Biochem. J. (1989) 262, 47–54 (Printed in Great Britain)

Rearrangements of the red-cell membrane glycophorin C (sialoglycoprotein β) gene

A further study of alterations in the glycophorin C gene

Stephen HIGH,* Michael J. A. TANNER,*§ E. Bruce MACDONALD† and David J. ANSTEE‡
*Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K., †Red Cross BTS, Queensland Division, G.P.O. Box 157, Brisbane, Qld. 4001, Australia, and ‡South West Regional Blood Transfusion Service, Bristol BS10 5ND, U.K.

We have cloned portions of the glycophorin C (sialoglycoprotein β) gene from individuals with red cells of normal, Gerbich and Yus phenotypes. The clones contain up to three exons of the glycophorin C gene (designated exons 2, 3 and 4). Analysis by restriction mapping and DNA sequencing confirmed that the deletions causing the Gerbich and Yus phenotypes are located entirely within the glycophorin C gene. Sequencing of the normal gene showed that not only do exon 2 and exon 3 have related DNA sequences, but also that both the 5′ and 3′ flanking intronic DNA sequences are almost identical. The two variant genes each lack a different exon: the Yus type gene lacks exon 2, whereas the Gerbich-type gene lacks exon 3. We suggest that the observed deletions are due to recombination between the regions of homologous intronic repeats. We also provide evidence that an unequal cross-over mechanism may be responsible for a number of observed glycophorin C gene rearrangements, including an insertion mutation in Lewis II (Ls^a)-type red cells that has not previously been reported.

INTRODUCTION

There are four readily detectable sialoglycoproteins in the human red-cell membrane [see Anstee & Tanner (1986) for a review], which are denoted glycophorins A, B, C and γ [sialoglycoproteins α, δ, β and γ in the nomenclature of Anstee *et al.* (1979)]. Glycophorins C and γ are minor components of this group. There is evidence that glycophorins C and γ are involved in the control of red-cell shape. Thus glycophorin C is known to interact with the red-cell cytoskeleton via protein band 4.1 (Mueller & Morrison, 1981; Sondag *et al.*, 1987). Both glycophorins C and γ are enriched in cytoskeleton preparations from normal red cells (Anstee *et al.*, 1984*b*), but are completely absent in cytoskeleton preparations from red cells of homozygous band-4.1-deficient individuals (Allosio *et al.*, 1985). Red cells of the Leach phenotype completely lack glycophorins C and γ and have been found to be elliptocytic (Anstee *et al.*, 1984*a*) and have decreased membrane stability and deformability (Reid *et al.*, 1987*a*).

Red cells of individuals of the Leach phenotype also lack the common Gerbich (Ge) blood-group antigens (Anstee *et al.*, 1984*a*), which are known to be carried on glycophorins C and γ [see Reid (1986) for a review]. A second group of Ge-negative individuals have been described whose red cells lack both glycophorins C and γ, but contain an abnormal sialoglycoprotein which is immunochemically related to glycophorin C. Two types of abnormal protein are found: the Ge type, which resists trypsin treatment of red cells, and the Yus type, which is sensitive to trypsin (Anstee *et al.*, 1984*b*). Normal glycophorin C is cleaved by trypsin at Arg-48 (Dahr *et al.*, 1982; Dahr & Beyreuther, 1985). The *N*- and *C*-termini of the normal, Ge- and Yus-type proteins appear to be immunochemically identical (Anstee *et al.*, 1984*b*; Reid *et al.*, 1987*b*).

cDNA clones for glycoprotein C have been isolated (Colin *et al.*, 1986; High & Tanner, 1987). Southern-blotting experiments on genomic DNA have shown that the Ge and Yus phenotypes are due to deletions of about 4 kb within the glycophorin C gene, with the 5′ and 3′ ends of the gene remaining intact (Le Van Kim *et al.*, 1987; Tanner *et al.*, 1988). Studies with an oligonucleotide probe suggested that, although the deletions are of the same size, different portions of DNA are deleted in the Ge and Yus phenotypes (Tanner *et al.*, 1988).

In the present work we have further characterized the nature of the alterations in the Ge- and Yus-type genes. Genomic clones for the normal and variant genes have been analysed by restriction mapping and sequence analysis. Our data show that the Ge and Yus phenotypes result from deletion of different exons within the normal glycophorin C gene. We provide evidence that non-homologous recombination may be responsible for these deletions.

MATERIALS AND METHODS

Materials

Restriction enzymes and DNA-modifying enzymes were from USB Corp., Cleveland, OH, U.S.A., Amersham International, Amersham, Bucks., U.K., and

Abbreviations used: nt, nucleotide(s); Ge, Gerbich; Ls^a, Lewis II.

§ To whom correspondence and reprint requests should be addressed.

The sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession nos. X13889, X13890, X13891, X13892 and X13893.

Gibco–BRL, Paisley, Renfrewshire, Scotland, U.K. EMBL 4 vector was from Amersham International, Gigapack Gold packaging extract was from Stratagene, San Diego, CA, U.S.A., and Sequenase was from USB Corp. Random priming kits were from BCL, Lewes, East Sussex, U.K. *Eco*RI restriction endonuclease was a gift from Dr. S. Halford.

Preparation and screening of genomic DNA libraries

Chromosomal DNA was prepared from Epstein–Barr-virus-transformed lymphocytes of normal and variant individuals as previously described (Tanner *et al.*, 1988). *Eco*RI-cleaved genomic DNA was ligated into *Eco*RI-cleaved EMBL 4 vector using the conditions described by Maniatis *et al.* (1982), supplemented with 4 mM-spermidine. The DNA was packaged using Gigapack Gold packaging extracts as described by the manufacturer. The libraries in EMBL 4 were plated on *Escherichia coli* Q359 to select for recombinants (Kaiser & Murray, 1985). The EMBL 4 genomic libraries were screened with the almost-full-length cDNA probe BET 2 (High & Tanner, 1987), using standard protocols (Maniatis *et al.*, 1982).

Purification, subcloning and sequencing of inserts

Recombinant bacteriophage DNA was prepared by the plate-lysate method (Davis *et al.*, 1980), and the inserts were subcloned into pUC 13. Plasmid recombinants were prepared as described by Maniatis *et al.* (1982) and were banded twice on CsCl gradients. Fragments of the original *Eco*RI inserts were also subcloned into pUC18 and pUC19 for further sequencing studies. In most cases a 'shotgun-cloning' approach was used, and the required recombinants were identified by analysis of restriction-enzyme digests. Sequencing of double-stranded DNA was by the dideoxy method of Sanger *et al.* (1977), using Sequenase as described by the manufacturer. Gradient gels were as described by Biggin *et al.* (1983). Sequence analysis made use of oligonucleotide primers constructed so that the sequence on both strands was obtained.

Southern-blot analysis

Southern-blot analysis with the cDNA probes was as described by Tanner *et al.* (1988), except that labelling was achieved by random priming. Where oligonucleotides were used as hybridization probes, they were 5′-end-labelled as described by Maniatis *et al.* (1982), and hybridization and washing conditions were those described by Thein & Wallace (1986).

RESULTS AND DISCUSSION

Isolation of genomic clones for glycophorin C from normal and variant individuals

Southern-blotting studies of normal human genomic DNA with a glycophorin C cDNA probe showed a major 15 kb *Eco*RI band, while similar analysis of Ge- and Yus-type DNA samples showed a 12 kb *Eco*RI band in each case (Le Van Kim *et al.*, 1987). EMBL 4 libraries were prepared from *Eco*RI-digested genomic DNA of normal, Ge- and Yus-type individuals. The libraries were screened with a glycophorin C probe (BET 2; High & Tanner, 1987). Positively hybridizing phage containing a 15 kb insert (designated 'N-GPC1') were obtained from the normal DNA library, whereas the Ge- and Yus-type DNA libraries yielded phage containing 12 kb inserts (designated 'G-GPC1' and 'Y-GPC1' respectively).

Characterization of the normal glycophorin C genomic clone

Synthetic oligonucleotide primers corresponding to portions of the coding sequence of normal glycophorin C cDNA were used to obtain the sequences around the exons of the glycophorin C genomic DNA clone. These studies indicated the presence of three exons in N-GPC1. No sequence was obtained from N-GPC1 using primers corresponding to the coding sequence on the 5′ side of nt 49 of the cDNA sequence (equivalent to the *N*-terminal side of Leu-16 of the protein). This suggested that the exon(s) corresponding to the *N*-terminal region of the protein was absent from N-GPC1.

The three exons present in N-GPC1 code for amino acid residues 17–35, 36–63 and from 64 to the *C*-terminus of the protein (including the 3′ non-coding region) respectively. The exons have been denoted exon 2, exon 3 and exon 4 respectively, as illustrated in Fig. 1. The sequences of the exons and the surrounding intronic regions are shown in Figs. 2(*b*), 2(*d*) and 2(*e*).

Sequence similarity between exon 2 and exon 3 of the normal glycophorin C gene

Computer analysis showed that the sequence of exon 2 and its flanking regions (Fig. 2*b*) and exon 3 and its flanking regions (Fig. 2*d*) are very similar. An alignment of the two sequences is shown in Fig. 3. It is clear that there is substantial similarity within the two exons and almost complete identity between the flanking regions. Indeed the 5′ and 3′ regions flanking exon 2 and exon 3 show more similarity than do the exons themselves. The above results show that the normal glycophorin C gene contains two regions which are almost identical over approx. 250 bp. The sequences at the two *Bam*HI sites to the 5′ of exon 2 and exon 3 are also very similar: 131 of 145 nt are identical (compare Figs. 2*a* and 2*c*), indicating that the repeated sequences to the 5′ side of the exons could be up to 1.5 kb in length (refer to Fig. 1). Fig. 3 shows that the sequences to the 3′ side of exon 2 and exon 3 are almost identical for 95 nt. Sequence data obtained on one strand showed this level of identity extended a further 200 nt (results not shown). It is also striking that the 3′ end of exon 3 (residues 111–139) shows a striking similarity to part of the intron sequence occurring just after exon 2 (residues 135–163).

By using double restriction-enzyme digests and probing Southern blots with oligonucleotide probes specific for exon 2, exon 3 and exon 4, the detailed restriction map shown in Fig. 1 was constructed. The DNA sequences presented in Figs. 2(*b*) and 2(*d*) show the presence of a *Sac*I site close to the 3′ boundary of both exon 2 and exon 3. Southern blotting of a *Sac*I digest of N-GPC1 and subsequent analysis using an exon-3-specific probe (ex3) showed exon 3 to be present on a 3.3 kb *Sac*I fragment as shown in Fig. 1. This confirms that exon 2 and exon 3 are separated by approx. 3.3 kb, as shown in Fig. 1.

*Bam*HI fragments of N-GPC1

*Bam*HI digestion of N-GPC1 yielded fragments of 9 and 3.3 kb, consistent with previous results from

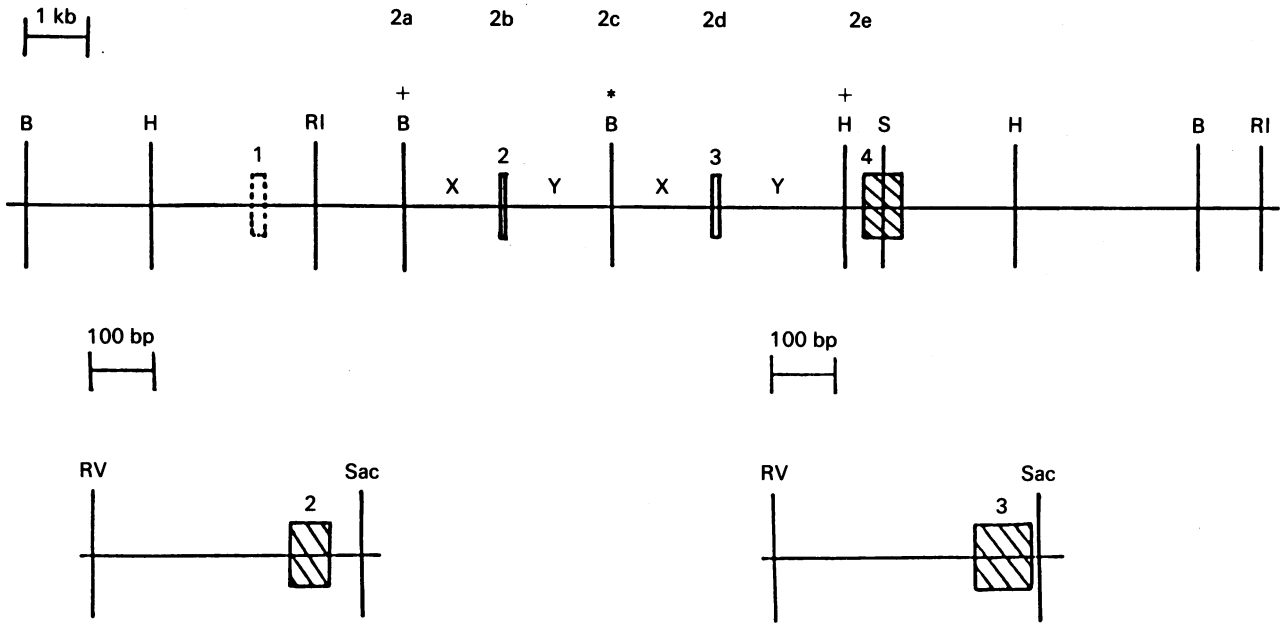


Fig. 1. Restriction map of the normal glycophorin C gene

The diagram shows the points of cleavage by restriction endonucleases in relation to the glycophorin C gene. Key to endonucleases: B, *Bam*HI; H, *Hin*dIII; RI, *Eco*RI; RV, *Eco*RV; Sac, *Sac*I; S, *Sma*I. The numbered hatched regions correspond to exons; the DNA sequence corresponding to a point *N*-terminal of Leu-16 was absent from the clone isolated, and this sequence is indicated by exon 1, the exact location of which is unknown. The clone N-GPC1 contained the DNA 15 kb of DNA between the two *Eco*RI sites marked. The regions to the 5′ and 3′ of exons 2 and 3, marked 'X' and 'Y' are the repeated regions referred to in the text. The *Bam*HI site marked '*' and the *Bam*HI and *Hin*dIII sites marked '+' are referred to in the text. The regions labelled 2a–2e above the Figure identify the regions for which sequence data are presented in Fig. 2.

Southern blotting, although the fragments had been assigned sizes of 12 and 3.8 kb respectively (Tanner *et al.*, 1988). Southern-blot analysis of the *Bam*HI fragments with oligonucleotide probes specific for exon 2 (ex2; see Fig. 2*b*) and exon 3 (ex3; see Fig. 2*d*) showed that exon 2 is present in the 3.3 kb fragment, while exon 2 is present on the 9 kb fragment (see Fig. 4).

These *Bam*HI fragments were subcloned and sequenced from their ends. The sequence of the 5′ end of the 3.3 kb fragment is shown in Fig. 2(*a*), and the sequence from the *Bam*HI site separating the 3.3 kb and 9 kb *Bam*HI fragments (* in Fig. 1) is shown in Fig. 2(*c*). The sequence of the 3′ end of the 9 kb *Bam*HI fragment and on the 5′ side of the *Bam*HI site separating the 3.3 kb and 9 kb *Bam*HI fragments was obtained on one strand only (results not shown).

Exon 2 and exon 3 of N-GPC1 are present on 420 bp and 400 bp *Eco*RV/*Sac*I fragments respectively (Fig. 1). The exon-2-specific probe (ex2) bound only to the 420 bp *Eco*RV/*Sac*I fragment derived from N-GPC1, whereas the exon 3 specific probe (ex3) bound only to the 400 bp *Eco*RV/*Sac*I fragment derived from N-GPC1 (see Fig. 4).

Characterization of the variant glycophorin C genomic clones

*Bam*HI digestion of G-GPC1 and Y-GPC1 both yielded 9 kb fragments, but no 3.3 kb fragments, consistent with previous Southern-blotting results (Tanner *et al.*, 1988) and with the data shown in Fig. 6(*a*) (below). The sequence of one end of both fragments proved to be identical with that of the 5′ end of the 3.3 kb *Bam*HI fragment of N-GPC1 shown in Fig. 2(*a*). The sequence of the other end of the 9 kb *Bam*HI fragments from G-GPC1 and Y-GPC1 was obtained on one strand only (results not shown) and was identical with that obtained from the 3′ end of the 9 kb *Bam*HI fragment from N-GPC1.

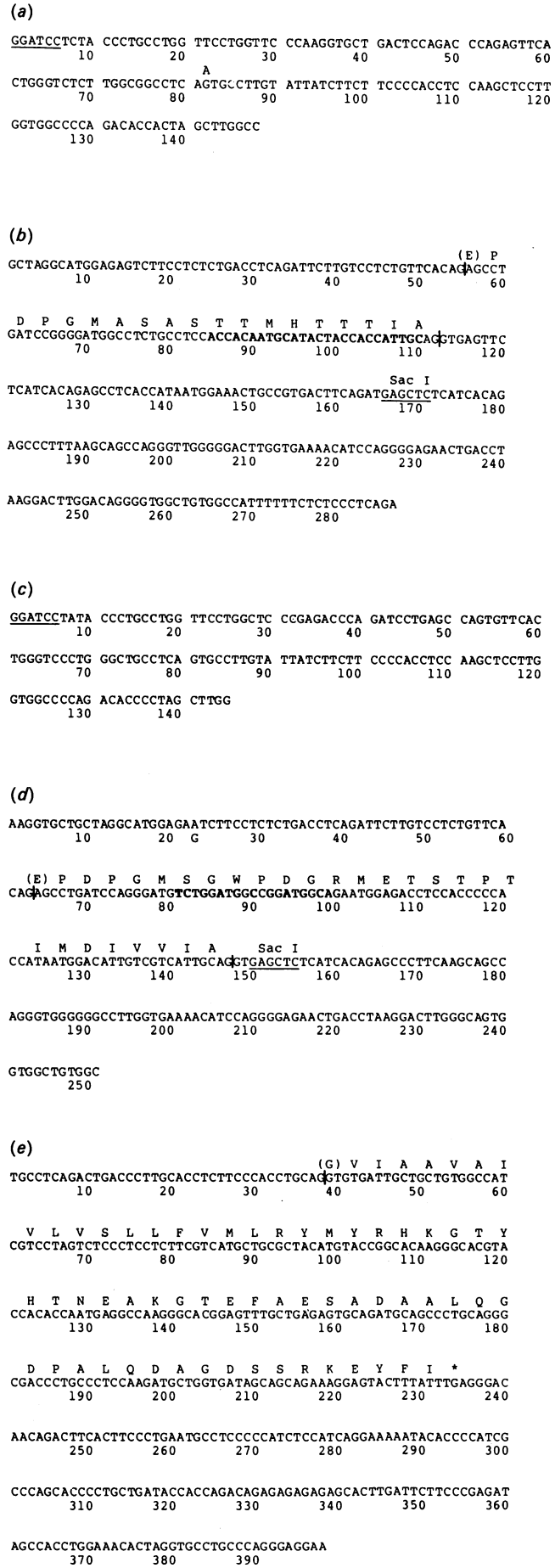
Southern-blotting analysis of the 9 kb *Bam*HI fragments from G-GPC1 and Y-GPC1 with the exon-specific oligonucleotide probes ex2 and ex3 showed that G-GPC1 lacks exon 3, but retained exon 2, whereas Y-GPC1 lacks exon 2, but retained exon 3 (see Fig. 4). As expected, only the 420 bp *Eco*RV/*Sac*I fragment was obtained from G-GPC1, and the fragment bound only the ex2 probe. The Y-GPC1 insert contained only the 400 bp *Eco*RV/*Sac*I fragments, which bound the ex3 probe (see Fig. 4).

We have previously shown that a 50 nt synthetic oligonucleotide probe P (corresponding to nt 61–110 of Fig. 2*b*) bound to fragments of the normal and Ge-type glycophorin C gene, but never to fragments of the Yus-type gene (Tanner *et al.*, 1988). It is clear that probe P is specific for exon 2, and this result is consistent with the absence of exon 2 in the Yus-type gene.

The deletions occurring in both the variants clearly encompass portions of the 3.3 kb and 9 kb fragments of N-GPC1 and result in the loss of the *Bam*HI site marked * in Fig. 1. In both variants the sequence present at the 5′ end of the 3.3 kb *Bam*HI fragment of N-GPC1 and the 3′ end of the 9 kb *Bam*HI fragment of N-GPC1 are also present on the 9 kb *Bam*HI fragment of the variant genes (G-GPC1 and Y-GPC1).

The variant genes have lost either exon 2 or exon 3 and flanking regions

The results of the Southern-blotting experiments using



probes specific for exon 2 and exon 3 (Fig. 4) show that G-GPC1 contains only exon 2 and Y-GPC1 contains only exon 3. Sequencing studies on G-GPC1 and Y-GPC1 showed that both contain exon 4. The sequence of the protein coding region and a large portion of the 3′ non-coding region was determined on both strands of the normal and variant DNAs (see Fig. 2*e*). The sequence of these regions in the variants was indistinguishable from the sequence in this region of N-GPC1. Thus the Ge phenotype is the result of a deletion of 3.3 kb of DNA, including exon 3, whereas the Yus phenotype is caused by a deletion of the same size, but which includes exon 2.

The nature of the deletions in the variants

The restriction map shown in Fig. 1 places exon 2 and exon 3 at approx. 3.3 kb apart, and thus recombination between either the homologous regions 5′ of exons 2 and 3 (X in Fig. 1) or the homologous regions 3′ of exons 2 and 3 (Y in Fig. 1) would result in the observed loss of 3.3 kb of DNA. Any such recombination between the homologous introns would result in the loss of exon 2 or exon 3 and would recreate an intron flanking the remaining exon which is similar in sequence to one of the original introns. Since the intron/exon boundaries at the 5′ and 3′ ends of exon 2 and exon 3 are identical (see Fig. 3), correct splicing of the primary transcript should still occur and result in a truncated mRNA lacking one or other of the two exons.

The protein products of the variant genes

The predicted protein products of the Yus- and Ge-type variants are illustrated in Fig. 5, together with the product of the normal glycophorin C gene. For simplicity the coding sequence on the *N*-terminal side of exon 2 is assumed to be located on one exon (exon 1 of Fig. 1). The predicted polypeptide sequences of the variant forms of glycophorin C are consistent with the data available

Fig. 2. Sequence data from the glycophorin C gene

(*a*) Sequence from the most 5′ *Bam*HI site of N-GPC1, G-GPC1 and Y-GPC1. The sequence was identical in all three cases except that residue no. 82 was an A in the Ge-type sample (G-GPC1). The *Bam*HI site is underlined. (*b*) Sequence from exon 2 and flanking intronic sequence of N-GPC1 and G-GPC1. The start and end of the exon are indicated by vertical lines, and the translation of the exon sequence is shown above. For the amino acid residue in parentheses, the first nucleotide of the codon is present on the exon 5′ of exon 2 (exon 1 of Fig. 1). The sequence in bolder type (i.e. nt 85–110) is that of the exon 2-specific oligonucleotide probe ex2. (*c*) The sequence from the *Bam*HI site separating the 3.3 and 9 kb *Bam*HI fragments of N-GPC1. The *Bam*HI site (indicated by '*' in Fig. 1) is underlined. (*d*) Sequence of exon 3 and flanking intronic sequence from N-GPC1 and Y-GPC1. The sequences were identical, except that nt 23 is a G in Y-GPC1 rather than the A in N-GPC1. The start and end of the exon are indicated by vertical lines, and the translation of the exon sequence is shown above. For the amino acid residue in parentheses the first nucleotide of the codon is present on exon 2. The sequence in bolder type (nt 81–100) is that of the exon-3-specific oligonucleotide probe ex3. (*e*) Sequence of part of exon 4 and 5′ intronic sequence from N-GPC1, G-GPC1 and Y-GPC1. The sequence was identical in all cases, stretching to the end of the protein-coding region and including the first 163 nt of the 3′ non-coding region.

```
                10        20        30        40        50
        GCTAGGCATGGAGAGTCTTCCTCTCTGACCTCAGATTCTTGTCCTCTGTTCA
        :::::::::::::: :::::::::::::::::::::::::::::::::::
        GCTAGGCATGGAGAATCTTCCTCTCTGACCTCAGATTCTTGTCCTCTGTTCA
                10        20        30        40        50

       60        70        80        90        100       110
CAGAGCCTGATCCGGGGATGGCCTCTGCCTCCACCACAATGCATACTACCACCATTGCAG
:::::::::::::: :::::: :         :         :   ::: :::    :
CAGAGCCTGATCCAGGGATGTCTGGATGGCCGGATGGCAGAATGGAGACCTCCACCCG--
       60        70        80        90        100       110

       120       130       140       150       160       170
GTGAGTTCTCATCACAGAGCCTCACCATAATGGAAACTGCCGTGACTTCAGATGAGCTCT
                         :::::::::::: : :: ::: : : ::: ::::::::
---------------------CACCATAATGGACATTGTCGTCATTGCAGGTGAGCTCT
                           120       130       140

       180       190 *     200       210       220       230
CATCACAGAGCCCTTTAAGCAGCCAGGGTTGGGGGACTTGGTGAAAACATCCAGGGGAGA
:::::::::::::: ::::::::::::: ::::: ::::::::::::::::::::::::
CATCACAGAGCCCTTCAAGCAGCCAGGGTGGGGGGCCTTGGTGAAAACATCCAGGGGAGA
 150       160       170       180       190       200

       240       250       260
ACTGACCTAAGGACTTGGACAGGGGTGGCTGTGGC
:::::::::::::::::: ::: ::::::::::::
ACTGACCTAAGGACTTGGGCAGTGGTGGCTGTGGC
 210       220       230       240
```

Fig. 3. An alignment of the DNA sequences of exon 2 and flanking regions and exon 3 and flanking regions

The upper sequence is that of exon 2 and flanking region, and the lower sequence is that of exon 3 and flanking region. The underlined sequences show the positions of the exons. The sequences in bolder type (nt 52–58 and 103–115 in sequence line 3, 52–58 in sequence line 4, and 130–142 in sequence line 6) illustrate that the 5′ and 3′ intron/exon boundaries of exon 2 and exon 3 are identical, as discussed in the text.

from studies on the protein products. Thus the Ge-type protein is smaller than the Yus type (Anstee *et al.*, 1984*b*; Dahr *et al.*, 1985). Exon 3 contains Arg-48, the site of extracellular cleavage of glycophorin C by trypsin (Dahr *et al.*, 1982). The sensitivity of the Yus protein to trypsin and resistance of the Ge protein to this proteinase (Anstee *et al.*, 1984*b*; Telen & Bolk, 1987) is consistent with the presence of exon 3 in the Yus gene and its absence in the Ge gene.

The deletion in Leach phenotype individuals

Glycophorins C and γ are both absent from Leach phenotype red cells, and no other glycophorin C-related protein is present (Anstee *et al.*, 1984*a*). The map of the glycophorin C gene (Fig. 1) allows a more precise interpretation of Southern-blotting studies on Leach phenotype DNA (Tanner *et al.*, 1988). Leach DNA lacked the normal 2.6 kb *Hin*dIII fragment containing exon 4, whereas the normal 8.3 kb *Hin*dIII fragment containing the 5′ end of the gene was reduced in size. This fragment bound the exon-2-specific probe P (nt 61–110 in Fig. 2*b*). When the *Bam*HI fragments of Leach DNA were studied, only the 6.2 and 3.3 kb fragments, which comprise the 5′ portion of the normal gene, were found. Probe P bound the 3.3 kb *Bam*HI fragment. Thus the Leach phenotype DNA retains the 5′ portion of the glycophorin C gene up to the *Bam*HI site on the 3′ side of exon 2 (marked * in Fig. 1). The deletion therefore extends from within the region marked X on the 5′ side of exon 3 to the 3′ side of exon 4. Exon 4 contains the membrane domain of glycophorin C, and this acts as a membrane-insertion signal as well as anchoring the protein in the membrane (High & Tanner, 1987). Thus if any transcripts of the Leach gene reach the protein-synthetic apparatus, the translation proteins will not be targeted to, or reside in, the red-cell membrane.

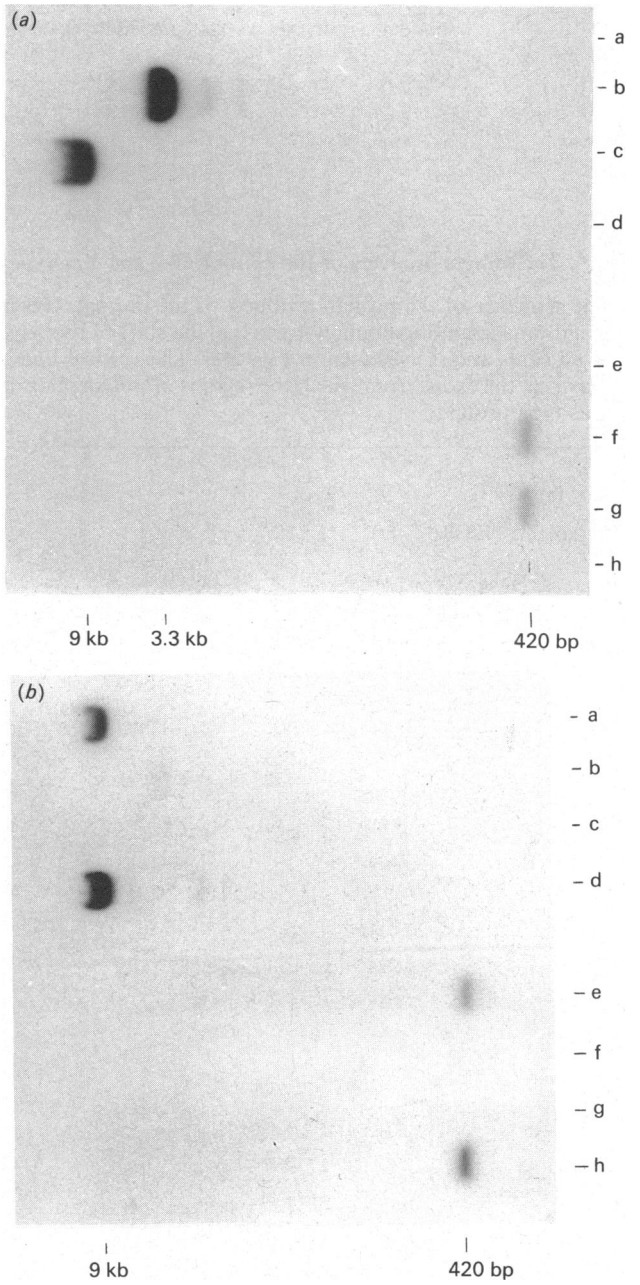


Fig. 4. Analysis of N-GPC1, G-GPC1 and Y-GPC1 with the exon-specific probes ex2 and ex3

(*a*) Shows a Southern blot of DNA fragments probed with the exon-2-specific probe ex 2. Tracks a–d refer to *Bam*HI digests of: a, 9 kb *Bam*HI fragment of N-GPC1 in pUC 18; b, 3.3 kb *Bam*HI fragment of N-GPC1 in pUC 18; c, 9 kb *Bam*HI fragment of G-GPC1 in pUC 18; d, 9 kb *Bam*HI fragment of Y-GPC1 in pUC 18. Tracks e–h were *Eco*RV/*Sac*I digests of: e, 9 kb *Bam*HI fragment of N-GPC1 in pUC 18; f, 3.3 kb *Bam*HI fragment of N-GPC1 in pUC 18; g, 9 kb *Bam*HI fragment of G-GPC1 in pUC 18; h, 9 kb *Bam*HI fragment of Y-GPC1 in pUC 18. (*b*) Shows a Southern blot of DNA fragments probed with the exon 3-specific probe ex3. The samples were exactly as in (*a*). In all cases 2 μg of cleaved DNA was loaded before electrophoresis and Southern blotting. The faint bands visible in some tracks are due to a small amount of star activity during the restriction-enzyme digestion and also caused by non-specific hybridization of the oligonucleotide probes to the large amounts of sample DNA used.

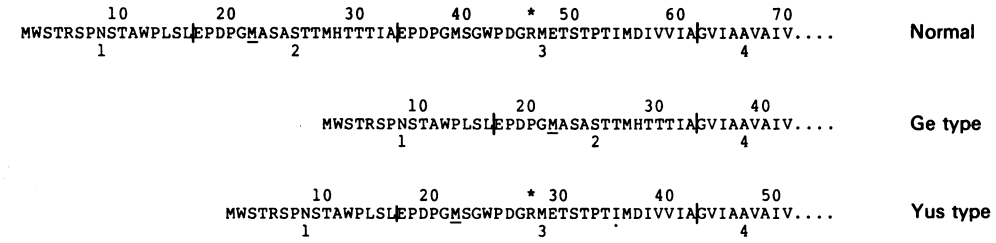


Fig. 5. The protein products of the normal, Ge- and Yus-type genes

The sequence of the protein products of the normal, Ge- and Yus-type glycophorin C genes is shown up to a point where the membrane-spanning domain starts (i.e. the start of the region coded for by exon 4). The rest of the protein sequence is identical in all cases and is as shown in Fig. 2(*e*). The vertical lines indicate the exon boundaries, and the numbers below the sequence illustrate the exons from which the regions of polypeptide are derived. *Indicates the point of trypsin cleavage in the normal and Yus-type protein.

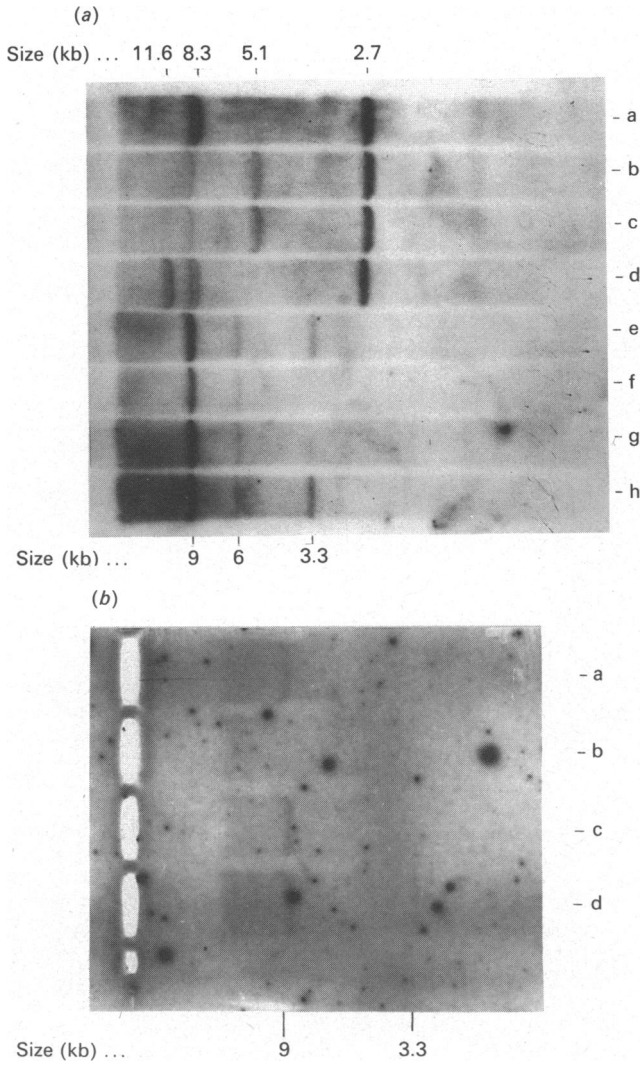


Fig. 6. (*a*) Southern blot of normal and variant DNA probed with BET 2, and (*b*) dried down gel of normal and variant DNA probed with ex2

(*a*) Tracks a–d were *Hin*dIII digests of: a, normal DNA; b, Ge-type DNA; c, Yus-type DNA; d, Ls^a-type DNA; tracks (*e*)–(*h*) were *Bam*HI digests of: e, normal DNA; f, Ge-type DNA; g, Yus-type DNA; h, Ls^a-type DNA. (*b*) The samples were *Bam*HI digests of 20 μg of: a, normal DNA; b, Ge-type DNA; c, Yus-type DNA; d, Ls^a-type DNA.

Evolution of the glycophorin C gene

It seems likely that exon 2 and exon 3 of the normal glycophorin C gene arose from a partial gene-duplication event, and a gene similar to the Ge- or the Yus-type gene may have been the evolutionary forerunner of the normal glycophorin C gene. The rarity in most populations of individuals expressing the Ge- or Yus-type protein (Reid, 1986) suggests that the partial gene duplcation may have invested some form of advantage to individuals of the normal phenotype, although the nature of this advantage is unclear. However, it is of interest that the Ge-type protein is relatively common in certain Melanesian populations (Booth & McLoughlin, 1972).

The abnormal glycophorin C in Ls^a red blood cells

The work of Macdonald *et al.* (1988) showed that the low-frequency Lewis II (Ls^a) antigen is due to an abnormal form of glycophorin C which is larger than normal glycophorin C. The Ls^a-type red cells also contain a form of glycophorin γ which is larger than normal. The genomic DNA of an individual heterozygous for this type was examined by Southern-blot analysis. The results of probing *Hin*dIII- and *Bam*HI-restricted samples of genomic DNA from an Ls^a-positive individual using the BET 2 cDNA probe are shown in Fig. 6(*a*). The Ls^a DNA sample digested with *Hin*dIII (see track d, Fig. 6*a*) clearly shows an additional band of 11.6 kb as well as the bands of 8.3 kb and 2.6 kb, which correspond to the bands seen in the *Hin*dIII digest of the normal sample (see track a, Fig. 6*a*). The faint band at 9 kb, which is invariant in the normal, Ge and Yus samples, has been noted previously (Tanner *et al.*, 1988), and although its origin remains unclear, it is apparent that it also increases by about 3 kb in the Ls^a sample (see Fig. 6*a*). The results of probing *Bam*HI-restricted DNA samples showed no differences in the bands present in the Ls^a and normal samples.

The above data suggested that the Ls^a DNA contains an insertion of 3.3 kb of DNA into the glycophorin C gene. A *Bam*HI digest of normal and variant DNA samples was probed with the exon-3-specific oligonucleotide probe ex3. The results show that a 3.3 kb band hybridizing ex3 was found only in the Ls^a sample. The normal, Yus and Ls^a samples also show a 9 kb band hybridizing with the ex3 probe (Figs. 6*b*). Exon 3 is located on the 9 kb *Bam*HI fragment, whereas exon 2 is located on the 3.3 kb *Bam*HI fragment of the normal glycophorin C gene. Our results are consistent with an

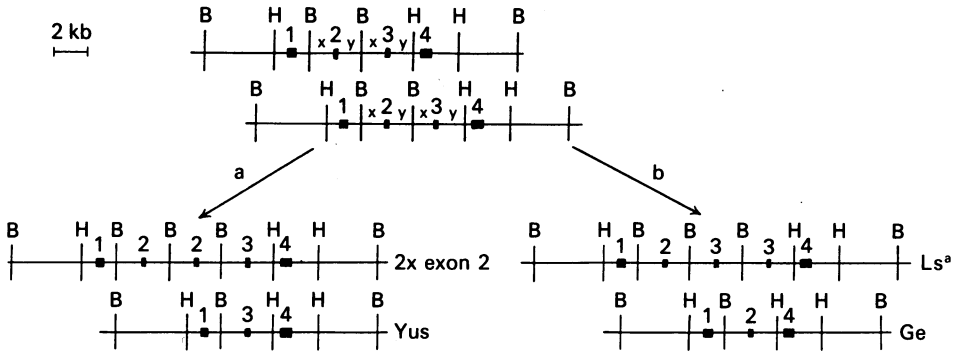


Fig. 7. Products of unequal crossing over between two misaligned glycophorin C genes

(*a*) Shows the results of an unequal cross-over between the repeated regions to the 5′ side of exon 2 and exon 3 (x), and (*b*) shows the result of an unequal cross-over between the repeated regions to the 3′ side of exon 2 and exon 3 (y).

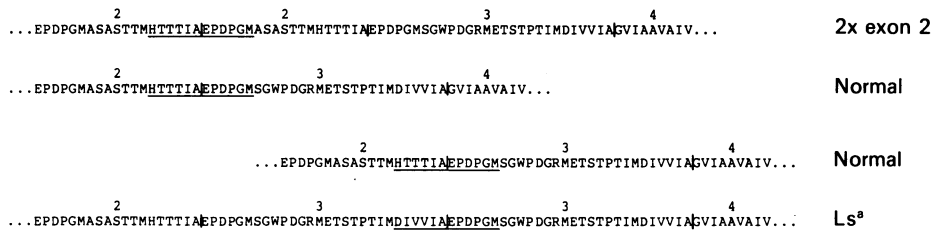


Fig. 8. Predicted protein products from insertion variants of the glycophorin C gene

The values above the sequence show the exon from which the polypeptide is derived. The vertical lines show the exon boundaries, and the underlined residues show where novel exon/exon junctions occur in the variants and compare them with the equivalent region of the normal protein.

insertion of 3.3 kb of DNA carrying exon 3 into the Ls^a gene. This would imply that the Ls^a DNA sample examined contains one copy of exon 2 and two copies of exon 3. Study of an individual homozygous for the Ls^a gene will be necessary to establish unequivocally this to be the case. The presence of an additional exon in the Ls^a gene is consistent with the increase in sizes of the Ls^a glycophorins C and γ observed by Macdonald *et al.* (1988). The additional DNA present in the Ls^a gene must be between the *Bam*HI and the *Hin*dIII sites marked + in Fig. 1, since the 6 kb *Bam*HI fragment on the 5′ side of this region and the 2.6 kb *Hin*dIII fragment containing exon 4 (i.e. on the 3′ side of this region) remain unaltered in the Ls^a DNA (Fig. 6*a*).

Origin of the variant glycophorin C genes

At least two possible mechanisms could generate the deletions occurring within the Ge- and Yus-type genes. The simplest model would involve looping out between the regions of similarity within the glycophorin C gene. Alternatively these deletions could result from misalignment of the internal repeating regions of two glycophorin C genes, followed by unequal crossing over. A similar situation is known to occur between the related genes for glycophorin A and glycophorin B [see Reid (1986) for a recent review]. The second mechanism (see Fig. 7) would result not only in the generation of genes lacking exon 2 or exon 3 (i.e. the Ge or Yus type), but would also generate a reciprocal gene which would have an additional copy of exon 2 or exon 3 inserted (Fig. 7).

The Ls^a variant appears to be an example of a reciprocal-insertion mutation, where one copy of exon 2 and two copies of exon 3 are present. Fig. 7(*b*) shows that, in the product of unequal crossing over, where an extra copy of exon 3 is present, this additional exon would be located on a 3.3 kb *Bam*HI fragment, as is found to be the case for the Ls^a gene. If, as we suggest, the Ls^a gene and the Ge-type gene are the reciprocal products of an unequal cross-over event, then an insertion mutation containing two copies of exon 2 and one copy of exon 3 would also be expected to occur as the reciprocal product of the Yus-type deletion (labelled '2x exon 2' in Fig. 7*a*).

Predicted protein product of the Ls^a gene

Fig. 8 shows the amino acid sequence predicted for the Ls^a-type glycophorin C. The amino acid sequence of the junction of the two copies of exon 3 differs by four of 12 residues compared with the exon 2/exon 3 junction present in normal glyophorin C (the underlined residues in Fig. 8). This abnormal region of sequence may give rise to the Ls^a antigen associated with the abnormal protein product. It is noteworthy that, in the alternative insertion mutation, which would contain two copies of exon 2 (2x; exon 2; Fig. 7*a*), the sequences around the exon junctions do not differ from those found in normal glyophorin C (see Fig. 8). This protein product would not be expected to give rise to a novel potentially antigenic sequence. Red cells containing this variant protein would be very difficult to detect, since it is unlikely that they will have any abnormal serological properties.

Unlike the Yus-type protein, the Ge type of glycophorin C contains a novel sequence at the junction of exon 2 and exon 4 (see Fig. 5). However, the Ge-type

protein is not known to give rise to any new antigen. Hydropathy analysis suggests that the polypeptide of normal glycophorin C enters the membrane at Ile-59, and it seems reasonable to suppose that the Ge-type polypeptide enters the membrane at a similar point (Thr-31), since the intracellular *C*-terminal boundaries of the membrane domains are identical in the two proteins. If this is the case, the novel protein sequence in the G-type protein would be buried within the membrane and would not be expected to be antigenic.

We thank Dr. P. A. Judson for the preparation of DNA samples, and Dr. M. E. Reid for helpful discussion of results. This work was supported in part by a grant from the Medical Research Council.

REFERENCES

Allosio, N., Morle, L., Bachir, D., Guetarni, D., Colonna, P. & Delaunay, J. (1985) Biochim. Biophys. Acta **816**, 57–62

Anstee, D. J. & Tanner, M. J. A. (1986) Br. J. Haematol. **64**, 211–215

Anstee, D. J., Mawby, W. J. & Tanner, M. J. A. (1979) Biochem. J. **183**, 193–203

Anstee, D. J., Parsons, S. F., Ridgwell, K., Tanner, M. J. A., Merry, A. H., Thomson, E. E., Judson, P. A., Bates, S. & Fraser, I. D. (1984*a*) Biochem. J. **218**, 615–619

Anstee, D. J., Ridgwell, K., Tanner, M. J. A., Daniels, G. L. & Parsons, S. F. (1984*b*) Biochem. J. **221**, 97–104

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965

Booth, P. B. & McLoughlin, K. (1972) Vox Sang. **22**, 73–84

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. & Beyreuther, K. (1985) Biol. Chem. Hoppe-Seyler **366**, 1067–1070

Dahr, W., Beyreuther, K., Kordowicz, M. & Kruger, J. (1982) Eur. J. Biochem. **125**, 57–62

Received 21 December 1988/10 March 1989; accepted 15 March 1989

Dahr, W., Moulds, J. J., Baumeister, G., Moulds, M., Keidrowski, S. & Hummel, M. (1985) Biol. Chem. Hoppe-Seyler **366**, 202–211

Davis, R. W., Botstein, D. & Rother, J. R. (1980) Advanced Bacterial Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

High, S. & Tanner, M. J. A. (1987) Biochem. J. **243**, 277–280

Kaiser, K. & Murray, N. E. (1985) in DNA Cloning, vol. 1 (Glover, D. M., ed.), pp. 1–47, IRL Press, Oxford and Washington

Le Van Kim, C., Colin, Y., Blanchard, D., Dahr, W., London, L. & Cartron, J.-P. (1987) Eur. J. Biochem. **165**, 571–579

Macdonald, E. B., Condon, J., Ford, D., Fisher, B. & Gerns, L. M. (1988) Int. Congr. ISBT-BBTS, London, July 1988: Book Abstr. P.T.8.56, 156

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, NY

Mueller, T. J. & Morrison, M. (1981) Erythrocyte Membranes, Volume 2: Recent Clinical and Experimental Advances, 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, VA

Reid, M. E., Chasis, J. A. & Mohandas, N. (1987*a*) Blood **69**, 1068–1072

Reid, M. E., Anstee, D. J., Tanner, M. J. A., Ridgwell, K. & Nurse, G. T. (1987*b*) Biochem. J. **244**, 123–128

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467

Sondag, D., Alloisio, N., Blanchard, D., Ducluzeau, M.-T., Colonna, P., Bachir, D., Bloy, C., Cartron, J.-P. & Delaunay, J. (1987) Br. J. Haematol. **65**, 43–50

Tanner, M. J. A., High, S., Martin, P. G., Anstee, D. J., Judson, P. A. & Jones, T. J. (1988) Biochem. J. **250**, 407–414

Telen, M. J. & Bolk, T. A. (1987) Transfusion **27**, 309–314

Thein, S. W. & Wallace, R. B. (1986) in Human Genetic Diseases: A Practical Approach (Davies, K., ed.), pp. 33–50, IRL Press, Oxford and Washington