	—	2.7	2.7
	P		11	6.7	6.7, 4.8	—
<i>Bam</i> HI	N		12, 6.2, 3.8	6.2, 3.8	12, 6.2, 3.8	12, 6.2
	C		12	—	12	12
	P		3.8	3.8	12, 3.8	—
<i>Pst</i> I	N		3.4, 1.6, 1.25	1.6, 1.25	1.6, 1.25	1.6, 1.25
	C		1.05, 0.9	—	1.05	1.05
	P		1.25	1.25	1.25	—

A weakly hybridizing band (12 kb) remained the same in all the DNA samples. This migrated very close to the 11-kb band in normal DNA in the blot shown in Fig. 2, but was resolved from the 11-kb band in normal DNA in other experiments (results not shown). The origin of this band is not clear.

*Bam*HI digests of DNA

The same experiment was done on *Bam*HI-digested DNA, and the results obtained using the BET2 cDNA probe are shown in Fig. 2. Normal DNA gave three bands of 12, 6.2 and 3.8 kb. The blots were also probed with N, C and P, and the results are given in Table 1. The 12-kb band represents the 3' end of the gene, whereas the 6.2- and 3.8-kb fragments contain the 5' portion of the gene with the 3.8-kb fragment containing the exon(s) detected by probe P. The Leach DNA lacked only the 12-kb fragment containing the 3' end of the normal gene. The Gerbich-type DNA from donor Gow and the Yus-type DNA from donor Cho lacked the 3.8-kb fragment. All the bands found in normal DNA were present in the

case of the Gerbich-type DNA from donor G.I. This is consistent with donor G.I. being heterozygous for both the Gerbich-type and Leach-type genotypes. Although none of the bands in the Yus-type DNA bound probe P, this probe bound both 12-kb and 3.8-kb bands in the heterozygous Gerbich-type DNA of G.I. In other experiments probe P was found to bind only the 12-kb band in the homozygous Gerbich-type DNA of donor Gow. Since, with normal DNA, probe P binds the 3.8-kb band, the 12-kb band in the Gerbich-type DNA (and presumably the Yus-type DNA) differs from the normal 12-kb band in that it probably spans the deletions in these variants and carries the 3' portion of the gene.

*Pst*I digests of DNA

Fig. 2 shows Southern blots of *Pst*I digests of normal and variant DNAs probed with BET2. Two different control Gerbich-positive samples were used. Both samples gave 3.4-, 1.6- and 1.25-kb bands, but they differed in that one sample (N19) gave a band of 1.05 kb, whereas the other (N6) showed a band of 0.9 kb. In other

experiments, a different control DNA sample gave both the 1.05- and 0.9-kb bands (results not shown). These results suggest the presence of a restriction-fragment-length polymorphism (RFLP) in the β gene of normal individuals. This RFLP in *Pst*I digests of normal DNA has also recently been reported by Arveiler *et al.* (1987), who also found 3.6-kb and 1.3-kb fragments in the *Pst*I digests. These authors did not detect the 1.6-kb fragment because their cDNA probe lacked the 5' untranslated region of the mRNA.

The Leach-type DNA did not give the 3.4-kb or 1.05-kb/0.9-kb bands. The other Gerbich-negative DNAs all gave similar patterns. These all lacked the 3.4-kb band, but contained the 1.6-, 1.25- and 1.05-kb bands. Other faint bands were present on the blot, but these were the same in all the different DNAs (Fig. 2). *Pst*I digests of DNA from donors P.L., Cho and G.I. were hybridized with probes N, C and P as well as BET2 (Table 1). The control normal DNA sample used in this experiment was heterozygous for the RFLP which gives rise to the 1.05- and 0.9-kb bands. The results show that the 3.4-, 1.6- and 1.25-kb fragments contain the 5' portion of the gene, the 1.25-kb fragment containing the region coding for nts 55–104 of the mRNA. The 3' end of the gene is contained in the 1.05-kb/0.9-kb fragments.

All of the DNAs tested contained a 1.25-kb fragment which hybridized with probe N. This fragment from normal, Leach and both the heterozygous and homozygous Gerbich-type DNAs of donors G.I. and Gow, but not the Yus-type DNA, also bound the oligonucleotide probe P. The normal and variant DNAs also differed in that the 3.4-kb band which bound probe N was absent from all the variant DNAs.

The Yus-type DNA differs from normal DNA in both the absence of the 3.4-kb fragment and the lack of binding of probe P to the 1.25-kb band. Making the reasonable assumption that the difference between the normal and Yus-type DNAs results from a single deletion in the Yus-type DNA, we can infer that the 1.25-kb and 3.4-kb fragments follow each other in the gene and that the deletion spans regions of both fragments. The elimination of one of the *Pst*I sites which gives rise to the 1.25-kb fragment would alter this fragment. The presence in the Yus-type DNA of the 1.25-kb band is therefore unexpected. It is likely that the 1.25-kb band in the Yus-type DNA is a new 1.25-kb fragment containing portions of both the normal 1.25-kb and 3.4-kb fragments which border the deletion in the Yus-type DNA.

The Leach-phenotype DNA retains the 1.25-kb fragment, but lacks the 3.4-kb fragment as well as the 1.1-kb band containing the 3' end of the gene. If, as the other evidence suggests, deletion of the 3' portion of the gene has occurred in the Leach phenotype, then the 1.25-kb fragment must be on the 3' side of the 1.6-kb fragment. Thus the order of the *Pst*I fragments in the normal gene is:

5' 1.6 kb, 1.25 kb, 3.4 kb, 1.05/0.9 kb 3'

The assignment of the 1.6-kb fragment to the 5' end of the gene is consistent with the inability of Arveiler *et al.* (1987) to detect the 1.6-kb fragment using a cDNA clone which lacked the 5' non-coding region of the mRNA.

Differences in the β gene of normal and variant individuals

Leach-phenotype individuals entirely lack the 3' portion of the β gene from a point on the 3' side of the

1.25-kb *Pst*I fragment. The 3' end point of this deletion is outside the β gene, and the deletion extends about 7 kb, as estimated from the sizes of the normal and Leach-phenotype *Hind*III-digest fragments. We can estimate from the *Pst*I digests that the deletion includes at least 4.5 kb of the β gene.

The *Hind*III digests suggest that the Gerbich-type and Yus-type DNA have deletions of similar size (about 6 kb), spanning different, but overlapping, regions. These deletions are entirely within the β gene. The 5' end points are closer to the 5' end of the gene than that in the Leach phenotype, since the 3.8 kb *Bam*HI fragment is present in the Leach-phenotype DNA but not the Gerbich- or Yus-type DNA. In the case of the Gerbich-type DNA the deletion occurs from the 3' side of the 1.25-kb *Pst*I fragment, whereas the Yus-type deletion starts further towards the 5' end of the gene and covers the exon(s) containing nts 55–104 of the coding sequence of the mRNA, which are detected by probe P. Clearly it will be necessary to use a genomic DNA probe for β to define the end points of the deletions which occur in the variants in more detail.

The Southern-blotting results suggest that the size of the exon-containing *Pst*I fragments of the normal β gene which we detect is approx. 7 kb. The gene itself may be significantly larger.

Binding of probe P to Southern blots

In order to obtain signals from the Southern blots, probe P had to be used under hybridization conditions much less stringent than would be expected from the size and composition of the probe (see the Materials and methods section). This suggests that the coding sequence in the gene complementary to this probe is interrupted by intervening sequences. Consistent with this, we were unsuccessful in attempts to use the 27-nt portion of probe P (oligonucleotide C1; see the Materials and methods section) in direct hybridization of agarose gels of the digest as described by Bos *et al.* (1984). Under the same conditions we could successfully detect the α gene in genomic DNA using a 27-nt oligonucleotide probe (results not shown).

Origin of the abnormal sialoglycoproteins in Gerbich-type and Yus-type red cells

The abnormal sialoglycoproteins found in Gerbich-type and Yus-type red cells are smaller than normal β , having apparent M_r values of 27000–35000 and 29000–36000 respectively (Anstee *et al.*, 1984b; Reid *et al.*, 1987). The broad spread in M_r of these molecules is due to the presence of *N*-glycosidically linked oligosaccharide-containing polylactosaminyl chains of variable length (Reid *et al.*, 1987). Both of these proteins react with monoclonal antibodies (BRIC 4 and BRIC 10) which react close to the extracellular *N*-terminus of normal β , suggesting that they have the amino acid sequence of normal β at their *N*-termini (Anstee *et al.*, 1984b). The cytoplasmic domains of these proteins also react with antibodies to the cytoplasmic domain of β (Reid *et al.*, 1987). The variant proteins thus differ from β in an internal portion of the polypeptide chain, which is consistent with the conclusions of the studies on genomic DNA. Since their membrane topology is similar to that of β , both proteins retain the hydrophobic transmembrane domain of β , which extends between amino acid residues 59 and 80. Thus the deletion in the

Fig. 3. Regions of internal homology in β

(a) DIAGON plot (Staden, 1982) of a portion of β cDNA against itself. Horizontal and vertical axes show nucleotide numbers of part of the coding region of sialoglycoprotein β cDNA. (b) Details of the three most extensive internal homologies in the coding region of β cDNA shown in (a). The numbers show the positions of the sequences in the coding region of β cDNA. Identities are indicated by asterisks.

Gerbich-type variant probably occurs between a point around the 3' end of the region detected by probe P (about amino acid 35) and amino acid 59, whereas in the Yus-type DNA the deletion encompasses at least a substantial portion of the region between amino acids 19 and 35.

Normal β contains the sequence EPDPGM repeated at residues 17–22 and 36–41 (see Fig. 4 below). Since repeated sequences in the DNA could give rise to internal deletions, the sequence of β mRNA was examined for regions of internal homology by using the DIAGON method of Staden (1982). Fig. 3(a) shows the presence of three regions of extended homology in the DNA sequence. In the repeated regions 1 and 2 (Fig. 3b) the

matched regions are in the same reading frame for the protein, whereas in homology 3 (Fig. 3b) the protein reading frames differ so that a deletion due to crossing-over or looping out would change the reading frame of the C-terminal portion of the protein and would also result in the loss of the extended hydrophobic amino acid sequence that forms the membrane-traversing domain of the protein.

Homology 1 (Fig. 3b) results from the protein-sequence repeat EPDPGM at residues 17–22 and 36–41 (Fig. 4), and at the DNA level only one mismatch occurs in 18 bp. Deletion between these repeats would result in the loss of nts 67–105 of the mRNA as well as one of the repeats. Most of the region detected by the oligonucleotide probe P would be absent. This change is similar to the alteration we have found to occur in the Yus-type DNA. In this case the 13 amino acids between residues 22 and 36 would be deleted, as well as one of the repeats, and the deleted amino acids would include seven serine or threonine residues which carry *O*-glycosidically linked oligosaccharides (Fig. 4).

Homology 2 (Fig. 3b) contains four mismatches in a region of 20 bp and reflects homology between the sequences ASASTTM (residues 23–29) and TSTPTIM (residues 51–57). In this case deletion would result in the loss of one of these repeats together with amino acids 30–50, but a substantial portion (33 nts) of sequence complementary to the oligonucleotide probe P would be retained in the deleted gene. This would be consistent with ability of the Gerbich-type DNA to bind probe P under the low-stringency conditions used in our Southern-blotting experiments. The serine and threonine residues in the homology sequence TSTPTIM (residues 51–57) are not glycosylated (Dahr *et al.*, 1985b). This deletion would result in the loss of 28 amino acids and between four and eight *O*-linked oligosaccharides, depending on the exact sequence and *O*-glycosylation of the homology region remaining in the deleted gene (Fig. 4).

A further difference in the structure of the protein resulting from deletion due to homology 1 and that due to homology 2 is that the former would retain Arg-48 of the normal protein, whereas this amino acid residue would be absent in the protein product of the deletion from homology 2 (Fig. 4). Normal β is readily cleaved at Arg-48 on trypsin treatment of red cells (Dahr *et al.*, 1982), and this is the only point of trypsin cleavage in the extracellular portion of the protein (Dahr *et al.*, 1982). The Yus-type protein is also cleaved by trypsin treatment of the red cells, whereas the Gerbich-type protein is resistant to this enzyme (Anstee *et al.*, 1984b; Anderson *et al.*, 1986). This supports the notion that the Yus-type variant results from deletion due to homology 1, whereas the Gerbich-type variant is a consequence of deletion caused by the presence of homology 2.

Certain human antisera (described as 'anti-Ge 3') appear to react with an antigenic determinant located on β between amino acid residues 40 and 55 (Kiedrowski *et al.*, 1985). These antisera react with the abnormal β in Yus-type red cells, but not that in Gerbich-type red cells (Reid, 1986), suggesting that at least part of the sequence containing these residues of β is present in the Yus-type protein but absent from the Gerbich-type protein, which is consistent with our assignment.

Both normal β and the abnormal proteins in Gerbich-type and Yus-type red cells contain *N*-glycosidically

Fig. 4. Suggested structures of variant sialoglycoproteins and γ

The regions of internal homology are shown by broken underlining (-----) (homology 1) or continuous underlining (——) (homology 2). The site of trypsin cleavage (Arg-48) in normal β and that suggested for the Yus-type protein are both shown by a double underlining (=). The homologous sequences suggested to lead to the deletion in the Gerbich-type protein are shown above each other, since the precise sequence and *O*-glycosylation in this region cannot be predicted. The exact position of the C-terminus of any Leach-phenotype polypeptide product is not known. + Indicates an *O*-glycosylation site and * indicates a site of *N*-glycosylation.

linked oligosaccharides, but the normal and variant oligosaccharides differ in size and structure (Reid *et al.*, 1987). When the *N*-glycan chains are removed, the M_r values of the Yus and Gerbich-type proteins are respectively 8000 and 10000 less than that of normal β (Reid *et al.*, 1987). The structures proposed for the variant proteins would yield products which can be calculated to be smaller than β by similar magnitudes if the *N*-glycan chains are not taken into account.

Fate of the product of the Leach-type gene

There is no evidence that Leach-phenotype red-cell membranes contain any β -related protein. The DNA from these individuals lacks the 3' portion of the normal β gene, including the exon which gives rise to the C-terminal portion of the protein and the 3' non-coding region of the mRNA. If any transcripts of the gene are made, these would lack the 3' non-coding region of the mRNA, including the site of polyadenylation, and may be unstable or not migrate out of the nucleus. Further, if the transcripts did reach the protein-synthesizing apparatus the polypeptide product would probably lack all, or most, of the membrane-traversing domain (see Fig. 4), since the 5' endpoint of the Leach deletion is close to the 3' endpoint of the Gerbich-type deletion, which is in the extracellular portion of the protein. Normal β lacks a cleaved *N*-terminal signal sequence for membrane insertion, and the membrane traversing domain probably carries out this function (High & Tanner, 1987). Thus any protein product of the Leach gene would not be inserted into the membrane or secreted from the cell but would remain within the cell and probably be degraded.

Origin of sialoglycoprotein γ

The presence of the abnormal sialoglycoproteins in Gerbich-type and Yus-type red cells is associated with the absence of both sialoglycoproteins β and γ . The structural similarities between these abnormal sialoglycoproteins and β and γ led us to suggest that the abnormal components may be hybrid molecules with their *N*-termini derived from β and C-termini derived from γ (Anstee *et al.*, 1984b; Reid *et al.*, 1987). Such hybrids are known to arise in the case of α and δ as a

result of unequal crossing-over between the α and δ genes (Anstee *et al.*, 1982). If the abnormal- β -related sialoglycoproteins in Gerbich-type and Yus-type were hybrid molecules resulting from the above mechanism, β and γ would have to be derived from separate genes. The results obtained in the present study suggest that the abnormal sialoglycoproteins in Gerbich-type and Yus-type red cells result from deletions that occur totally within the β gene. If γ were the product of a separate gene, it would be difficult to explain its absence in these variant red cells.

Colin *et al.* (1986) found a single β -related mRNA species of 1.4 kb in erythroid cells. Another study suggested that small amounts of a 0.6-kb mRNA were present in erythroid cells in addition to a major 1.2-kb species (High & Tanner, 1987). The 1.2-kb mRNA clearly corresponds to the β mRNA, but it is not clear whether the smaller RNA represents the mRNA for γ or a transcript of another β -related gene. If the smaller mRNA is indeed the mRNA for γ , this could arise by differential splicing from the β gene.

However, an alternative mechanism could also account for the occurrence of γ . The RNA sequence CCAGGAUGU around the initiator methionine codon for β (High & Tanner, 1987) does not conform closely with the consensus sequence CC(A/G)CCAUGG found in eukaryotic mRNAs (Kozak, 1984, 1986a). In the case of some viral mRNAs, initiation occurs at different methionine residues to give two translation products in the same reading frame (Kozak, 1986b). Here it is often the case that the sequence around the methionine codon nearest the 5' end of the mRNA does not closely fit the consensus initiation sequence. Although initiation occurs predominantly from this methionine codon, significant amounts of translation product initiate from the next methionine codon downstream in the mRNA. This is suggested to be due to 'leaky' scanning of the mRNA in a 5'-to-3' direction by the 40 S ribosome subunit (Gupta & Kingsbury, 1985; Kozak, 1986b).

In the case of β mRNA the methionine codon downstream from that which would give rise to β is at amino acid 22 and is within the sequence

CCGGGGAUGG. This sequence fits the consensus initiation sequence better than the upstream initiation codon and contains the G residue after the methionine codon which is thought to be important in efficient initiation. The sequence preceding this methionine codon is also similar to the sequence preceding the upstream methionine codon and matches it in four of the six nucleotides before the ATG. It is possible that γ arises from leaky initiation at this downstream methionine, which gives rise to Met-22 of β . Compared with β this truncated product (Fig. 4) would lack the *N*-terminal 21 amino acids, which includes the site of *N*-glycosylation (Asn-8), and six sites of *O*-glycosylation and have an M_r of 27000, similar to that of γ .

The sequence of β mRNA suggests that β is not translated with an *N*-terminal signal sequence which is cleaved during the insertion of the protein into the membrane (High & Tanner, 1987) but contains an uncleaved internal signal sequence which is part of the membrane-associating hydrophobic domain. These membrane-associating domains have been shown to act as signal sequences in several cases (Lipp & Dobberstein, 1986; Speiss & Lodish, 1986). If γ results from leaky initiation of β mRNA, it would still retain the membrane domain and be inserted into the red-cell membrane. In contrast, α is translated with a cleaved *N*-terminal signal sequence (Siebert & Fukuda, 1986). Translation products from leaky initiation at downstream methionine codons in α mRNA would lack the signal sequence and could not be inserted into the erythrocyte membrane.

If γ is indeed a truncated form of β , then clearly the amino acid sequence of γ must be identical with that of amino acid residue 22 onwards of β , providing that it does not undergo any other post-translational processing. Although amino-acid-sequencing studies will be required to establish this unequivocally, immunochemical evidence indicates that β and γ are very similar in structure (Reid *et al.*, 1987). β and γ share antigenic determinants in both extracellular and intracellular domains, and the only immunological differences between the two molecules are in the *N*-terminal extracellular domain. Monoclonal antibodies which react with the extracellular domain of β but not with γ have been described (Anstee *et al.*, 1984a,b). Other monoclonal antibodies of similar specificity have been reported to react with epitopes in the region of amino acid residues 1–4 of β (Dahr, 1986). Most human anti-Gerbich sera recognize an epitope that is present on the extracellular domain of γ but not of β (Reid *et al.*, 1987). These antibodies could react with new epitopes which become exposed in the truncated form of β . These immunological differences can be readily reconciled with the hypothesis that γ is a truncated form of β .

We thank Dr. B. Cedegren, Dr. I. Fraser, Dr. M. Mohammed and Ms. M. E. Reid for their help in obtaining blood samples, and Dr. B. Bradley in preparing transformed cell lines. This work was supported in part by a grant from the Medical Research Council, and S. H. was the recipient of a Medical Research Council Studentship.

REFERENCES

- Anderson, S. E., McKenzie, J. L., McLoughlin, K., Beard, M. E. J. & Hart, D. N. (1986) *Pathology* **18**, 407–412
- Anstee, D. J. & Tanner, M. J. A. (1986) *Br. J. Haematol.* **64**, 211–215
- Anstee, D. J., Mawby, W. J. & Tanner, M. J. A. (1982) in *Membranes and Transport* (Martonosi, A., ed.), vol. 2, pp. 427–433, Plenum Press, New York
- Anstee, D. J., Parsons, S. F., Ridgwell, K., Tanner, M. J. A., Merry, A. H., Thomson, E. E., Johnson, P., Bates, S. & Fraser, I. D. (1984a) *Biochem. J.* **218**, 615–619
- Anstee, D. J., Ridgwell, K., Tanner, M. J. A., Daniels, G. L. & Parsons, S. F. (1984b) *Biochem. J.* **221**, 97–104
- Arweiler, B., Le Van Kim, C., Colin, Y., Mandel, J. L. & Cartron, J. P. (1987) *Nucleic Acids Res.* **15**, 1880
- Bidwell, J. L. & Jarrold, E. A. (1986) *Mol. Immunol.* **23**, 1111–1116
- Bos, J., Velkaan de Vries, M., Jansen, A. M., Viemrsmas, G. H., van Boom, J. H. & van de Eb, A. J. (1984) *Nucleic Acids Res.* **12**, 9155–9163
- Colin, Y., Rakael, C., London, J., Romeo, P. H., D'Auriol, L., Galibert, F. & Cartron, J. P. (1986) *J. Biol. Chem.* **261**, 229–233
- Dahr, W. (1986) in *Recent Advances in Blood Group Biochemistry* (Vengelen-Tyler, V. & Judd, W. J., eds.), pp. 23–65, American Association of Blood Banks, Arlington
- Dahr, W., Beyreuther, K., Kordowicz, M. & Kruger, J. (1982) *Eur. J. Biochem.* **125**, 57–62
- Dahr, W., Moulds, J., Baumeister, G., Moulds, M., Keidrowski, S. Z. & Hummel, M. (1985a) *Biol. Chem. Hoppe-Seyler* **366**, 201–211
- Dahr, W., Blanchard, D., Hummel, M., Cartron, J. P., & Beyreuther, K. (1985b) in *Glyconjugates: Proceedings of the Eighth International Symposium*, Houston (Davidson, E. A., ed.), pp. 21–22, Praeger Publishers, New York
- Daniels, G. L., Shaw, M.-A., Judson, P. A., Reid, M. E., Anstee, D. J., Colpitts, S., Cornwall, S., Moore, B. P. L. & Lee, S. (1986) *Vox Sang.* **50**, 117–121
- Doyle, A., Jones, T. J., Bidwell, J. L. & Bradley, B. A. (1985) *Hum. Immunol.* **13**, 199–209
- Gupta, K. C. & Kingsbury, D. W. (1985) *Biochem. Biophys. Res. Commun.* **131**, 91–97
- High, S. & Tanner, M. J. A. (1987) *Biochem. J.* **243**, 277–280
- Kiedrowski, S., Dahr, W., Moulds, J., Moulds, M., Blanchard, D. & Cartron, J. (1985) *Biol. Chem. Hoppe-Seyler* **366**, 810–811
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872
- Kozak, M. (1986a) *Cell* (Cambridge, Mass.) **47**, 283–292
- Kozak, M. (1986b) *Cell* (Cambridge, Mass.) **47**, 481–483
- Lipp, J. & Dobberstein, B. N. (1986) *Cell* (Cambridge, Mass.) **46**, 1103–1113
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Mueller, T. J. & Morrison, M. (1981) in *Erythrocyte Membranes 2: Recent Clinical and Experimental Advances* (Krukeberg, W. C., Eaton, W. C. & Brewer, G. J., eds.), pp. 95–112, A. R. Liss, New York
- Reid, M. E. (1986) in *Recent Advances in Blood Group Biochemistry* (Vengelen-Tyler, V. & Judd, W. J., eds.), pp. 67–104, American Association of Blood Banks, Arlington
- Reid, M. E., Martynewycz, M.-A., Walford, F. E., Crawford, M. N. & Miller, L. H. (1986) *Transfusion* **27**, 213–314
- Reid, M. E., Anstee, D. J., Tanner, M. J. A., Ridgwell, K. & Nurse, G. T. (1987) *Biochem. J.* **244**, 123–128
- Siebert, P. D. & Fukuda, M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1665–1669
- Speiss, M. & Lodish, H. F. (1986) *Cell* (Cambridge, Mass.) **44**, 177–185
- Staden, R., (1982) *Nucleic Acids Res.* **10**, 2951–2961