The mechanism of production of multiple mRNAs for human glyophorin A

Jawed Hamid and Alfred T.H. Burness*

Faculty of Medicine, Memorial University of Newfoundland, St John's, Newfoundland A1B 3V6, Canada

Received April 16, 1990; Revised and Accepted August 29, 1990

EMBL accession no. X51798

ABSTRACT

The major sialoglycoprotein in the human red cell surface membrane, glyophorin A is encoded by a single gene. However, this gene gives rise to three species of glyophorin A mRNA of sizes about 1.0, 1.7 and 2.8 kilobases in reticulocytes, foetal liver cells and erythroleukaemic K562 cells. In an investigation of how the three mRNAs originated, we showed by primer extension analysis that all three mRNAs in K562 cells had identical 5' termini and, by nucleotide sequencing of correlated cDNAs, that they had identical coding regions, except for the well-known glyophorin A^M-A^N polymorphism. However, we found also by sequencing the cDNAs that the mRNAs apparently differed from each other in the lengths of their 3' untranslated regions. This was confirmed by Northern blot analysis which also provided evidence that the three mRNAs originated by use of different polyadenylation signals of which seven were found in the longest cDNA we analyzed.

INTRODUCTION

Human erythrocytes contain at least four sialoglycoproteins known under a variety of nomenclatures such as glyophorins A, B, C and D (ref. 1). A number of biological properties have been ascribed to glyophorin A which is the most abundant of the red cell sialoglycoproteins (2). These properties include MN blood group activity (3) and receptor activity for several entities such as wheat germ agglutinin (4), the malarial parasite, *Plasmodium falciparum* (5) and several viruses, for example, influenza virus (6, 7), encephalomyocarditis virus (8, 9), reovirus (10) and bluetongue virus (11).

Glyophorin A has been extensively studied as a model membrane protein (12). It contains 131 amino acids organized into an extracellular, transmembrane and intracellular domain containing about 72, 20 and 39 amino acids, respectively (13). The extracellular domain is glycosylated containing one N-linked complex carbohydrate side chain and fifteen O-linked units which are predominantly tetrasaccharides (13, 14).

Glyophorin A is encoded by a single copy gene on chromosome 4 at q²⁸-q³¹ (Ref. 15, 16) but three different species of glyophorin A mRNA have been found in normal human

reticulocytes (17), in human foetal liver (16) and in human erythroleukaemic K562 cells (18) which express glyophorin A on their surface (19). The sizes of these three mRNAs were reported to be about 2.8, 1.7 and 1.0 kb (Ref. 18). Such multiple mRNAs could arise from a single gene in eukaryotic cells by utilization of multiple initiation or termination sites, differential processing at the 3' end of the pre-mRNA or by alternate splicing (20). To investigate in what ways the three glyophorin A mRNAs differ from each other and to decide which mechanism(s) gave rise to them, we compared the primary structures of several glyophorin A cDNAs and used Northern blot analysis to examine the structures of the three mRNAs. Based upon these observation, a mechanism for production of the three different mRNAs is proposed.

MATERIALS AND METHODS

Materials

Oligodeoxynucleotides were from the following sources: we synthesized the GPA-C oligonucleotide mixture manually; GPA-N1 and GPA-N2 were purchased from the Biotechnology Service Centre, The Hospital for Sick Children, Toronto, Ontario; and sequences GPA-MS, GPA-ML and GPA-L were obtained from the DNA Synthesis Laboratory, University of Calgary, Alberta. The complementary nucleotide sequences to these oligonucleotides are shown in Fig. 2a. Ultra pure agarose and restriction endonucleases were from Bethesda Research Laboratories; Ficoll, polyvinylpyrrolidone, bovine serum albumin, RNase A and ethidium bromide were from Sigma Chemical Company; salmon sperm DNA and nucleoside triphosphates were from Pharmacia; yeast tRNA was from Boehringer Mannheim; reverse transcriptase was from Life Sciences Inc.; nylon membranes (Hybond-N), [γ -³²P]ATP (> 5000 Ci/mMol) and [α -³²P]dCTP (> 3000 Ci/mMol) were from Amersham Ltd.; and T4 DNA ligase was from New England Biolabs. The source of other materials is indicated in the text.

cDNA Cloning

A K562 cell, oligo (dT)-primed, λ gt10 cDNA library (Clontech Laboratories) grown in *E. coli*, strain C600 Hfl (ref. 21) was screened using ³²P end-labelled (22) oligonucleotide GPA-N2.

* To whom correspondence should be addressed

Hybridization was performed at 50°C in a solution containing 6× SSC (1×SSC is 0.9 M NaCl in 0.03 M Na citrate), 10×Denhardt's solution (1×Denhardt's solution is 0.02% bovine serum albumin, 0.02% Ficoll and 0.02% polyvinylpyrrolidone), 0.1% SDS and yeast tRNA (100 µg/ml) as described previously (23). After hybridization for about 18 h, the duplicate filters were washed for 5 min in 6×SSC at 60°C and then exposed to Kodak X-Omat RP film for about 18 h in cassettes with 2 intensifying screens at -70°C.

DNA Sequencing

The *Eco*RI-generated cDNA fragments isolated from six positive clones were subcloned into the *Eco*RI site of pUC19 (Bethesda Research Laboratories) in early experiments, or later in the Bluescript vector (Stratagene). Plasmid DNA isolated from Bluescript subclones, using either caesium chloride density gradient centrifugation (24) or a quick plasmid DNA preparation method (25), was subjected to the exonuclease III-mung bean nuclease deletion procedure following the instructions of the manufacturer of the deletion kit (Stratagene), then religated and used to transform competent *E. coli*, XL1 Blue cells prepared by a modified method of Hanahan (26). Double-stranded and single-stranded plasmid DNA was nucleotide sequenced by the chain termination method of Sanger (27) using Sequenase (United States Biochemical Corporation). The sequencing strategy is shown in Fig. 2(c).

Southern blotting

Southern blotting (28) was performed as described by Maniatis *et al.* (24). Hybridizations with oligonucleotide GPA-N2 were performed as described above for cDNA cloning. The oligonucleotides GPA-N1 and GPA-C were hybridized in solutions of the same composition as used for GPA-N2 but at 37°C and 52°C, respectively. Following hybridization, the blots were washed for 5 min at 60°C, 52°C and 39°C for GPA-N2, GPA-N1 and GPA-C, respectively, and exposed to film as described for cDNA cloning. When the 0.8 kbp *Eco*RI fragment from λ-gpa6 was used as a probe, hybridization was performed for about 18 h at 42°C in a solution containing 6×SSC, 0.1% SDS, 200 µg/ml sonicated salmon sperm DNA and 5×Denhardt's solution. The membranes were then washed successively with 6×SSC in 0.1% SDS for 15 min then with 2×SSC in 0.1% SDS for 15 min and finally with 0.1×SSC in 0.1% SDS at 55°C throughout and autoradiographed as for cDNA cloning.

Northern blotting

About 10 µg poly A⁺ RNA prepared from K562 cell total RNA (19) using an oligo (dT) cellulose column (29), was electrophoresed on a denaturing 1.0% agarose gel containing 2.2 M formaldehyde (24) and transferred to a Hybond-N nylon membrane following the manufacture's instructions. Hybridization with oligonucleotide GPA-N2 was performed as described for screening the cDNA library. Oligonucleotide GPA-MS, GPA-ML and GPA-L hybridization was performed at 50°C in a solution containing 5×SSPE (1×SSPE is 0.15 M sodium chloride, 1 mM sodium phosphate and 1 mM EDTA), 5×Denhardt's solution, yeast tRNA (100 µg/ml), sonicated salmon sperm DNA (50 µg/ml) and 0.1% SDS. The nylon membranes were washed in 6×SSC containing 0.01% SDS for 5 min at 60°C for oligonucleotide GPA-N2. For oligonucleotides GPA-MS, GPA-ML and GPA-L the washing was performed for

1 h at ambient temperature in 2×SSC containing 0.1% SDS followed by 15 min at 53°C (about 20°C below their T_d). The membranes were then autoradiographed as described above for screening cDNA clones.

Primer extension analysis

K562 cell RNA (50–100 µg total or 5–10 µg poly A⁺ RNA) was mixed with 0.5–1.0 pmole ³²P-labelled GPA-N2 and denatured by heating at 80°C for 3 min followed by annealing at 42°C for 1 h. Seven units of reverse transcriptase and the four deoxynucleoside triphosphates (20 mM each) were added to the annealed RNA-primer complex which was then incubated at 42°C for 1 h, treated with DNase-free RNase A (20 µg) at 37°C for 20 min and extracted with phenol:chloroform (1:1). The product was then analyzed by electrophoresis on a sequencing gel containing 5% polyacrylamide and 7 M urea.

RESULTS

cDNA Cloning

Glycophorins A and B have identical amino acid residues in positions 1 to 26 (glycophorin A numbering is used) and are highly homologous in the region containing residues 59 to 72, but differ from each other in that residues 27 to about 58 are missing from glycophorin B (1). To avoid selecting glycophorin B cDNAs, a K562 cell cDNA library was screened with oligonucleotide GPA-N2 which was complementary to the sequence coding for amino acids 30 to 40 in glycophorin A (Fig. 2a). This resulted in the isolation of six clones containing glycophorin A sequences (designated λ-gpa1 and λ-gpa3 to λ-gpa7). Analysis by agarose gel electrophoresis of *Eco*RI-digested

Fig. 1. Agarose gel electrophoresis of DNA isolated from glycophorin A cDNA clones and digested with restriction endonuclease *Eco*RI. The DNA fragments were detected (a) by staining the gel with ethidium bromide, or (b) by hybridization with ³²P end-labelled oligonucleotide GPA-N2 after transfer by Southern blotting to a nylon membrane. Numbers above the lanes indicate clones λ-gpa1 and λ-gpa3 to λ-gpa7. Lane M contains *Hind* III-generated phage λ DNA fragments as molecular size markers run on the same gel. The position and sizes in kbp of the various cDNA fragments is also indicated.

(a)

1 CAGGAACCAGCTCATGATCTCAGG

25 ATGTATGGAAAAATATCTTTGTATTACTATTGTCAGCAATTGTGAGCATATCAGCATCAAGTACCCTGGTGGCAATGCACACTTCA
M Y G K I I F V L L L S A I V S I S A S S T T G V A M H T S

115 ACCTCTTCTTCAGTCACAAAGAGTTACATCTCATCAGACAAATGATACGACAAACGGACACATATGCAGCTCTCTAGAGTCAT
T S S S V T K S Y I S S Q T N D T H K R D T Y A A T P R A H

205 GAAGTTTCAGAAATTTCTGTAGAACTGTTACCCCTCAGAAGAGGAAACGGAGAAAGGGTACAACCTGCCATCATTCTCTGAACCA
E V S E I S V R T V Y P P E E E T G E R V Q L A H H F S E P

295 GAGATAACACTCATTATTTTGGGGTGATGGCTGGTATTGGAACGCTCTTAATTTCTACGGTATTCGCCGACTGATAAGAAAA
E I T L I I F G V M A G V I G T I L L I S Y G I R R L I K K

385 AGCCCATCTGATGTAACCTCTCCCTCACCTGACACAGACGTGCCTTTAAGTCTGTTGAAATAGAAATCCAGACACAAGTGATCAA
S P S D V K P L P S P D T D V P L S S V E I E N P E T S D Q

475 TGAGAACTCTGTTCAACCAACCAATGTGGAAAGAACACAAAGACATAAGACTTCAGTCAAGTGAAAAATTAACATGTGGACTGGACA
END

#A1

565 CTCCAATAAATTATATACCTGCCTAAGTTGTACAATTTGCAATGCAATTTTCATTATAATGAGTCCAGTGACTCAATGATGGGAAAA

655 AAATCTCTGCTCATTAAATTTCAAGATAAAGAACAAATGTTTCCTTGAATGCTTGTGTTGTAGCATAATTTTAGAATGTTT

745 GAGAATCTGATCCAAACTTTAGTTGAATTCATCTACGTTTGTGTTAATTAACCTTAACCTATTCTATTGTATTATAATGATGTTCTG

835 TCAATGAAGGCTGAAATACCTAGATGAAGTTTAGATTTTCTCTATTGTAACCTTTTGAAGTCTGGTTTCATTGTTTAAATAAATT

#A2

925 AAGGGGACACTAAAGCTATCATTCTCTTCACTTGAACAGGCAAGATATAATATTACATGAATGATTACTATATTTGTTTACAC

#A3

1015 TAATAAGCTTATGCTCAGAAATGCCATACACACACACACACATTTATCATTAAATGCATAAATCAACACAAAGGTTTCCCA

1105 TTAATGAAATATTACATATATATAAGTGCCATATTTAAATAATTTGCTCAACAGTAGAAGTATGTCGGAGCACTCACTGAAGCTTCG

1195 ATTTCCCACTGAAGAGTATTGTTGTTAGTAGAGTTATCCCGGAGAGGAAAAAGAACTTACGACCTTCTTTATAACAGAAAGCTCA

1285 ACTCTAAATCAACAAGATGTGCAACCGGACATGCAGGTGAATTTTTAATAGTTACTATAAGGTTCTCAATTAATCTTTAATCTG

#A4

1375 TCCAGTCCCAGTTTCTCTATTAAATAAATTTGGAAATGCTTTAAACCTTTAAAGGAAATTTCTAGATATAGAACTAAAGGACTGT

1465 GACTATACAGTGTCACTCATTGTTAGTAAACTTAAAGCAAAAAACAAAAAGACCTTCTGTGATCTTTATTTCCGAAT

#A5

1555 AATAAAATCTATAGCTTTTATTATTGTTGATACCAAGTAAATGTTTCTATTTCGATATTTTCAAGCATGGTAACAGAAATTT

#A6

1645 ACCTTTTAAATAATTAATAATCTAAATTTAACTACTGTATGTTCCGAGAGTGTGTTGTTACTATATTGACTACTTAAATAGAGAA

GPA-NL

1735 TGAGACTAAGAAGGGAACATTTCTGTGATACATGTTTTTAAAGTAATTTTTAAGAGCATTATTAGGTTAATTTAATCCAATTAAATGA

#A7

1825 CCCAATGCCAGGTAATTTTAAATTTACATTTTAAATAAGCAACATGTTGAAACAAGAGGGTGAGATTAACTTTTGTCTAAAGT

GPA-L

1915 AATTTACAGTCAAGACAGGAAGAGATCAGAGTGAATGTGCTTCTTAACAGAGCTACAGAAATTTAGTGAATAATTAAGTACAACT

2005 GCTTTGACCTCCTTGAACCTTTTCAAGCAATTTCTGTACTTCTATATATGAATGCTTAGCAATTTTCTGCTACTATAACAGAATAC

2095 GACAGACTGCC

(b)

(c)

Fig. 2. (a) Nucleotide sequence of glycophorin A cDNA derived from clones λ -gpa3 and λ -gpa6, together with the predicted amino acid sequence represented by the single letter code. Nucleotide sequences in bold and marked # A1 to A7 above the line indicate polyadenylation signals. The regions recognized by the various oligonucleotides used for screening cDNA clones, Southern and Northern blotting and for primer extension are underlined and in bold; the names of the oligonucleotides are given above the appropriate nucleotide sequences. Note that the 3' end of oligonucleotide GPA-N1 and the 5' end of oligonucleotide GPA-N2 overlap by three nucleotides. Also note that oligonucleotides GPA-N1 and GPA-C are mixtures whereas the others are exact sequence oligonucleotides. (b) Diagrammatic representation of the size and location of the various cDNA clones sequenced. Abbreviations: E1 and E2, *Eco*R1 sites; 5' U, 5' untranslated region; S, signal peptide; C, coding sequence; and 3' U, 3' untranslated region. The bold vertical lines represent the location of polyadenylation signals. (c) Horizontal arrows indicate the direction and approximate position of the regions sequenced.

DNA from all six clones revealed a common cDNA fragment of about 0.8 kbp in size (Fig. 1a). In addition, clones λ -gpa3 and λ -gpa5 contained a fragment of size about 0.9 kbp while clone λ -gpa6 also contained an additional fragment but of about 1.3 kbp in size (Fig. 1a). Thus, three of the clones contained 0.8 kbp inserts, two had 1.7 kbp inserts and one a 2.1 kbp insert.

When *Eco*R1-digested DNA fragments were transferred by blotting from an agarose gel to a nylon membrane which was then probed with GPA-N2, the 0.8 kbp fragments from all of the cDNA clones retained the signal (Fig. 1b). An identical result to this was obtained (result not shown) when similar blots were probed with oligonucleotides GPA-N1 or GPA-C which were complementary to regions encoding amino acids 24–30 and 122–127, respectively, of the 131 amino acids in glycophorin A (Fig. 2a). These results indicated that the 0.8 kbp fragment from each cDNA clone probably contained the full coding sequence in addition to some nucleotides representing either the 5' or the 3' untranslated regions, or both. In contrast, the 0.9 and 1.3 kbp fragments did not hybridize with oligonucleotide GPA-N2 (Fig. 1b) nor with GPA-N1 nor GPA-C (result not shown) demonstrating that neither fragment contained coding region sequences and suggesting that they corresponded to either the 3' or the 5' untranslated regions of glycophorin A mRNA.

Sequencing the cDNA clones

The nucleotide sequence of the 0.8 and 1.3 kbp fragments from cDNA clone λ -gpa6 was determined in both directions. Although it was evident that the 0.8 kbp fragment corresponded to the coding region based on published amino acid (13) and partial nucleotide sequences (18), it was not clear whether the 1.3 kbp fragment lay 3' or 5' with respect to the coding region. Partial sequencing of 0.8 and 0.9 kbp fragments from our other five clones revealed no overlapping sequences which would have resolved the problem. However, while our work was in progress, the structures of two glycophorin A cDNA clones which overlapped the sequences we had determined were published (16, 17) enabling us to conclude that the 1.3 kbp fragment lay 3' with respect to the 0.8 kbp fragment. Similarly, since nucleotide sequences found in the 0.9 kbp fragments were present in the 1.3 kbp fragment, it was concluded that the 0.9 kbp fragments in cDNA clones λ -gpa3 and λ -gpa5 also lay 3' of the coding region.

The 0.8 kbp fragments from each of the cDNA clones (Fig. 1a) differed slightly from each other in size. In their 5' untranslated regions, λ -gpa1, λ -gpa5 and λ -gpa6 had 23 bases, λ -gpa7 had 36 bases and λ -gpa3 and λ -gpa4 had 42 bases. In addition, λ -gpa3 was 24 bases longer at the 3' end than the other 0.8 kbp fragments. Further, the 0.8 kbp fragment from λ -gpa5 had the complete coding sequence for the A^N polymorphic form of glycophorin in which leucine and glycine are found in positions 1 and 5 (numbered from the N-terminus), respectively, whereas the other five 0.8 kbp fragments had the coding sequence for glycophorin A^M which contains serine and glutamic acid in these positions (1). Apart from these differences, the six 0.8 kbp fragments were identical.

The longest available sequence for glycophorin A^M cDNA is shown in Fig. 2a. This structure was derived from that for the 0.8 and 1.3 kbp fragments from λ -gpa6, and the 24 bp fragment located between the two internal *Eco*RI sites at nucleotide numbers 747 and 771 in λ -gpa3. The approximate positions of the 0.8, 0.9 and 1.3 kbp fragments from the six cDNA clones we isolated are shown diagrammatically in Fig. 2b.

The sequence we found for nucleotides 1 to 488 in λ -gpa6 (Fig. 2a), which encompasses the coding region, was identical to that reported by Kudo and Fukuda (30) for glycophorin A genomic DNA also from K562 cells. Similarly, our sequence for nucleotides 90 to 935 was identical to that found by Rahuel *et al.* (16) for a glycophorin A cDNA clone prepared from a human foetal liver cDNA library. The λ -gpa6 sequence differed from the first partial glycophorin A cDNA sequence from K562 cells reported by Siebert and Fukuda (18) but agreed with the revised sequence published by the same authors (31), except that we found T not A in position 228 (Fig. 2a); however, this difference would result in a silent mutation.

Tate and Tanner (17) have also sequenced a glycophorin A cDNA clone named ALP1 which was obtained from a human reticulocyte cDNA library. The sequence was 32 nucleotides longer than λ -gpa6 at the 3' end and terminated at nucleotide 935 (Fig. 2a), excluding a 16 base long remnant of the poly (A) tail. The following nucleotide positions in λ -gpa6 differed from those in ALP1 which contained the bases given in parentheses: 62(A), 83(T), 95(A), 96(G), 117(T) and 552(C). The bases in positions 83, 95 and 96 in ALP1 would result in glycophorin A with leucine and glutamic acid at amino acid residues 1 and 5, instead of serine and glycine our sequence predicts (Fig. 2a), demonstrating that ALP1 and λ -gpa6 code for the polymorphic N and M forms of glycophorin A, respectively (1). The difference at nucleotide 62 would result in glutamic acid at amino acid position -7 in the signal sequence in ALP1 and alanine in λ -gpa6, while differences at positions 117 and 552 represent silent mutations. The sequence we obtained for cDNA clone λ -gpa5, like ALP1, had bases T, A and G in positions 83, 95 and 96 and, therefore, codes for glycophorin A^N. Unlike ALP1, however, λ -gpa5 was found to have the same bases as λ -gpa6 in positions 62, 117 and 552.

Relationship between cDNA clones and mRNAs

To determine if the three mRNAs reported for glycophorin A (16–18) were represented by the six cDNA clones we had isolated, K562 cell poly A⁺ RNA was subjected to Northern blot analysis using oligonucleotide GPA-N2 which recognized all six cDNAs (Fig. 1b). Two broad bands of sizes slightly above 1.0 and 1.7 kb and a narrower band of about 2.8 kb were detected in the approximate proportions 30%, 60% and 10%, respectively, as determined by densitometry of the autoradiogram (Fig. 3, lane N2). These mRNAs obviously correspond to those described previously by Siebert and Fukuda (18) whose nomenclature is used in the present report.

The realization that the cDNAs we had isolated could conveniently be grouped into three size classes of 0.8, 1.7 and 2.1 kbp gave rise to the notion that each class was derived from the appropriate glycophorin A mRNA of sizes 1.0, 1.7 and 2.8 kb, respectively. Certainly, cDNA clone λ -gpa6 of size 2.1 kbp could not have been derived from the 1.0 or 1.7 kb mRNAs and, therefore, must have originated from the 2.8 kb mRNA.

cDNA clones λ -gpa3 and λ -gpa5 of size 1.7 kbp could also have been reverse transcribed from the 2.8 kb mRNA. However, this is unlikely since it would imply at least three of the six cDNAs isolated were derived from the least abundant mRNA. Since these two cDNA clones could not have been derived from the 1.0 kb mRNA, the most reasonable conclusion is that they were derived from the 1.7 kb mRNA.

The three 0.8 kbp cDNA clones (λ -gpa1, λ -gpa4 and λ -gpa7) were not necessarily derived from the 1.0 kb mRNA. They could

Fig. 3. Northern blotting of K562 cell poly A⁺ RNA using as probes: lane N2, oligonucleotide GPA-N2; lane MS, oligonucleotide GPA-MS; lane ML, oligonucleotide GPA-ML; and lane L, oligonucleotide GPA-L. Lane A⁻ contained K562 cell poly A⁻ RNA which was probed with oligonucleotide GPA-ML. The sizes given in kb were determined by comparison with the position of molecular weight markers run in parallel lanes.

have been produced from the larger cDNAs by cleavage at one of the two *Eco*RI sites located at positions 747 or 771 (Fig. 2a). Nevertheless, it seems reasonable to postulate that at least one of the three 0.8 kbp cDNAs was derived from the 1.0 kb mRNA, the second most abundant of the three mRNAs.

Primer Extension Analysis

To determine if the three glycoporphin A mRNAs differed from each other in size because they had 5' untranslated regions of different length, poly A⁺ and total RNA from K562 cells were subjected to primer extension analysis using oligonucleotide GPA-N2 which codes for amino acids 30 to 40 unique to glycoporphin A. Sequencing showed that clones λ -gpa1, λ -gpa5 and λ -gpa6 contained 5' untranslated regions of 24 nucleotides long which were shorter than those for the other three clones (see above). In addition, there were 144 nucleotides of translated sequence to the start of the region coded by GPA-N2 which was, itself, 33 nucleotides long. Therefore, fragments produced during primer extension analysis by reverse transcription to the 5' terminus of glycoporphin A using GPA-N2 as a primer should be at least 201 nucleotides long. Fragments shorter than this could arise by premature termination and should be ignored.

Another consideration was that the mRNAs were present approximately in the proportions 30%, 60% and 10%, respectively, as described above. Thus, if each mRNA had its own unique 5' terminus, three bands in approximately these proportions would be expected during primer extension analysis.

In fact primer extension analysis with GPA-N2 revealed a single, major product greater in size than 201 nucleotides at a

Fig. 4. Primer extension analysis of K562 cell RNA. Oligonucleotide GPA-N2 was used as a primer to extend 10 and 5 μ g K562 cell poly A⁺ RNA (lanes 1 and 3, respectively), or 50 and 100 μ g total RNA (lanes 2 and 4, respectively). The numbers on the left indicate the positions that products containing 150 and 200 bases would be found based on sequencing reactions (lanes G, A, T and C) which used oligonucleotide GPA-N2 as a primer. The suggested position of the major, full-length, primer-extended product is indicated by '5' end'.

position of about 230 nucleotides long with only trace amounts of other components (Fig. 4a, Lanes 1–4). The size of the major product was deduced by running nucleotide sequencing reactions of the Bluescript plasmid- λ -gpa6 boundary in parallel lanes (Fig. 4a, Lanes G, A, T, C) adjacent to the primer extension lanes. Thus, the result of primer extension analysis was consistent with a single initiation site for synthesis of all three glycoporphin A mRNAs ruling out the possibility that differences at the 5' end of glycoporphin A mRNAs contributed to the large size differences seen in the three glycoporphin A mRNAs. The results also suggested that the 5' untranslated region in all three glycoporphin A mRNAs was about 53 nucleotides long.

A similar size was reported by Tate and Tanner (17) for the 5' untranslated region of a cDNA derived from the smallest glycoporphin A mRNA. Further, Kudo and Fukuda (30) established by S1 nuclease mapping that initiation of glycoporphin A transcription begins at the 5'-most nucleotide of the cDNA sequence found by Tate and Tanner.

Searching for Differences in the Coding Region

As discussed above, it was likely that all three glycoporphin A mRNAs were reverse transcribed and were represented in the three size classes of cDNA we isolated. Determination of the sequences for the six 0.8 kbp fragments revealed that they all contained the same number of nucleotides in the segment corresponding to the coding region for glycoporphin A. In addition, the number of nucleotides in the coding region of glycoporphin A cDNAs sequenced by others was identical (16, 17), or very nearly so (18) to the number we found. Assuming that all three mRNAs were represented in the cDNAs we and others have sequenced, it appears that the mRNAs have

identically-sized coding regions thus excluding this region as the basis for the differences in mRNA size observed.

Differences at the 3' Ends of the mRNAs

Nucleotide sequencing the various glycophorin A cDNA clones isolated in the present study showed that they differed in the lengths of their 3' terminal regions (Fig. 2b). To determine if these dissimilarities reflected size differences in the mRNAs, themselves, oligonucleotides GPA-MS, GPA-ML and GPA-L were synthesized to probe Northern blots of K562 cell poly A⁺ RNA.

Oligonucleotide GPA-MS which was complementary to nucleotides 1623 to 1647 in the cDNA (Fig. 2a) revealed a broad band of about 1.7 kb and a less abundant component of 2.8 kb but no 1.0 kb mRNA (Fig. 3, lane MS).

Oligonucleotide GPA-ML which was complementary to nucleotides 1814 to 1838 hybridized to the 2.8 kb mRNA and, in addition, to less abundant 2 kb and 5 kb components (Fig. 3, lane ML). The latter two were possibly 18S and 28S ribosomal RNAs, respectively, since similar sized components were also detected by Northern blot analysis of poly A⁺ RNA probed with GPA-ML (Fig. 3, lane A⁺). Hybridization to these additional components could have occurred because oligonucleotide GPA-ML was AT-rich and, therefore, required low stringency conditions for hybridization. Another possible explanation for the detection of low levels of RNA of about 2 kb in length by oligonucleotide ML was that 1.7 kb mRNA was in fact heterogeneous and contained a component of about 2 kb in size. Consistent with this was the observation that 1.7 kb mRNA was present as a broad band (Fig. 3, lanes N2 and MS).

Oligonucleotide GPA-L which was complementary to nucleotides 2056 to 2081 in the cDNA (Fig. 2a) hybridized to the 2.8 kb mRNA species and to two much less abundant components of about 2 and 5 kb in size (Fig. 3, lane L). The latter two components did not quite line up with the 2 and 5 kb components seen in lanes ML and A⁺ (Fig. 3) because lane L was from a different electrophoresis run. Nevertheless, enough runs have been made to conclude that the components of about 2 kb and 5 kb in size were the same and were probably 18S and 28S ribosomal RNAs, respectively, since similar sized components were also detected by Northern blot analysis of poly A⁺ RNA probed with GPA-L (result not shown). As for oligonucleotide GPA-ML, detection of these additional components could have occurred because oligonucleotide GPA-L was AT-rich and, therefore, required low stringency conditions for hybridization, or in the case of the 2 kb band represented a minor component of this size.

From these results it is concluded that: (i) the 1.0 kb mRNA did not reach as far as the region encompassed by nucleotides 1623 to 1647 (GPA-MS) (Fig. 2a); (ii) the 3' region of the bulk of 1.7 kb mRNA extended beyond nucleotides 1623 to 1647 (GPA-MS) but not as far as 1814 to 1838 (GPA-ML), although a minor 2.0 kb mRNA recognized by GPA-ML might be present; and (iii) the 3' region of the 2.8 kb mRNA extended beyond nucleotide 2105 (i.e. beyond the end of clone λ -gpa6). It seems self-evident that a 1.0 kb mRNA would not hybridize to an oligonucleotide containing bases 1623 to 1647 in our clone λ -gpa6, that a 1.7 kb mRNA extends beyond base 1647 but not as far as base 1814 and that a 2.8 kb mRNA must extend beyond base 2105. However, these conclusions were obvious only once we had found that transcription of these mRNAs began at the same site and that λ -gpa6 extended within about 50 bases of the 5' terminus.

DISCUSSION

Three glycophorin A mRNAs of size 1.0, 1.7 and 2.8 kb could differ from each other in the lengths of their 5' untranslated regions produced, for instance, by use of three different initiation sites for transcription. We showed in this report that this was not the explanation for the existence of the three mRNAs since the evidence suggested that all were found to contain identically sized 5' untranslated regions of about 53 nucleotides long based on primer extension analysis.

The three mRNAs could differ from each other internally, the smaller species containing deletions resulting from alternative splicing, for example. We found that the complete nucleotide sequences present in the small and medium sized cDNAs were also present without interruptions in the largest cDNA. This indicated that the smaller mRNA species did not differ from the larger ones by containing deletions, assuming that the sequences in the cDNAs were representative of those in the three mRNAs.

Messenger RNAs could differ from each other in the size of their 3' untranslated regions and this is where we found the glycophorin mRNAs were unlike. Such differences could theoretically arise by differential splicing, the smaller mRNA species lacking terminally located exons. While our work was in progress, Kudo and Fukuda (30) published details of the gene organization and intron-exon junctions for glycophorin A. These authors found that about 4 amino acids of the coding region and all of 3' untranslated region are encoded by a single exon of about 2.1 kb in length. Accepting that the 2.8 kb mRNA contains this exon, an RNA lacking it would be 0.7 kb in size and not 1.0 nor 1.7 kb as found for the two smaller mRNA species. Clearly, differential splicing at the 3' end of a gene of this structure does not explain the origin of the three mRNAs seen.

The most reasonable explanation of the origin of the three glycophorin A mRNAs is that either the single copy gene is transcribed into three different sized pre-mRNAs each of which

Fig. 5. Proposed mechanism of production of 3 mRNAs from a single glycophorin A gene. The hatched boxes labelled E1-E6 represent the 6 glycophorin A exons encoding the 5' untranslated region, the signal peptide and protein sequence; the open box with the bold vertical bars represents the exon containing the last 3 amino acids of glycophorin A sequence, the termination signal and the complete 3' untranslated region (30). The bold vertical bars indicate the positions of polyadenylation signals numbered A1 to A7 which were identified in the present investigation and the polyadenylation signal 'An' present in the sequence reported by Kudo and Fukuda (30). The continuous line between the exon boxes represents the introns.

is processed to its own unique mRNA, or alternatively, one pre-mRNA is produced which is processed into three mRNAs (Fig. 5). Either mechanism would utilize different polyadenylation signals (AAUAAA) of which there are seven in the 3' untranslated region of our longest cDNA (Fig. 2a).

Tate and Tanner (17) and Rahuel *et al.* (16) isolated and sequenced glycophorin A cDNAs which included a 3' untranslated region of approximately 460 bases and, in addition, a poly A segment, the presence of which indicated that both cDNAs contained an authentic 3' mRNA terminus. The similarity in size of these cDNAs to that of the smallest glycophorin mRNA suggests that the clones represent the complete sequence of the 1.0 kb mRNA. Both cDNA clones examined contain the two polyadenylation signals we also found starting at nucleotide positions 569 and 917 (Fig. 2a). It would appear that the second of these signals is that recognised by the endonuclease which cleaves pre-mRNAs prior to polyadenylation to produce the smallest glycophorin A mRNA.

We suggest that the 1.7 kb mRNA is produced similarly principally through use of polyadenylation signal 6 (nucleotides 1652 to 1657) (see Figs. 2 and 5). This suggestion is based on the observation that Northern blot analysis showed that nucleotide sequence (1623 to 1647, GPA-MS) immediately upstream of polyadenylation signal 6 was present in the intermediate sized mRNA whereas nucleotide sequence (1814 to 1838, GPA-ML), which lies between polyadenylation signals 6 and 7 was absent at least from the bulk of 1.7 kb mRNA. The qualification that it was the bulk of 1.7 kb mRNA rather than all of the 1.7 kb mRNA which was produced by utilization of polyadenylation signal 6 was added since this mRNA species was possibly heterogeneous and our results did not exclude the possibility that polyadenylation signals 4, 5 and 7 were also used to some extent. The suspicion that more than one polyadenylation signal might be used to generate the 1.7 kb mRNA species was raised not only by the broad appearance of the 1.7 kb mRNA on Northern analysis (Fig. 3, lane N2) but also by the finding that oligonucleotides GPA-ML and GPA-L, in addition to detecting the 2.8 kb mRNA, also revealed a component slightly larger than 1.7 kb (Fig. 3, lanes ML and L). Oligonucleotide GPA-ML could be expected to detect an mRNA generated by employing polyadenylation signal 7 (Figs. 2 and 5) which would produce a component close to 1.9 kb in size not including a poly A sequence. However, it was also possible that oligonucleotides GPA-ML and GPA-L were hybridizing non-specifically to 18S ribosomal RNA (Fig. 3, lane A⁻).

The 2.8 kb mRNA could have originated by cleavage from its own unique pre-mRNA or more likely from a common pre-mRNA. However, the use of the most 3' terminally located polyadenylation signal that we found (number 7) (Fig. 2a, Fig. 5) would produce an mRNA of considerably less than 2.8 kb in size so that a polyadenylation signal beyond the 3' end of our cDNA clone λ -gpa6 must be used to generate the 2.8 kb mRNA species. In fact, in the partial sequence for the 3' untranslated region of glycophorin A reported by Kudo and Fukuda (30) a polyadenylation signal is present which we have termed 'n' (Fig. 5); use of this signal would produce an mRNA of about 2.8 kb.

The 1.7 kb glycophorin mRNA was more abundant than the 1.0 or 2.8 kb species (Fig. 3, lane N2). If our explanation that the three mRNAs are generated by exploiting different polyadenylation signals is correct, a possible reason for the greater abundance of the 1.7 kb species is that signal 6 is used more frequently for polyadenylation than are signals 2 and 'n' which in turn are used in preference to signals 1, 3, 4, 5 and 7.

AAUAAA elements recognised for polyadenylation purposes often have GU- and U-rich sequences within the adjacent 30 residues or so downstream (32, 33). Consistent with this observation, we found that sequences following polyadenylation signal 6 were richer in GU and U than were those following signal 2; this is particularly so if 50 residues rather than 30 residues downstream were considered (Fig. 2a). However, although signals 2 might be used less frequently than signal 6, it is nevertheless recognised for polyadenylation purposes and yet sequences following it showed no preponderance of GU or U compared with signals 1 and 3 to 5 which are apparently used more rarely, if at all. Similarly, signal n, although probably used in polyadenylation, is not followed by GU/U rich sequences (30).

An alternative explanation for the difference in the abundance of the three mRNAs is that they differ in stability. There are several reports that (A + U)-rich regions just upstream of the 3' poly (A) tail decrease mRNA stability (34–36). If the same is true for glycophorin A mRNAs, it would be predicted that the less abundant 1.0 and 2.8 kb components would be richer than the 1.7 kb mRNA in (A + U) residues close to their poly (A) segment. Polyadenylation signals 2 and 6 (Fig. 2a) and signal n (ref. 30) contained 72, 76 and 71 (A + U) residues, respectively, within the 100 residues immediately upstream. Thus, although the regions adjacent to these signals were noticeably (A + U)-rich, sequences upstream of signal 6 if anything were actually richer than signals 2 and n in (A + U).

It would appear, therefore, that the excess of the 1.7 kb mRNA over the 1.0 and 2.8 species is not related to the content of (A + U) residues upstream nor to GU/U residues downstream of the polyadenylation signal and that the mechanism which regulates glycophorin A mRNA abundance remains to be determined.

The structure of the longest glycophorin A mRNA, in particular, needs further comment. It is interesting to note that clone λ -gpa6 contained the equivalent of a 3' untranslated region almost three times the length of the coding region. Even more remarkable was that cDNA clone λ -gpa6 of 2.1 kb size did not fully represent the largest glycophorin A mRNA of size 2.8 kb. Therefore, the actual size of the 3' untranslated region in the large glycophorin A mRNA must be even longer than that represented in this cDNA clone.

ACKNOWLEDGEMENTS

We thank Ms Ingrid Pardoe for excellent assistance and Dr. H.B. Younghusband for much valuable advice during the course of this work which was supported by grants from the Canadian Medical Research Council and the Canadian Diabetes Association.

REFERENCES

1. Dahr, W. (1986) In Vengelen-Tyler, V. & Judd, W.J. (eds.) Recent Advances in Blood Group Biochemistry. American Association of Blood Banks, Arlington, VA, pp. 23–65.
2. Furthmayr, H. (1978) J. Supramolec. Struct. **9**, 79–95.
3. Springer, G.F., Nagai, Y. and Tegtmeier, H. (1966) Biochem. USA **5**, 3254–3272.
4. Adair, W.L. and Kornfeld, S.J. (1974) J. Biol. Chem. **249**, 4696–4704.
5. Pasvol, G., Jungery, M., Weatherall, D.J., Parsons, S.F., Anstee, D.J. and Tanner, M.J.A. (1982) Lancet **2**, 947–950.
6. Kathan, R.H., Winzler, R.J. and Johnson, C.A. (1961) J. Exptl. Med. **113**, 37–45.
7. Marchesi, V.T. and Andrews, E.P. (1971) Science **174**, 1247–1248.
8. Enegren, B. and Burness, A.T.H. (1977) Nature (London) **268**, 536–537.
9. Allaway, G.P. and Burness, A.T.H. (1986) J. Virol. **59**, 768–770.
10. Paul, R.W. and Lee, P.W.K. (1987) Virology **159**, 94–101.

11. Eaton, B.T. and Crameri, G.S. (1989) *J. Gen. Virol.* **70**, 3347–3353.
12. Welsh, E.J. and Thom, D. (1985) *Biopolymers*, **24**, 2301–2331.
13. Tomita, M. and Marchesi, V.T. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2964–2968.
14. Fukuda, M., Lauffenburger, M., Sasaki, H., Rogers, M.E. and Dell, A. (1987) *J. Biol. Chem.* **262**, 11952–11957.
15. Cook, P.J.L., Lindenbaum, Salonen, R.H., De La Chapelle, A., Dakar, M.G., Buckton, K.E., Noades, J.E. and Tippet, p. (1981) *Ann. Hum. Genet.* **45**, 39–47.
16. Rahuel, C., London, J., d'Auriol, L., Mattei, M.G., Tournamille, C., Skrzynia, C., Lebouc, Y., Galibert, F. and Cartron, J.P. (1988) *Eur. J. Biochem.* **172**, 147–153.
17. Tate, C.G. and Tanner, M.J.A. (1988) *Biochem. J.* **254**, 743–750.
18. Siebert, P.D. and Fukuda, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1665–1669.
19. Gahmberg, C.G. (1983) *Meth. Enzymol.* (1983) **96**, 281–289.
20. Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1989) *Ann. Rev. Genet.* **23**, 527–577.
21. Benton, W.D., and Davis, R.W. (1977) *Science* **196**, 180–181.
22. Sgarbetta, V., and Khorana, H.G. (1972) *J. Mol. Biol.* **72**, 427–444.
23. Meinkoth, J., and Wahl, G. (1984) *Anal. Biochem.* **138**, 267–284.
24. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor, NY.
25. Budowle, B. (1985) *Biotechniques* **93**, 98–101.
26. Hanahan, D. (1988) In Glover, D.M. (ed.), *DNA Cloning-A Practical Approach*. IRL Press, Oxford, Vol. 1, pp. 109–135.
27. Sanger, F., Nicklen, S.D. and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **76**, 5463–5476.
28. Southern, E.M. (1975) *J. Mol. Biol.* **98**, 503–517.
29. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
30. Kudo, S. and Fukuda, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4619–4623.
31. Siebert, P. and Fukuda, M. (1986) *J. Biol. Chem.* **261**, 12433–12436.
32. McLauchlan, J., Gaffney, D., Whitton, L. and Clements, J.B. (1985) *Nucl. Acids Res.* **13**, 1347–1368.
33. Levitt, N., Briggs, D., Gil, A. and Proudfoot, N.J. (1989) *Genes Dev.* **3**, 1019–1025.
34. Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimmer, S. and Cerami, A. (1986) *Proc. Soc. Natl. Acad. Sci. USA* **83**, 1670–1674.
35. Shaw, G. and Kamen, R. (1986) *Cell* **46**, 659–667.
36. Wilson, T. and Treisman, R. (1988) *Nature* **336**, 396–399.