Isolation of cDNA clones for human erythrocyte membrane sialoglycoproteins α and δ

Christopher G. TATE and Michael J. A. TANNER*

Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K.

We have isolated almost full-length cDNA clones corresponding to human erythrocyte membrane sialoglycoproteins α (glycophorin A) and δ (glycophorin B). The predicted amino acid sequence of δ differs at two amino acid residues from the sequence determined by peptide sequencing. The sialoglycoprotein δ clone we have isolated contains an interrupting sequence within the region that gives rise to the cleaved *N*-terminal leader sequence for the protein and represents a product that is unlikely to be inserted into the erythrocyte membrane. Comparison of the cDNA sequences of α and δ shows very strong homology at the DNA level within the coding regions. The two mRNA sequences are closely related and differ by a number of clearly defined insertions and deletions.

INTRODUCTION

The sialoglycoproteins are a group of integral membrane proteins of the human erythrocyte membrane, all of which are extensively glycosylated at their *N*-termini. Four sialoglycoproteins have been characterized, denoted α , β , γ and δ in order of decreasing apparent M_r on SDS/polyacrylamide-gel electrophoresis (for nomenclature see Anstee *et al.*, 1979). The functions of α (glycophorin A) and δ (glycophorin B) have yet to be established; individuals with erythrocytes lacking both these proteins (M^k/M^k) are apparently normal (Tokunaga *et al.*, 1979), as are individuals with erythrocytes that carry genetic variants of α and δ (Anstee *et al.*, 1982a; Moulds, 1983). Unlike α and δ , both β and γ are associated with the erythrocyte skeleton (Owens *et al.*, 1980), and there is evidence that suggests that β is connected to band 4.1 protein (Mueller & Morrison, 1981).

The complete protein sequence of α has been determined (Tomita *et al.*, 1978), and more recently the sequence of δ has been reported (Blanchard *et al.*, 1987a). α can carry either the blood-group-M or -N antigens, which are determined by amino acid substitutions at residues 1 and 5. The first 26 amino acid residues of the blood-group-N antigen form of α are identical with the *N*-terminal 26 amino acid residues of δ . Further homology is also found in the membrane-spanning region of the two proteins (Dahr *et al.*, 1980, 1985). In addition to bearing blood-group-N antigen activity, δ also carries the blood-group-S and -s antigens, and these are determined by amino acid substitutions at residue 29 (Dahr *et al.*, 1980).

Siebert & Fukuda (1986a) isolated and sequenced a partial cDNA clone for the blood-group-M form of α that contained the bulk of the protein-coding region. This clone hybridized to four different mRNA species in erythroid tissue. On the basis of experiments with a synthetic oligonucleotide probe representing the *N*-terminal region common to α and δ and another probe representing the *C*-terminal sequence specific for α , they

suggested that one of these mRNAs (600 nts) was the mRNA for sialoglycoprotein δ , whereas the other three mRNAs all behaved as if they coded for α or α -like proteins (Siebert & Fukuda, 1986a).

We have isolated almost full-length cDNAs for the blood-group-N form of α and for what may be a variant of the blood-group-s form of δ . Both cDNAs encompass the complete 3' non-coding region and coding regions as well as substantial portions of the 5' non-coding regions. The protein sequence of δ predicted from the δ cDNA shows two amino acid differences from that determined by Blanchard *et al.* (1987a) by peptide sequencing methods; it also suggests that α and δ have almost identical *N*-terminal leader sequences. The cDNAs for the two proteins show very strong homology in the protein-coding regions and 5' non-coding regions with long regions of identity at the DNA level, which suggests that they result from a very recent gene duplication of a blood-group-N-active form. However, the 3' non-coding regions appear to be completely unrelated. The major differences in the protein-coding portions of the cDNAs lie in well-defined insertions and deletions.

MATERIALS AND METHODS

Materials

Restriction enzymes and DNA-modifying enzymes were from Amersham International (Amersham, Bucks., U.K.), Gibco BRL (Paisley, Strathclyde, Scotland, U.K.) and Pharmacia Biochemicals (Milton Keynes, Bucks., U.K.). [³²P]dATP and ATP were from Amersham International. K562 cells were a gift from Dr. D. J. Anstee and Dr. P. Judson (Regional Blood Transfusion Centre, Bristol, U.K.).

Isolation of cDNA clones

A human reticulocyte cDNA library in λ gt11 (Conboy *et al.*, 1986) was screened with polynucleotide kinase-labelled (Maniatis *et al.*, 1982) synthetic oligonucleotide

Abbreviation used: nt, nucleotide residue.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession nos. X08054 and X08055.

* To whom correspondence should be addressed.

probes NTA and CTA by using the method described by Siebert & Fukuda (1986b).

Oligonucleotides were synthesized by using a Dupont Coder 300 DNA synthesizer. The sequences of the oligonucleotides synthesized were:

NTA²³¹ ATTTGTCTGTGATGAGATGTAACCTCTT
 CTA⁸¹³ AGAACTTAAAGGCACGTCTGTGTCAGG
 AN1⁵⁹⁷ GTGTCCAGTCCACGTGTAA
 AC2¹⁶⁹ GTTACATCTCATCACAGAC
 AC3⁶⁶³ CCAGTGA CTCAATGATGGGG
 DC1⁸¹⁰ CTATTCGCCGACTGAT
 DN1⁸⁵¹ TAGGCAAGATCAGGCAGC
 DC2⁵⁶² AGGGAAGAGGAAGAGC

Oligonucleotides NTA and CTA correspond to the complement of the coding sequence of the α cDNA

(Siebert & Fukuda, 1986a). Oligonucleotides AC2 and AC3 correspond to the coding strand, and oligonucleotide AN1 to the complement of the Δ LP1 cDNA sequence. Oligonucleotides DC1 and DC2 correspond to the coding strand of δ , and oligonucleotide DN1 to the complement of the DEL1 cDNA sequence.

Nucleotide sequence analysis

The cDNA inserts were excised from λ gt11 and inserted into the Bluescript vector (Northumbria Biologicals, Cramlington, Northumbria, U.K.). The whole cDNA inserts and restriction fragments were sequenced from Bluescript by the dideoxy method (Sanger *et al.*, 1977). Oligonucleotides were used as internal primers to complement the use of the T3 and T7 primers (corresponding to the bacteriophage RNA polymerase promoter sites in the vector) for sequencing the intact cDNAs and restriction fragments of them from their ends in the vector (see Fig. 1). Double-stranded DNA sequencing of the insert-containing bacteriophage was carried out with the use of the λ gt11 reverse sequence primer. The bacteriophage DNA and primer were heated at 90 °C for 1 h and immediately placed on ice before carrying out sequencing by the dideoxy method. Both strands of each sequence were determined.

Fig. 1. Structures of α and δ clones and sequencing strategies

(a) Sialoglycoprotein α cDNA, ALP1; (b) sialoglycoprotein δ cDNA, DEL1. The open boxes represent the coding regions for the proteins. The hatched box in (b) represents the region of DEL1 homologous to the α -1-globin sequence (Michelson & Orkin, 1980). Abbreviations used for the restriction endonucleases shown on the partial map are: B, *Bcl*I; E, *Eco*RI; S, *Sac*I; E*, *Eco*RI*; D, *Dde*I; Sp, *Sph*I; T, *Taq*I. Unlabelled arrows represent sequence obtained from restriction fragments and from intact cDNA by the use of primers corresponding to T3 and T7 RNA polymerase promoter sites in the sequencing vector. Sequence obtained by the use of internal primers is shown as an arrow with the oligonucleotide primer name. Details of the oligonucleotides used are in the Materials and methods section. The sequence primed by the λ gt11 reverse primer is indicated by λ .

Northern-blot analysis

K562-cell RNA was prepared by the guanidinium thiocyanate method of Chirgwin *et al.* (1979). Polyadenylated RNA was prepared by using oligo(dT)-cellulose chromatography (Aviv & Leder, 1972). Electrophoresis was carried out in formaldehyde-containing gels (Maniatis *et al.*, 1982) with 5 μ g of polyadenylated RNA per track. The probe used to detect the interrupting sequence in DEL1 was derived by *Dde*I digestion of DEL1 followed by blunt-ending of the total digest fragments with mung-bean nuclease (Pharmacia) as recommended by the manufacturers (with 1 unit of enzyme for 10 min at 37 °C). The fragments were then ligated into the *Eco*RV site of the Bluescript vector. The probe was identified by DNA sequencing, which showed that it contained only nts 118–193 of DEL1 and resulted from chew-back by the mung-bean nuclease preparation from the *Dde*I sites. The probe used to detect the 3' non-coding region of δ was isolated by *Sph*I digestion of DEL1 followed by blunt-ending of the gel-purified fragment with mung-bean nuclease as described above and was then ligated into the *Eco*RV site of the Bluescript vector. The subclone containing the probe sequence was shown to contain only nts 511–608 of DEL1 and resulted from chew-back of the original *Sph*I fragment by the mung-bean nuclease preparation. The sites of *Dde*I and *Sph*I digestion predicted by the δ DNA sequence were confirmed by gel electrophoresis of restriction digests of DEL1 with *Dde*I and with *Sph*I. These probes were used in the vector in the experiment shown in Fig. 4, and the purified ALP1a insert was used in the same experiment.

RESULTS AND DISCUSSION

Isolation of sialoglycoprotein α and δ cDNA clones

A human reticulocyte λ gt11 cDNA library was screened in duplicate with oligonucleotide probes NTA and CTA. NTA corresponds to a portion of the *N*-terminal amino acid sequence of α that is identical with that of δ ; CTA corresponds to a portion of the *C*-terminal amino acid sequence of α that is unique to this protein (Siebert & Fukuda, 1986a). Two clones were detected by the use of the oligonucleotide probe NTA, only one of which cross-hybridized with CTA. The latter cDNA (ALP1) was presumed to correspond to α , whereas the cDNA that hybridized only to NTA (DEL1) was assumed to correspond to δ .

Sequence of α cDNA

ALP1 was excised from the λ gt11 vector with *Eco*RI, and two fragments of approx. 800 bp (ALP1a) and 200 bp (ALP1b) were isolated. These fragments were sequenced separately using the strategy shown in Fig. 1(a). The overlap between the two fragments was established by excising the insert from the bacteriophage at the *Kpn*I and *Sac*I sites that border the cloning site in λ gt11 (Huynh *et al.*, 1985). The insert gave rise to two fragments owing to the presence of an internal *Sac*I site in the cDNA (see Fig. 1a), but we were unsuccessful in subcloning the *Kpn*I–*Sac*I fragment that contains the *Eco*RI sites. However, we were successful in subcloning the partial *Kpn*I–*Sac*I fragment that contained the entire ALP1 insert. Sequencing of this showed that ALP1 in fact contained two closely separated *Eco*RI sites. We were unable to detect the smallest (24 bp) *Eco*RI fragment

of ALP1 in digests of the insert-containing bacteriophage. The sequence of the other strand of this region was established by direct sequencing of the insert-containing λ gt11 bacteriophage DNA.

The DNA sequence of ALP1 is shown in Fig. 2(a). The cDNA contained 967 nts excluding the poly(A) tail and included the polyadenylation signal AATAAA, 19 nts from the start of the poly(A) tail. The translated cDNA sequence was clearly the same as the blood-group-N form of α (Tomita *et al.*, 1978; Wasniowska *et al.*, 1977), with the codon corresponding to the *N*-terminal residue of the mature protein being at nt 114 of the ALP1 cDNA, while the methionine codon that initiates translation of the *N*-terminal cleaved leader sequence is located at nt 57. The cDNA isolated by Siebert & Fukuda (1986a) contained the bulk of the coding sequence for the blood-group-M form of α , corresponding to nts 47–492 of ALP1. ALP1 also contains 56 nts of the 5' non-coding region and all the 3' non-coding region of α mRNA. As expected, ALP1 and the cDNA clone described by Siebert & Fukuda (1986a) differ at the codons representing amino acid residues 1 and 5 of the processed protein. In ALP1, Leu (TTA) and Glu (AGG) are present at amino acid residues 1 and 5, whereas Ser (TCA) and Gly (GGT) are present in the clone described by Siebert & Fukuda (1986a). These amino acid changes correspond to the known amino acid sequence differences between the blood-group-M and -N alleles of α (Wasniowska *et al.*, 1977; Dahr & Uhlenbruck, 1978). The nucleotide sequence of ALP1 predicts a further amino acid difference in the *N*-terminal leader sequence of the blood-group-M and -N proteins since a Glu at amino acid –7 is predicted in the blood-group-N form instead of Ala in the blood-group-M form. Five other nucleotide differences in the coding region of the two sequences give rise to silent mutations (see Fig. 3).

Sequence of δ cDNA

The cDNA insert DEL1 that was assumed to correspond to δ was also sequenced as described in Fig. 1(b). The insert proved to be a head-to-head concatenate with a portion of α -1-globin cDNA (Michelson & Orkin, 1980). The 1034 nt insert contained 418 nts corresponding to α -1-globin. The sequence of the remaining 616 nts is shown in Fig. 2(b). The sequence contained two polyadenylation signals, AATAAA, in tandem 29 and 23 nts from the poly(A) tail. The protein sequence predicted from DEL1 shows near identity with the sequence of sialoglycoprotein δ determined by peptide sequencing (Blanchard *et al.*, 1987a), with the *N*-terminus of the mature protein starting at nt 194, but differs in two respects. The cDNA sequence predicts Cys at amino acid 50 as opposed to Ser, and an additional amino acid residue (Ala) at the *C*-terminus of the protein. Blanchard *et al.* (1987a) noted that their assignment of Ser to residue 50 was tentative and were also unable to establish the *C*-terminal residue of δ . The sequence of DEL1 predicts a Thr residue at amino acid 29 of the mature protein, showing that it represents the blood-group-s form of δ .

An interrupting sequence in DEL1 cDNA

The 5' end of the DEL1 cDNA (nts 1–89) is identical in sequence with nts 5–93 of the ALP1 sialoglycoprotein α cDNA (see Fig. 3), and includes the codon for the methionine residue that initiates translation of the cleaved

Fig. 2. Sequence of ALP1 and DEL1 cDNA clones

The numbering below the DNA sequences refers to the nucleotide sequences. The protein sequence is shown in single-letter code, and the numbering is shown above this sequence with amino acid residue 1 being the *N*-terminal residue of the mature protein. (a) Sequence of ALP1. (b) Sequence of DEL1. The cDNA sequence corresponding to sialoglycoprotein δ is preceded by 422 nts of α -1-globin cDNA (sequence not shown) from Lys-40 to the poly(A) tail. The interrupting sequence in DEL1 cDNA is underlined.

present in the centre of the region that gives rise to the *N*-terminal cleaved leader sequence of α , and contains an in-phase termination codon at nt 107. Although there are other methionine codons on the 5' border of this sequence, these are all closely followed by in-frame termination codons. The next downstream methionine

Fig. 3. Homology between ALP1 and DEL1

Homology between ALP1 and DEL1 was determined by using the alignment computer program NUCALN (Wilbur & Lipman, 1983). The known nucleotide differences between the blood-group-M and -N forms of α cDNA are also indicated.

codon in-phase with the δ coding sequence is at nt 215, which gives rise to Met-8 of the expressed protein. This product would lack an *N*-terminal leader sequence and is unlikely to be incorporated into the membrane. Even if it were inserted into the membrane [for example by the mediation of the membrane-spanning sequence, as is the case of sialoglycoprotein β (High & Tanner, 1987)], the product would lack the *N*-terminal seven amino acids of δ , which carry blood-group-N antigen activity. Northern-blot analysis of polyadenylated RNA from the K562 erythroleukaemic cell line was carried out with the use of a probe containing nts 118–193 of DEL1 that contained 56 nts of the interrupting sequence and 20 nts of the downstream coding sequence. Under high stringency conditions no transcripts binding the probe were detected, although the RNA contained all four sialoglycoprotein α -related mRNA species (see Fig. 4). This suggests that the RNA containing this sequence is not present at significant abundance in polyadenylated RNA from this source.

The origin of this sequence interrupting two adjacent regions of sialoglycoprotein δ cDNA is not clear. One possibility suggested by the Northern-blot analysis was that the cDNA could represent a partially processed

mRNA and the interrupting sequence is an intron that would normally be excised from the mRNA and thus be present in very small amounts in the polyadenylated RNA. However, although the 3' end of the interrupting sequence contains the invariant AG found at 3' splice sites, the 5' end of the sequence contains GA rather than the invariant GT found at 5' splice sites. Aebi *et al.* (1987) have shown that the presence of GA in the consensus 5' splice sequence abolishes splicing, with inefficient cleavage at the 5' splice site and the accumulation of lariat intermediates.

It is known that artifacts can arise during the production and cloning of cDNA, and the interrupting sequence may also well be the result of an artifact of this type. However, an alternative and interesting possibility also presents itself. The human reticulocyte RNA used as the source for the cDNA library was obtained from sickle-cell-anaemia patients undergoing exchange transfusion (Conboy *et al.*, 1986). Because of the well-known prevalence of sickle-cell anaemia among black individuals, it is very likely that material from members of this population was present in the RNA preparation. Individuals with erythrocytes that do not carry sialoglycoprotein δ (S-s- erythrocytes) are also not

Fig. 4. Northern-blot analysis of K562 mRNA with sialoglycoprotein α and δ cDNA probes

Lanes a–c, autoradiographs of Northern blots of K562 mRNA probed with: lane a, part of the 3' non-coding region (nts 511–608) of the δ cDNA clone DEL1; lane b, nts 1–783 of the α cDNA (ALP1a); lane c, as lane b but with longer exposure of the autoradiograph to show the 3200-nt band. An additional Northern blot was probed with the intervening sequence (nts 118–193) of DEL1, but no bands were detected (results not shown).

uncommon in the black population, and the homozygous condition occurs with frequencies ranging from 1 to 35% in different Negroid populations (Lowe & Moores, 1972). It is possible that the mRNA corresponding to DEL1 is the product of the variant sialoglycoprotein δ gene of one of these individuals who do not express sialoglycoprotein δ on their erythrocytes. If this is the case the condition could result from a splicing defect that does not allow the normal excision of the interrupting sequence in DEL1 that corresponds with an intron in the normal sialoglycoprotein δ gene. It should be noted that this could arise by a T $\rightarrow$ A mutation at nt 91 in the donor splice site of this putative intron. The presence of T at this point would generate the sequence CAG/GTGAAT, which is very close to the consensus sequence for 5' splice

sites, CAG/GTGAGT (reviewed by Padgett *et al.*, 1986). If normal δ precursor RNA does contain a T residue at this position, splicing out of this putative intron at this 5' splice site with the AG at the 3' end of the sequence (nts 172–173) as the 3' splice site would lead to an mRNA and protein sequence in the region of the N-terminal leader that is identical with that of blood-group-N type α . Alternatively, the AG at nts 169–170 may act as a 3' splice site, since the sequence around these residues fits the 3' splice site consensus sequence CAG/G. In this case splicing would result in the insertion of a Gly residue within the N-terminal leader, yielding the sequence ...VLLSGEIVS..., which may still allow membrane insertion of δ . Although the 5' sides of these putative 3' splice sites are not, as is often the case, particularly rich in pyrimidine residues, a short pyrimidine tract does occur in this region. Clearly further studies will be

needed to establish whether DEL1 does indeed represent the RNA product of a variant δ gene. Taking account of the interrupting sequence, the region of DEL1 representing normal δ mRNA can be calculated to be 523 or 526 nts.

After the completion of this study Siebert & Fukuda (1987) reported the sequence of a cDNA clone for δ that contains the coding region and most of the 3' non-coding region of what appears to be normal δ mRNA. This sequence shows that δ mRNA corresponds to nts 1–89 and nts 174 onwards of our DEL1 cDNA, with δ and the blood-group-N form of α having identical N-terminal sequences. The coding sequence of their δ cDNA is identical with that reported in the present paper except for the inserted region within our sequence (nts 90–173) and a nucleotide change at nt 387. The latter predicts a Thr residue at amino acid 69 of δ , whereas our sequence predicts Ser at this residue. The protein sequence determined by Blanchard *et al.* (1987a) is in agreement with our prediction. A further difference between the two cDNA sequences is in the 3' non-coding regions between nts 435 and 447. The sequence reported by Siebert & Fukuda (1987) suggests the presence of a 16-nt sequence repeated in this region. It is not clear at this stage which sequence represents the mRNA structure in this region.

Sialoglycoprotein α and δ mRNAs

Siebert & Fukuda (1986a) found four sialoglycoprotein- α -related mRNAs in erythroid cells and suggested on the basis of indirect evidence that the smallest of these (600 nt) was the mRNA for δ . Northern-blot analysis was carried out on polyadenylated RNA from the K562 erythroleukaemic cell line (Fig. 4). When ALP1a cDNA, which contains the entire coding region of sialoglycoprotein α , was used as a probe four mRNAs of 3200, 2000, 1200 and 700 nts were detected (Fig. 4), which were clearly the same as those found by Siebert & Fukuda (1986a). A probe containing nt 511 to the polyadenylation site of DEL1 was used to search for the sialoglycoprotein δ mRNA (Fig. 4). This probe detected only the smallest (700 nt) mRNA. Although this result clearly shows that the 700-nt mRNA codes for sialoglycoprotein δ , the relationship between the three larger mRNAs remain unclear. Siebert & Fukuda (1986a) have shown that they all contain at least part of the coding region of α , including the region giving rise to the part of the N-terminal amino acid sequence represented by oligonucleotide NTA and the C-terminal sequence represented by oligonucleotide CTA. The similarity in size between ALP1 cDNA (967 nts) and the 1200-nt mRNA makes it likely that the cDNA represents this mRNA.

Relationship between sialoglycoprotein α and δ genes

Alignment of the ALP1 and the DEL1 cDNA sequences (Fig. 3) shows very strong homology at the DNA level within the coding regions of the two cDNAs, with the non-homologous regions clearly originating from insertions and deletions. Omitting consideration of the interrupting sequence in DEL1 discussed above, the 5' non-coding regions of the cDNAs are identical, and this identity extends to nt 192 of the α clone with the exception of a single silent nucleotide change at nt 149. The α cDNA then contains a 96-nt insertion relative to

the δ clone. A 40-nt region of homology follows, and this contains seven nucleotide changes, each of which results in an amino acid difference between the two proteins. This region carries the Ss antigen activity of sialoglycoprotein δ , and is located in the extracellular portion of the proteins adjacent to the membrane-spanning domains. A short (9-nt) insertion in the δ cDNA follows before an 85-nt region of strong homology that codes for the membrane-spanning domains of the two proteins. This region contains six nucleotide changes, which result in four amino acid differences between the proteins. The remaining coding region corresponding to the C-terminal cytoplasmic domain of α is not present in sialoglycoprotein δ . Finally, the 3' non-coding regions of the two mRNAs are quite unrelated. The different homologous regions of the two proteins show different degrees of conservation, and the near-identity at the DNA level of nearly 200 nts at the 5' ends of the cDNA that contain the 5' non-coding region and the region coding for the N-terminal leader sequence and the N-terminal 26 amino acid residues of the mature proteins is very striking.

There is strong genetic evidence that α and δ are coded by separate but closely linked genes (Race & Sanger, 1975; Mawby *et al.*, 1981), and this has been confirmed by Southern-blotting studies (Siebert & Fukuda, 1986b). It is clear that the two proteins resulted from a recent

gene duplication. It is noteworthy that the 9-nt insertion in δ cDNA together with the amino acid changes in the short region of homology 5' to this insertion has the effect of extending the hydrophobic membrane-spanning domain of sialoglycoprotein δ by eight amino acid residues. The conserved sequence RRLIK, which probably marks the cytoplasmic boundary of the membrane domain and substantial conservation of the homologous membrane-bound segment, suggests that this portion of the two proteins has a similar location with respect to the membrane and implies that the hydrophobic extension may protrude from the extracellular side of the membrane

for a substantial distance (1.2 nm if in a helical conformation). It seems unlikely that this distance could be accommodated by local changes in the thickness of the lipid bilayer, and this hydrophobic extension could interact with other membrane proteins. There is some evidence (Race & Sanger, 1975) for an association of sialoglycoprotein δ with the Rhesus-blood-group

antigens, since there is a diminished expression of Ss antigens in Rh_{null} erythrocytes (which lack the proteins carrying Rhesus antigens). Dahr *et al.* (1987a) have suggested that the blood-group-U antigen, which is located on the portion of sialoglycoprotein δ close to the membrane surface, might be formed by combination of δ and the Rhesus protein. The hydrophobic extension on δ may form the site of association of δ and the Rhesus protein.

Hybrids of sialoglycoprotein α and δ

The erythrocytes of certain rare individuals contain sialoglycoproteins which are hybrids of α and δ (reviewed by Anstee *et al.*, 1982a) and probably arise from chromosomal misalignment with unequal crossing-over between the α and δ genes (Mawby *et al.*, 1981). This mechanism is similar to that which gives rise to the Lepore and anti-Lepore types of hybrid δ/β globins. The Lepore-like (α - δ) hybrids [for example, Mi^v (Anstee *et al.*, 1979)] have their N-termini derived from α and C-termini derived from δ , and are associated with the deletion of normal α and δ . Although the amino acid sequences of the hybrid sialoglycoproteins have not been established, studies on antigenic determinants present on the Mi^v hybrid suggest that the N-terminal portion comprises a region of blood-group-M type α from residues 1 to approx. 54 (Anstee *et al.*, 1982b), and this is joined together with the portion of sialoglycoprotein δ including the s antigen (specified by Thr-29) to the C-terminus. Inspection of the alignment in Fig. 3 suggests that the hybrid results from cross-over within the six homologous nucleotides on the 5' side of nt 295 of ALP 1 (within the coding for Arg-61 of α) and nt 279 of DEL 1 (within the codon for Thr-29 of δ), or possibly an homologous intron immediately 5' to this region. The amino acid sequence of the hybrid around the cross-over point would be:

residue no. of α	54	58	29	33	residue no. of δ
	P	E	E	T/G	E/T
	G	E	T	G	Q
	L	V			

where / encloses the region of the cross-over. The product retains the sequence Gly-Glu on the N-terminal side of the Thr residue determining s antigen activity in δ (Thr-29) and this may be important in the expression of the s antigenic determinant.

The anti-Lepore-like (δ - α) hybrid sialoglycoproteins are of two types. One is found in Ph (or Dantu) erythrocytes (Tanner *et al.*, 1980) and the other is found in St^a-positive erythrocytes (Anstee *et al.*, 1982b). The amino acid sequence of the St^a hybrid has been determined (Blanchard *et al.*, 1987b) and the sequence at the cross-over:

residue no. of δ	22	26	61	65	residue no. of α
	S	S	Q	T	N/G
	E	R	V	Q	L
	A				

suggests that it originates from unequal cross-over in the same location as that giving rise to the Mi^v Lepore-like (α - δ) hybrid. It is possible that the Mi^v and St^a sialoglycoproteins are the Lepore- and anti-Lepore-like products of the same cross-over event. The sequence at the junction in the Dantu (δ - α) hybrids has also been determined (Dahr *et al.*, 1987b):

residue no. of δ	34	38	72	76	residue no. of α
	H	R	F	T	V/P
	E	I	T	L	I

This would be consistent with the result of an unequal cross-over occurring within the 5-nt homology between nts 323–327 of ALP 1 and nts 307–311 of DEL 1 (Fig. 3) or a homologous intron at this point.

Since the present paper was submitted, Rahuel *et al.* (1988) have reported the sequence of an α cDNA clone corresponding to the blood-group-M form of α . Their clone corresponds to nts 123–967 of ALP1. Differences

occur at nts 127 and 128 owing to the different antigenic forms encoded by the two cDNAs. There is also a silent mutation at nt 149 and a single base change at nt 584 in the 3' non-coding region.

We are grateful to Dr. N. Mohandas for the reticulocyte cDNA library and to Dr. D. J. Anstee and Dr. P. Judson for supplying K562 cells. C.G.T. is the recipient of an S.E.R.C. Studentship.

REFERENCES

- Aebi, A., Hornig, H. & Weissmann, C. (1987) *Cell* **50**, 237–246
- Anstee, D. J., Mawby, W. J. & Tanner, M. J. A. (1979) *Biochem. J.* **183**, 193–203
- Anstee, D. J., Mawby, W. J. & Tanner, M. J. A. (1982a) in *Membranes and Transport* (Martonosi, A., ed.), vol. 2, pp. 427–433, Plenum Press, New York
- Anstee, D. J., Mawby, W. J., Parsons, S. F., Tanner, M. J. A. & Giles, C. M. (1982b) *J. Immunogenet.* **9**, 51–56
- Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408–1412
- Blanchard, D., Dahr, W., Hummel, M., Latron, F., Beyreuther, K. & Cartron, J.-P. (1987a) *J. Biol. Chem.* **262**, 5808–5811
- Blanchard, D., Dahr, W., Beyreuther, K., Moulds, J. & Cartron, J.-P. (1987b) *Eur. J. Biochem.* **167**, 361–366
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9512–9516
- Dahr, W. & Uhlenbruck, G. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* **359**, 835–843
- Dahr, W., Beyreuther, K., Steinbach, H., Gielen, W. & Kruger, J. (1980) *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 895–906
- Dahr, W., Blanchard, D., Hummel, M., Cartron, J.-P. & Beyreuther, K. (1985) *Proc. Int. Symp. Glycoconjugates 8th* (Davidson, E. A., Williams, J. C. & Di Ferrante, N. M., eds.), pp. 21–22, Praeger Scientific, New York
- Dahr, W., Kordowicz, M., Moulds, J., Gielen, W., Lebeck, L. & Kruger, J. (1987a) *Blut* **54**, 13–24
- Dahr, W., Beyreuther, K., Moulds, J. & Unger, P. (1987b) *Eur. J. Biochem.* **166**, 31–36
- High, S. & Tanner, M. J. A. (1987) *Biochem. J.* **243**, 277–280
- Huynh, T. V., Young, R. A. C. & Davis, R. W. (1985) in *DNA Cloning* (Glover, D. M., ed.), vol. 1, pp. 49–78, IRL Press, Oxford
- Lowe, R. F. & Moores, P. (1972) *Hum. Hered.* **22**, 344–350
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 202–203, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Mawby, W. J., Anstee, D. J. & Tanner, M. J. A. (1981) *Nature (London)* **291**, 161–162
- Michelson, A. M. & Orkin, S. H. (1980) *Cell* **22**, 371–377
- Moulds, J. (1983) in *Immunochemistry of the MNSs Blood Group System in Red Cell Membrane Glycoconjugates and Related Genetic Markers* (Cartron, J.-P., Rouger, P. & Salmon, C., eds.), pp. 17–26, Librairie Arnette, Paris
- Mueller, T. J. & Morrison, M. (1981) *Erythrocyte Membranes*, vol. 2: Recent Clinical and Experimental Advances, pp. 95–112, A. R. Liss, New York
- Owens, J. W., Mueller, T. J. & Morrison, M. (1980) *Arch. Biochem. Biophys.* **204**, 247–254
- Padgett, R. A., Grabowski, T. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150
- Race, R. R. & Sanger, R. (1975) *Blood Groups in Man*, 6th edn., Blackwell Scientific Publications, Oxford
- Rahuel, C., London, J., d'Auriol, L., Mattei, M.-G., Tournamille, C., Skrzynia, C., Lebouc, Y., Galibert, F. & Cartron, J.-P. (1988) *Eur. J. Biochem.* **172**, 147–153
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Siebert, P. D. & Fukuda, M. (1986a) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1665–1669
- Siebert, P. D. & Fukuda, M. (1986b) *J. Biol. Chem.* **261**, 12433–12436
- Siebert, P. D. & Fukuda, M. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6735–6739
- Tanner, M. J. A., Anstee, D. J. & Mawby, W. J. (1980) *Biochem. J.* **187**, 493–500
- Tokunaga, E., Sasakawa, S., Tamaka, K., Kawamata, H., Giles, C. M., Iken, E. W., Poole, J., Anstee, D. J., Mawby, W. J. & Tanner, M. J. A. (1979) *J. Immunogenet.* **6**, 383–390
- Tomita, M., Furthmayr, H. & Marchesi, V. T. (1978) *Biochemistry* **17**, 4756–4770
- Wasniowska, K., Drzeniek, Z. & Lisowska, E. (1977) *Biochem. Biophys. Res. Commun.* **76**, 385–390
- Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 726–730

Received 6 November 1987/1 February 1988; accepted 25 May 1988