Biochem. J. (1987) 243, 277–280 (Printed in Great Britain)

Human erythrocyte membrane sialoglycoprotein β

The cDNA sequence suggests the absence of a cleaved *N*-terminal signal sequence

Stephen HIGH and Michael J. A. TANNER
Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K.

We have isolated cDNA clones corresponding to the human erythrocyte membrane sialoglycoprotein β. The clones encompass the coding region for the protein, 120 residues of the 5′ non-coding region and the 3′ non-coding region. The cDNA sequence suggests that sialoglycoprotein β is not translated with the cleaved *N*-terminal signal sequence usual in a membrane protein of this type. Sialoglycoprotein β or a closely related homologue is present in human kidney as well as erythroid cells.

INTRODUCTION

The human erythrocyte membrane contains four readily detected sialoglycoproteins [denoted α, β, γ and δ; see Anstee *et al.* (1979) for nomenclature]. The most abundant are sialoglycoproteins α (glycophorin A) and δ (glycophorin B), and these have related amino acid and DNA sequences (Furthmayr, 1978; Siebert & Fukuda, 1986*b*). Sialoglycoproteins β (glycophorin C or glycoconnectin) and γ form a distinct group and may be related to each other (Anstee *et al.*, 1984). Unlike sialoglycoproteins α and δ, sialoglycoproteins β and γ are associated with the erythrocyte cytoskeleton, and there is evidence that sialoglycoprotein β is associated with this structure through band 4.1 protein (Mueller & Morrison, 1981). The observation that individuals with erythrocytes that lack both sialoglycoproteins β and γ (Leach phenotype) have elliptocytosis (Anstee *et al.*, 1984) suggests an involvement of these proteins in the maintenance of erythrocyte shape.

Sialoglycoprotein β is orientated with *N*-terminus extracellular and *C*-terminus in the cytoplasm in a manner similar to sialoglycoprotein α. The amino acid sequence of the heavily glycosylated *N*-terminal portion has been determined by Dahr *et al.* (1982). This sequence was obtained from a fragment isolated by trypsin treatment of intact erythrocytes, thus confirming its extracellular location. The only tyrosine residues in the protein are intracellular, since they can only be labelled from the cytoplasmic side of the erythrocyte membrane (Mueller & Morrison, 1974). These tyrosine residues are *C*-terminal to the single membrane-spanning segment in the protein (Colin *et al.*, 1986). Colin *et al.* (1986) used the partial amino acid sequence to construct suitable oligonucleotide probes and isolated a cDNA for sialoglycoprotein β. They determined the complete amino acid sequence of the protein from the sequence of a 380 bp cDNA clone. This cDNA lacked the first two codons of the mature protein and thus contained no 5′ non-coding sequence, and also contained very little of the 3′ non-coding sequence.

We have isolated and sequenced cDNA clones that represent 1036 nt of the mRNA for sialoglycoprotein β (for definition of nt see Fig. 2 legend). The clones extend from 120 nt to the 5′ side of the start of the coding region of the protein to the polyadenylation site of the mRNA. Inspection of this sequence suggests that sialoglycoprotein β, unlike sialoglycoprotein α (Siebert & Fukuda, 1986*a*), lacks the *N*-terminal cleaved signal sequence for membrane insertion usually found on a protein of this type.

MATERIALS AND METHODS

Materials

Restriction enzymes and DNA-modifying enzymes were from Amersham International, Amersham, Bucks., U.K., Pharmacia Biochemicals, Milton Keynes, Bucks., U.K., and Gibco BRL, Paisley, Strathclyde, Scotland, U.K. [^{32}P]dATP was from Amersham International. Reticulocyte-rich blood was a pool of samples from individuals with sickle-cell anaemia obtained from the Central Middlesex Hospital, London, U.K. Human kidney tissue that was unsuitable for transplantation was obtained from the U.K. Transplant Service, Bristol, U.K. K562 and HEL 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 foetal liver cDNA library in λgt11 (Clontech Laboratories, Palo Alto, CA, U.S.A.) was screened with polynucleotide kinase-labelled synthetic oligonucleotide probes (Maniatis *et al.*, 1982). The screening followed the method of Mason & Williams (1985) except that the stringent washes were at 9 °C below the estimated T_d. [T_d was calculated as described by Meinkoth & Wahl (1984).]

Oligonucleotides C1, C2, NC1, NC2 and NC3 were synthesized manually by the method of Sproat & Gait (1984). Oligonucleotides C3 and NC4 were synthesized by using a Dupont Coder 300 DNA synthesizer. The sequences of the oligonucleotides synthesized were:

C1	55	GATCCGGGGATGGCCTCTGCCTCCACC
C2	82	ACAATGCATACTACCACCATTGC
C3	505	CCCGAGATAGCCACCTGG
NC1	24	GTTGGGGCTTCTCGTCGACCACAT
NC2	89	TGCATTGTGGTGGAG
NC3	378	GTACTCCTTTCTGCTGCTATCACC
NC4	645	CTCCCACCTCCGTGGAGC

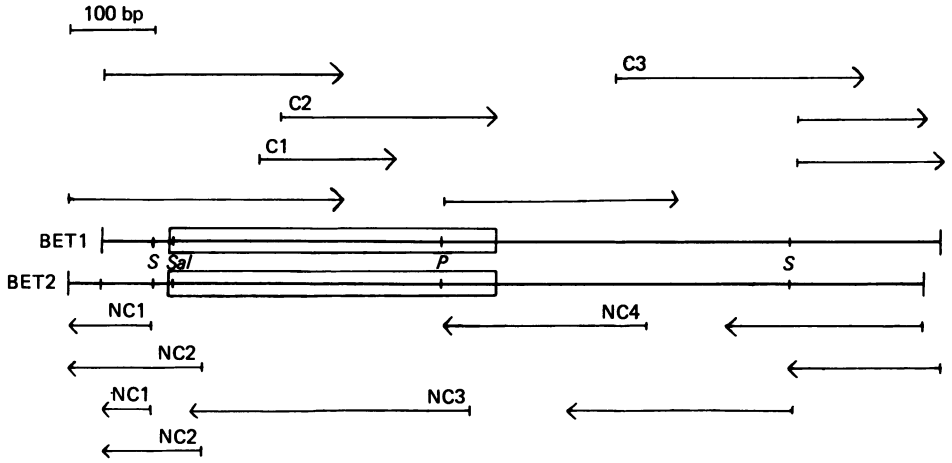

Fig. 1. Restriction map and sequencing strategy for sialoglycoprotein β clones

A partial restriction map of BET1 and BET2 for commonly used restriction enzymes is shown (*S* = *Sma*I, *Sal* = *Sal*I and *P* = *Pst*I). The arrows indicate the direction of sequencing. Unlabelled arrows represent sequences derived from intact cDNAs and restriction fragments sequenced from their ends by using M13 universal primer. The labelled arrows represent sequences obtained by priming with the oligonucleotide indicated. Details of the oligonucleotides are given in the Materials and methods section. The open box shows the position of the coding region.

Oligonucleotides C1–C3 correspond to the coding strand of the cDNA, and oligonucleotides NC1–NC4 to the complement of the cDNA sequence.

Nucleotide sequence analysis

The cDNA inserts were excised from λgt11 with *Eco*RI and inserted into plasmid pUC9. The whole cDNA inserts and restriction fragments of them were subcloned into M13mp10 or M13mp11 for DNA sequencing by the dideoxy method of Sanger *et al.* (1977). Oligonucleotides were used as internal primers to complement the use of the M13 universal primer on the intact cDNAs and restriction fragments of them (see Fig. 1). The sequence was determined at least once and on average 2.7 times on each strand of the DNA.

```
CCCAGGTGCCGCTTCCTCTCGCCGCCGAGGGTCAGGAGCCCGGGAGCGCGACCCTCCCCCGGCCCGGCCTGGCCCGGCCTGGCCAGTCCCCGCGGTCTCTGCCCGGGCTGACGCCCAGGA
      -110      -100      -90       -80       -70       -60       -50       -40       -30       -20       -10

 M  W  S  T  R  S  P  N  S  T  A  W  P  L  S  L  E  P  D  P  G  M  A  S  A  S  T  T  M  H  T  T  T  I  A  E  P  D  P  G
ATGTGGTCGACGAGAAGCCCCAACAGCACGGCGTGGCCTCTCAGCCTCGAGCCTGATCCGGGGATGGCCTCTGCCTCCACCACAATGCATACTACCACCATTGCAGAGCCTGATCCAGGG
       10        20        30        40        50        60        70        80        90       100       110       120

 M  S  G  W  P  D  G  R  M  E  T  S  T  P  T  I  M  D  I  V  V  I  A  G  V  I  A  A  V  A  I  V  L  V  S  L  L  F  V  M
ATGTCTGGATGGCCGGATGGCAGAATGGAGACCTCCACCCCCACCATAATGGACATTGTCGTCATTGCAGGTGTGATTGCTGCTGTGGCCATCGTCCTAGTCTCCCTCCTCTTCGTCATG
      130       140       150       160       170       180       190       200       210       220       230       240

 L  R  Y  M  Y  R  H  K  G  T  Y  H  T  N  E  A  K  G  T  E  F  A  E  S  A  D  A  A  L  Q  G  D  P  A  L  Q  D  A  G  D
CTGCGCTACATGTACCGGCACAAGGGCACGTACCACACCAATGAGGCCAAGGGCACGGAGTTTGCTGAGAGTGCAGATGCAGCCCTGCAGGGAGACCCTGCCCTCCAAGATGCTGGTGAT
      250       260       270       280       290       300       310       320       330  C    340       350       360

 S  S  R  K  E  Y  F  I  *
AGCAGCAGAAAGGAGTACTTTATTTGAGGGACAACAGACTTCACTTCCCTGAATGCCTCCCCCATCTCCATCAGGAAAAATACACCCCATCGCCCAGCACCCCTGCTGATACCACCAGAC
      370       380       390       400       410       420       430       440       450       460       470       480

AGAGAGAGAGCACTTGATTCTTCCCGAGATAGCCACCTGGAAACACTAGGTGCCTGCCCAGGGAGGAACGGAGGAGGACTCGCGCTACAAGAGGCCACTCCCAGGGACCCAGGGAGGC
      490       500       510       520       530       540       550       560       570       580       590       600

GATGGCCACCCCAGAGGCCACCTTTTGCTCCACCGGAGGTGGGAGAAAATCTGGGCACGATCGGGCCCCCTGGGCAGTGCAGGACAACATCAGCTCACTGGCAGGAAAGTCCTTGTTGAG
      610       620       630       640       650       660       670       680       690       700       710       720

GGTGAGGGGGTGCTGGGGTACCCGGGGGCTGGGGAAGCAAGGAAATAAGTCATCTGTATGCTGACTGGGGATAATGGCATCAAATGTCAGTCCTTGACATTTGGGGGGAACAGCAGGTGC
      730       740       750       760       770       780       790       800       810  A    820       830       840

CAGAGCTAAAAGGTACCTTTGTCTGCCATTGATCCAGCTCAGAACGATTGGAAATAAATTTGAAATGTAACCGAGCAAAAA
      850       860       870       880       890       900       910       920
```

Fig. 2. Sequence of cDNA for sialoglycoprotein β

The sequence shown is a composite of the clones BET1 and BET2. The numbering refers to the nucleotide sequence, with nt 1–3 being the position of the *N*-terminal methionine of the mature protein. The protein sequence is shown in the single-letter code. The polymorphisms at nt 333 and nt 813 are indicated. The nucleotides found at these positions in BET1 are shown in the body of the sequence, and those found in BET2 are shown below the main sequence.

Northern blot analysis

Reticulocyte RNA was prepared by using the LiCl/urea method of Clemens (1984). K562 cell, HEL cell and human kidney RNAs were 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 done with formaldehyde-containing gels (Maniatis *et al.*, 1982) and 5 μg of polyadenylated RNA per track.

RESULTS AND DISCUSSION

Isolation and sequencing of sialoglycoprotein β cDNA clones

A human foetal liver cDNA library was screened with an oligonucleotide probe, NC1, corresponding to the nucleotide sequence of the *N*-terminal sequence of the protein (Colin *et al.*, 1986). Two cDNA clones were isolated (BET1 and BET2) and shown to correspond to sialoglycoprotein β by their hybridization to oligonucleotides C1, C2 and NC3, which correspond to other portions of the coding sequence of sialoglycoprotein β determined by Colin *et al.* (1986). Restriction mapping indicated that these two clones overlapped each other, and they were sequenced independently by using the sequencing strategy shown in Fig. 1. The composite sequence of 1041 nt obtained from BET1 and BET2 is shown in Fig. 2. BET1 extends from nt −79 to the poly(A) tail at the 3′ end of the mRNA, the polyadenylation recognition signal AATAAA being present 24 nt from the poly(A) tail. BET2 extends from the 5′ side of BET1, from nt −120 to nt 901. The DNA sequences of BET1 and BET2 were found to differ at two points, at nt 333 and at nt 813, but the change at nt 333, although in the coding region, is a silent mutation. The BET1 and BET2 cDNA clones probably represent polymorphic forms of sialoglycopotein β mRNA. The sequence of both clones contains an A residue in the 3′ non-coding region at nt 391. Colin *et al.* (1986) found a T residue at this point in their cDNA sequence.

Structure of sialoglycoprotein β mRNA

Northern blots of polyadenylated RNA from human reticulocytes and from the human erythroleukaemic cell lines K562 and HEL probed with the BET2 cDNA showed a major mRNA species of 1.2 kb (Fig. 3, lanes a–c). This result is similar to that reported by Colin *et al.* (1986), who obtained a 1.4 kb band on hybridization with their probe. Fig. 3 also shows the presence of a much weaker band at 0.6 kb. It is possible that this represents the mRNA of a related but less abundant sialoglycoprotein, such as sialoglycoprotein γ.

The *N*-terminus of sialoglycoprotein β is a methionine residue (Dahr *et al.*, 1982), and the codon for this starts 121 nt from the 5′ end of the sequence shown in Fig. 2. It is preceded by an in-phase termination codon, TGA, located 12 nt earlier. No other methionine codon is present in any reading frame upstream of this methionine. This strongly suggests that the methionine residue at the *N*-terminus of the mature protein is the point of initiation of translation even though the mRNA sequence does not conform to the consensus sequence AXXATGG that is usually found around initiation codons (Kozak, 1981).

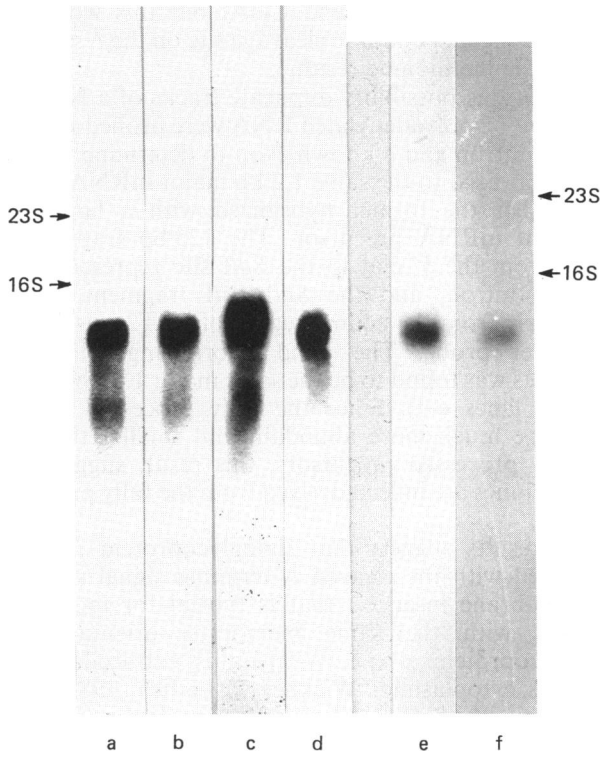

Fig. 3. Northern blot analysis of tissue mRNA with sialoglycoprotein β cDNA probes

Lanes a–d, autoradiograph of a Northern blot probed with BET2 and washed at high stringency (15 mM-NaCl/1.5 mM-sodium citrate buffer, pH 7.0, containing 1% SDS at 65 °C). The polyadenylated RNA samples are from: lane a, human reticulocytes; lane b, HEL cell line; lane c, K562 cell line; lane d, human kidney. Lanes e–f, polyadenylated RNA from K562 cells separated on a different gel to that shown in lanes a–d. Autoradiograph of Northern blot probed with: lane 3, the *Sal*I–*Pst*I fragment containing the coding region of sialoglycoprotein β; lane f, the 125 bp fragment from the 5′ end of BET2 to the *Sal*I site containing the 5′ non-coding region. The blot was washed in 15 mM-NaCl/1.5 mM-sodium citrate buffer, pH 7.0, containing 1% SDS at 60 °C.

Examination of the sequence on the 5′ side of the ATG at nt 1 showed a possible splice-junction acceptor site CCCAG/GAATG at nt −2 (Mount, 1982). No consensus splice-donor-site sequence $\begin{matrix}C\\A\end{matrix}AG/GT\begin{matrix}A\\G\end{matrix}AGT$ is present 5′ to this acceptor-site sequence in the cDNA clones. However, this splice-donor sequence could be present in the intact RNA from which the cDNA clones originated, since our cDNA sequence probably does not extend to the 5′ end of the RNA. If this donor sequence were present, splicing at this junction could generate a glycine codon at amino acid residue −1. A glycine residue is often found on the *N*-terminal side of the point of cleavage of signal peptides (von Heijne, 1983). Thus the possibility remained that both BET1 and BET2 were cDNA copies of a partially processed mRNA species with the signal sequence in an exon 5′ to the sequence shown in Fig. 2. In this case the in-phase stop codon would be spliced out in the fully processed message. Alternatively, splicing could lead to the generation of the

arginine codons AGA and CGA, but this would not provide a signal peptidase cleavage site on the *N*-terminal side of the methionine residue.

To test this possibility duplicate tracks of a Northern blot of K562 polyadenylated RNA were probed with this possible intron and a known exon to determine whether they hybridized to the same 1.2 kb major mRNA species or whether the former hybridized with a larger less abundant mRNA precursor. The 125 bp fragment of BET2 from the 5′ end to the *Sal*I site represented the putative intron, and the *Sal*I–*Pst*I fragment derived exclusively from the coding region of BET2 was used as the other probe. The band hybridizing with both fragments was found to be the same major 1.2 kb mRNA (Fig. 3, lanes e–f). Since the fully processed mRNA should be much more abundant and smaller than any partially processed precursors, this result suggests the cDNA clones are indeed derived from the fully processed RNA.

Our results suggest that sialoglycoprotein β is not translated with the cleaved *N*-terminal signal sequence for membrane insertion that is typical for monotopic proteins with the same membrane orientation as sialoglycoprotein β (i.e. *N*-terminus extracellular, *C*-terminus cytoplasmic; Wickner & Lodish, 1985). This contrasts with the case of sialoglycoprotein α, in which the cDNA sequence predicts the presence of a cleaved *N*-terminal signal sequence of 19 amino acid residues (Siebert & Fukuda, 1986*a*).

When the cDNA clone BET2 was used to probe a Northern blot of human kidney mRNA it was found to hybridize with a 1.2 kb mRNA species even under conditions of full stringency (Fig. 3, lane d). This suggests that sialoglycoprotein β or a very close homologue is present in human kidney, and is consistent with the observation by Hawkins *et al.* (1985) that a monoclonal antibody against sialoglycoprotein β binds to human renal capillary endothelium. It is clear that sialoglycoprotein β is not an erythroid-cell-specific molecule. The less abundant 0.6 kb mRNA found in erythroid cells was not detected in kidney mRNA even under low-stringency hybridization conditions.

We are grateful to Dr. B. Bradley for supplying human kidney tissue, Dr. Sally Davies for samples of reticulocyte-rich blood and Dr. D. Anstee and Dr. P. Judson for K562 and HEL cells. S.H. was the recipient of a Medical Research Council Studentship.

REFERENCES

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., Johnson, P., Bates, S. & Fraser, I. D. (1984) Biochem. J. **218**, 615–619
Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 1408–1412
Chirgwin, J. M., Przbyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
Clemens, M. J. (1984) in Transcription and Translation: A Practical Approach (Hames, B. D. & Higgins, S. J., eds.), pp. 211–230, IRL Press, Oxford
Colin, Y., Rahuel, 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., Kordowicz, M. & Kruger, J. (1982) Eur. J. Biochem. **125**, 57–62
Furthmayr, H. (1978) Nature (London) **271**, 519–524
Hawkins, P., Anderson, S. E., McKenzie, J. L., McLoughlin, K., Beard, M. E. J. & Hart, D. N. J. (1985) Transplant. Proc. **17**, 1697–1700
Kozak, M. (1981) Nucleic Acids Res. **9**, 5223–5252
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, pp. 202–203, Cold Spring Harbor Laboratory, Cold Spring Harbor
Mason, P. J. & Williams, J. G. (1985) in Nucleic Acid Hybridisation: A Practical Approach (Hames, B. D. & Higgins, S. J., eds.), pp. 113–137, IRL Press, Oxford
Meinkoth, J. & Wahl, G. (1984) Anal. Biochem. **138**, 267–284
Mount, S. M. (1982) Nucleic Acids Res. **10**, 459–472
Mueller, T. J. & Morrison, M. (1974) J. Biol. Chem. **249**, 7568–7573
Mueller, T. J. & Morrison, M. (1981) Erythrocyte Membranes, vol. 2: Recent Clinical and Experimental Advances, pp. 95–112, A. R. Liss, New York
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Siebert, P. D. & Fukuda, F. (1986*a*) Proc. Natl. Acad. Sci. U.S.A. **83**, 1665–1669
Siebert, P. D. & Fukuda, F. (1986*b*) J. Biol. Chem. **261**, 12433–12436
Sproat, B. S. & Gait, M. J. (1984) in Oligonucleotide Synthesis: A Practical Approach (Gait, M. J., ed.), pp. 83–115, IRL Press, Oxford
von Heijne, G. (1983) Eur. J. Biochem. **133**, 17–21
Wickner, W. & Lodish, H. F. (1985) Science **230**, 400–407

Received 24 November 1986/13 January 1987; accepted 23 January 1987