MECHANISM OF PRODUCTION OF MULTIPLE mRNAs FOR THE MAJOR SIALOGLYCOPROTEIN OF HUMAN ERYTHROCYTES, GLYCOPHORIN A

JAWED HAMID, B.Sc.(Honours), M.Sc.

ISBN 0-315-73327-6

Canada

MECHANISM OF PRODUCTION OF MULTIPLE mRNAs FOR THE MAJOR SIALOGLYCOPROTEIN OF HUMAN ERYTHROCYTES, GLYCOPHORIN A

By

Copyright (C) Jawed Hamid, B.Sc.(Honours), M.Sc.

A thesis submitted to the School of Graduate Studies

in partial fulfilment of the requirements

of the degree of

Doctor of Philosophy

Faculty of Medicine

Memorial University of Newfoundland

November, 1991

St. John's Newfoundland

ABSTRACT

Glycophorins are a family of human erythrocyte membrane glycoproteins. At least four different sialoglycoproteins glycophorins, A (or α), B (or δ), C (or β) and D (or γ) have been detected in humans. Glycophorin A is the major sialic acid-containing protein of human erythrocyte membranes. It consists of 131 amino acids distributed in three structural domains. Glycophorin A is encoded by a single gene which gives rise to three different mRNAs, large (2.8 kb), medium (1.7 kb) and small (1.0 kb) in reticulocytes and in K562, a human erythroleukaemia cell line expressing glycophorin A on its surface.

Six clones were isolated from a cDNA library constructed with K562 cell mRNA in λgt10 phage using as a probe a synthetic oligonucleotide (GPA-N2) encoding amino acid numbers 30 to 40 of glycophorin A. Nucleotide sequencing of the six clones revealed that all contain an identical protein coding region except for the well known glycophorin A^M-A^N polymorphism and essentially identical 5′ untranslated regions. In contrast, clones differ substantially in the length of their 3′ untranslated regions. Examination of the 3′ untranslated region of the largest clone revealed seven poly(A) addition signals (AATAAA). To study how the single gene encoding glycophorin A generates three different mRNAs, primer extension analysis and Northern blotting experiments were performed. These experiments supported the

findings of the cDNA sequencing and revealed that the three glycophorin A mRNAs differ in the length of their 3´ untranslated region. The primary structure of the three glycophorin A mRNAs is deduced based upon the nucleotide sequence of various cDNAs, primer extension analysis and Northern blotting experiments. A mechanism is proposed for the generation of the three glycophorin A mRNAs from a single glycophorin A gene that involves differential processing of the 3´ end of glycophorin A pre-mRNA utilizing multiple poly(A) addition signals.

[**Key words:** human; erythrocyte; erythrocyte membrane; K562 cells; sialoglycoprotein; glycophorins; glycophorin A (α); cDNA cloning; mRNA; mRNA processing; polyadenylation].

ACKNOWLEDGEMENT

I am grateful to Dr. A.T.H. Burness, the chairman of my supervisory committee for his advice, help, encouragement and understanding throughout this programme. I am also thankful to Dr. Burness for his extensive review and criticism throughout the writing of this dissertation. I wish to thank Dr. H.B. Younghusband for his willingness to have very helpful discussions during this research project. I would also like to thank Dr. A. Pater for his advice and suggestions. I am thankful for the help and assistance provided by Ms. I. U. Pardoe throughout this programme. I wish to thank Mr. M.P. Baldeh for helping me with various computer programmes. Thanks are due to School of Graduate Studies and Faculty of Medicine for financial support throughout this programme.

Finally, I am very grateful to my family, my brothers, sisters and especially my mother without their love and encouragement this work was not possible.

Dedicated to the memories of my mother

and

Dr. Alfred T. H. Burness

LIST OF PUBLICATIONS

PAPERS

Hamid, J. and Burness, A.T.H. (1990). The mechanism of production of multiple mRNAs for human glycophorin A. Nucl. Acids Res. *18*, 5829-5836.

Pardoe, I.U., Grewal, K.K., Baldeh, M.P., Hamid, J. and Burness, A.T.H. (1990). Persistent infection of K562 cells by encephalomyocarditis virus. J. Virol., *64*, 6040-6044.

ABSTRACTS

Grewal, K.K., Hamid, J., Pardoe, I.U., and Burness, A.T.H. (1991). Transfection of bovine cells with glycophorin A cDNA induces susceptibility to encephalomyocarditis virus infection. VIIIth European Study Group on the Molecular Biology of Picornaviruses, Canterbury, U.K. Abstract A11.

Burness, A.T.H., Grewal, K.K., Hamid, J., and Pardoe, I.U. (1991). Biological properties of a receptor for mouse encephalomyocarditis virus expressed on human and bovine cells. Federation of American Societies for Experimental Biology, Atlanta, Georgia, Abstract 6881.

Grewal, K.K., Pardoe, I.U., Hamid, J., Baldeh, M.P., and Burness, A.T.H. (1990). The role of glycophorin A in the infection K562 cells by encephalomyocarditis virus. VIIIth International Congress of Virology, Berlin.

Hamid, J. and Burness, A.T.H. (1989). Molecular cloning and characterization of mRNA for glycophorin A. Canadian Federation of Biological Societies meeting, Calgary, abstract 228.

Pardoe, I.U., Hamid, J., Baldeh, M.P., Grewal, K.K. and Burness, A.T.H. (1989). Persistent infection of K562 cells by encephalomyocarditis virus.
2nd International Symposium on Positive strand RNA viruses, Vienna, Austria, Abstract 880.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

BSA	Bovine serum albumin
CF	Cleavage factor
CPF	Cleavage and polyadenylation factor
cpm/μg	Counts per min/μg
CSF	Cleavage specificity factor
Cstf	Cleavage stimulation factor
DTT	Dithiothreitol
DEP	Diethylpyrocarbonate
DNase	Deoxyribonuclease
EDTA	Ethylenediaminetetra acetic acid
EMC	Encephalomyocarditis
FMMP	Formylmethionyl sulfone methyl phosphate
Fuc	Fucose
Gal	Galactose
GalNAc	N-acetyl galactosamine
GlcNAc	N-acetyl glucosamine
Hfl	High frequency lysogenization
hnRNA	Heterogeneous nuclear RNA
IPTG	Isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
Man	Mannose
Mi	Miltenberger
MOPS	Morpholinopropanesulfonic acid
mRNP	Messenger ribonuleoprotein particles
NeuNAc	N-acetyl neuraminic acid (sialic acid)
OD	Optical density
PAS	Periodic acid-Schiff
PHA	Phytohaemagglutinin
PF	Polyadenylation factor
PFU	Plaque forming units
RNase	Ribonuclease
RPMI 1640	Rosewell Park Memorial Institute 1640
SDS	Sodium dodecyl sulphate
SF	Specificity factor
snRNP	Small nuclear ribonucleoprotein particle
SSC	Saline-sodium citrate
SSPE	Saline-sodium phosphate-EDTA
STE	Sodium chloride-tris-HCl-EDTA buffer
TCA	Trichloroacetic acid
TE	Tris-HCl-EDTA buffer
TF	Transcription factors
Tris	Tris hydroxymethyl (aminomethane)
tRNA	Transfer RNA

CHAPTER 1

INTRODUCTION

Glycophorins are a group of sialoglycoproteins present in erythrocyte membranes of humans and many animal species. At least four different sialic acid-rich glycoproteins termed glycophorins A (or α), B (or δ), C (or β) and D (or γ) have been detected in humans. Members of this group of erythrocyte membrane proteins have been implicated in diverse phenomena such as regulation of membrane structure by controlling membrane fluidity, deformability and rigidity, erythrocyte senescence, and erythroid cell differentiation. Blood group antigens, MNSs are also found to be associated with glycophorins. Glycophorins, particularly glycophorin A, probably act as receptor for a number of entities including viruses such as, encephalomyocarditis virus (EMC), influenza virus, blue tongue virus and reovirus; possibly the malarial parasite, *Plasmodium falciparum*; and some bacterial strains.

Since their detection in the late 1930s as blood group MN antigens glycophorins have served as model transmembrane proteins. A lot of information has been accumulated over past few decades about the biochemical characteristics and genetics of these sialoglycoproteins. The complete amino acid sequence has been determined for glycophorin A, B and C.

Previous studies have shown that glycophorin A, the major sialoglycoprotein

of human erythrocyte membranes has a transmembrane orientation. It consists of 131 amino acids, distributed in three structural domains. The extracellular NH_2-terminal domain is composed of about 72 amino acids. The extracellular domain is followed by an intramembranous domain of about 23 amino acid residues inserted into the lipid bilayer. A short region comprising about 40 amino acid residues is exposed to the interior of the cell.

At the time when this work was initiated it was known that glycophorin A is encoded by a single copy gene on chromosome 4 at q_{28}-q_{31}. There was no information available about the primary structure of the mRNA for glycophorin A. Therefore, this work was initiated with an objective to clone cDNA for glycophorin A. It was hoped that information thus generated would be useful in the study of glycophorin A generally and also in its role as a virus receptor (particularly for EMC). However, while this work was in progress a partial cDNA sequence was published and the presence of three different sized mRNAs for glycophorin A were reported in human reticulocytes as well as in K562 cells. The sizes of these three mRNAs are about 2.8, 1.7 and 1.0 kb. Similar multiple RNAs were later shown to be present in human spleen erythroblast cells. Therefore, further efforts were concentrated in determining the primary structure of the three mRNAs.

In eukaryotic cells production of multiple mRNAs from a single gene can involve multiple initiation sites, alternate splicing, differential termination or

processing at 3′ ends. One way to distinguish between these mechanisms is to determine the primary structure of the mRNAs. cDNA cloning of glycophorin A mRNAs could provide some information about the structural differences between the three different glycophorin A mRNAs. The information thus generated could probably be used to predict the mechanism of production of these mRNAs from a single glycophorin A gene.

I initiated the task of cloning and characterizing glycophorin A mRNAs with an objective of investigating the mechanism of generation of multiple mRNAs from a single gene. To achieve this goal I have isolated cDNAs encoding glycophorin A mRNAs from a K562 cell cDNA library using synthetic oligonucleotides as probes. I determined the nucleotide sequences of these cDNAs and used this information to predict the structure of the three glycophorin A mRNAs. Oligonucleotides were synthesized based upon the nucleotide sequence of various cDNAs encoding glycophorin A and then used as probes to check the validity of the prediction of primary structure of the three different glycophorin A mRNAs by Northern blotting of K562 mRNA. These studies have shown that the three glycophorin A mRNAs have similar 5′ untranslated and coding region but they differ in the length of their 3′ untranslated regions. I have proposed a mechanism for the production of the three different mRNAs for glycophorin A that differ in their 3′ untranslated region.

Review of Literature

1.1. Scope of the review

In this review I will present a brief description of glycophorins concentrating more on human glycophorin A. This will be followed by a description of various biological activities in which glycophorins specially glycophorin A has been implicated to have a role. In the latter part of this project efforts were concentrated on investigating the mechanism of generation of the three glycophorin A mRNAs that varied in the length of their 3´ untranslated region; the second part of this review is concerned with the biosynthesis of mRNAs with special emphasis given on processing of the 3´ ends of pre-mRNAs.

1.2. Membranes

The concept of membranes was first developed in the last century in an investigation of the penetration of pigments into damaged and undamaged plant cells (reviewed in Harrison and Lunt, 1975). In this early investigation it was observed that plant cells were osmotically sensitive and could change their volume according to the osmotic strength of the surrounding medium. Based upon this observation the term *plasma membrane* was coined to describe a boundary surrounding the cell. We have come a long way since these early observation regarding the existence of such a boundary to the present day understanding of the composition, structure, organization and function of not only the plasma membrane but other cellular membranes

important for normal operation of a cell (Harrison and Lunt, 1975). It is now well established that all eucaryotic cells are surrounded by plasma membranes. These membranes contain a lipid bilayer formed mainly of phospholipids in such a way that the hydrophobic side chains of these lipids are sequestered in the hydrophobic interior of the bilayer and the charged groups are exposed to the aqueous environment at both the intracellular and extracellular space (Karp, 1984).

The fluid mosaic model presented by Singer and Nicholson (1971, 1972) is generally accepted to depict the arrangement found in all biological membranes. According to this model the bulk of phospholipids are arranged in the form of a discontinuous bilayer with their polar heads in contact with water and the associated proteins are either bound to the charged surface of the lipid bilayer mainly by electrostatic interactions (peripheral proteins) or intercalated to varying degrees in the hydrophobic interior of the bilayer (integral proteins). This latter type of protein is arranged in an amphipathic structure, that is, with the polar groups protruding from the membrane into the aqueous phase and the non-polar groups mostly intercalated in the hydrophobic interior of the membrane. A membrane in this model is envisaged as a two dimensional solution of proteins in a viscous phospholipid bilayer. Both lipids and proteins are capable of lateral mobility in the plane of the membrane (reviewed in Malhotra, 1980).

Although the presence of a plasma membrane was first demonstrated in plant

cells much of the work has dealt with animal cell membranes. The human erythrocyte membrane has served as a model in the determination of the structural organization and function of membranes due to the ease with which it can be isolated without the contamination with other cellular membranes.

1.3. Erythrocyte membrane proteins

Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis of proteins extracted from human erythrocyte membranes followed by coomassie blue staining revealed multiple bands; six of these bands arbitrarily designated I, II, III, IV, V and VI comprised about 60% of total human erythrocyte membrane proteins (Fairbanks *et al.*, 1971). Periodic acid-Schiff's (PAS) staining of a gel containing separated erythrocyte membrane proteins revealed four bands designated PAS-1, PAS-2, PAS-3 and PAS-4 (Fairbanks *et al.*, 1971). Lenard (1970), also by SDS-polyacrylamide gel electrophoresis of human erythrocytes membrane proteins detected multiple bands which were stained with coomassie blue. This author (Lenard, 1970) further showed that human erythrocyte membranes contained at least 14 distinct proteins which were not degradation products of higher molecular weight components. Many of these erythrocyte membrane proteins have been identified and include α and β spectrin (bands 1 and 2), ankyrin (band 2.1), anion channel (band 3), actin (band 5), and glyceraldehyde 3-phosphate dehydrogenase (band 6, Cohen, 1983). In addition to the major proteins other erythrocyte membranes proteins involved in a number of functions have been detected. For example, at least 200 different proteins have been

detected by two dimensional polyacrylamide gel electrophoresis analysis of human erythrocyte membranes (Rubin and Milikowski, 1978).

The structure, organization and functions of the erythrocyte membrane proteins have been reviewed a number of times (Steck, 1974; Marchesi *et al.*, 1976; Marchesi, 1979; Steck and Hainfeld, 1977; Bennet, 1985; Cohen, 1983; Chasis and Shohet, 1987). With the exception of sialoglycoproteins, most of the proteins of human erythrocyte membranes have been assigned a function, such as a role in contact with other constituents of the human erythrocyte membrane at the cell surface, a catalytic function or a capacity to support or stabilize the erythrocyte membrane structure (Marchesi, 1979). Some of the important characteristics of the major human erythrocyte membrane proteins will be described very briefly in the next section.

1.3.1. Spectrin

It is the most abundant skeletal protein of erythrocytes (about 10^5 tetramers/cell), composed of two types of subunits, α and β, of molecular weight 240,000 and 225,000 daltons, respectively. These subunits are capable of self association to form heterodimers ($\alpha\beta$), tetramers $(\alpha\beta)_2$ and high order oligomers. In addition to these self associations, the spectrins are also found to interact non-covalently with a number of other erythrocyte cytoskeleton proteins including, ankyrin, actin and protein 4.1 (Chasis and Shohet, 1987).

1.3.2. Protein 4.1

The protein 4.1, a globular protein, has two components, 4.1a and 4.1b of molecular weight 80,000 and 78,000 daltons, respectively. The dimer binds to the tail of the spectrin tetramer and plays an important role in the formation of a complex with spectrin and actin. Protein 4.1 may also act as a link between the cytoskeleton and the erythrocyte membrane (Chasis and Shohet, 1987). It is suggested that protein 4.1 forms a link with glycophorin A (Anderson and Loverin, 1984; Mueller and Morrison, 1981), glycophorin C and possibly D (Reid *et al.*, 1987a and b).

1.3.3. Actin

The monomeric form of human erythrocyte actin has a molecular weight of 43,000 daltons and there are about 5×10^5 copies per cell. Most of the erythrocyte actin is present in the filamentous form, F-actin, which is composed of oligomers containing 12 to 17 subunits (Bennet, 1985). Erythrocyte actin associates weakly with spectrin and this association is enhanced by protein 4.1 (Chasis and Shohet, 1987).

1.3.4. Ankyrin

Ankyrin is a phosphoprotein of molecular weight 215,000 daltons that is present in 10^5 copies/cell. It binds to spectrin as well as the cytoplasmic part of band 3 (anion transport channel). Ankyrin also binds to tubulin (Bennet, 1985).

1.3.5. Tropomyosin

About 6 x 10^5 copies/cell of tropomyosin are present. The erythrocyte tropomyosin is similar to the non-muscle tropomyosin in many properties including, amino acid composition, and having two subunits of molecular weight, 29,000 and 27,000 daltons. The monomeric units associate to form heterodimers which have the capability to interact with F-actin. Erythrocyte tropomyosin functions perhaps to stabilize actin filaments (Bennet, 1985).

1.3.6. Myosin

Erythrocyte myosin which contains Mg^{2+} ATPase activity is present in about 6000 copies/cell, is composed of a heavy chain of molecular weight 210,000 daltons and two light chains of molecular weight 25,000 and 19,500 daltons. (Steck and Hainfeld, 1977).

1.3.7. Anion transport channel

This is the most abundant protein of the human erythrocyte membrane as revealed by coomassie blue staining (band III, Fairbanks *et al.*, 1970). This protein is also detected by PAS staining (Tanner, 1978). There are about 10^6 copies per erythrocyte (Steck, 1974). This protein runs as a characteristic diffused band on polyacrylamide gel with an apparent molecular weight of 86,000 to 90,000 daltons. It is an integral membrane protein that passes through the membrane many times to form a pore or channel through which passage of anions is controlled (Tanner, 1978).

1.4. Early studies on erythrocyte sialoglycoproteins

A number of histochemical methods in combination with light or electron microscopy demonstrated the presence of carbohydrates associated with plasma membranes (reviewed in Winzler, 1970, 1972). The method most commonly employed for carbohydrate staining involves the periodic acid-Schiff's (PAS) reaction. This reaction depends on the periodate oxidation of vicinal hydroxyl groups in carbohydrates and the resultant production of aldehydes which react with aldehyde reagents such as acid fuchsin to give a pink product (Winzler, 1972). This staining method has been used extensively to characterize carbohydrate-containing molecules. Various studies have shown that carbohydrate-containing molecules of membranes include glycoproteins and glycolipids in addition to some free polysaccharides. Another important finding regarding carbohydrate was that the carbohydrate containing portions of these molecules in membranes were invariably distributed asymmetrically, i.e. mainly on the extracellular face (Winzler, 1970).

The sialoglycoproteins now known as glycophorins were first characterized as blood group antigens, MN, the second blood group system discovered by Landsteiner and Levine (1927) following the discovery of the ABO system. Unlike the ABO blood group system which was identified accidentally by the presence in humans of antibodies reactive against erythrocytes of different individuals, the MN blood group system was discovered by reaction of erythrocytes with antisera which were raised in animals with the assumption that there must be more than one blood group system

in humans. To explain the serological relationship between M and N antigens these authors (Landsteiner and Levine, 1927) concluded that most likely these antigens were the products of co-dominant alleles (M and N) that resided at a single genetic locus (reviewed in Issitt, 1981).

The erythrocyte sialoglycoproteins were subsequently shown to be involved in the phenomenon of haemagglutination. Hirst (1942) in his studies of influenza virus and chicken erythrocyte interaction (haemagglutination) demonstrated the presence of sialic acid on chicken erythrocyte receptors. It was demonstrated later that this binding was due to sialoglycoproteins (Gottschalk, 1960). Since then a number of other viruses have also been shown to attach to human erythrocytes and cause haemagglutination. Most of these viruses probably use the sialoglycoproteins, especially glycophorin A as their erythrocyte receptor (reviewed in Burness, 1981).

The receptor for influenza virus isolated from human erythrocytes has been characterized repeatedly by many investigators and shown to be a virus haemagglutination inhibitor (Howe *et al.*, 1957; Kathan *et al.*, 1961; Kathan and Winzler, 1963; Burness and Pardoe, 1983). Kathan and Winzler (1963) showed that the M and N antigens were on the same molecule as the haemagglutination inhibitor.

1.4.1. Sialopeptides

The biochemical characterization of human erythrocyte glycoproteins began

with the observation that treatment of erythrocytes with proteolytic enzymes resulted in marked decrease in electrophoretic mobility of the erythrocytes. Investigations on the mechanism underlying this phenomenon revealed that the decrease in the electrophoretic mobility was due to a decrease in net surface electrical charge of erythrocytes brought about by the release of sialoglycopeptides. Many investigators have isolated and characterized the released sialoglycopeptides following proteolytic enzyme treatment of either intact erythrocytes or proteins isolated from erythrocyte membranes (reviewed in Winzler, 1970). For instance, Cook *et al.* (1960) isolated a sialoglycopeptide by treating washed human erythrocytes with trypsin; based upon the absorption spectrum of the peptide they concluded that the polypeptide contained sialic acids (Cook *et al.*, 1960). Furthermore, release of this sialopeptide was correlated with about 20% decrease in net negative charge of erythrocytes which resulted in the concurrent decreased electrophoretic mobility of the erythrocytes (Cook *et al.*, 1960).

A glycopeptide which contained 30 to 50% of the sialic acid content of the intact erythrocytes and had a molecular weight of 10,000 daltons was isolated from human erythrocytes by trypsin treatment (Winzler *et al.* 1967). In addition to containing sialic acid the glycopeptide also contained galactose, N-acetylglucosamine, and N-acetylgalactosamine. The carbohydrate moiety was found to be linked with proteins by an O-glycosidic linkage between N-acetyl galactosamine and hydroxyl groups of serine and threonine. In addition to alkali-labile O-glycosidic linked

oligosaccharide, the sialopeptide also contained carbohydrate in an N-glycosidic linkage (Winzler *et al.*, 1967).

A glycopeptide was released from isolated erythrocyte membrane glycoprotein preparation by trypsin or pronase treatment; this peptide was identical with the one isolated by trypsin treatment of intact erythrocytes (Lisowska, 1968). In addition to this carbohydrate-containing peptide another peptide, which contained an abundance of hydrophobic amino acids and no carbohydrate was also detected (Lisowska, 1968) most probably representing the intramembranous portion of glycophorin A. Trypsin treatment of human erythrocytes destroyed their M and N antigenicity with the release of sialic acids and hexosamines (Makela *et al.*, 1960). Similarly, Ohkuma and Shinohara (1967) isolated a sialopeptide from pronase treated M, MN and N erythrocytes and concluded that these glycopeptides had structural similarities to each other. Burness and Pardoe (1983) isolated several sialoglycopeptides by treatment of glycophorin preparations with chymotrypsin.

Characterization of the glycopeptides released from human erythrocytes or from proteins extracted from them established the existence of sialoglycoproteins in human erythrocyte membranes. Although human erythrocyte membranes contain at least four different sialoglycoproteins (also see section 1.5) glycophorin A is the major sialoglycoprotein of human erythrocytes and hence a major proportion of the glycopeptides released from the erythrocyte glycoproteins in various studies discussed

above must have been derived from glycophorin A.

1.4.2. Characterization of glycoproteins of erythrocytes

As mentioned above (section 1.3) human erythrocytes proteins fractionated by SDS-polyacrylamide gel electrophoresis followed by PAS-staining separated into four distinct bands (Fairbanks *et al.*, 1971). Out of these four PAS bands the fastest moving band, just behind the tracking dye, was attributed to erythrocyte membrane glycolipid while the three slower moving bands labelled PAS-1, -2 and -3 were credited to erythrocytes sialoglycoproteins. With the exception of the band 3 protein which stained slightly with PAS (a minor band by PAS-staining), none of the other PAS-staining bands (PAS-1 to -4) were visible by stains for protein such as coomassie blue (Fairbanks *et al.*, 1971). The proportion of the bands seemed to vary depending upon the conditions of isolation of the proteins. Therefore, it was speculated, that the PAS-2 and -3 were due to either contamination by other glycoproteins or were degradative products of PAS-1 generated during the isolation procedures (Marchesi *et al.*, 1976). However, this speculation was not correct. In fact it had already been shown that the resolution of the human erythrocyte membrane glycoproteins depended upon the buffers and conditions used during isolation and electrophoresis (Teuch and Morrison, 1974) and indeed as will be discussed later (section 1.5) the erythrocyte membranes contain more than one sialoglycoprotein.

Many of the minor PAS-stainable bands do not comprise unique glycoproteins,

but rather are aggregates of lower molecular weight proteins (Tanner, 1978). Using the buffer system of Fairbanks *et al.* (1971) at least seven PAS-stainable bands were reported by Tanner (1978). The proportion of PAS-1 and PAS-2 varied depending upon the solubilization conditions of the glycoproteins (Teuch and Morrison, 1974). In the phosphate buffer system, PAS-1 was the predominant form, whereas if the glycoproteins were dissolved in solutions containing Tris buffer more of the PAS-2 was obtained (Teuch and Morrison, 1974). Furthermore, using the technique of two dimensional gel electrophoresis with ^{125}I-labelled proteins of human erythrocyte membranes, it was shown that when the sialoglycoproteins were dissolved in Tris buffer with SDS and heated at 100°C before polyacrylamide gel electrophoresis, the PAS-2 form was favoured over PAS-1 (Teuch and Morrison, 1974).

The observation that the two PAS bands (-1 and -2) represent the major sialoglycoprotein of human erythrocytes, termed glycophorin A (Marchesi *et al.*, 1972), was confirmed by investigation of PAS bands in En(a$^-$) cells which lack glycophorin A (Tanner and Anstee, 1976b). PAS-1 and PAS-2 have a dimer-monomer relationship (Furthmayr and Marchesi, 1976) and appear to be in equilibrium in erythrocytes where the dimeric form of glycophorin A, stabilized by hydrophobic interactions between the two monomers, is predominant (Welsh *et al.*, 1985). A synthetic peptide corresponding to the transmembrane domain of glycophorin A formed a specific reversible complex with native glycophorin A, glycoproteins of erythrocytes and of K562 cells in the absence of denaturants (Bormann *et al.*, 1989).

Under similar conditions synthetic peptides containing amino acid sequence from transmembrane domains of glycophorin C or interleukin 2-receptor Tac antigen did not interact with the transmembrane domain of glycophorin A nor did these peptides inhibit the binding of the synthetic glycophorin A transmembrane polypeptide to native glycophorin A. Therefore, it was proposed that the membrane spanning domain of glycophorin A contained all the information necessary for membrane insertion and anchoring as well as for specific binding sites that mediate interaction between membrane glycoproteins (Bormann *et al.*, 1989).

The PAS-3 band is produced by glycophorin B (Furthmayer *et al.*, 1975) which also shows up in SDS-polyacrylamide gel in the form of a heterodimer with glycophorin A (PAS-4) (Tanner, 1978). En(a$^-$) cells which lack glycophorin A also do not stain for PAS-4 (Tanner and Anstee, 1976b) and the S-s-U- erythrocytes which lack PAS-3 glycoprotein, also do not show the PAS-4 band (Dahr *et al.*, 1975b). In addition to glycophorin A and B, at least two other sialoglycoproteins, C and D, are also found in human erythrocytes (Anstee, 1990; Cartron *et al.*, 1990).

1.5. Nomenclature of various glycophorins

It is now generally agreed that human erythrocyte membranes contain at least four different sialoglycoproteins (Anstee, 1990; Cartron *et al.*, 1990). These proteins were designated as α, β, γ and δ in order of their increasing electrophoretic mobility (Anstee *et al.*, 1979; Anstee and Tanner, 1986). Three of these sialoglycoproteins, α,

δ and β, were also termed as A, B and C, respectively, in the order of their decreasing amount in erythrocyte membranes (Furthmayr *et al.*, 1975). Dahr *et al.* (1978) used the designation which was based upon the blood group antigenicity of these sialoglycoproteins; MN for glycophorin A, Ss for glycophorin B and component D for glycophorin C. In recent years use of the designation by Furthmayr *et al.*(1975) has been more prevalent. Beside the four glycophorin bands, the PAS-staining of SDS-polyacrylamide gel of human erythrocytes shows other minor components some of which are due to homo- and hetro-dimers of glycophorin A and B (Anstee *et al.*, 1979; Furthmayr *et al.*, 1975). The present perception of the original PAS bands (Fairbanks *et al.*,1971) is as follows: PAS-1 is a dimer of glycophorin A, PAS-2 and PAS-3 are monomers of glycophorin A and B, respectively. Additionally, yet another minor PAS band, PAS-2´ is glycophorin C (component D) (Anstee *et al.*, 1979) and Component E is glycophorin D (Anstee and Tanner, 1986).

1.6. Isolation procedures

Membrane glycoproteins such as glycophorins are associated firmly with the membrane lipids. Therefore, to isolate these glycoproteins it is necessary to disrupt and dissolve the membranes while still maintaining the structures of the proteins. Various methods have been used to isolate intact glycoproteins from human erythrocytes membranes. Most of the earlier methods were based upon phenol extraction of erythrocyte membranes (Kathan *et al.*, 1961; Springer *et al.*, 1966; Winzler, 1969). In an attempt to isolate the haemagglutination inhibitor from the

human erythrocytes a glycoprotein preparation was recovered by extraction with phenol followed by chloroform-methanol mixture (Kathan *et al.*, 1961). Since no further purification steps were performed this glycoprotein preparation was almost certainly contaminated with other glycoproteins. However, this preparation must contain glycophorin A as the predominant component, since the molecular weight of this component calculated from its sedimentation constant by analytical ultracentrifugation was found to be 31,000 daltons, a value similar to that obtained on the basis of amino acid sequence and carbohydrate content of glycophorin A (Tomita and Marchesi, 1975).

Sialoglycoproteins from human erythrocyte membranes were isolated in an aqueous solution using pyridine-water mixture at 4°C and removing other membrane proteins by precipitation (Blumenfeld *et al.*, 1970). Butanol extraction of human erythrocyte ghosts under high ionic condition also resulted in the extraction of these sialoglycoproteins in high yield as judged by estimation of MN activity; these sialoglycoproteins were further purified by gel filtration of the aqueous phase on sepharose 6B (Anstee and Tanner, 1974). Although this resulted in the isolation of the major sialoglycoprotein (glycophorin A) in fairly pure state, loading of samples at high concentration in SDS-polyacrylamide gel showed that the preparation still contained traces of other PAS-positive material (Anstee and Tanner, 1974). Extraction of human erythrocyte membranes with a mixture of chloroform-methanol followed by gel filtration on Sephadex G-100 resulted in the isolation of various

sialoglycoproteins which retained their MNSs activity and were free from other contaminating membrane proteins (Hamaguchi and Cleve, 1972). Another method for the isolation of the sialoglycoproteins used lithium diiodosalicylate (Marchesi and Andrews, 1971). Further modifications of this method incorporating lithium diiodosalicylate-phenol extraction followed by gel filtration in the presence of Zwitterionic detergent Ammonyx-Lo resulted in the isolation of glycophorin A (Furthmayr and Marchesi, 1983). This preparation was fairly free from other sialoglycoprotein (B, C and D) which could then be isolated from the remaining preparation (from which glycophorin A had already been isolated) by preparative SDS-polyacrylamide gel electrophoresis. High-performance liquid chromatography in combination with lectin-affinity and metal-interaction chromatography was also used to separate glycophorins A, B and C from other proteins and to fractionate into individual components (Corradini *et al.*, 1988)

1.7. Glycophorin A

Glycophorin A is the major sialoglycoprotein of human erythrocytes since it comprises about 85% of the total PAS-stainable material (Cartron *et al.*, 1990) and contains about 60% by weight of the total sialic acid of human erythrocytes membranes (Moulds and Dahr, 1989). A single erythrocyte contains about 2×10^5 to 10^6 copies of glycophorin A (Anstee, 1990).

1.7.1. Molecular weight

As discussed above (section 1.5) PAS-1 and -2 are the dimer and monomer forms of glycophorin A. The monomer runs with a mobility corresponding to a molecule of size at 36,000 daltons in SDS-polyacrylamide gel (Cartron *et al.*, 1990). In an earlier study of human erythrocyte membrane sialoglycoproteins a molecular weight as high as 10^6 was proposed (Baranowski *et al.*, 1959). Another study yielded a slightly lower value, about 6×10^5 daltons (Springer *et al.*, 1966). These erroneous values were caused by aggregation of the sialoglycoproteins (Blumenfeld *et al.*, 1970). Furthermore, estimation of molecular weight by SDS-polyacrylamide gel electrophoresis varies with the change in the concentration of acrylamide; for example, values of 53,000 and 40,000 daltons have been obtained in gels of 5 and 8%, respectively (Tanner and Boxer, 1972). The differing extent of binding of the detergents to carbohydrate rich and hydrophobic portions of the sialoglycoproteins in comparison to the non-membranous polypeptides used as standards for molecular weight calibration also contributed to these faulty values (Grefrath and Reynolds, 1974; Tanford and Reynolds, 1976). However, as early as the 1960s values of about 31,000 daltons were obtained from the sedimentation coefficient (Kathan *et al.*, 1961; Morawieki, 1964). This is close to the value for molecular weight of glycophorin A deduced on the basis of amino acid and carbohydrate composition (Tomita and Marchesi, 1975).

1.7.2. Molecular organization, three structural domains

The amphiphilic nature of the sialoglycoproteins of human erythrocytes first recognized by Morawieki (1964) was further clarified by Winzler (1969) on the basis of investigations on the sialopeptides released by trypsin treatment of human erythrocytes or glycoprotein extracted from erythrocyte membranes. Two peptides of very distinct characteristics were identified. One of these peptides was very rich in carbohydrate and the other in hydrophobic amino acids (Winzler, 1970). It was proposed that the sialoglycoprotein was anchored in the lipid bilayer through the hydrophobic polypeptide whereas the carbohydrate rich portion of the protein was exposed to the cytoplasm (Morawieki, 1964; Winzler, 1970). This prediction of orientation of glycophorin A has been found to be correct and is supported by the complete amino acid sequence of glycophorin A (Tomita and Marchesi, 1975) and various labelling methods discussed below. However, the original model of membrane anchoring is modified slightly in that the hydrophobic portion spans the lipid bilayer and the COOH-terminus passes completely through the membrane and is exposed to the cytoplasm (Tomita and Marchesi, 1975; Welsh *et al.*, 1985).

1.7.3. Amino acid sequence

The complete sequence of the 131 amino acids in glycophorin A suggested the presence of three structural domains, namely extracellular (residues 1 to 64), intramembranous (residues 65 to 96) and cytoplasmic (residues 97 to 131) (Tomita and Marchesi, 1975; Tomita *et al.*, 1978). The NH_2-terminal carbohydrate rich region

which is exposed to the outside of the cell membrane is followed by a stretch of amino acids comprised mainly of uncharged residues (a hydrophobic region rich in leucine and isoleucine) inserted into the membrane. There is a clustering of charged residues at both ends of the hydrophobic region. The presence of a hydrophobic region enclosed within charged residues is topographically similar to a cross section of a phospholipid bilayer. The small COOH-terminal cytoplasmic domain is hydrophilic and rich in proline (Segrest *et al.*, 1972). Although the presence of three structural domains of glycophorin A has been confirmed by many investigators the actual numbers of amino acids present in each of these domains has shown to be different from those originally proposed by Tomita and Marchesi (1975) and will be discussed in the next section).

Based upon the studies involving various physical methods and conformational analysis it is suggested that glycophorin A contains a number of α-helical regions: 1) arginine-39 to tyrosine-52; 2) glutamine-63 to glutamic acid-70; 3) glutamic acid-72 to leucine-89; 4) isoleucine-95 to lysine-101; and 5) leucine-118 to aspargine-125. With the exception of the structure of sequence "1", which occurs only at low pH and is possibly stabilized by non-covalent interaction with O-linked oligosaccharides, the other helical conformations occur in the dimeric form of glycophorin A at physiological pH and ionic strength (Welsh *et al.*, 1985). The α-helical conformation of sequence "3" is also in agreement with the generally observed conformation for membrane spanning domains of transmembrane proteins.

1.7.4. Labelling studies

Various labelling studies performed *in situ* on sialoglycoproteins of human erythrocyte membranes are consistent with the amino acid sequence regarding their structural organization. In a study of the orientation of erythrocyte membrane proteins using the membrane impermeable reagent, formylmethionyl sulfone methyl phosphate (FMMP), it was observed that when intact ghosts were used, only a part of glycophorin A was labelled, whereas in leaky ghosts the whole of the glycophorin A was labelled (Bretscher, 1971). This was taken as an evidence that a part of glycophorin A is exposed to the outside of the cell membrane while another region is either located within the membrane or passes completely through it and is exposed to the cytoplasm thus becoming accessible to reagent only in leaky ghosts.

Since FMMP is a low molecular weight reagent, it is possible that the membrane might be permeable to such a reagent in some conditions causing faulty labelling (Tanner, 1978). A study of the orientation of glycophorin A resulted in a similar conclusion using high molecular weight labelling reagent, lactoperoxidase which selectively labels tyrosine residues and can not cross cell membranes (Segrest *et al.*, 1973). Only the residues present in the NH_2-terminal domain of glycophorin A were labelled when lactoperoxidase was used with intact erythrocytes. The presence of label only in the NH_2-terminal domain in intact erythrocytes indicated that this domain was exposed to the outside of the erythrocyte membrane; the COOH-terminal domain which contains one tyrosine residue, was not labelled on intact cells

indicating it is intramembranous or exposed to the inside of the cells (Segrest *et al.*, 1973; Morrison *et al.*, 1974).

Marchesi *et al.* (1972) arrived at similar conclusions using phytohaemagglutinin (PHA)-conjugated ferritin. When observed in electron microscope, PHA-conjugated ferritin was localized to the sites on the membrane that corresponded to the pattern of distribution of the electron dense particle (intramembranous particles) which were embedded between the two faces of the lipid bilayer. This suggested that the PHA receptors on the extracellular surface were connected physically to the underlying intramembranous particles. Further, trypsin digestion which modifies glycophorin A also modified the pattern of intramembranous particles, again indicating that the glycophorin A is embedded in the erythrocyte membrane. Other supporting evidence for a transmembrane orientation for glycophorin A was provided by immunochemical labelling methods. For instance, Cotmore *et al.* (1977) found that antibodies raised against ferritin conjugated to the region containing amino acids 102 to 118 of glycophorin A did not react with intact or trypsinized erythrocytes. However, the same authors reported that in fixed sections of these cells only the inner surface was labelled which confirmed the transmembrane orientation of glycophorin A and ruled out the possibility of faulty labelling due to rearrangement of membrane proteins during the labelling procedure. Bretscher (1975) provided further proof that the COOH-terminal lies in the cytoplasm by isolating an FMMP-labelled polypeptide (C2) by cyanogen bromide treatment of a glycophorin A preparation. When the

COOH-terminal peptide (C2) was subjected to finger print analysis using trypsin and chymotrypsin it yielded two main labelled peptides. These two peptides were identical to two other peptides (G1 and G2) previously shown to come from that part of glycophorin A which is on the cytoplasmic side of the membrane (Segrest *et al.*, 1972). Hence, it was concluded that the COOH-terminal region of glycophorin A lies in the red cell cytoplasm (Bretscher, 1975).

The earlier prediction of the three structural domains in glycophorin A were further explored using the hydrophobic labelling reagents, such as 1-azido-4-iodo[^{3}H]benzene (Wells and Findlay, 1979). Most of the radioactivity was found to be incorporated in a chymotryptic peptide (alanine-65 to arginine-98) which indicated that this region of glycophorin A lies predominantly within the membrane. Closer examination of the labelled products showed very little radioactivity up to histidine-66, suggesting that the point of entry into the lipid membrane environment was to the COOH-terminal side of histidine-66. When a photosensitive carbene attached covalently to fatty acids was used to label glycophorin A in reconstituted lipid vesicles a high level of incorporation was in glutamic acid-70 and the residues immediately adjacent to it in the COOH-terminal direction (proline-71 and glutamic acid-72 (Ross *et al.*, 1982); this was interpreted by Welsh *et al.* (1985) as demonstrating that glutamic acid-70 to glutamic acid-72 were close to the membrane boundary. Therefore, it was proposed that glutamic acid-72 and tyrosine-93 define, respectively, the outer and inner limit of the membrane domain of glycophorin A *in vivo*.

Furthermore, it was suggested that the residues that span the membrane (isoleucine-73 to serine-92) fold into an α-helical conformation with a length consistent with the dimensions of lipid bilayer (Welsh *et al.*, 1985). These labelling studies have also defined the limits of the extracellular and intracellular domains. The extracellular NH_2-terminal domains consists of amino acid residue 1 to about 72 and the COOH-terminal domain consists of amino acid residues 93 to 131 (Welsh *et al.*, 1985).

1.7.5. Carbohydrate structure

The NH_2-terminal domain of glycophorin A contains a very high proportion of serine and threonine. On average fifteen of these serine or threonine residues have simple oligosaccharide units linked through O-glycosidic bonds (Dahr, 1986). In addition to these simple O-linked oligosaccharides there is a single complex oligosaccharide unit attached through an N-glycosidic bond to the aspargine residue at amino acid position 26 (Tomita and Marchesi, 1975).

1.7.5.1. O-linked oligosaccharides

The structure of the alkali-labile oligosaccharide units attached to serine and threonine residues in glycophorins was investigated by alkaline borohydride treatment (Winzler *et al.*, 1967; Thomas and Winzler, 1969 a and b; Adamany and Kathan, 1969). It was shown that most of the oligosaccharide was linked to the peptide chain through O-glycosidic bonds from the anomeric carbon of N-acetylgalactosamine to the hydroxyl groups of serine and threonine. Further, the low molecular weight

oligosaccharides released by alkaline borohydride treatment were mostly tetrasaccharides containing N-acetylgalactosamine, galactose and N-acetyl neuraminic acid (sialic acid) with a molar ratio of 1:1:2, respectively [Fig. 1.1 (A)] (Thomas and Winzler, 1969 a and b; Adamany and Kathan, 1969). In addition to the tetrasaccharides alternate oligosaccharide structures were also proposed (Springer and Desai, 1975; Fukuda *et al.*, 1987; Adamany and Blumenfeld, 1983). For example, an alternative structure [Fig. 1.1 (B)] of a pentasaccharide containing two galactose molecules per oligosaccharide unit was also proposed (Springer and Desai, 1975). However, it is believed that the tetrasaccharide structure shown in Fig. 1.1 (A) is the predominant structure (Fukuda *et al.*, 1987; Adamany and Blumenfeld, 1983). Notwithstanding this, there is some evidence to support the presence of heterogeneity in the O-linked oligosaccharides. Fukuda *et al.* (1987) reported that in addition to the tetrasaccharide [Fig. 1.1 (A)] which contains two sialic acid residues, and trisaccharides containing one sialic acid residue (monosialylated trisaccharides), glycophorins may also possess novel structures containing three sialic acid residues (trisialylated pentasaccharides). These authors found that the tetrasaccharide was the predominant form (78%); whereas 17% and 5% of the O-linked oligosaccharides were in the form of monosialylated trisaccharide and novel trisialylated pentasaccharide structures, respectively. Adamany and Blumenfeld (1983) reported a family whose individual members in addition to the common tetrasaccharide contained substitutions of varying chain lengths on N-acetylgalactosamine containing N-acetylglucosamine and galactose.

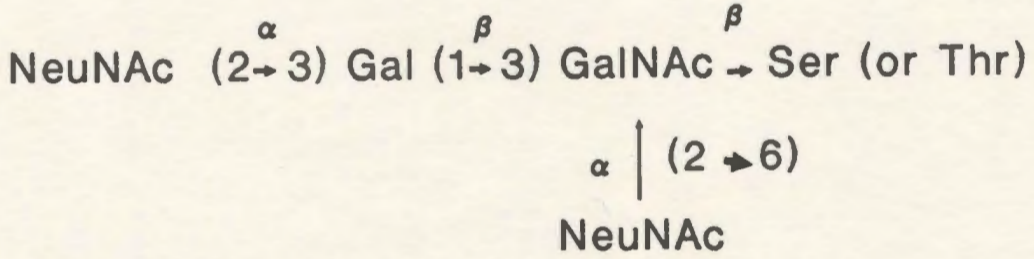
Fig. 1.1. Structures of O-linked oligosaccharides of glycophorin A.

(A) As proposed by Thomas and Winzler (1969a, b) and

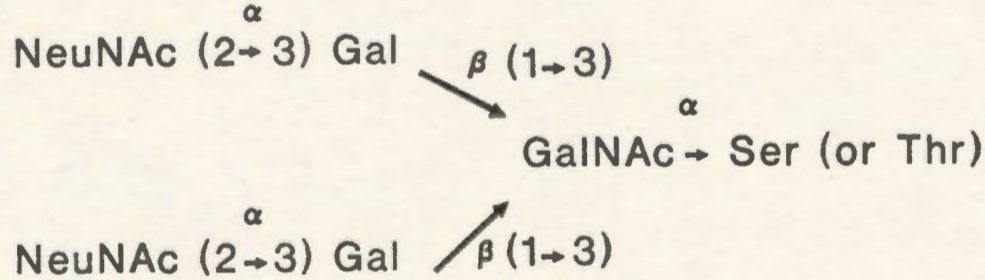
(B) As proposed by Springer and Desai (1975)

The abbreviation used are: NeuNAc, N-acetyl neuraminic acid (sialic acid); Gal, galactose and GalNAc, N-acetyl galactosamine.

(A)


NeuNAc (2→3) Gal (1→3) GalNAc → Ser (or Thr)
α
β
β
α (2→6)
NeuNAc

(B)


NeuNAc (2→3) Gal
β (1→3)
GalNAc → Ser (or Thr)
NeuNAc (2→3) Gal
β (1→3)

1.7.5.2. N-linked oligosaccharides

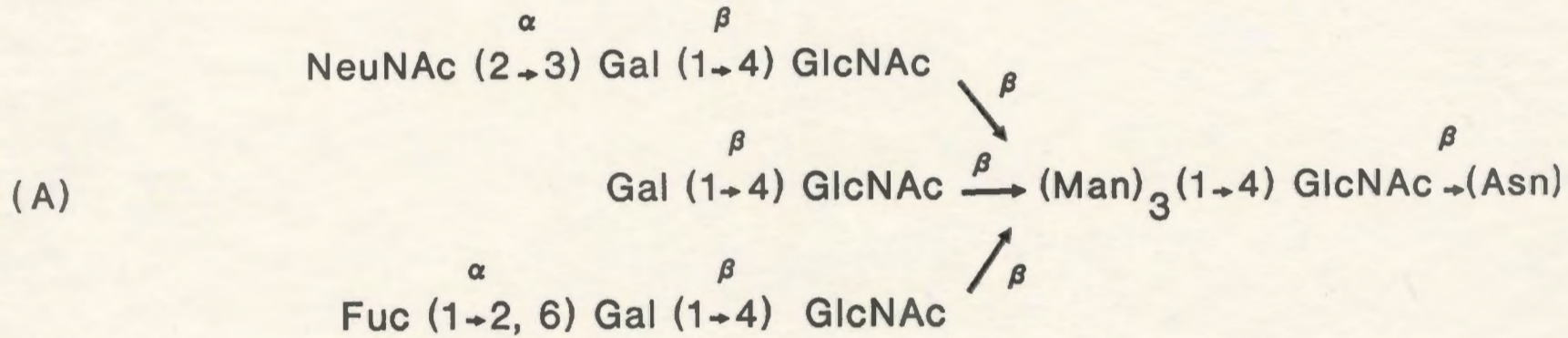
Glycophorin A contains a single, alkali-stable, complex type of oligosaccharide unit attached to the aspargine-26 through an N-glycosidic linkage. Two different structures were proposed (Kornfeld *et al.*, 1970; Thomas and Winzler, 1971 [Fig. 1.2 (A) and Fig. 1.2 (B)]. However, both of these structures represent a carbohydrate composition that is very different (Marchesi *et al.*, 1976) from that reported for a pure preparation of glycophorin A (Tomita and Marchesi, 1975). The reason for this discrepancy could be the difficulty in isolating the intact oligosaccharide unit. The aspargine residue to which the complex oligosaccharide unit is attached is directly linked to an aspartic acid residue. This aspargine-aspartic acid bond is relatively resistant to proteolytic cleavage, so that most of the glycopeptides derived from this region might have two aspartic acid residues, and calculation of the molar ratios of the carbohydrate residues would be misleading (Tomita and Marchesi, 1975). Furthermore, the earlier glycophorin A preparations used to analyze the oligosaccharide structure were almost certainly contaminated with other glycoproteins (Marchesi *et al.*, 1976). Two alternate structures for N-linked oligosaccharide of glycophorin A have been proposed which were determined after the isolation of intact oligosaccharide units by hydrazinolysis (Irimura *et al.*, 1981; Yoshima *et al.*, 1980). These structures have a carbohydrate composition more closely related to those isolated from a pure glycophorin A preparation also used for amino acid sequencing (Tomita and Marchesi, 1975). The two structures proposed by Yoshima *et al.* (1980) varied in their sialic acid composition [Fig. 1.3 (A) and (B)]. It was

Fig. 1.2. Structures of N-linked oligosaccharide of glycophorin A.

(A) As proposed by Kornfeld *et al.* (1970) and

(B) As proposed by Thomas and Winzler (1971)

The abbreviation used are: NeuNAc, N-acetyl neuraminic acid (sialic acid); Gal, galactose; GlcNAc, N-acetyl glucosamine; Man, mannose and Fuc, fucose.


(A)
NeuNAc (2→3) Gal (1→4) GlcNAc
Gal (1→4) GlcNAc
(Man)3 (1→4) GlcNAc →(Asn)
Fuc (1→2, 6) Gal (1→4) GlcNAc

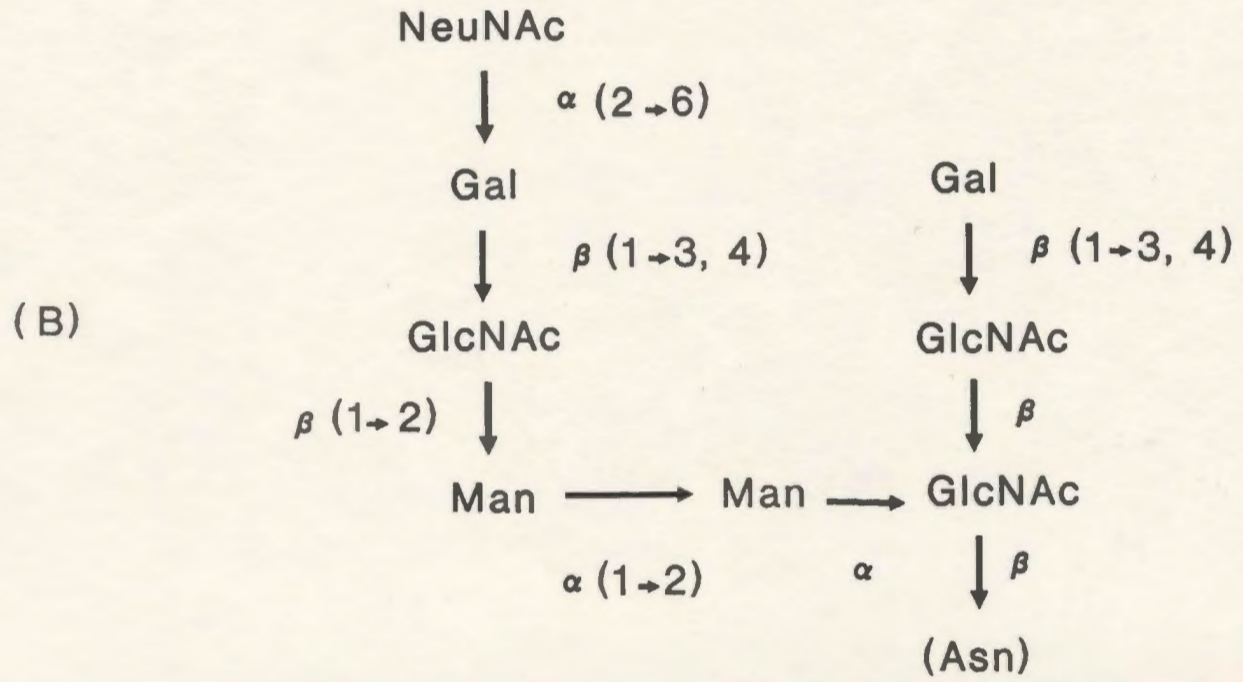
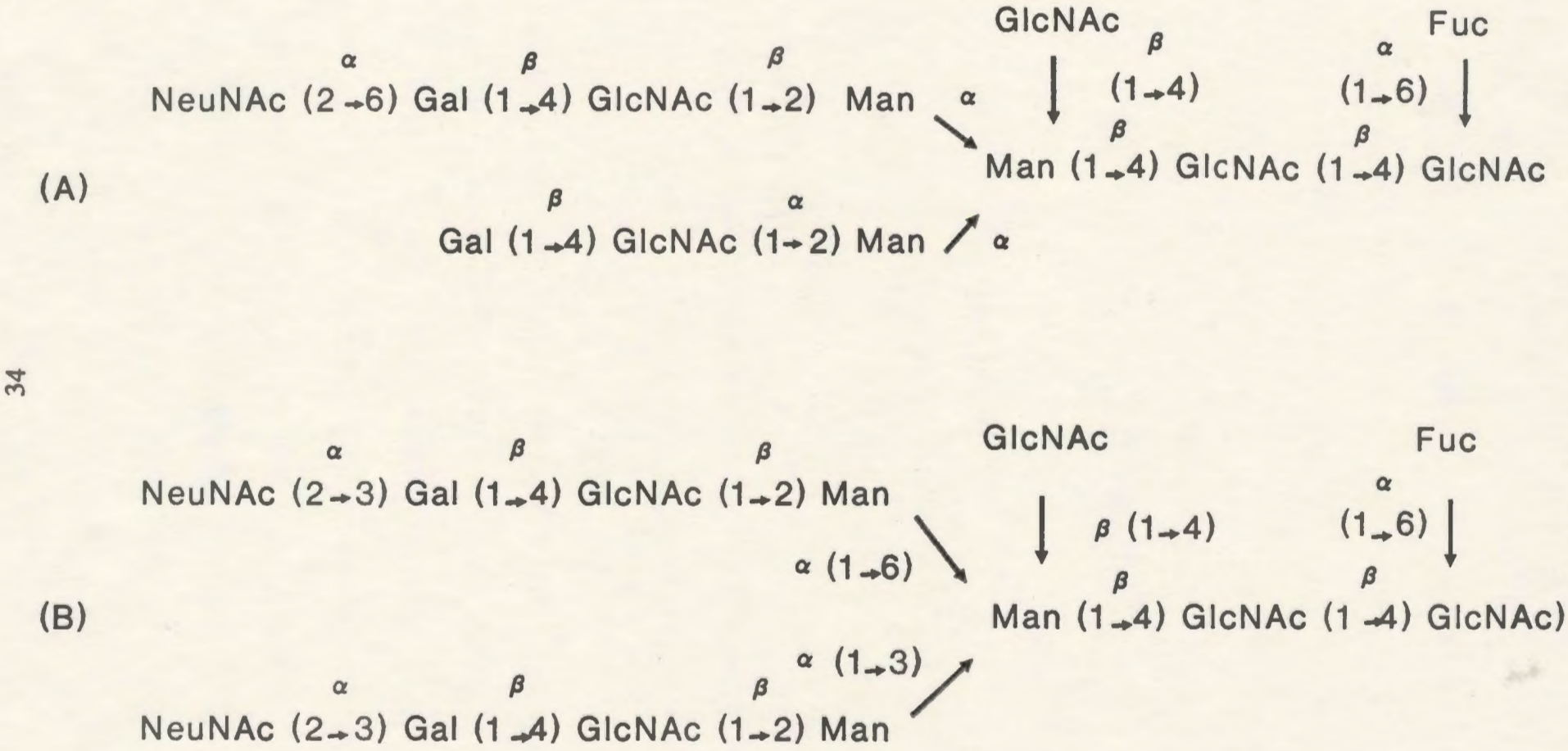

(B)
NeuNAc
α (2→6)
Gal
β (1→3, 4)
GlcNAc
β (1→2)
Man
α (1→2)
Man
α
Gal
β (1→3, 4)
GlcNAc
β
GlcNAc
β
(Asn)

Fig. 1.3. Structures of N-linked oligosaccharide of glycophorin A as proposed by Yoshima et al. (1980).

The abbreviation used are: NeuNAc, N-acetyl neuraminic acid (sialic acid); Gal, galactose; GlcNAc, N-acetyl glucosamine; Man, mannose and Fuc, fucose.


(A)
NeuNAc (2→6) Gal (1→4) GlcNAc (1→2) Man
α β β α
GlcNAc
β
(1→4)
Fuc
α
(1→6)
Man (1→4) GlcNAc (1→4) GlcNAc
β β
Gal (1→4) GlcNAc (1→2) Man α
β α
(B)
NeuNAc (2→3) Gal (1→4) GlcNAc (1→2) Man
α β β
α (1→6)
α (1→3)
NeuNAc (2→3) Gal (1→4) GlcNAc (1→2) Man
α β β
GlcNAc
β (1→4)
Fuc
α
(1→6)
Man (1→4) GlcNAc (1→4) GlcNAc)
β β

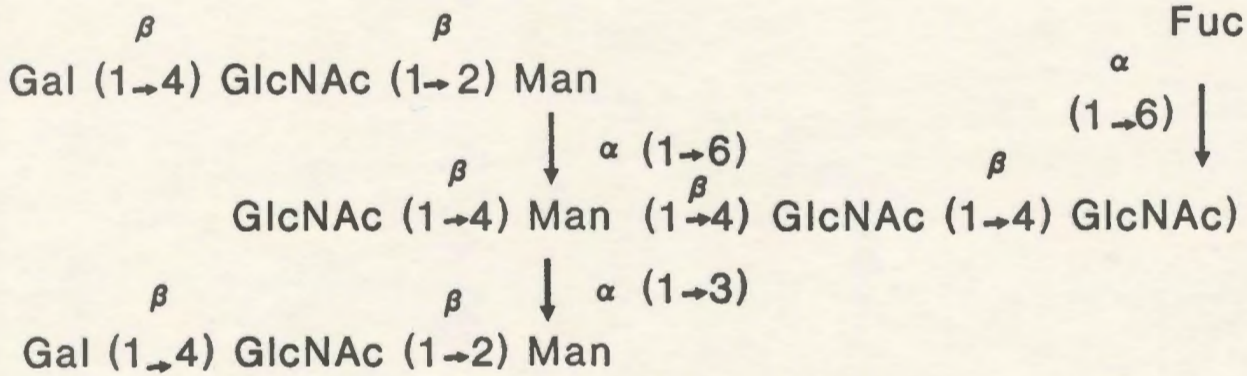
suggested that the structure which contained less sialic acid [Fig. 1.3 (A)] was a partially desialylated product of other form [Fig. 1.3 (B)], or alternatively the two structures had product-intermediate relationship in the biosynthesis of glycophorin A (Yoshima *et al.*, 1980). An alternate structure for the N-linked oligosaccharide was also proposed (Irimura *et al.*, 1981), which was similar in composition to those proposed by Yoshima *et al.* (1980) except that it contained no sialic acids (Fig. 1.4). Despite the variations these structures had a carbohydrate composition similar to those obtained for purified glycophorin A and hence could represent average structures of N-linked oligosaccharides of glycophorin A (Anstee, 1990). The minor variations could be a reflection of heterogeneity of this oligosaccharide unit similar to that observed for O-linked oligosaccharides (Fukuda *et al.*, 1987).

1.7.6. Biosynthesis of glycophorin A

Glycophorin A is expressed only in erythroid cells (also see section 1.7.7.1). In addition to normal bone marrow cells, a few cell lines derived from human leukaemias also express glycophorin A on their surface, for example, K562 cells (Lozzio and Lozzio, 1975; Villeval *et al.*, 1983), HEL cells (Papayannopoulou *et al.*, 1987) and OCIM1 and OCIM2 cells (Papayannopoulu *et al.*, 1988). The K562 cell glycophorins have been extensively studied and evidence so far collected shows that, with some minor differences, K562 glycophorins are similar to those found on mature erythrocytes (Gahmberg *et al.*, 1979). For instance, the O-linked oligosaccharides of glycophorin A on K562 cell have less sialic acid than that of mature erythrocytes

Fig. 1.4. Structure of N-linked oligosaccharide of glycophorin A as proposed by Irimura et al. (1981).

The abbreviation used are: Gal, galactose; GlcNAc, N-acetyl glucosamine; Man, mannose and Fuc, fucose.


Fuc
α
(1→6)
β
β
Gal (1→4) GlcNAc (1→2) Man
α (1→6)
β
β
β
GlcNAc (1→4) Man (1→4) GlcNAc (1→4) GlcNAc)
α (1→3)
β
β
Gal (1→4) GlcNAc (1→2) Man

(Gahmberg *et al.*, 1979). There are about 1.1 x 10^6 molecules of glycophorin A on each K562 cell (Gahmberg *et al.*, 1979). The glycophorin A found on erythrocytes is relatively immobile and not susceptible to ligand-induced redistribution as observed in K562 cells. Further, it was noted that there was lower activity of MN antigenic determinants in K562 cells, possibly a reflection of decreased sialylation of K562 cell glycophorin A (Gahmberg *et al.*, 1984).

The biosynthesis of glycophorin A has been studied in K562 cells (Jokinen *et al.*, 1979; Gahmberg *et al.*,1980, 1983; Morrow and Rubin, 1987). Two different pathways have been proposed (Gahmberg *et al.*, 1980, 1983; Morrow and Rubin, 1987). Cell free translation of 16 to 18S poly(A)$^+$ RNA from K562 cells revealed that glycophorin A was synthesized as a single polypeptide of molecular weight 19,500 daltons, which was about 5000 daltons more than that is expected based upon the 131 amino acids of mature glycophorin A (Tomita and Marchesi, 1975); it was proposed that glycophorin A is synthesized with a cleavable signal peptide (Jokinen *et al.*, 1981).

Cell free translation of poly(A)$^+$ RNA from K562 cell in the presence of dog pancreas membranes (Jokinen *et al.*, 1981) resulted in the synthesis of a precursor with apparent molecular weight, 37,000 daltons, which was similar to that obtained *in vivo* (Jokinen *et al.*, 1979; Gahmberg *et al.*, 1983). The peptide synthesized *in vitro* bound to a lentil lectin-Sepharose column which binds α-D-mannoside residues,

indicating the presence of N-glycosidic oligosaccharides. The same peptide showed a decrease in the molecular weight of only about 2000 daltons after endoglycosidase H treatment which cleaves N-glycosidically linked oligosaccharides. The decrease in apparent molecular weight after endoglycosidase treatment to 35,000 daltons and not to 19,500 daltons suggested that it also contained O-linked oligosaccharides (Jokinen *et al.*, 1981).

Gahmberg *et al.* (1980) reported that tunicamycin which inhibits N-glycosylation, without affecting O-glycosylation, did not inhibit the incorporation of glycophorin A into the membranes of K562 cells. This indicated that the N-glycosidic carbohydrate was not necessary for intracellular transport and proper membrane localization. Jokinen *et al.* (1985) detected two new precursors of apparent molecular weight of 24,000 and 30,000 daltons. These precursors were found to contain O-glycosidic oligosaccharides, but no N-glycosidic carbohydrate moiety. Jokinen *et al.* (1985) concluded that a fraction of glycophorin A molecules became O-glycosylated at an early stage and another fraction at a later stage of biosynthesis.

In contrast to the above pathway, a simpler pathway for glycosylation of glycophorin A in K562 cells was proposed (Morrow and Rubin, 1987). In this study a unique precursor was identified which contained both the N- and O-linked oligosaccharides. In addition, two precursors of apparent molecular weight 27,000 and 31,000 daltons were identified. The smaller of these contained one N-linked high

mannose type oligosaccharide and rapidly converted to a transient 31,000 intermediate by addition of N-acetylgalactosamine residues to serine/threonine hydroxyl groups. Subsequent maturation involved conversion of the high mannose N-linked oligosaccharide to a complex type and addition of a galactose and terminal sialic acid residues to the O-linked N-acetylgalactosamine residues.

The discrepancy between these two studies (Jokinen *et al.*, 1985; Morrow and Rubin, 1987) could be due to presence of two different populations of glycophorins in K562 cells having differences in the structure of their N-linked oligosaccharide (Silver *et al.*, 1987). On the other hand, the two populations of glycophorin A precursors might reflect some heterogeneity in cell lines which may have arisen during the numerous *in vitro* passages of these cells since they were first isolated from a patient (Lozzio and Lozzio, 1975).

1.7.7. Physiological functions of glycophorin A

Glycophorins have been characterized extensively since they were discovered as blood group antigens, MN, in 1930s. Although there is some information on the function of glycophorins C and D (section 1.12), that for glycophorins A and B remains unknown. Nevertheless, many different functions have been suggested for glycophorins and some of these will be discussed in the following section.

1.7.7.1. Differentiation antigen

Many studies have shown that glycophorins are expressed only on the surface of cells belonging to erythroid lineage and not expressed in other haemopoetic cells. For instance, Fukuda and Fukuda (1984) by cell surface labelling using periodate oxidation followed by reduction with $NaB[^3H]_4$ or by endo-β-galactosidase digestion (which oxidizes C-6 of galactose or N-acetylgalactosamine) followed by $NaB[^3H]_4$ reduction showed that glycophorins were present as major components in erythroid precursor cells, proerythroblasts and erythroblasts and absent from the surface of granulocytes and monocytes. Similarly, in another study using metabolic labelling and immunofluorescence, glycophorin A expression was also detected in the proerythroblast stage (Yurchenco and Furthmayr, 1980). Gahmberg *et al.* (1978), also detected glycophorin A expression only in erythroid cells using bone marrow cells with a glycophorin A specific antiserum in a staphylococcus protein rosseting assay.

In a comparative study of O-linked oligosaccharides attached to a group of closely related proteins, termed leukosialins, from various leukaemic cell lines it was shown that proteins of each haemopoetic cell lineage (erythroid, myeloid and lymphoid) contained a characteristic set of O-linked oligosaccharide attached to their glycoproteins (Carlsson *et al.*, 1986). Differences were found in the core structure as well as in the extent and linkage of sialylation. The structure of the O-linked oligosaccharides in K562 leukosialin was similar to that of glycophorin A. When expressed in other cell lines such as HL-60 (a promyelocytic) or HSB-2 (a T-

lymphoid) leukosialin depicted a carbohydrate structure that was different from that found on K562 cells. Therefore, it was concluded that each cell line shows a specific carbohydrate structure that is characteristics of that particular cell line (Carlsson *et al.*, 1986).

It has been suggested that glycophorin A which does not seem to have an important physiological role to play during the circulatory phase of the erythrocytes life cycle might have an important function at earlier stages of erythroid cell development (Fukuda and Fukuda, 1984). While undergoing different stages of development, the erythroid precursor cells, in addition to showing the developmental stage specific expression of globin genes, also show a number of changes in their surface molecules (Fukuda and Fukuda, 1984). For example, there is change in the carbohydrate structure linked to band 3 protein; as a result the fetal i antigen is converted into adult I antigen (Fukuda and Fukuda, 1984). Similarly glycophorin A shows stage specific changes during maturation of precursor cells. For instance, the M and N antigen activity of glycophorins was either very weak or completely absent before the polychromatic normoblast stage even though glycophorin A molecules could be detected at the earliest recognizable erythroid precursor, the proerythroblast (Ekblom *et al.*, 1985). Further, cells at various stages of erythroid differentiation showed very strong reaction with a polyclonal anti-glycophorin A antiserum. In contrast, a glycophorin A specific monoclonal antibody, R10 reacted very poorly with cells from earlier stages, namely, normoblast and basophilic normoblasts, but the cells

from polychromatic normoblast and later stages including mature erythrocytes showed strong reaction with the antibody R10. This increased level of reaction of monoclonal antibody with erythroid cells at later stages of development was due to the increase O-glycosylation and not due to an increase in glycophorin A content (Tonkonow *et al.*, 1982). It was further reported that the human erythroleukaemic cell line, K562 exhibited increased O-glycosylation on induction with haemin (Gahmberg *et al.*, 1984). These results were in contrast to the observation that increased globin gene expression in these cells resulted from increased mRNA levels induced by haemin. Glycophorin levels seemed to be constitutive and K562 cells exhibited no apparent increase in content after haemin induction (Tonkonow *et al.*, 1982). However, there is one report suggesting that differentiation of erythroid cells was not accompanied by increased O-glycosylation (Loken *et al.*, 1987). These workers showed by immunofluorescence using two monoclonal antibodies, one dependent and the other independent of glycosylation, that glycophorin A expression in normal bone marrow was accompanied by coordinate glycosylation.

Comparing the expression of glycophorin A with other glycoproteins, Fukuda and Fukuda (1984) noted that the glycophorin A content increased with maturation of erythroid cells and was a major component in erythroblasts generated *in vitro*. In contrast, band 3 was barely detectable at the erythroblast stage but seemed to appear slightly later, while two other glycoproteins, GP 105 and GP 95 appeared only in immature cells (Fukuda and Fukuda, 1984). These authors further noted that band

3 protein and spectrin expression coincided with a stage in which they had a functional role, i.e. in mature erythrocytes. Therefore, it is possible that glycophorin A may also serve as an erythroid cell specific marker and take part in some cellular recognition process in later stages of erythroid development, while GP105 and GP95 serve the same role in earlier stages.

Glycophorin A may also have a role in the circulating erythrocyte. This would be consistent with the finding that it is found at a much higher level in mature erythrocytes (Fukuda and Fukuda, 1984; Gahmberg *et al.*, 1984). The reported increased O-glycosylation of glycophorin A (Gahmberg *et al.*, 1984) possibly takes place in preparation for its function in the circulatory phase. Other supporting evidence for a role in mature erythrocytes can be drawn from a report which showed that as maturation occurs, proteins which are destined to be an important part of the erythrocyte membrane, such as spectrin are excluded from coated pits and remain at the cell surface. Other proteins such as the transferrin receptor present abundantly at earlier stages are endocytosed in coated pits and degraded. (Marshal *et al.*, 1984). These authors reported that when K562 cells were induced to differentiate, they produced globin and redistributed glycophorin and spectrin to one pole of the cell. These apparent changes in glycosylation and redistribution of glycophorin A on cell surface during erythroid maturation seems to point towards the importance of these molecules at these stages (Fukuda and Fukuda, 1984).

1.7.7.2. Role in the removal of senescent erythrocytes from circulation

It has been suggested that glycophorins and especially their carbohydrate moieties have a role in aging and sequestration of erythrocytes from the circulation (Aminoff, 1988). This speculation is based upon the observation that young and old erythrocytes in circulation have different sialic acid contents, the former having 10-15% more. It was suggested that progression of desialylation of glycophorin A takes place as the erythrocytes age. The complete desialylation of glycophorin A results in the appearance of multiple galactose-beta-(1,3) N-acetylgalacotosamine disaccharide residues on erythrocytes. It is proposed that these residues are recognized by the reticuloendothelial system, resulting in the sequestration of these desialylated senescent erythrocytes from circulation. Aminoff (1988) has reviewed the evidence accumulated for or against this hypothesis but the matter is not resolved.

1.7.7.3. Role in membrane function

In its life span of 120 days each erythrocyte travels a long distance passing through narrow capillaries where it might undergo the stress of shearing forces (Chasis *et al.*, 1985). Therefore, to keep the integrity of its membrane while performing its normal functions each erythrocyte must undergo drastic changes in its conformation, and be capable of regaining its shape. The transmembrane location of glycophorins, and especially of glycophorin A has led to the suggestion that glycophorin A may have a role in membrane deformability (Chasis *et al.*, 1985). Binding of ligands specific for glycophorin A but not for blood group A and B nor

band 3 glycoprotein, with erythrocytes or resealed membranes resulted in a marked decrease in membrane deformability. This indicated that glycophorin A molecules were probably involved in the process of membrane deformability. Erythrocyte membranes or cytoskeletons prepared from erythrocytes after glycophorin A specific ligand binding contained glycophorin A. This partitioning of glycophorin A in the insoluble cytoskeleton fraction was not observed in the cells deficient in skeletal proteins such as spectrin or protein 4.1. Based upon these observations, Chasis *et al.* (1985) suggested that there is a ligand-induced interaction between glycophorin A and skeletal proteins and that this interaction can directly influence membrane deformability. This ligand-induced rigidity may have a significant role in shape changes during passage of erythrocytes through narrow capillaries. However, in a later report Chasis and Schrier (1989) pointed out that some rigid cells are capable of shape change, and, therefore, membrane deformability is not predictive of impaired capacity for shape change.

It has been suggested that the interaction of glycophorin A with the cytoskeleton may have a role in signal transduction across the membrane similar to that of hormones and neurotransmitters receptors (Butterfield *et al.*, 1983). This suggestion was based upon finding that modification of sialic acid on the outer surface of erythrocyte membranes induced alterations in the physical state of proteins on the opposite side of lipid bilayer. Such phenomena may be important in the normal functioning of erythrocytes. Haemin, a haemoglobin breakdown product is

elevated in aged erythrocytes. It also accumulates in the erythrocyte lipid bilayer and the protein skeletal network in individuals with sickle-cell anaemia and β-thalassaemia. It has been suggested that haemin disrupts skeletal protein-protein interaction (Wyse and Butterfield, 1989). Since there is some evidence that glycophorin A interacts with the cytoskeleton (Anderson and Lovrien, 1984), it is possible that the haemin-induced protein:protein disruption results in conformational changes in glycophorin A on the extracellular face of membrane which in turn results in changes in the conformation of cell surface sialic acid. The change in conformation of sialic acid may have some significance in the normal role of glycophorin A in maintaining the membrane integrity (Anderson and Lovrien, 1984).

There is some support for the suggestion that O-linked carbohydrates of glycophorin A may have a role in stabilization of the membrane (Pinnaduwage and Huang, 1989). *In vitro* studies with dioleoylphosphatidyl-ethanolamine showed that glycophorin A stabilized bilayers. This activity was dependent upon sialic acid content since neuraminidase treatment of glycophorin A before incorporation into dioleoylphosphatidyl-ethanolamine bilayer reduced its stabilizing ability. There is some evidence, supplied again by *in vitro* studies that incorporation of glycophorin A into large unilamellar vesicles of dioleoylphosphotidyl-choline increases bilayer permeability (Van Hoogvest *et al.*, 1984). However, the physiological significance of such studies remains to be determined.

There is at least one report which suggests that glycophorins interact with cytoskeletal proteins, especially protein 4.1 of erythrocytes (Anderson and Lovrien (1984). It was shown that protein 4.1 associated specifically with the cytoplasmic end of glycophorin A on inside-out erythrocyte membrane vesicles and also with glycophorin A reconstituted into phosphotidylcholine vesicles. The evidence was that the specific association was inhibited with antibodies raised against an affinity purified cyanogen bromide fragment from the cytoplasmic domain of glycophorin A. This report was in contrast to other reports (Mueller and Morrison, 1981; Alloisio *et al.*, 1985) providing evidence that glycophorin C and D and not A interacted with cytoskeletal proteins (see section 1.12).

Further evidence against interaction of glycophorin A and cytoskeletal proteins is that no abnormal phenotype such as elliptocytosis has been so far reported in the individuals with for instance En (a-) cells which show a complete lack of glycophorin A (Tanner and Anstee, 1976b). In contrast, as discussed in section 1.12 lack of glycophorin C and possibly of D does result in the abnormal phenotype of erythrocytes. It is possible that the interaction between glycophorin A and band 4.1 only takes place in non-physiological conditions, such as in phosphotidylcholine vesicles and in inside out erythrocyte membrane ghosts but not *in vivo*. It is also possible that even though glycophorin A does interact with band 4.1 *in vivo*, this interaction is not essential for the maintaining the erythrocyte shape. Comparative analysis of the interactions of glycophorins A, B, C and D revealed that glycophorin

C and D bind strongly to the membrane skeleton and are difficult to solubilize whereas glycophorin A and B bind to a lesser extent and are mostly recovered in the soluble fraction (Bloy *et al.*, 1987).

Another role for the carbohydrate part of glycophorin A may be to decrease unwanted interaction with other cells or molecules (Vitala and Jarnefelt, 1989). This may help in decreasing the adherence of erythrocytes to capillaries during circulation. Glycophorin A is heavily glycosylated and has many negatively charged sialic acids which enables this glycoprotein to fit this role very well by causing negative charge repulsion. The carbohydrate component can have many conformations; in one extreme conformation they may be stretched out whereas in another extreme they may remain very close to the membrane. These conformational extremes, or any variation in between, may have a significance in acting as an obstacle for protein molecules in the plasma or in the surface membranes of other cells to get into close contact with erythrocytes (Vitala and Jarnefelt, 1989).

Although glycophorin A has been implicated in many physiological functions, as reviewed above its normal function is still far from clear and much more work is needed to resolve this issue.

1.7.8. Blood group activity of glycophorin A

1.7.8.1. MN blood group activity

The earliest investigations to characterize the MN antigens revealed that

neuraminidase treatment of erythrocytes destroyed their ability to react with anti-M and anti-N antibodies. This led Springer *et al.* (1976) to attempt to isolate the oligosaccharide structures that carry the determinants that are subjected to enzymic degradation. They concluded that the M and N antigens were associated with a tetra- and pentasaccharide structure, respectively, present on erythrocyte membrane sialoglycoproteins, the major one of which is glycophorin A. It was further noted that the oligosaccharide side chain contained N-acetyl galactosamine, galactose and had terminal sialic acid residues. Further, since neuraminidase which removes sialic acids also inhibits the MN antigen activity of the sialoglycoproteins, sialic acid must be involved and could have an important role in determining the structure of, and differences between M and N antigens. Springer *et al.* (1976) suggested that difference between M and N was that M contained an extra sialic acid residue. These authors believed that they could take M and degrade it to N by removal of one of the terminal sialic acid residues or add sialic acids to N antigen and convert it to M antigen. Evidence believed to support this assumption was the finding by many workers that the erythrocytes of type M+N- carried a small amount of antigen that reacted with anti-N (Baranowski and Lisowska, 1963; Lisowska and Jeanloz, 1973). Thus it seemed that the N might be a precursor to M. This weak N activity is actually associated with glycophorin B which has a very similar NH_2-terminal structure to glycophorin A (Dahr *et al.*, 1975a). The activity of this cryptic 'N' antigen is revealed only after the erythrocytes are treated with trypsin which on intact cells or on membranes releases the NH_2-terminal peptide from glycophorin A but not from

glycophorin B making the 'N' antigen of glycophorin B capable of reacting with anti-N antibodies (Dahr *et al.*, 1975a). The results of many investigators (reviewed in Issitt, 1981) do not support the explanation of the nature of MN antigens by Springer *et al.* (1976). As discussed above (section 1.7.5.1), the pentasaccharide structure as proposed by Springer *et al.* (1976) is not found in glycophorin A to a great extent and both glycophorin A^M and A^N contain tetrasaccharide as the predominant form in addition to a small proportion of trisaccharides and pentasaccharides (Fukuda *et al.*, 1980). Furthermore, no difference in sialic acid content has been observed in the oligosaccharide chains isolated from M+N- and M-N+ erythrocytes as would be expected if the hypotheses put forward by Springer *et al.* (1976) was correct (Dahr *et al.*, 1975a; Blumenfeld and Adamany, 1978).

Sadler *et al.* (1979) showed that purified sialyltransferases from porcine submaxillary glands restored M antigen activity to neuraminidase-treated intact human erythrocytes that were M+ prior to removal of sialic acid and not to the neuraminidase-treated erythrocytes that were M- prior to the treatment with neuraminidase. This enzyme could not convert N+ cells to M+ cells. This provided the evidence that the N antigen was not a precursor of M antigen.

The importance of the polypeptide backbone in determining the structure of the MN antigen was provided by amino acid modification studies (Morawieki, 1967; Lisowska and Duk, 1972, 1975). Acetylation of free amino groups of glycophorin A

by acetic anhydride treatment on intact erythrocytes or on an isolated NH_2-terminal peptide derived from it resulted in the loss of serologically defined M and N activities suggesting that free amino groups were involved in the expression of the activity of these antigens (Lisowska and Duk, 1975).

These observations were further supported by the amino acid sequence of glycophorin A (Tomita and Marchesi, 1975). This sequence was determined from glycophorin A isolated from pooled blood samples which contained M+N- and M-N+. However, the sequence was found to be quite homologous except that a few sites including position one and five of the NH_2-terminus showed amino acid variation. Serine/leucine and glycine/glutamic acid were found to occupy positions one and five, respectively. It was later shown that this variation is actually due to amino acid polymorphism at these positions and defines the antigenic determinants of MN blood group system. The M glycophorin A contains serine and glycine at positions one and five, respectively, whereas N glycophorin A contains leucine and glutamic acid at these positions (Dahr *et al.*, 1977; Wasniowska *et al.*, 1977; Blumenfeld and Adamany, 1978; Furthmayr, 1978b; Lisowska and Wasniowska, 1978).

Although the above finding of amino acid polymorphism revealed the major difference between M and N antigenicity it is appropriate to assume that carbohydrate residues are also an important part of the MN antigen system. Amidation of carboxyl groups of glycophorin causes loss of MN antigen suggesting

that sialic acids are involved in the expression of MN antigen (Ebert *et al.*, 1972). Studies using periodate oxidation and a variety of MN specific lectins suggested that one of the sialic acid residues present at the end of O-linked oligosaccharides was part of the MN antigen (Dahr *et al.*, 1975a).

The above discussion is consistent with the hypothesis that inherited differences of M and N reflect differences in the conformation of glycophorin A which results from amino acid differences found at the NH_2-terminus of the molecule (Anstee, 1981).

Use of synthetic peptides showed that the primary determinants of M and N antigenicity were contained in an octapeptide segment derived from NH_2-terminal amino acids (residue 1 to 8) of glycophorin A^M and A^N, respectively (Pederson *et al.*, 1990). It was shown that the M-specific octapeptide without any carbohydrate could cause a significant decrease in the agglutination of erythrocytes of M antigenicity by monoclonal anti A^M antibodies. Similarly, the N-specific octapeptide could also decrease agglutination of erythrocytes of corresponding antigenicity. It was further noted that leucine-1 of glycophorin A^N and serine-1 of glycophorin A^M were the primary determinant of their corresponding antigens (Pederson *et al.*, 1990).

1.7.8.2. Other blood groups of glycophorin A

In addition to the major MN blood groups, there is a large body of literature

describing a number of other antigens associated with glycophorin A. Some of these are very well defined biochemically while others are still in their earlier stages of evaluation and their structures have not been worked out completely (reviewed in Issitt, 1981; Lisowska, 1988; Mould and Dahr, 1989; Anstee, 1981, 1990). Many variant antigens, usually occurring with low frequency in the population have been defined by antibodies found in human or animal immune sera. On the other hand, rare individuals with variant erythrocytes such as En(a-), lacking glycophorin A (Tanner and Anstee, 1976b), or S-s-U-, containing defective glycophorins (Dahr *et al.*, 1975b), or M^k, lacking both glycophorin A and B (Anstee and Tanner, 1978) usually have in their sera alloantibodies directed against common antigens absent in their erythrocytes.

The naturally occurring variant forms of glycophorins fall into three categories (reviewed in Lisowska, 1988): 1) those with changes in oligosaccharide structures, such as TF, Tn, Cad; 2) those with amino acid replacements in the polypeptide chain, for example, Mc, Mg, He, MiI, MiII; and 3) hybrid sialoglycoproteins produced from fused genes encoding parts of various glycophorins, especially of A and B, for example, MiV, Dantu and Sta. Some additional cells, as mentioned above, are also known which completely lack glycophorin A [En(a-) cells] or lack glycophorin B (S-s- cells) or lack both glycophorin A and B (M^k cells) (Lisowska, 1988). The description of all of these variant forms of MN antigens is beyond the scope of the present review and will not be discussed further.

1.7.9. Virus receptor activity of glycophorin A

Ever since it was reported that influenza virus caused agglutination of chick erythrocytes (Hirst, 1942), erythrocytes have been used extensively to study the mechanism of these interactions and the erythrocyte glycoproteins have served as model receptors for a number of viruses. These studies have shown that sialic acids are important for attachment of many viruses to erythrocytes (reviewed in Burness, 1981).

Hirst (1942) showed that the chick erythrocytes which were agglutinated by the influenza virus were modified in such a way that they no longer remained capable of interacting with fresh virus. It was suggested that this loss of haemagglutination activity was due to the presence of an enzyme (termed receptor destroying enzyme) in the virus coat capable of destroying the virus receptors on the erythrocyte surface (Hirst, 1942). The mechanism of this inhibition was first explained by Gottschalk (1960). It was shown that the haemagglutinating sites of the virus specifically bind to the sialic acid-containing receptors on adjacent erythrocytes causing them to agglutinate. An active neuraminidase in or close to the agglutinating sites in the virus removes sialic acids from the receptors. Hence the receptors no longer bind the virus, elution of the virus from the cells results and further agglutination of the cells can no longer take place.

Most of the earlier studies on virus-erythrocyte interaction dealt with

erythrocytes, their membranes or with crude preparations of human erythrocytes sialoglycoproteins isolated from human erythrocytes which contained all four glycophorins (A, B, C and D), the major proportion of which is glycophorin A. Out of these four glycophorins, glycophorin A has been reported to be specific receptor on erythrocyte for a number of viruses such as encephalomyocarditis virus, influenza virus (Energren and Burness, 1977; Allaway *et al.*, 1986), reovirus (Paul and Lee, 1987), Blue tongue virus (Eaton and Crameri, 1989) and mengovirus (Anderson and Bond, 1987).

The receptor activity of glycophorin A for encephalomyocarditis virus has been studied extensively by Burness and co-workers (reviewed in Allaway *et al.*, 1986). These studies have shown that encephalomyocarditis virus attaches to erythrocytes of a number of species (Angel and Burness, 1977). The encephalomyocarditis virus apparently attaches specifically to the major sialoglycoprotein, glycophorin A (Energren and Burness, 1977). Treatment of erythrocytes with neuraminidase reduced encephalomyocarditis virus attachment to about 10% of the untreated control cells. This suggested that the receptor for encephalomyocarditis virus contained sialic acids. The evidence that this sialic acid containing human erythrocyte receptor site was glycophorin was provided by haemagglutination inhibition of encephalomyocarditis virus by glycophorin prepared from human erythrocytes (Energren and Burness, 1977). As mentioned before, such preparations must have contained other sialoglycoproteins (glycophorins B, C and D) in addition to glycophorin A. To

distinguish among different sialoglycoproteins, the attachment of encephalomyocarditis virus was investigated with En(a-) cells which lack glycophorin A (Tanner and Anstee, 1976b) and S-s-U+ cells which lack glycophorin B (Dahr *et al.*, 1975b). The lack of binding of encephalomyocarditis virus to En(a-) cells and normal binding to S-s-U+ cells suggested that glycophorin A, not B, is the human erythrocyte receptor for encephalomyocarditis virus (Allaway and Burness, 1986).

The location of the binding site for encephalomyocarditis virus on glycophorin A was investigated by treatment of glycophorins with proteolytic enzymes such as trypsin, chymotrypsin and ficin on intact erythrocytes, on isolated erythrocyte membranes or glycophorin preparations isolated from these cells (Allaway and Burness, 1986). When human erythrocytes were treated with chymotrypsin, which removes amino acid residues 1 to 34 from glycophorin A, there was very little or no effect on binding of encephalomyocarditis virus to these cells. This indicated that amino acid residues 1 to 34 with 12 O-linked oligosaccharide units were not important for encephalomyocarditis attachment. Proteolytic enzyme treatment of human glycophorin A *in situ* (Allaway and Burness, 1986) or of isolated glycophorin A (Burness and Pardoe, 1983) revealed the possible involvement of amino acid residues 35 to approximately 72, the possible point of membrane insertion of glycophorin A. Four of the amino acids in this region, residues 37, 44, 47 and 50 are glycosylated. A polypeptide, CH-O (comprising amino acid residues 35 to 118) released from isolated glycophorin A by chymotrypsin treatment and purified by gel

filtration inhibited haemagglutination by encephalomyocarditis virus whereas other chymotryptic peptides released from glycophorin A did not (Burness and Pardoe, 1983). Therefore, these binding studies localized the binding site for encephalomyocarditis virus to the chymotryptic peptide (CH-O) comprising amino acids 34 to 118 of glycophorin A.

Trypsin treatment of erythrocytes which removes four more amino acids from glycophorin A than does chymotrypsin, i.e. residues 1 to 39, had a drastic affect on encephalomyocarditis virus attachment and reduced it to about 30% to 50% of control untreated cells (Allaway and Burness, 1986). These observation suggested that the receptor site for encephalomyocarditis virus on glycophorin A includes amino acid residues 35 to 39 and the oligosaccharide attached to threonine-37 may be particularly important. The residual 30 to 50% attachment after trypsin treatment also indicated that the segment containing amino acid residues 40 to about 72 (the segment exposed to the cell surface), together with the oligosaccharides on serine-44, serine-47 and threonine-50 may also be involved in encephalomyocarditis virus-erythrocyte interaction.

Ficin is a proteolytic enzyme whose exact cleavage site on glycophorin A is not known, except that it cleaves glycophorin A closer to the cell surface than does chymotrypsin or trypsin. It removes about 50 amino acids from glycophorin A and releases all of the 16 oligosaccharide units. Treatment of human erythrocytes with

ficin completely destroyed the encephalomyocarditis virus binding to these cells. This provided further supporting evidence for the conclusion that the binding site for encephalomyocarditis possibly involves amino acid residue 35 to about 50 (Allaway *et al.*, 1986).

As mentioned above, oligosaccharides with their sialic acids are important for encephalomyocarditis virus attachment to glycophorin A (Burness and Pardoe, 1981; Energren and Burness, 1977). Removal of about 40% sialic acid from erythrocytes prevented haemagglutination by encephalomyocarditis virus, or attachment to glycophorin A. It was also indicated in these investigations that not all sialic acids of glycophorin A were important for encephalomyocarditis virus attachment (Burness and Pardoe, 1981). Further investigation on the role of sialic acids revealed that polyhydroxy side chains in sialic acid were not required for encephalomyocarditis virus attachment but carboxyl groups were involved in the attachment (Tavakkol and Burness, 1990).

1.7.10. Malarial parasite invasion

Several studies have shown that invasion of erythrocytes by the malarial parasite *Plasmodium falciparum* merozoite is dependent upon sialic acids (reviewed in Hadley *et al.*, 1986 and Hermantin, 1987). Treatment of erythrocytes with neuraminidase reduces invasion; and erythrocytes with variant oligosaccharides such as Tn which have reduced amount of sialic acid and galactose, or Cad cells which

have an additional N-acetylgalactosamine residue are all resistant to invasion (Hermantin, 1987). Similarly En(a-) which completely lack glycophorin A and M^k cells which lack both glycophorin A and B are partially resistant to invasion (Hadley *et al.*, 1987). As a result of these studies glycophorins particularly glycophorin A have been implied to be the primary erythrocyte receptor site for *Plasmodium falciparum*.

Further evidence for the involvement of glycophorins in malarial parasite binding to human erythrocytes has been provided by the following competitive inhibition studies with glycophorins A or B: (i) glycophorin A embedded in liposomes; (ii) erythrocyte extracts containing variable amounts of sialoglycoproteins; (iii) enzyme-generated fragments of glycophorin A and Fab fragments of antibody prepared against glycophorin A; (iv) monoclonal antibodies directed against different portion of glycophorin A; and (v) human sera containing antibodies directed against glycophorin A, or various lectins; all of these studies support the suggestion that glycophorin A is the erythrocyte receptor required for *Plasmodium falciparum* invasion (reviewed in Hermentin, 1987).

It has been suggested that the Wrb antigen located on glycophorin A, close to the membrane insertion site is involved in invasion by malarial parasites (Pasvol *et al.*, 1982a, b; Ridgwell *et al.*, 1983). Later studies using fresh Wrb+ and Wrb- cells showed that both types of cell are fully susceptible to invasion and hence perhaps the Wrb antigen does not have a specific role in malarial parasite invasion (Hermentin

et al., 1987). However, as mentioned before many observations supporting the role of sialic acids in O-linked oligosaccharides of glycophorin A in malarial parasite binding have been made. It has been suggested that the sialic acid residues may be involved as the initial point of recognition and attachment in this process (Hermentin, 1987).

Pasvol *et al.* (1982a, b) put forward the hypothesis for the role of glycophorins A and B in invasion of erythrocytes by *Plasmodium falciparum*. According to the hypothesis, invasion takes place in two steps. The initial attachment of the merozoite surface coat to the erythrocyte may reflect a lectin-ligand-like interaction in which the parasite binds in a specific manner to a cluster of oligosaccharides present on glycophorin A, B or both. Once attachment has occurred and the apical end of the merozoite with its specialized organelle has orientated to the membrane, further specific conformational alterations in the erythrocyte may occur which trigger the process of erythrocyte deformation and parasite entry, possibly involving the Wrb antigen of glycophorin A. However, as mentioned earlier, a role for the Wrb antigen is open to doubt (Hermantin, 1987).

Two studies (Okoye and Bennet, 1985; Friedman *et al.*, 1985) provided evidence that another erythrocyte membrane protein, band 3 glycoprotein could be a receptor during invasion of human erythrocytes. However, a role for glycophorin A was not excluded since recently it has been shown that band 3 and the Wrb antigen

of glycophorin A interact. In any case, glycophorin A seems to be involved in the initial attachment in a relatively weak and relatively non-specific interaction; this is followed by transfer of the information to the cytoskeleton and since glycophorin A may interact with underlying cytoskeleton proteins such as protein 4.1 (section 1.7.7.3) it is a good candidate for this function of information transfer.

1.8. Glycophorin B

1.8.1. Biochemical characterization

The first indication that the erythrocyte sialoglycoproteins were a mixture was supplied by PAS-staining of the erythrocyte proteins separated by SDS-polyacrylamide gel electrophoresis (Fairbanks et. al, 1971). A sialoglycoprotein fraction prepared by a lithium diiodosalicylate-phenol procedure (Marchesi and Andrews, 1971) when subjected to gel filtration in Ammonyx-Lo separated into two fractions (Furthmayr *et al.*, 1975). The major fraction was glycophorin A while the other fraction ran at the position of PAS-3 and revealed an apparent molecular weight of 25,000 daltons when analyzed by SDS-polyacrylamide gel electrophoresis. It comprised about 10% of the total PAS-stainable material of erythrocytes. Based upon the differences in tryptic digestion pattern and amino acid composition between the two fractions it was concluded that the minor fraction represented another sialoglycoprotein which had some features, such as sugar content and capacity to form high molecular weight aggregates in common with glycophorin A (Furthmayr *et al.*, 1975). Some differences in structures of the two proteins were also noted. For example, unlike glycophorin A,

glycophorin B did not seem to possess any significant length of cytoplasmic COOH-terminal domain (Furthmayr *et al.*, 1975).

The NH_2-terminal peptide of glycophorin B contains sialic acid, galactose and N-acetylgalactosamine, supporting the notion that O-glycosidically linked oligosaccharides of glycophorin B are of the type found in glycophorin A (Furthmayr, 1978a; Dahr *et al.*, 1980a, b). The similarities of O-linked oligosaccharide of glycophorin A and B were further supported by the binding of lectins, such as those from *Maclura aurantaca* and *Arachis hypogaea* specific for O-glycosidically linked oligosaccharides to glycophorin B (Tanner and Anstee, 1976a). The lack of binding with *Phaseolus vulgaris* and *Ricinus communis* lectins suggested that glycophorin B did not contain N-glycosidically linked oligosaccharides (Robinson *et al.*, 1975; Tanner and Anstee, 1976a; Anstee *et al.*, 1977). This was further supported by analysis of a crude preparation of PAS-3 (Furthmayr *et al.*, 1975).

The NH_2-terminal tryptic peptide (residue 1 to 35) of glycophorin B was isolated by gel filtration in the presence of detergents (Furthmayr, 1978a; Dahr *et al.*, 1980a, b). The amino acid sequence of this peptide (Furthmayr, 1978), revealed that its first 26 residues were identical with those of glycophorin A^N; the threonine and serine residues present in positions similar to those found in the NH_2-terminal peptide of glycophorin A were also found to be glycosylated. Unlike in glycophorin A, aspargine-26 was not glycosylated in glycophorin B (Furthmayr, 1978a). The

absence of N-linked oligosaccharide was consistent with the absence of mannose and N-acetylglucosamine in glycophorin B (Furthmayr, 1978a). In view of the well established rule of glycosylation of aspargine residues which requires the sequence - aspargine-X-serine or aspargine-X-threonine, glycophorin B can not be glycosylated at aspargine-26 due to lack of serine or threonine residue at the position 28 which is glutamic acid in glycophorin B (Lisowska, 1988).

The complete amino acid sequence of glycophorin B has been determined by polypeptide sequencing (Blanchard *et al.*, 1987). A fragment of glycophorin B comprising residues 36 to 71 and containing the intramembranous domain was isolated by high performance ion exchange and gel filtration in the presence of Triton X-100. The amino acid sequence of the intramembranous domain of glycophorin B was found to be similar to that of glycophorin A. The accuracy of the amino acid sequence was confirmed by isolation and sequencing of a cDNA clone (Siebert and Fukuda, 1987), except that glycophorin B was predicted to contain an extra amino acid (alanine-72) not detected by peptide sequencing. The 72 amino acids of glycophorin B were predicted to be organized into three structural domains, a hydrophillic domain (residues 1 to 44), an intramembranous region (residues 45 to 64) and a very short cytoplasmic tail (residues 65 to 72)(Anstee, 1990). In addition to the complete identity of the first 26 amino acids and the associated O-linked oligosaccharides to that of glycophorin A, nucleotide and peptide sequencing predicted that glycophorin B has a similar sequence in the membrane spanning

domain. An important difference is that the region comprising amino acids 27 to about 57 of glycophorin A are missing from glycophorin B. However, residues 58 to 100 of glycophorin A are also similar in sequence to residues 37 to about 72 of glycophorin B.

1.8.2. Blood group antigens of glycophorin B

Another blood group antigen associated with the MN sialoglycoproteins was detected by reaction with an antibody, anti-S, whose antigen was distinct from MN (Sanger and Race, 1947). These authors noted that S was produced by a gene at a locus very close to that at which M and N were situated. Levine *et al.* (1951) reported the discovery of anti-s, the antibody that defines the antigen antithetical to S. Using anti-S and anti-s, the different phenotypes: S+s-; S+s+ and S-s+ were recognized as representing the genotypes: SS; Ss and ss, respectively. Another antibody causing severe haemolytic reaction detected yet another antigen, termed U antigen which was also associated with the sialoglycoproteins of human erythrocytes (Wiener *et al.*, 1953). It was further shown by Greenwalt *et al.* (1954) that U is an MN system antigen, since two individuals with U- erythrocytes which lacked glycophorin B were also found to be S- and s-.

The biochemical characterization of the nature of the Ss and U antigens revealed that they were associated with glycophorin B (Dahr *et al.*, 1975a; Tanner and Anstee, 1976a; Anstee *et al.*, 1979). Similar to the MN blood group system, an

amino acid polymorphism represents the structural differences between the S and s antigens. S+s- glycophorin B has Met at 29, whereas glycophorin B from S-s+ contains a threonine at this position (Dahr *et al.*, 1980a). Apart from these amino acid differences, the receptor sites of anti-S antibodies and anti-s have only been characterized partially (Dahr *et al.*, 1982). Amino acid residues 28, 34 and 35, as well as the oligosaccharide attached to threonine-25, might also be involved in defining the S and s antigens (Dahr *et al.*, 1982).

The U antigen is a labile structure located within residues 33-40 of glycophorin B and it requires lipids for optimal expression of activity (Dahr and Kruger, 1983). The U antigen is closely associated with the Rh blood group complex and is weakened or altered qualitatively in several erythrocytes that possess alteration in Rh blood group antigen. It has been suggested that glycophorin B and proteins encoded by the Rh locus form a complex during membrane biosynthesis (Dahr *et al.*, 1987b). The formation of such an aggregate might be necessary for U activity.

As mentioned above, the NH_2-terminal structure of glycophorin B is identical with that of glycophorin A^N. This is why glycophorin B also carries an additional N-antigen, the cryptic antigen N, denoted 'N', to distinguish it from the N antigen of glycophorin A (Dahr *et al.*, 1975a). The 'N' antigen is only weakly reactive in intact erythrocytes. However, it can be easily detected by agglutination tests with erythrocytes treated with trypsin, which cleaves the NH_2-terminal 39 residues of

glycophorin A on intact erythrocytes or erythrocyte membranes and has no affect on glycophorin B in intact erythrocytes (Dahr *et al.*, 1975a; Tomita *et al.*, 1978).

In addition to the common antigens, some other rare blood group activities such as He and Mv are also associated with glycophorin B and localized in its NH_2-terminal sequence (Dahr and Longster, 1984). The rare Miltenbergers (Mi) Mi-III, Mi-IV and Mi-VI are variants representing serologically related antigens of glycophorin B and generated by an insertion of a segment of a glycophorin A gene into the one for glycophorin B (Laired-Fryer *et al.*, 1986).

1.9. Glycophorin C

SDS-polyacrylamide gel electrophoresis revealed that the band previously thought to be glycophorin A monomer, PAS-2 was actually composed of two components (PAS-2 and PAS-2´) running very close together (Mueller and Morrison, 1974). Furthmayr *et al.* (1975) called the PAS-2´ protein glycophorin C which makes up about 4% of the total PAS-staining bands (Cartron *et al.*, 1990). There are about 50,000 to 100,000 glycophorin C molecules/erythrocyte (Anstee *et al.*, 1984; Anstee, 1990). Studies on En(a-) variant erythrocytes which lack glycophorin A (Tanner and Anstee, 1976b), and on erythrocytes which are defective in glycophorin B, S-s-u- (Dahr *et al.*, 1975b; Tanner *et al.*, 1977) revealed that they had a normal PAS-2´ confirming the distinct identity of glycophorin C.

Binding experiments showed that glycophorin C associates with *Phaseolus vulgaris* lectin indicating the presence of mannose and N-acetylglucosamine, the sugar components found in N-glycosidically linked oligosaccharide similar to that found at position 26 in glycophorin A (Tanner *et al.*, 1977; Furthmayr, 1978a). Similarly, binding of *Maclura arantiaca* and *Arachis hypogaea* lectins to glycophorin C and the presence of galactose, N-acetylgalactosamine and sialic acid in oligosaccharides of glycophorin C also suggested that it contained O-linked oligosaccharide side chains similar to those found in glycophorin A (Anstee, 1981).

Recently more detailed biochemical characterizations have been performed on glycophorin C preparations (Dahr *et al.*, 1982; Blanchard *et al.*, 1987). The amino acid sequence of an NH_2-terminal tryptic glycopeptide comprising 47 amino acid residues (Dahr *et al.*, 1982) and that of a hydrophobic peptide comprising residue 49-88 (Blanchard *et al.*, 1987) was determined by peptide sequencing. The remaining amino acid sequence was deduced from cDNA clones (Colin *et al.*, 1986). Glycophorin C is composed of 128 amino acids distributed in three structural domains (Colin *et al.*, 1986). The extracellular NH_2-terminal domain (amino acid residues 1 to about 57) contains 13 O-linked oligosaccharides which resemble those found on glycophorin A in containing N-acetylgalactosamine, galactose and sialic acids and one N-linked oligosaccharide unit (Dahr and Beyreuther, 1985). About 24 amino acids (residues 58 to 81) are predicted to be inserted into the membrane bilayer (Colin *et al.*, 1986). The intramembranous domain is followed by a cytoplasmic domain

(residues 82 to 128) comprising many charged residues (Colin *et al.*, 1986).

1.10. Glycophorin D

The structure of glycophorin D is not known completely because it makes up a very small percentage of the human erythrocyte sialoglycoprotein preparation as revealed by SDS-polyacrylamide gel electrophoresis (Anstee *et al.*, 1979). Some information available from immunological and biochemical studies suggests that glycophorin D is immunologically and structurally related to glycophorin C (Anstee, 1990; Cartron *et al.*, 1990). Both glycophorin C and D carry the blood group Gerbich antigen (Anstee and Tanner, 1986) and polyclonal antibodies raised against purified glycophorin C or D react with antigenic determinants common to the COOH-terminus or intramembranous regions of both glycoproteins. Furthermore, when the peptides generated by enzymatic and acid cleavage of glycophorin C and D were compared most of the peptides from glycophorin D seemed to be identical to those from the intramembranous and COOH-terminal domains of glycophorin C, but some were present exclusively in glycophorin C digest (El-Maliki *et al.*, 1989). Further, since murine monoclonal antibodies reacting with the NH_2-terminal region of C but not with glycophorin D have been described this indicated that glycophorin D might be an abridged version of glycophorin C in its NH_2-terminal region (Dahr *et al.*, 1987a). Preliminary amino acid sequence analysis of glycophorin D is consistent with this assumption (El-Maliki *et al.*, 1989).

1.11. Blood group antigens of glycophorin C and D

Glycophorin C and D are associated with blood group antigens collectively known as Gerbich (Reid *et al.*, 1987a, b). Gerbich 3 (Ge:3) is located in a region of glycophorin C (amino acid residue 42 to 49) that surrounds a tryptic cleavage site at position 48 and a glycosylated residue (serine-42) (Dahr *et al.*, 1987a). Antisera raised against Gerbich 3 antigen of glycophorin C also bound to glycophorin D indicating that glycophorins C and D have common structural features recognized by anti-Gerbich 3 antibodies (Dahr *et al.*, 1987a). The Gerbich 2 (Ge:2) antigen was localized to a tryptic peptide on glycophorin D comprising about 20 to 30 amino acid residues (Dahr *et al.*, 1987a). It is interesting to note that even though both glycophorin C and D share identical amino acid sequence in this region only glycophorin D reacts with Gerbich 2 antisera and the reason for this strange behaviour is not known. However, other investigators (Kuczmarski *et al.*, 1987) reported that their anti-Gerbich 2 antisera reacted exclusively with glycophorin C. The resolution of this discrepancy must await further characterization of structural features of Gerbich 3 antigen.

1.12. Functions of glycophorins C and D

The normal physiological function of glycophorin A and B is not known (as discussed in sections 1.7.7). In contrast, there is evidence that both glycophorin C and D interact in normal erythrocytes through their cytoplasmic domains with skeletal protein 4.1 on the internal face of the phospholipid bilayer and this association is important for maintaining the normal discoid shape of erythrocytes (reviewed in

Anstee, 1990). Glycophorin C which is found in normal erythrocyte skeletal preparations was absent from skeletal preparations from an individual with hereditary elliptocytosis who had a total deficiency of band 4.1 (Mueller and Morrison, 1981). The glycophorin C content in the erythrocyte ghosts of individuals with homozygous 4.1(-) hereditary elliptocytosis was sharply reduced and the glycophorin D content was also decreased, whereas in heterozygotes, the glycophorin C content seemed to be slightly reduced while that for glycophorin D seemed to be normal (Alloisio *et al.*, 1985). Many other individuals showing complete absence of glycophorin C (such as variant blood group, Leach type) or showing abnormal glycophorin C (such as Ge:-1, -2, -3 and Yus type) have been reported whose erythrocytes show an abnormal erythrocyte phenotype (i.e. elliptocytosis)(reviewed in Anstee, 1990). Therefore, it seems at least glycophorin C and possibly D has a functional importance, in the maintenance of human erythrocyte membrane integrity which is brought about by its interaction with cytoskeleton protein 4.1.

1.13. Glycophorins of animal erythrocytes

The presence of "glycophorin" like molecules on the erythrocyte surface has been described in many animal species, including: non-human primates such as common chimpanzees (Rearden, 1986), monkey (Murayama *et al.*, 1989), in other mammals such as pig (Honma *et al.*, 1980), horse (Murayama *et al.*, 1981), cow (Murayama *et al.*, 1982), dog (Murayama *et al.*, 1983), sheep (Lutz *et al.*, 1976), goat (Fletcher *et al.*, 1976), rabbit (Honma *et al.*, 1982), rat (Laing *et al.*, 1988) and mouse

(Sarris and Palade, 1982) and in chicken, the only avian species examined (Weise and Ingram, 1976). However, with the exception of monkey, pig, horse, bovine and mouse, in most of these species only very limited studies have been performed (reviewed in Krotkiewski, 1988). Nevertheless, all of the animal glycophorins characterized so far reveal at least one integral protein associated with the erythrocyte membrane. These proteins share a number of biochemical properties with human glycophorins, such as the presence of sialic acid and having the capability of forming high molecular weight aggregates (homo- or heterodimers or larger oligomers) when extracted from membranes. The molecular weights of these proteins range between 19,000 to 34,0000 daltons, about 20-68% of which is due to carbohydrate. Frequently these glycophorins contain as much as 50% sialic acid by weight with the exception of rabbit erythrocyte glycophorin which has very little sialic acid. Most of these glycophorins contain multiple units of O-linked oligosaccharide, usually in the form of tetrasaccharides but with some trisaccharides and pentasaccharides. A variety of sialic acids such as N-acetylneuraminic acid, N-glycolylneuraminic acid and O-acetyl-N-acetylneuraminic acid are found in these sialoglycoproteins. Most of the animal glycophorins contain at least one, and some, such as pig contain even two N-linked oligosaccharides. Bovine erythrocytes contain two glycophorins, GP-1 and GP-2 with most unusual oligosaccharides containing a wide variety of O-linked oligosaccharides, up to deca and undecasaccharide units (Krotkiewski, 1988).

To date the complete amino acid sequence has been determined by amino acid sequencing for glycophorins from erythrocytes of monkey (Murayama *et al.*, 1989), horse (Murayama *et al.*, 1982) and pig (Honma *et al.*, 1980). Recently, the complete amino acid sequence of a mouse glycophorin has been deduced from a cDNA clone (Matsui *et al.*, 1989). Comparison of the amino acid sequences of these animal glycophorins with those of humans shows significant similarity only in the membrane spanning domains. There is one exception, however, the monkey glycophorin MK, described in more detail next shows additional similarity in the NH_2-terminal as well as in the COOH-terminal domain (Murayama *et al.*, 1989).

1.13.1. Monkey glycophorin (MK)

This glycophorin contains 144 amino acids distributed in three domains, namely, extracellular, intramembranous and intracellular. The monkey glycophorin has 18 oligosaccharides attached O-glycosidically to the peptide backbone through serine or threonine residues; no N-linked oligosaccharide is present (Murayama *et al.*, 1989). The molecular weight of monkey glycophorin MK is 35,000 daltons calculated on the basis of amino acid and carbohydrate content (Murayama *et al.*, 1989). This glycophorin possesses both M and N blood group activity as determined by haemagglutination inhibition using commercially available anti-M and anti-N sera (Murayama *et al.*, 1989).

1.13.2. Horse glycophorins HA and HB

Two glycophorins HA and HB have been isolated from horse erythrocyte membranes with lithium diiodosalicylate:phenol extraction followed by preparative SDS-polyacrylamide gel electrophoresis and ion exchange chromatography in the presence of Ammonyx-Lo (Murayama *et al.*, 1981). Glycophorin HA, the major component of horse glycophorin, has an apparent molecular weight of 20,000 daltons and contains 120 amino acids. There is no N-linked oligosaccharide in horse glycophorin, HA, but there are 10 O-linked oligosaccharides units mostly of the type found in human glycophorins with a preponderance of tetrasaccharides, with some disaccharides and trisaccharides (Fukuda *et al.*, 1980).

1.13.3. Porcine glycophorin

This glycophorin contains 133 amino acids, with 10 O-linked oligosaccharide units giving a molecular weight of about 27,000 daltons (Honma *et al.*, 1980). The carbohydrate structure of porcine glycophorin is quite different from that of humans. The major O-linked oligosaccharide of porcine glycophorin is a trisaccharide which contains N-acetyl-glucosamine not found in humans in their O-linked oligosaccharides. In addition, other oligosaccharides including disaccharides, hexasaccharides and even some decasaccharides are present (Kawashima *et al.*, 1982; Honma *et al.*, 1980).

1.13.4. Mouse glycophorin

In mouse two glycophorins SGP 2.1 and SGP 3.1 with molecular weights of

46,000 and 31,000 daltons, respectively, have been isolated using the lithium diiodosalisylate:phenol extraction method (Sarris and Palade, 1982). The amino acid sequence of one of these mouse glycophorins was deduced by sequencing a cDNA clone and shown to consist of 168 amino acids (Matsui *et al.*, 1989). Like other mammalian glycophorins, the mouse glycophorin shares amino acid sequence homology with human glycophorins, only in the intramembranous domain. The nucleotide sequences of the cDNA clone for mouse glycophorin revealed that unlike human glycophorins A and B, mouse glycophorin is synthesized without a cleavable signal peptide. Some murine erythroid cell lines also express two related glycophorins, termed Gp2 and Gp3 of molecular weight, 44,000 and 29,000 (Dolci and Palade, 1989). It was noted that these glycoproteins show developmental stage specific changes in their oligosaccharides.

1.14. Biosynthesis of mRNA

Cells regulate many steps in the pathway from DNA to protein synthesis to control gene expression. Most genes are regulated at multiple levels beginning at the stage of transcription initiation, which is the predominant step in control of gene expression. In contrast, some other genes are transcribed at constant rates and their protein levels are regulated solely by processes that affect mRNA abundance. These post-transcriptional controls include alternate splice-site choice, 3´ end formation by cleavage and polyadenylation, translation initiation and regulated mRNA degradation (or mRNA stability). Most of the control processes require either specific sequences

or secondary structures in the mRNAs or both (Alberts *et al.*, 1989). Differential processing of pre-mRNA may also result in the generation of multiple products from a single gene. Two types of protein-coding transcription unit exist: simple transcription units that encode one protein; and complex transcription units that may encode two or more proteins produced by differential processing of the primary transcript, or in some rare cases produced by more than one primary transcript (Darnell *et al.*, 1990).

In eucaryotic cells mRNA synthesis is a complex process that takes place in the nucleus, and begins with the transcription of DNA by RNA polymerase II. The primary product of transcription, pre-mRNA, undergoes extensive modifications before it is transported to the cytoplasm to act as a template for protein synthesis (Blackburn and Gait, 1990).

1.14.1. Transcription

In eucaryotes most of the protein-coding genes contain a sequence, TATRAAR (R = A or T), the TATA box, about 25 to 35 nucleotides upstream from the RNA polymerase II start site. In addition to the TATA box other upstream promoter elements, for example the CAAAT box are also present in many eucaryotic genes. Additional sequences termed enhancers also affect gene transcription and function even when located thousands of nucleotide upstream or downstream of the promoters and in either orientation with respect to the gene transcription. These

enhancer elements bind regulatory proteins, some of which are found only in a restricted cell type whereas others involved in the regulation of house keeping genes are found in large number of cells (reviewed in Alberts *et al.*, 1989).

Briefly, three general transcription factors ($TF_{II}B$, $TF_{II}D$ and $TF_{II}E$) are required to begin transcription. The transcription factor $TF_{II}D$ (tata-factor), a large protein complex, binds to the TATA box and forms a stable, "committed", pre-initiation complex. This is followed by binding of RNA polymerase II. Neither $TF_{II}B$ nor $TF_{II}E$ can bind to DNA on its own. However, $TF_{II}B$ perhaps binds to the RNA polymerase II and may contact the DNA after binding to RNA polymerase II. $TF_{II}E$ can act as an ATPase and the energy released by ATP hydrolysis probably brings about conformational changes in some protein(s) in the pre-initiation complex. Alternatively the energy released by ATP hydrolysis probably assists in melting the DNA to make an open initiation complex which sets the stage for transcription to begin. An additional factor, an elongation factor, $TF_{II}S$ is also required for efficient transcription (reviewed in Darnell *et al.*, 1990).

1.14.2. Pre-mRNA processing

The primary product of transcription, pre-mRNA is found in the heterogeneous nuclear RNA (hnRNA) fraction which contains components of various size from less than 2.0 to more than 200 kb. Only a small fraction of hnRNA is converted into mRNA (Darnell *et al.*, 1990). In this process most of the hnRNA is

degraded so that even though the rate of production of hnRNA is about 58% of total RNA synthesis in a cell, the mRNA produced is only about 3% of the steady-state quantity of RNA in a cell. Most of this RNA loss is accounted for by removal of extra nucleotides from the pre-mRNA molecule (introns and 3´ end flanking sequences) together with some RNA loss due to premature termination of transcription or to a portion of hnRNA that is not processed further (Alberts *et al.*, 1989). Gene transcription generally continues beyond the 3´ end of mature mRNAs. These extra nucleotides at 3´ ends of pre-mRNAs are removed by endonucleolytic cleavage, followed by addition of a poly(A) tail to the 3´ hydroxyl group of the terminal nucleotide (Darnell *et al.*, 1990, see below, section 1.14.6).

1.14.3. Capping

As soon as about 10 to 20 nucleotides are synthesized, the primary transcript is capped at its 5´ end (Blackburn and Gait, 1990). To achieve this, the 5´-triphosphate of the primary transcript is hydrolysed to a disphosphate and a guanosine 5´-monophosphate is then transferred from GTP to the 5´ end to give a 5´-5´ pyrophosphate linkage. The N_7 position of the terminal guanine is then methylated by the transfer of a methyl group from S-adenosylmethionine. In some cases the cap structure (7 Me G [5´-5´]-ppp X_pY....) is further methylated at the 2´ hydroxyl position of nucleotide "X" or both nucleotides "X and Y" (Blackburn and Gait, 1990). The 5´ cap structure is essential for initiation of translation and probably protects the mRNAs from the action of nucleases (Mizumoto and Kaziro, 1987).

1.14.4. Splicing

Transcription of most genes encoding proteins in eucaryotic cells results in the formation of pre-mRNA containing introns which are then removed by the process of splicing (reviewed in Sharp, 1987; 1988). Nucleotide sequence analysis of many genes containing introns has revealed that the pre-mRNAs contain moderately conserved short sequences at the intron-exon boundaries and a tendency for a pyrimidine-rich region just upstream of the 3′ splice site. The first two nucleotides, GU and last two nucleotides, AG in introns are very well conserved. For splicing to occur, the 5′ cap structure is essential. Splicing occurs in a large complex particle termed a spliceosome which is composed of a large number of proteins and several snRNPs including U1, U2, U4, U5 and U6 (Darnell *et al.*, 1990). A novel RNA form, a lariat RNA which consists of a circular component with an extended tail is generated during splicing of pre-mRNA. Lariat RNAs are formed by a branch about 20 to 50 nucleotides upstream of the 3′ splice site, by linking an adenosine through a 2′-5′phosphodiester bond to the 5′ end of the intron. The released intron is degraded very rapidly. A kinetic intermediate in the splicing of pre-mRNAs is generated by cleavage at the 5′ splice site. This intermediate consists of the 5′ exon, and the lariat RNA form of the intron linked by a normal phosphodiester bond to the 3′ exon (Padgett *et al.*, 1987). Analysis of splicing intermediates of chick oviduct ovalbumin mRNA revealed that splicing of various exons occur in a preferred pathway as opposed to an obligatory or completely random pathway (Lewin, 1987).

1.14.5. Transcription termination

Once transcription by RNA polymerase II begins, it continues beyond the 3´ end of the mature mRNA, sometimes very far past the poly(A) addition site (reviewed in Birnstiel *et al.*, 1985; Platt, 1986; Proudfoot, 1989; and Darnell *et al.*, 1990). This phenomenon first observed in the analysis of pulse-labelled RNAs of the adenovirus major late transcription unit (Nevins and Darnell, 1978) has since been shown by the nuclear run-off transcription experiments to occur in a variety of eucaryotic genes, including globin genes from various animals, the dihydrofolate reductase gene of mouse, the ovalbumin gene of chicken and the calcitonin/calcitonin gene-related peptide gene of rat (Proudfoot, 1989). In these genes, transcription termination seemed to occur heterogeneously over a kb or more of 3´ region sequences. For example, in the mouse α-amylase gene, transcription termination occurs over 2 to 4 kb past the 3´ end of the mature mRNA. In contrast, in human α-globin gene, transcription termination occurs only 100 to 300 bases beyond the 3´ end of mature mRNA. These above examples of termination of transcription give rise to the belief that instead of being a specific process the termination of transcription by RNA polymerase II occurs through a random process (Proudfoot, 1989).

Linkage between transcription termination and 3´ end formation was established by investigations on the α2-globin gene of a patient with α-thalassaemia. A single change of a base in the gene's poly(A) addition signal, AATAAA (a conserved sequence that is necessary for 3´ end processing of pre-mRNA, see section

1.14.6.2.1 below) to AATAAG not only inactivated 3´ processing, but also caused an increased level of transcription past the 3´ flanking region where transcription termination normally occurs (Whitelaw and Proudfoot, 1986).

Transfection with recombinant plasmids containing the adenovirus major late promoter and a poly(A) addition site of simian virus 40 revealed that when mutated poly(A) addition sites were used, a large amount of nuclear RNA complementary to the recombinant plasmid construct accumulated indicating that transcription termination was not taking place efficiently. In this way, it was established that sequences required for efficient cleavage/polyadenylation are also needed for efficient transcription termination (Connelly and Manley, 1988).

Two models that involve active 3´ end processing have been proposed to explain the mechanism of transcription termination. The first model was based upon investigations of nucleotide sequences required for efficient transcription termination within the mouse β-major globin gene (Logan *et al.*, 1987). Two DNA elements in the 3´ flanking region were required for efficient termination: an upstream sequence that included two poly(A) addition signals; and a downstream region previously shown to be involved in the transcription termination of RNA polymerase II in the mouse β-major globin gene. It was proposed by Logan *et al.* (1987) that the cleavage of the nascent transcript at the poly(A) addition site, occurring before the transcription by RNA polymerase II is complete, alters the conformation of the nascent RNA in such

a way that the transcription complex is destabilized and terminated. It was further suggested that either the cleavage event generates an uncapped 5′ end that would be rapidly degraded, again destabilizing the transcription complex, or the polymerase II complex carried an anti-termination factor with it. The factor could leave the complex at the polyadenylation site, possibly marking the nascent RNA for cleavage and simultaneously altering the transcription complex preparing it to terminate at the next appropriate site (Logan *et al.*, 1987).

In contrast to the model of termination presented above, an alternate model was proposed by Connelly and Manley (1988) implicating the 3′ end processing mechanism directly in transcription termination. It was observed by these authors that the sequences required for efficient polyadenylation of simian virus 40 early mRNA were also required for transcription termination. It was proposed that as soon as RNA polymerase II passes an appropriate poly(A) addition site, 3′ processing activities begin their action to produce a 3′ end and polyadenylation takes place. Meanwhile the polymerase II continues transcription past the poly(A) addition site, still attached to a growing nascent transcript with the 5′ end recently released from the 3′ end of the mRNA sequence. This newly formed 5′ terminus is unprotected, for instance, by a 5′ cap structure and must therefore, be subjected to degradation by a 5′ to 3′ exonuclease activity which eventually may lead to termination of transcription beyond the 3′ end of mature mRNA. This model of transcription termination predicted the existence of two factors: a helicase activity required to

unwind the nascent RNA (beyond the 3′ end of mRNA) which presumably base pairs with the anti-sense strand while being synthesized; and a 5′ to 3′ exonuclease activity to degrade the nascent RNA transcript comprising the sequences beyond the poly(A) addition site of mature mRNA. Existence of such an exonuclease activity has been reported but no helicase activity has been cited as yet (Proudfoot, 1989).

This second model accounts for often observed length heterogeneity seen in termination and the fact that, in different genes, termination of transcription seems to occur at very different distances downstream of poly(A) addition sites, characteristics of the particular gene in question. A gene having an inefficient poly(A) addition site would allow polymerase II to transcribe a considerable distance past it before cleavage occurs and hence the helicase and the exonuclease would have a lot more nascent transcript to act on before catching up with the RNA polymerase II (Proudfoot, 1989).

The above discussion regarding transcription termination hints at a role for transcription termination in the control of gene regulation. For example, in the first model (Logan *et al.*, 1987), the elongation factor can be subjected to regulation. Similarly a 5′ to 3′ exonuclease activity and (or) helicase activity could also be regulated. The passage of RNA polymerase II along the DNA could be blocked by the presence of protein factors. For example, in the adenovirus major late protein promoter a CAAT-box-binding protein associates with the CAAT box. Deletion of

the CAAT sequence from an adenovirus major late protein promoter-simian virus 40 poly(A) addition site construct resulted in read through of transcription supporting the notion that the presence of the protein factor blocked the transcription (Connelly and Manley, 1988).

Many cultured mammalian cells produce approximately equal amounts of full-length mRNA transcripts and short 5´ capped RNAs of only a few hundred nucleotides (Darnell *et al.*, 1990). This phenomenon of premature termination may point towards the inherent inefficiency of the transcription apparatus or indicate a normal regulatory process involved in the control of gene expression (Bentley and Groudine, 1986). These investigators showed that in undifferentiated growing HL-60 cells, about half the transcripts from the cellular oncogene, c-myc, were full length and presumably gave rise to mature mRNA. In contrast, when the cells were induced to differentiate with retinoic acid, only 2 to 3% of c-myc transcripts were full length even though the transcription rate of the gene remained the same in induced and uninduced cells. Nuclear run off transcription experiments showed that this decrease in c-myc mRNA content was at the level of elongation (Bentley and Groudine, 1986). The site responsible for premature termination was mapped to two stretches of T near the first exon-intron junction, about 300 nucleotide downstream of the transcription start site of the c-myc gene. By injection in *Xenopus* oocytes the sequence required for premature termination was further mapped to within a 95 base region located -130 to -35 relative to the exon-1/intron-1 boundary, was found to be

orientation dependent and was also active downstream of a heterologous promoter (Bentley and Groudine, 1988).

This type of regulation is not unique to the c-myc gene; other examples of such regulation were noted in viral late transcription units, such as in the polyoma virus, simian virus 40 and adenovirus major late and also in cellular genes, such as *Drosophila* heat shock protein 70, human c-fos, mouse β-globin and immunoglobulin μ and δ constant regions (Proudfoot, 1989). In mouse immunoglobulin μ and δ constant region genes, termination of transcription between μ and δ exons prevents expression of δ constant region polypeptides in certain early B cells. In later stages of B cell development, both μ and δ polypeptides are expressed together by alternate splicing and in these B cells termination does not occur between μ and δ (Mather *et al.*, 1984).

Some information is being generated regarding the specific sequence requirement for either regulated or normal transcription termination but the mechanism is not completely understood. Some of the features of a termination site are as follows (Manley *et al.*, 1989). It requires a functional poly(A) addition signal (AAUAAA) and a sequence located at or near the actual site of termination. In the human α-globin gene, a 50-bp sequence located about 300 bp downstream of the poly(A) addition site was identified. This orientation-dependent sequence could facilitate termination not only at its natural position, but also within an intron, so long as a functional poly(A) addition site was inserted upstream. In adenovirus, an

inverted CCAAT box sequence just upstream of the late promoter functioned as a terminator sequence (Manley *et al.*, 1989). *In vitro* transcription studies have shown that in poxviruses, the sequence TTTTTNT signals termination approximately 50 nucleotides downstream (Earl *et al.*, 1990). In vaccinia virus the signal is actually recognized in RNA as UUUUUNU. The presence of this element near the ends of many early poxvirus genes and its infrequent occurrence near the start or middle of those genes, suggested that this signal is also used *in vivo* (Earl *et al.*, 1990). However, despite the common occurrence of this sequence in the middle of late genes, late mRNAs are long, vary in size and are apparently not terminated prematurely near these signals. This led to the suggestion that this TTTTTNT motif is utilized only in the early transcription system or, in late infection, requires a yet unrecognized sequence in addition to the termination signal, TTTTTNT (Earl *et al.*, 1990).

1.14.6. Polyadenylation

Most eucaryotic mRNAs, with the exception of replication-dependent histone gene transcripts and some viral mRNAs, contain a stretch of A residues at their 3′ ends. The poly(A) tail is not encoded in the DNA but added post-transcriptionally to the 3′ hydroxyl group generated by cleavage of pre-mRNA (Darnell *et al.*, 1990). Alkaline hydrolysis and end group analysis provided strong evidence for the localization of poly(A) at the 3′ end of mRNAs (Brawerman, 1976). This was further supported by the use of highly purified exonucleases specific for the 3′ end of mRNAs (reviewed in Brawerman, 1976).

The presence of a poly(A) tail in eucaryotic mRNAs remained unnoticed until the 1960s even though there were many reports of mammalian RNAs rich in AMP (Hoyer *et al.*, 1963; Salzman *et al.*, 1964). For instance, an RNA component rich in adenine content was isolated from rat liver cytoplasm (Hadjivassiliou and Brawerman, 1967). This RNA fraction labelled far more rapidly than ribosomal RNA with [^{14}C]orotic acid, was heterogeneous in size and was very effective in stimulation of protein synthesis by cell free extract from *E.coli* cells, properties suggesting that it probably represented an mRNA fraction. Lim and Canellakis (1970) provided the evidence that globin mRNA contained a poly(A) tail by showing that a 9S RNA component from rabbit reticulocyte polysomes bound to polystyrene beads which bind purine-rich polyribonucleotides.

The first estimate for the size of poly(A) tails was obtained by rate zonal sedimentation of labelled poly(A) released from polysomes by ribonuclease treatment and was found to be about 80 nucleotide long (Lee *et al.*, 1971). Another estimate based upon calibration curves relating poly(A) size to sedimentation coefficient was of about 220 (Jefferey and Brawerman, 1974). This latter value falls well within the range (200 to 250) of that found in most of mammalian cells and probably in all vertebrate cells (Darnell *et al.*, 1990). In cells of lower animals and plants, the length of poly(A) tail may be shorter than in mammalian cells and the size is characteristic of that particular group of organisms. Once the mRNA reaches the cytoplasm, the poly(A) segment is decreased in size gradually as the mRNA ages and at steady state,

its size ranges between about 30 and 250 nucleotides (Darnell *et al.*, 1990).

1.14.6.1. Functions of poly(A)

The wide existence of the poly(A) tail in mammalian mRNAs led many investigators to suggest that it helps to transport mRNAs from the nucleus to the cytoplasm (reviewed in Brawerman, 1976), controls mRNA stability (reviewed in Ross, 1988; Jackson and Standart, 1990) and aids in translation of mRNAs (reviewed in Bernstein and Ross, 1989; Munroe and Jacobson, 1990). Following is a brief review of these functions.

1.14.6.1.1. Transport

The inhibition of polyadenylation by the adenosine analog, cordycepin (3´-deoxyadenosine) decreased the transport of mRNA from the nucleus to the cytoplasm. This observation led to the suggestion that nuclear polyadenylation is an essential step in the production of mRNAs for its transfer to the cytoplasm (Brawerman, 1981). Adenovirus mRNA produced *in vitro* in the presence of cordycepin lacking poly(A) tail was correctly spliced but did not enter the cytoplasm, further suggesting a role for poly(A) in mRNA transport (Zeevi *et al.*, 1981). It appears, however, that the presence of a poly(A) segment is not an absolute requirement for mRNA transport to the cytoplasm, since histone mRNAs lacking poly(A) tails are efficiently transported to the cytoplasm (Perry *et al.*, 1974)

1.14.6.1.2. mRNA stability

The steady-state levels of many mRNAs are determined to a significant extent by their turnover rates (Ross, 1988). Transcriptional regulation seems to be important for the accumulation of cell type-specific mRNAs like globin in red cell precursors; a similar correlation between mRNA levels and transcription rates does not exist for housekeeping genes. Some genes are transcribed at a relatively low rate, but their mRNAs are fairly abundant. Other genes are transcribed at relatively high rates, but their mRNAs are rare. In such cases, the steady state mRNA level correlates with its stability. The turnover rate of mRNAs from many genes can affect the regulation of the gene (Ross, 1988). There are examples of the regulation of mRNA turn over rate influenced by sequences in the 5′ or 3′ untranslated regions (Ross, 1988).

There is plenty of evidence implicating a role of poly(A) in mRNA stability (reviewed in Jackson and Standart, 1990; Bernstein and Ross, 1989). Several observations indicate that there is a correlation between the length of poly(A) and mRNA degradation: 1) The addition of poly(A) to some mRNAs stabilizes them; 2) some, but not all, mRNAs probably lose their poly(A) tail before they are degraded, suggesting that removal of poly(A) is a prerequisite for degradation and that an mRNA is more stable with its poly(A) tail; and 3) certain hormones and sugars that stabilize specific mRNAs induce elongation of poly(A) (Ross, 1988) (also see section 1.14.6.1.4).

When injected into *Xenopus* oocyte the half life of rabbit globin mRNAs, with or without varying lengths of poly(A) was found to be dependent upon the poly A length (Nudel *et al.*, 1976). This RNA had stability equal to that observed for normal mRNA [with a poly(A) tail length of about 150 As] as long as the length of poly(A) remained at least 32 nucleotides long. In contrast, the stability of an mRNA preparation with about 16 A residues was at least 10 times less than normal mRNA (Nudel *et al.*, 1976). Histone mRNAs that normally lack a poly(A) tail had a longer half life when polyadenylated before injection into oocytes (Huez *et al.*, 1978).

The poly(A) of all mammalian cell mRNAs with the possible exception of mRNAs in oocytes are associated with a number of proteins. A common 78,000 dalton protein is the predominant factor binding to all mRNAs (Bernstein and Ross, 1989; Jackson and Standart, 1990). This poly(A) binding protein seems to be the component that is responsible for providing mRNA stability. It has been observed that once the poly(A) tract has reached some minimum length (about 30-40) further poly(A) shortening and subsequent degradation of mRNA takes place very rapidly (Bernstein and Ross, 1989). This observation is consistent with the view that poly(A) is associated with a poly(A) binding protein and organized into a nucleosome-like ribonucleoprotein complex (Bernstein and Ross, 1989).

Although the poly(A) binding protein is found to be associated with a stretch of about 27 adenosine residues (Baer and Kornberg, 1983), the actual length of the

contact region is only about 12 nucleotides (Sachs *et al.*, 1987). The association of poly(A) tails with poly(A) binding protein protects the mRNA from the action of non-specific nucleases resulting in increased stability of the mRNA. For instance, in a cell free mRNA decay system, Bernstein *et al.* (1989) investigated β-globin mRNA stability as a function of its association with poly(A) binding protein and observed that in the presence of a competitor poly(A) but not of poly(G), poly(U) or poly(C), the β-globin mRNA degradation rate increased greatly. There was no affect on the stability of histone mRNA which is normally unpolyadenylated. These observations indicated that the high concentration of competitor poly(A) in the *in vitro* system exhausted the limited supply of poly(A) binding protein resulting in little or no retention of the binding protein to the poly(A) tract of β-globin mRNA which as a result became more prone to nuclease attack (Bernstein *et al.*, 1989).

The above discussion complies with the conclusion that the poly(A) binding protein protects the poly(A) tail and the mRNA attached to it. This implies that the poly(A) tail of mRNAs such as c-fos and c-myc undergoing very rapid shortening must be naked for a higher proportion of the time than stable mRNAs such as globin. A major determinant of the rapid turnover of these mRNAs as well as lymphokine mRNAs is multiple copies of an AU-rich motif with a consensus sequence, UAUUUAU common to the 3´ untranslated regions of these mRNAs (Caput *et al.*, 1986; Wilson and Treisman, 1988; Jones and Cole, 1987). It has been suggested that poly(A) binding protein readily migrates from the poly(A) tail to these

3´ untranslated region instability sequences, leaving the poly(A) tract naked for a high proportion of the time thus making the mRNA vulnerable to degradation (Bernstein and Ross, 1989).

1.14.6.1.3. Translation

Results of some early studies involving *in vitro* translation systems and various methods of poly(A) tail removal led to the conclusion that poly(A) tails of mRNA do not aid in translation (Bard *et al.*, 1974; Munoz and Darnell, 1974; Sippel *et al.*, 1974; Soreq *et al.*, 1974; Williamson *et al.*, 1974; Spector *et al.*, 1975). For instance, it was observed that artificial deadenylation or blockage of poly(A) tail with poly(U) did not significantly reduce the translatability of mRNA (Doel and Carey, 1976). However, this conclusion was somewhat erroneous since it was based upon the use of translation systems that were later shown to reinitiate translation very poorly and thus were unable to detect little differences in mRNA translational efficiencies (Doel and Carey, 1976). These investigators showed that native ovalbumin mRNA containing a poly(A) tail was translated more efficiently than its deadenylated equivalent in reticulocyte lysates and no such difference was observed in a less active wheat germ extract. Many investigators have provided evidence that the presence of a poly(A) tail aids in translation. For instance, Deshpande *et al.*, (1979) reported that native $\alpha 2\mu$-globulin mRNA with an average length of poly(A) tail of 175 nucleotides was translated more efficiently and reached peak translatability faster than its poly(A)-poor counterpart with an average length of about 40 A residues. Similarly other

poly(A)-containing mRNAs such as chicken lysozyme and *Xenopus* β-globin mRNAs were translated more efficiently in *Xenopus* oocytes than their deadenylated counterparts (Drummond *et al.*, 1985; Galili *et al.*, 1988).

It has been shown that, like enhancement of mRNA stability, the increased translatability of poly(A)-containing mRNAs also involves its association with poly(A) binding protein (Jackson and Standart, 1990; Munroe and Jacobson, 1990). Jacobson and Favreau (1983) found that exogenously added poly(A) inhibited translation of only the poly(A)$^+$ mRNAs but not of poly(A)$^-$ mRNAs. Further, this competitive inhibition of translation was dependent upon the size of competitor poly(A) and could be overcome by translating messenger ribonucleoprotein partcles (mRNPs) which presumably had bound poly(A) binding protein, instead of mRNA preparations devoid of any bound protein. Another important finding was that the poly(A)-mediated inhibition did not affect the average size of the polypeptide synthesized, suggesting that the inhibition occurred at the level of translation initiation. These observations have been confirmed in a variety of translation systems (reviewed in Munroe and Jacobson, 1990). Other experiments with a variety of reagents used separately to block different steps of the protein synthetic pathway revealed that in the absence of a poly(A) tail, the rate of 40S ribosome-mRNA complex joining with a 60S ribosome subunit to yield an 80S ribosome complex was reduced (Munroe and Jacobson, 1990).

1.14.6.1.4. Developmental regulation of polyadenylation

In the early development of many animal species, translational regulation of several maternal mRNAs is correlated with changes in poly(A) length; mRNAs that undergo poly(A) elongation are subsequently translated, while others that lose their poly(A) tails dissociate from ribosomes (Wickens, 1990a). Similarly other mRNAs undergo a change in poly(A) tail length as a result of fertilization. In somatic cells, an increase in poly(A) tail length results in enhanced translation of insulin, vasopressin and growth hormone mRNAs (Wickens, 1990a). It was demonstrated in frog oocytes that the nuclear and maturation-specific polyadenylation activities were distinct in substrate specificity and subcellular localization (Fox *et al.*, 1989). The maturation-specific activity which did not seem to involve any nuclear activity, was perhaps cytoplasmic in origin and was dependent upon the AAUAAA sequence and another element UUUUUAU located a few nucleotides upstream from the poly(A) addition signal. Poly(A) tails of about 50 to 300 nucleotides were added to the 3´ ends of maturation specific mRNA. Based upon these observations, Fox *et al.* (1989) proposed that this maturation-specific polyadenylation is responsible for translational activation of these mRNAs. Translational activation of dormant tissue-type plasminogen activator mRNA by polyadenylation during meiotic maturation of mouse oocyte has also been demonstrated (Vassalli *et al.*, 1989). In these cells, as in the *Xenopus* oocytes, both elements (AAUAAA and UUUUUAU) were required for polyadenylation.

In contrast to maturation specific polyadenylation, deadenylation did not seem to require any cis-acting sequences (Varnum and Wormington, 1990). It was shown by investigation of various mRNAs that the deadenylation and translational inactivation of maternal mRNAs during *Xenopus* oocyte maturation occurred by a default pathway in which transcripts lacking the cytoplasmic polyadenylation element (UUUUUAU) underwent poly(A) removal (Fox and Wickens, 1990; Varnum and Wormington, 1990).

1.14.6.2. Sequence requirement for polyadenylation

1.14.6.2.1. AAUAAA Sequence

Comparison of the primary structure of many mRNAs revealed that the hexanucleotide, AAUAAA located about 20 nucleotide upstream from the 3´ end of the mRNAs was important in polyadenylation (Proudfoot and Brownlee, 1976). After this first report of the presence of a sequence involved in the polyadenylation process, almost all mRNAs of higher eucaryotes have been shown to contain this hexanucleotide sequence at about 10 to 30 bases upstream to the 3´ end of mature mRNAs. Berget (1984) analyzed 61 vertebrate sequences and detected AAUAAA sequence in about 90% of the cases, or a variant AUUAAA sequence in about 10% of the mRNAs. With the exception of the invariable U at position three of the hexanucleotide sequence, very rare variants of other nucleotides were also detected in many mRNAs (Manley *et al.*, 1988). Recently, Sheets *et al.* (1990) analyzed 269 vertebrate cDNA sequences and detected this hexanucleotide AAUAAA or a very

close variant, the most common of which was AUUAAA, between 4 and 50 nucleotides upstream from the poly(A) tail; 75% had the sequence at a distance of 15 to 25 nucleotides upstream of the poly(A) tail.

The importance of the AAUAAA sequence for polyadenylation was established by studies of mutants of this hexanucleotide, occurring both naturally as well as created by recombinant DNA techniques. For instance, a patient with α-thalassaemia had a point mutation converting the AAUAAA sequence to AAUAAG resulting in a reduced level of α-globin mRNA (Higgs *et al.*, 1983). Similarly, a patient with β-thalassaemia had AAUAAA converted to AAUAAG and also produced reduced levels of β-globin mRNA (Orkin *et al.*, 1985).

Fitzgerald and Shenk (1981) studied the effect of deletion of a small number of bases in the vicinity of the AAUAAA sequence of simian virus 40 late transcription unit and showed that polyadenylation was prevented when these deletions included the AAUAAA sequence. It was further noted that the mutants carrying small deletions between the AAUAAA and the normal poly(A) addition site of simian virus 40 mRNA produced mRNAs polyadenylated at new downstream sites; this indicated that sequences downstream of AAUAAA were also important for accurate selection of the polyadenylation site (Fitzgerald and Shenk, 1981). Natural revertants of some of these artificially created mutants producing normal polyadenylated mRNAs revealed that they had regained the AAUAAA or a close

analogue of this polyadenylation signal (Swimmer and Shenk, 1981). Investigation with the adenovirus early 1A gene revealed that the hexanucleotide sequence (AAUAAA) was essential for cleavage of the primary transcript to produce the 3´ end of mature mRNA (Montell *et al.*, 1983). In this work it was shown that a single point mutation converting the invariant U to G in the AAUAAA sequence of the adenovirus early 1A gene mRNA decreased the efficiency of cleavage without having any effect on polyadenylation of those RNA which did get cleaved (Montell *et al.*, 1983).

The earlier studies on cleavage and polyadenylation were complicated by the apparent coupling of these processes *in vivo*, mainly because cleaved but non-polyadenylated RNAs were never detected, making it impossible to determine which of the two reactions were affected by a particular mutation. However, development of *in vitro* cleavage and polyadenylation systems have begun to shed some light on the requirement of AAUAAA sequence for cleavage, or polyadenylation or both (reviewed in Manley, 1988). These *in vitro* studies showed that a single base substitution in AAUAAA sequence reduces the efficiency of cleavage (Sheets *et al.*, 1990). However, the studies also showed that mutations in the hexanucleotide sequence also interferes with polyadenylation of the natural poly(A) addition site (Sheets *et al.*, 1990).

Sheets *et al.* (1990) did a systematic study of the effect of single base

substitutions in AAUAAA sequence to resolve the sequence requirement for cleavage and polyadenylation. These workers mutated all six nucleotides (one at a time) and created all 18 possible single base variations of the hexanucleotide sequence. With the exception of AUUAAA (the most common natural variant of AAUAAA) all of the single base substitutions caused a great reduction in the addition of poly(A) to RNAs that ended at the poly(A) addition site and also prevented the cleavage of RNAs that extended beyond the poly(A) addition site. It was further noted that both the efficiency of cleavage and polyadenylation varied with different point mutations in the hexanucleotide sequence and that the least effect was by the AUUAAA mutation. However, for a given point mutation, the extent of decrease in polyadenylation and cleavage was the same. This analysis showed that every base in the hexanucleotide sequence is required and provided evidence that the AAUAAA sequence is necessary for both of the 3´ end processing reactions, implying that these two processing reactions required the same factor(s) (Sheets *et al.*, 1990).

1.14.6.2.2. Downstream sequences

The discussion so far has been concentrated on the role of the AAUAAA sequence in cleavage/polyadenylation. In addition to the hexanucleotide, other sequences have also been implicated in 3´ end processing of pre-mRNAs. Lai *et al.* (1979) first noted the presence of a downstream sequence involved in processing the 3´ end of the chicken ovomucoid gene mRNA. An AUGUGUUGGA element located 20 to 50 nucleotides downstream of the AAUAAA sequence has been

detected in several genes (Taya *et al.*, 1982). The presence of multiple poly(A) addition signals in many genes which apparently were not always used, led to the suggestion that the almost invariant hexanucleotide sequence could not be the only sequence responsible for accurate cleavage/polyadenylation. (McDevitt *et al.*, 1984). Therefore, it was suggested that since transcription proceeds beyond the poly(A) addition site, sequences distal to the poly(A) addition site may have a role in this process (McDevitt *et al.*, 1984). These authors generated a series of deletion mutants in the adenovirus E2A transcription unit. Sequences downstream from the 3´ end of the mature RNA were deleted and the transcript produced from these deleted genes were assayed for their ability to produce functional E2A mRNA. The analysis showed that about 35 nucleotides downstream to the poly(A) addition site were required for the formation of E2A mRNA (McDevitt *et al.*, 1984). An extensive search of about 200 mammalian and eucaryotic viral RNAs detected a sequence, YGUGUUYY (Y = pyrimidine), closely related to the AUGUGUUGGA element (Taya *et al.*, 1982) in about 66% of those examined (McLauchlan *et al.*, 1985).

Birnstiel *et al.* (1985) compared various downstream sequences of the genes known at the time and pointed out that most showed an over representation of TGT, found sometimes repeated in conjunction with oligo-T stretches, in a region about 30 base pairs downstream of the AATAAA sequence. Therefore, these sequences which contained an over representation of G and T residues were termed "GT clusters" (Birnstiel *et al.*, 1985). By comparing 74 genes Renan (1987) detected an element

located 5 to 20 bases downstream of the poly(A) addition site with the sequence TTGPNNNTTTTTT (P = A or G) in about half of the sequences. In some cases, this sequence overlapped with the conserved sequence YGTGTTYY (Y = C or T) reported by McLauchlan *et al.* (1985). There is considerable evidence that these downstream elements are important for the 3´end processing of polyadenylated RNAs from a wide variety of genes such as the hepatitis B virus surface protein gene, simian virus 40 early and late transcription units, the rabbit β-globin gene, the herpes simplex virus thymidine kinase gene, human α1, α2 and γ interferon genes, chicken ovomucoid and chicken ovalbumin genes (reviewed in Birnstiel *et al.* 1985; Manley, 1988).

Downstream sequences located within about 50 bases 3´ to the RNA cleavage site have been shown to be required for efficient and accurate polyadenylation of a number of mRNAs including those for simian virus 40 late proteins, rabbit β-globin, bovine growth hormone and herpes simplex virus type 1 thymidine kinase gene *in vivo* (Ryner *et al.*, 1989), and *in vitro* for pre-mRNA for adenovirus-2 L3 (Skolnik-David *et al.*, 1987a, b), simian virus 40 late proteins and herpes simplex virus type 1 thymidine kinase genes (Ryner *et al.*, 1989). Unlike the AAUAAA sequence, the downstream sequences do not seem to be conserved to a great extent in different pre-mRNAs. However, there have been some reports of homologous sequence conservation (Ryner *et al.*, 1989).

Gil and Proudfoot (1984) showed that accurate 3′ end formation for the rabbit β-globin gene required a 35 bp region extending from three nucleotides 5′ to 31 nucleotide 3′ of the poly(A) addition site and that both of the GU- and U-rich elements were required for efficient processing. It was further noted that removal of either of these elements or changing the distance between the two elements decreased the efficiency of cleavage indicating that correct spacing between the two sequences is also necessary (Gil and Proudfoot (1987).

Efficient use of the poly(A) addition site in simian virus 40 early mRNA *in vivo* required sequences between 5 and 18 bp downstream of the cleavage site and the same sequences were also needed for *in vitro* cleavage (Hart *et al.*, 1985). Further experiments using a series of point mutants of this poly(A) addition site and of the GU-rich downstream element revealed that certain base changes decreased the efficiency of cleavage while others increased it (McDevitt *et al.*, 1986). Moreover, the position of the downstream element relative to the AAUAAA and cleavage site was important since moving the element 40 nucleotide downstream, or inversion abolished the function (McDevitt *et al.*, 1986). The adenovirus E2A downstream sequence also had similar properties with one difference and that was that it was U-rich. Mutagenesis studies recognized two different downstream elements both of which were shown to be equally active in restoring the cleavage activity to pre-mRNAs defective in cleavage (McDevitt *et al.*, 1986). These viral studies supported the findings on rabbit β-globin gene pre-mRNA processing (Gil and Proudfoot, 1984;

1987). However, one difference noted was that the β-globin gene had both the GU- and U-rich sequences and both of these sequences were required for efficient cleavage (Gil and Proudfoot, 1987).

Evidence from chemical modification of nucleotides in various sequences involved in 3´ end processing of a synthetic pre-mRNA suggested that although each base in the hexanucleotide was required for cleavage, no single modification in the nucleotides downstream of the poly(A) addition site had any effect on cleavage. On the basis of these experiments, Conway and Wickens (1987) inferred that the critical features of the downstream elements were either diffuse or redundant. However, recently Levitt *et al.* (1989) constructed a synthetic poly(A) addition site based upon the highly efficient poly(A) addition site of rabbit β-globin and demonstrated that for efficient polyadenylation the minimum sequence required was AAUAAA and a GU/U clusters with a correct spacing of about 22 nucleotides between them.

In addition to the GU/U clusters, other sequences have been implicated in having a role in cleavage/polyadenylation reactions. A pentanucleotide sequence, CAYUG (Y= pyrimidine) is found to be present in close proximity to the poly(A) addition site of many genes (Berget, 1984; Benoist *et al.*, 1980). However, many genes lack this sequence and like GU/U clusters this sequence does not seem to be conserved (Manley, 1988).

In most mRNAs an adenosine residue is almost invariably present at the poly(A) addition site and is usually preceded by a cytosine residue (Sheets *et al.*, 1990). Therefore, the first A in the poly(A) tail is in fact part of the pre-mRNA, rather than the first residue added to the cleaved pre-mRNA. Hence, a cytosine is present at the 3´ end of about 59% of mRNAs (Sheets *et al.*, 1990). The terminal adenosine residue seems to have a role in the formation of the precise 3´ end because substitution with other nucleotides influences the precise point of poly(A) attachment (Sheets *et al.*, 1990).

The above discussion regarding the role of downstream elements in 3´ end processing illustrates the complexity of the process. However, as discussed above at least two different sequences are required for the formation of the 3´ ends of mRNAs. Firstly, the almost invariable sequence (AAUAAA) which is present about 25 nucleotide upstream from the 3´ end of the mRNAs (Sheets *et al.*, 1990). Secondly, the downstream element which is reported to be present anywhere between 5 and 50 nucleotides downstream from the poly(A) addition signal (Birnstiel *et al.*, 1985; Renan, 1987: McLauchlan *et al.*, 1985); many different consensus motifs have been proposed for the downstream sequence but it seems to be less conserved except that it shows an over representation of G and U. Additionally, the CA dinucleotide is present at the poly(A) addition sites of many mRNAs.

1.14.6.3. Biochemistry of polyadenylation

The apparent *in vivo* coupling of cleavage/polyadenylation further adds to the complexity of various elements (section 1.14.6.2.1) implicated in having a role in the processing of 3´ ends of pre-mRNAs. For instance, due to the apparent coupling of the two reactions it seems to be difficult to distinguish which sequence is required for each of the two reactions (Manley, 1988). The other question is whether the two processes are obligatorily coupled to each other or can be separated from each other and lastly what is the mechanism of cleavage/polyadenylation and what are the factors involved in these reactions. The development of various *in vitro* systems has provided some answers to these questions (Manley, 1983; Moore and Sharp, 1984; 1985). Some of the findings utilizing these system are discussed below.

The first successful *in vitro* reaction for polyadenylation that was specific for pre-mRNAs was performed using a whole-cell lysate of HeLa cells (Manley, 1983). It was shown that this lysate contained a poly(A) polymerase activity that efficiently added poly(A) tails to exogenously added pre-mRNAs. This addition of poly(A) was controlled since, after a length of poly(A) tail similar to that obtained *in vivo* was achieved, no more nucleotides were added. However, this *in vitro* reaction system was not capable of performing the cleavage reaction to create the correct *in vivo* poly(A) addition sites and the poly(A) tail was added to the 3´ ends of the added RNAs. Nonetheless, this *in vitro* system was specific enough to add poly(A) tails to the 3´ ends of only those pre-mRNAs which had 3´ termini lying downstream of authentic

poly(A) addition sites. This polyadenylation of only the pre-mRNAs with authentic poly(A) addition sites implied that specific nucleotide sequences located near the 3′ terminus of the pre-mRNA may be required for polyadenylation (Manley *et al.*, 1985).

To resolve the specific nucleotide sequences requirement for polyadenylation Manley *et al.* (1985) constructed a number of rearranged and deleted DNA templates encoding simian virus 40 early pre-mRNA, transcribed them in a separate *in vitro* transcription system and incubated the resulting RNAs with the HeLa cell lysate mentioned above. It was observed that, as *in vivo* the AAUAAA sequence was necessary for polyadenylation *in vitro*. Furthermore, it was also shown, that in this *in vitro* system, efficient polyadenylation could take place even in the absence of cleavage to create authentic 3′ ends. Therefore, it provided the evidence that cleavage/polyadenylation reactions which seem to be coupled *in vivo* could also occur separately *in vitro* and hence are not necessarily dependent upon each other (Manley *et al.*, 1985).

Another *in vitro* system for polyadenylation, also from HeLa whole-cell lysates developed by Moore and Sharp (1984), could accurately polyadenylate at the L3 poly(A) addition site of adenovirus. However, *in situ* synthesis of the pre-mRNA in the same extract was essential for the poly(A) addition site specific polyadenylation to take place (Moore and Sharp, 1984). Further improvement of this system

eliminated the need for *in situ* synthesis of pre-mRNAs, and exogenously added pre-mRNA with the L3 polyadenylation site was accurately cleaved as well as polyadenylated (Moore and Sharp, 1985). The latter system was capable of reproducing both of the cleavage and polyadenylation reactions accurately and hence resembled more the *in vivo* situation. The reaction system could be further modified to perform only the cleavage reaction simply by the addition of the ATP analog, α-β-methylene-adenosine 5´ triphosphate. The separation of the two reactions indicated that cleavage and polyadenylation reactions were not coupled to the synthesis of pre-mRNA. Zarkower *et al.* (1986) assessed the role of the AAUAAA sequence in cleavage/polyadenylation by creating point mutations in the simian virus 40 late pre-mRNA and using the above *in vitro* system. These investigators provided evidence that efficient polyadenylation of the precleaved RNAs [RNAs with 3´ termini at the poly(A) addition site] was dependent upon the AAUAAA sequence present upstream from the poly(A) addition site (Zarkower *et al.*, 1986).

Development of these *in vitro* cleavage/polyadenylation systems have opened up new avenues to study the mechanism and the regulation of polyadenylation, which was not possible to study with *in vivo* systems. The *in vitro* systems were manipulated to gain a better understanding of the role of various sequences in pre-mRNAs and the various protein factors involved in cleavage/polyadenylation as described next.

1.14.6.3.1. Protein factors

In recent years biochemical fractionation of the *in vitro* systems has provided further understanding of the mechanism of 3´ end processing. These investigations have revealed the complexity of the reactions which require multiple factors (reviewed in Humphrey and Proudfoot, 1988; Manley, 1988; Wickens, 1990b). A description of some of the important features of these protein factors now follows.

Polyadenylation of a pre-mRNA containing the AAUAAA sequence and having a 3´ terminus requiring cleavage to produce a poly(A) addition site has been shown to require at least two activities; an AAUAAA specificity factor and an enzyme that adds poly(A) to the 3´ hydroxyl group of pre-mRNA (reviewed in Humphrey and Proudfoot, 1988; Wickens, 1990b).

1.14.6.3.2. Poly(A) polymerase

A poly(A) polymerase activity detected about 30 years ago in crude cellular extracts was capable of adding poly(A) to any RNA such as tRNA (reviewed in Edmonds and Winter, 1976) and was partially purified (Bardwell *et al.*, 1990a, b). The apparent non-specificity of the poly(A) polymerase seemed to indicate that it was perhaps not involved in polyadenylation of mRNAs (Wickens, 1990b). However, it has been shown that indeed the classical poly(A) polymerase is the enzyme that adds poly(A) tails to pre-mRNAs by acquiring a specificity factor (see below). This poly(A) polymerase shares many biochemical properties with the classical poly(A) polymerase

including antigenicity (Terns and Jacob, 1989) and a molecular weight of 40,000 to 60,000 daltons (Christofori and Keller, 1989; Gilmartin and Nevins, 1989).

This enzyme was found to be very non-specific and polyadenylated any RNA in the presence of manganese and in the absence of a specificity factor (Christofori and Keller, 1989). However, the enzyme became very specific and polyadenylated only the AAUAAA-containing RNAs on the substitution of manganese with magnesium and the addition of a specificity factor (Christofori and Keller, 1989). Further, it was noted that the cleavage activity was part of poly(A) polymerase activity since even after extensive purification of the poly(A) polymerase the cleavage activity could not be separated from it (Christofori and Keller, 1989; Terns and Jacob, 1989). However, the last finding is not supported by others (Takagaki *et al.*, 1988, 1989) who showed that the cleavage activity and poly(A) polymerase were two separable entities as described in the next section.

1.14.6.3.3. Specificity factor(s)

It has been recognized that the cleavage/polyadenylation reaction is a very complex process and requires poly(A) polymerase and a number of factors including a specificity factor. Many investigators have characterized the factors involved in 3´ end processing (Christofori and Keller, 1988; Gilmartin *et al.*, 1988; McDevitt *et al.*, 1988; Takagaki *et al.*, 1988; Gilmartin and Nevins, 1989). Some of the salient features of these factors and their role in 3´ end processing will be discussed here briefly.

Takagaki *et al.* (1988; 1989) fractionated HeLa cell nuclear extract by ammonium sulphate precipitation and DEAE-Sepharose chromatography into two major fractions. One of these fractions contained a non-specific poly(A) polymerase which became AAUAAA-specific when mixed with another fraction containing a cleavage specificity factor (CSF). The cleavage specificity factor was required to confer AAUAAA dependence on the polymerase and contained specific cleavage activity (Takagaki *et al.*, 1988). The cleavage specificity factor was further fractionated into three fractions: a) the cleavage stimulation factor (Cstf) which stimulated endonucleolytic cleavage several fold; b) the cleavage factor (CF) which was required for endonucleolytic cleavage, was actually composed of two factors (CF1 and CF2); and c) the specificity factor (SF) which was required for both cleavage and AAUAAA-dependent polyadenylation. It was shown that these above mentioned fractions together with a very highly purified poly(A) polymerase could perform accurate and efficient cleavage and polyadenylation (Takagaki *et al.*, 1989).

It has been shown by a UV cross linking procedure to various polyadenylation substrates containing the AAUAAA sequence that a 64,000 daltons protein reacts with the AAUAAA sequence (Wilusz and Shenk, 1988). It was suggested by Wilusz *et al.* (1990) that the 64,000 dalton protein bound specifically to the AAUAAA sequence in the RNA polymerase II nascent transcripts and interacted with other factors to initiate 3´ end processing of pre-mRNAs. This protein was found to cross-link to substrate RNAs only when the fractions containing cleavage stimulation factor

(Cstf) and specificity factor (SF) were mixed together in the presence of substrate RNA (Wilusz *et al.*, 1990). This finding was consistent with the conclusion that this 64,000 dalton protein was a constituent of the multicomponent complex required for the AAUAAA dependent polyadenylation of RNAs (Wilusz *et al.*, 1990). It was shown that the 64,000 dalton protein was indeed one constituent of Cstf (Takagaki *et al.*, 1989) which was composed of three distinct polypeptide chains of 77,000, 64,000 and 50,000 daltons (Takagaki *et al.*, 1990).

Christofori and Keller (1988) fractionated a HeLa nuclear extract but obtained three different components involved in 3´ end processing of pre-mRNAs; these were poly(A) polymerase, cleavage factor (CF) and cleavage and polyadenylation factor (CPF). The last component was a ribonucleoprotein complex containing U11 snRNP. These workers showed that for polyadenylation of precleaved RNA, poly(A) polymerase and CPF were sufficient, whereas for uncleaved pre-mRNAs all three fractions were required. It was suggested that the CPF corresponds to the cleavage specificity factor (CSF) characterized by Takagaki *et al.* (1988) as described above. However, further characterization of the cleavage and polyadenylation factor (CPF) revealed that it had a molecular weight of 180,000 to 200,000 daltons (Christofori and Keller, 1988), which was lower than 360,000 daltons reported for the cleavage specificity factor (CSF) characterized by Takagaki *et al.* (1988).

Gilmartin and Nevins (1989) described a number of factors involved in 3´ end

formation of mRNAs. These factors were: a) polyadenylation factor 1 (PF1) which had a poly(A) polymerase activity; b) polyadenylation factor 2 (PF2) a complex of multiple proteins, some of which interacted directly with the AAUAAA sequence and provided specificity to PF1; c) cleavage factor 1 (CF1); and d) cleavage factor 2 (CF2). The CF1 and CF2 were shown to be required along with PF2 for cleavage activity. Gilmartin and Nevins (1989) presented a pathway of complex formation for the initiation of cleavage that was followed by polyadenylation. According to this proposed pathway PF1 and PF2 were required for specific polyadenylation of the cleaved RNA; both of these factors together with two additional factors CF1 and CF2 were required for the endonucleolytic cleavage of the pre-mRNAs. The PF1 was suggested to be equivalent to the cleavage polyadenylation factor (CPF) characterized by Christofori and Keller (1988). It was further suggested by Gilmartin and Nevins (1989) that PF2 recognized the AAUAAA sequence and formed an unstable binary complex which in turn facilitated the binding of CF1 that required a downstream GU/U cluster for efficient binding. Once this ternary complex was formed, the complex remained stable and could be viewed as committed for the processing of the poly(A) addition site. This was followed by binding of CF2 and PF1. The cleavage of the pre-mRNA could then take place and CF2 dissociated from the complex and polyadenylation at the correct poly(A) addition site could continue until about 200 A residues were added (Gilmartin and Nevins, 1989).

The mechanism of addition of poly(A) has been studied in some detail by

Sheet and Wickens (1989) who showed that this process occurred in two phases: in the first phase addition of each adenosine residue was dependent on the AAUAAA sequence and about 10 A residues were added one at a time; and in the second phase addition of A residues did not require AAUAAA but needed the oligo-(A) primer synthesized during the first phase. It was further noted that both reactions were catalysed by the same poly(A) polymerase and that the regulation of final length of poly(A) tail required a factor present in the HeLa cell nuclear extract (Sheet and Wickens, 1989).

It is clear from the above discussion that cleavage/ polyadenylation reactions are very complex processes requiring a large number of factors. The interplay of these factors brings about two important results: cleavage of pre-mRNA to form the 3´ end; and addition of a poly(A) tail. The requirement of many factors makes this process a good candidate for post-transcriptional control of gene expression. Like alternate splicing, poly(A) addition site choice as a mode of gene expression control is well documented (Leff *et al.*, 1986). One level of control where the poly(A) addition site choice could be regulated may be in the binding of various factors especially CF1 to the downstream GU/U sequences. The other possibilities are that either secondary structure of pre-mRNA in the vicinity of the poly(A) addition site might regulate its selection (Brown *et al.*, 1991) or some protein factor(s) may be involved (Denome and Cole, 1988).

CHAPTER 2

MATERIALS AND METHODS

2.1. Materials

Most of the chemical reagents used were purchased either from Sigma Chemical Company (USA) or BDH Chemicals (Canada) unless stated otherwise. The oligo(dT)-primed λgt10 cDNA library constructed from K562 cell poly(A)$^+$ RNA was obtained from Clontech Laboratories Inc. (USA). Oligodeoxynucleotides were obtained from the following suppliers: I synthesized the mixture GPA-C manually at the Biochemistry Department of The University of British Columbia, Vancouver. Oligonucleotide GPA-N2 and the mixture GPA-N1 were obtained from the Biotechnology Service Centre, The Hospital for Sick Children, Toronto, Ontario; another oligonucleotide GPA-C2, same as coding strand and oligonucleotides complementary to the 3´ untranslated region of glycophorin A, GPA-MS, GPA-ML1, GPA-ML2 and GPA-L were obtained from the DNA Synthesis Laboratory, University of Calgary, Alberta.

Restriction endonucleases *Eco*RI, *Kpn*I, *Xho*I, *Hind*III, *Hpa*II, *Dra*I, *Stu*I, *Alu*I and *Bam*HI were purchased from Bethesda Research Laboratories (Canada), *Xmn*I, *Rsa*I and *Sac*I were from New England Biolabs (USA). Bluescript plasmid vectors, XL1-Blue cells and the Bluescript Exo/Mung kit were obtained from Stratagene

(USA). The Sequenase kit was purchased from United States Biochemicals (USA) and the Geneclean kit was from Bio 101 Inc. (USA).

Ultrapure cesium chloride, redistilled phenol, formamide, ultrapure agarose, RNA ladder molecular size markers, 5´ DNA terminus labelling system kit, *E.coli* DH5α cells and plasmid vector pUC19 were purchased from Bethesda Research Laboratories (Canada). Ficoll, polyvinylpyrrolidone, bovine serum albumin (BSA), maltose, glucose, polyethylene glycol (PEG, 6,000), RNase A, ethidium bromide, diethylpyrocarbonate (DEP) and the antibiotics, ampicillin and tetracycline hydrochloride were obtained from Sigma Chemical Company (USA). Bacto-yeast extract, Bacto-tryptone, Bacto-Agar and dimethyldichlorosilane were purchased from BDH chemicals (Canada). The tissue culture medium Roswell Park Memorial Institute 1640 (RPMI 1640) and fetal calf serum were purchased from Flow Laboratories (Canada). M13 universal sequencing primer, M13 reverse sequencing primer, salmon sperm DNA, T4 polynucleotide kinase, nucleoside triphosphates and Sephadex G-50 were purchased from Pharmacia Laboratories (Canada). Yeast tRNA and alkaline phosphatase were purchased from Boehringer Mannheim (Canada). Reverse transcriptase was obtained from Life Sciences Inc. (USA). The nylon membranes Hybond-N, [γ-^{32}P]ATP (>5,000 Ci/mmol), [α-^{32}P]dCTP (3,000 Ci/mmol) [α-^{35}S]dATP (1350 Ci/mmol), Nick Translation and the Multiprime DNA Labelling System kit were purchased from Amersham (Canada) Ltd. T4 DNA ligase was purchased from New England Biolabs (USA). Dialysis bags (molecular weight cut,

12,000-14,000) were purchased from Spectrum Medical Industries Inc. (USA). Tris (hydroxymethyl) aminomethane (Tris) and mixed bed resin AG 501-X8 were obtained from BioRad (USA). Oligo(dT)-cellulose (type 3) was purchased from Collaborative Research Inc. (USA).

Methods

2.2. General methods for the preparation of solutions

Most of the solutions to be used for recombinant DNA work were routinely made with double distilled water that was deionized by passage through a Millipore filtration unit and sterilized by autoclaving at 15 lb/square inch at 121°C for a time period appropriate for the volume of the solution to be autoclaved. The reagents which could not be autoclaved were usually prepared in pre-autoclaved distilled water and sterilized by filtration through a 0.22 μm cellulose nitrate membrane.

2.3. K562 cell culture

Human erythroleukaemic cells, K562 (Lozzio and Lozzio, 1975) were grown in an incubator at 37°C with CO_2:O_2 (5:95%) in the tissue culture medium RPMI 1640 which was prepared according to the manufacturer's instruction with 10% fetal calf serum. To prepare cells for RNA extraction, 500-1000 ml of the growth medium was distributed in 5-10 flasks and inoculated with K562 cells (5 x 10^4/ml). The cells were allowed to grow for 4-6 days before harvesting for RNA extraction (section 2.18.2).

2.4. Microbiological techniques

2.4.1. Luria-Bertani (LB) medium

The following components were dissolved in about 800 ml water in a 2 L flask

Bacto-yeast extract	5 g
Bacto-tryptone	10 g
NaCl	10 g

The solution was adjusted to pH 7.4 by the addition of 10 M NaOH solution and the volume was made up to 1 L. If needed, the medium was distributed into smaller volumes, sterilized by autoclaving, cooled to the room temperature prior to the addition of glucose or maltose (0.2%, w/v, final concentration, from a 20% stock solution sterilized by filtration through a 0.22 μm cellulose nitrate membrane) and (or) any other additional reagents (antibiotics, reagents for detection of β-galactosidase production) as required.

2.4.2. Luria-Bertani (LB) medium containing agar

For culture media requiring agar, 15 g Bacto-agar was added to 1 L of the liquid medium before autoclaving. To prepare the culture plates, the autoclaved agar medium was allowed to cool to 45-50°C before glucose and any other additional reagents as required were added. The contents of the flask were mixed by swirling gently and poured into the culture plates (100 x 15 mm or 150 x 15 mm). The agar was allowed to set at room temperature and the plates were then stored at 4°C till use.

2.4.3. Top agar/agarose medium plates

The medium was prepared as described (section 2.4.1) for the liquid culture medium (LB) except 0.75% (w/v) of either agar or agarose was used instead of 1.5% agar. To prepare this medium, 0.75% (w/v) of either agar (for plates to be used for growing transformants) or 0.75% agarose (to be used for plaque lifts) was dissolved in LB medium by heating, distributed in 2.5 ml quantities for 100 x 15 mm plates or 7.5 ml quantities for 150 x 15 mm plates and sterilized by autoclaving. To prepare the culture plates containing the top agar/agarose, the LB-agar plates stored at 4°C were pre-warmed by incubation at 37°C for a few hours. The top agar/agarose medium was liquefied by heating in a boiling water bath or a microwave oven and allowed to cool to about 48°C. Additional reagents (for e.g. glucose, maltose, antibiotics or reagents for the detection of β-galactosidase production) were added at this time and poured on top of the solid LB-agar plate swirling it very quickly for an even distribution of the agar/agarose solution in the form of a thin layer.

2.4.4. Media for the detection of β-galactosidase production

To prepare top agarose for screening bacterial colonies producing β-galactosidase, the plates were prepared as described (section 2.4.3.) except that the liquefied top agar/agarose solution also contained 50 μl Bluo-gal and 10 μl isopropyl-β-D-thiogalactopyranoside (IPTG)/2.5 ml top agar/agarose; Bluo-gal was stored as a 2 mg/ml stock solution in dimethylformamide and IPTG was stored as a 100 mM stock solution in water. IPTG induces β-galactosidase production in the cells

containing the gene turning the cells blue in the presence of the histochemical substrate (Bluo-gal).

2.4.5. Ampicillin solution

The stock solution containing 10 mg/ml ampicillin in water was sterilized by filtration through a membrane (0.22 μm), stored in small aliquots at -20°C and used at the final concentration of 100 μg/ml.

2.4.6. Tetracycline solutions

Stock solutions containing 15 mg/ml tetracycline hydrochloride in 50% ethanol was sterilized by filtration through a membrane (0.22 μm), stored in small aliquots at -20°C and used at the final concentration of 15 μg/ml.

2.4.7. Bacterial cell culture

The experiments involving bacteriological techniques were performed using standard aseptic techniques. The bacterial cells, unless stated otherwise, were grown either in the liquid culture medium, Luria-Bertani medium containing 0.2% (w/v) glucose and appropriate antibiotic(s) or the same medium with 1.5% (w/v) Bacto-agar. The culture was always initiated with an isolated colony that was grown on a culture medium plate containing appropriate antibiotic(s) to verify the bacterial strain.

2.4.8. Storage of bacterial cultures

About 5-10 ml of liquid medium (LB) containing the appropriate antibiotic(s) was inoculated with a single bacterial colony and allowed to grow overnight by incubation at 37°C. This overnight culture was then distributed in 0.85 ml aliquots in microfuge tubes. To each of these tubes, 0.15 ml sterile glycerol which was previously sterilized by autoclaving was added, mixed and the tubes were then stored at -20°C or at -70°C. Just before use, a tube was withdrawn from the freezer, thawed and an aliquot streaked on LB-agar plate to generate isolated colonies by incubation at 37°C for about 18 h. The plate with the isolated colonies was either used immediately or stored at 4°C for future use until discarded after a few weeks.

2.4.9. Storage of bacteriophage stocks

SM medium: This medium was used for preparing dilutions and (or) storage of bacteriophage. .

NaCl	5.8 g
$MgSO_4.7\ H_2O$	2 g
Tris-HCl (1 M, pH 7.5)	50 ml
Gelatin (2%, w/v)	5 ml
water	to 1 L
sterilized by autoclaving	

To prepare the bacteriophage stock, an isolated plaque was mixed with 0.3 ml

overnight culture of *E. coli* C600 Hfl and incubated at 37°C for 15 min. This phage-bacterial suspension was mixed with 7.5 ml top agarose (section 2.4.3), poured on an LB-agar plate with 0.2% maltose (section 2.4.2) and allowed to grow for about 18 h at 37°C. After this incubation period 5 ml SM medium was added to the plate which was then left standing at room temperature to allow for the diffusion of the phage particles from the agar. The resulting phage suspension was transferred to a glass centrifuge tube (Corex, 15 ml). The plate was rinsed with more SM medium (5 ml), the suspension transferred to the centrifuge tube and the plate was discarded. The lysate was then centrifuged at 10,000 rpm for 15 min at 4°C in a Beckman centrifuge using a JS-13 rotor. Following centrifugation, the cleared plate lysate was transferred to a clean tube and stored with a few drops of chloroform at 4°C or at -70°C with dimethyl sulfoxide at a final concentration of 7% if storage for long periods was desired.

2.5. Phenol extraction of nucleic acids

Phenol-chloroform solution: Double-distilled phenol melted by immersing the bottle in a 65°C water bath was mixed with an equal volume of TE [10 mM Tris-HCl, 1 mM ethylenediaminetetra acetic acid (EDTA) pH 8.0]. After mixing for a few minutes, the aqueous and organic layers were separated by centrifugation in a bench top centrifuge (Damon/IEC model, HN-S) at 2,000 rpm. The equilibrated phenol was then mixed with an equal volume of chloroform and isoamyl alcohol (24:1, v/v).

An equal volume of the phenol:chloroform:isoamylalcohol (25:24:1) solution was added to the nucleic acids (DNA or RNA) solution, mixed and centrifuged at room temperature to separate the organic and aqueous layers. When small volumes of nucleic acid solution were to be purified, centrifugation was performed using a microfuge tube in an Eppendorf centrifuge (model 5414) whereas larger volumes were centrifuged in a bench top centrifuge at about 4,000 rpm for 15 min at room temperature. After the first extraction, the aqueous layer was transferred to a clean tube and rextracted with the phenol solution once or twice more. The aqueous layer was transferred to a clean tube and subjected to extraction with a chloroform:isoamyl alcohol (24:1) for a few times or until a clear interface was obtained. The clear aqueous solution was then subjected to ethanol precipitation described in the next section.

2.6. Ethanol precipitation of nucleic acids

About 0.1 volume of 3 M sodium acetate pH 4.8 and absolute ethanol prechilled to -20°C (two or two and half volumes for DNA or RNA solution, respectively) were added. The contents of the tubes were mixed together and the precipitates were allowed to form by storage at -20°C for about 18 h or for a short period (15 min) at -70°C if smaller volumes were involved. Precipitates were collected by centrifugation either in a Beckman centrifuge (model J-21B) at 10,000 rpm using a Beckman JA-20 or JS-13 rotor at 4°C or for smaller volumes in an Eppendorf centrifuge at the maximum speed. Precipitates were washed twice in 70% ethanol,

centrifuging for short time between each wash and then dried under vacuum using either an in-house vacuum outlet for very small pellet or a vacuum oven (at room temperature) connected to a vacuum pump for larger pellets. The dried pellets were then dissolved in TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

2.7. Quantitation of nucleic acids

To determine the quantity of nucleic acids, an aliquot of the aqueous solution was diluted in water and examined in an Hitachi spectrophotometer (model, 100-80 A). The optical density (OD) at 260, 280 and 320 nm was determined in a 1 cm path length quartz cuvette. The reading of OD_{260} allowed calculation of concentration of nucleic acid in the sample. An OD of 1 was taken to correspond to approximately 50 μg/ml for double-stranded DNA, 40 μg/ml for single-stranded DNA and RNA and 20 μg/ml for oligonucleotides (Maniatis *et al.*, 1982). The ratio of OD at 260/280 was also determined which provided an estimate for purity of nucleic acid solution. Pure DNA and RNA solutions have 260/280 ratio of 1.8 and 2.0, respectively (Maniatis *et al.*, 1982). In addition to the OD measurements, the nucleic acid solutions were also scanned in the same instrument between the wave length of 220-320 nm.

2.8. Restriction endonuclease digestion

Digestion with various restriction endonucleases was performed using appropriate buffer provided by the supplier of the restriction endonuclease at the recommended temperature for a few hours (usually for 1 h). When the plasmid

DNAs were prepared by the quick method (section 2.16.2.), 1 μl of DNase-free RNase from a 10 mg/ml solution was also included in the reaction mixture to digest possible contaminating RNAs. Following the restriction digestion the DNA samples were analyzed by agarose gel electrophoresis as described in the next section.

2.9. Agarose gel electrophoresis

Agarose gel electrophoresis was performed according to the instruction provided by Maniatis *et al.* (1982) using either a submerged gel electrophoresis apparatus, BRL Model H4, 20 x 25-cm gel bed or H3, 11 x 14-cm gel bed or a Bio-Rad Mini-Sub cell, 7 x 10 cm-gel bed. A 0.6-1.5% agarose gel depending upon the size of DNA fragment to be analyzed, in Tris-borate-EDTA buffer (0.089 M Tris and boric acid, 0.002 M disodium EDTA pH 8.3) containing ethidium bromide (final concentration, 0.5 μg/ml, w/v) was prepared. The gel was also submerged in the Tris-borate-EDTA buffer with 0.5 μg/ml ethidium bromide. The DNA samples prepared by mixing with the gel loading dye (0.25% bromophenol blue and xylene cyanol in 15% Ficoll, type 400), loaded in the slots and electrophoresed at a constant voltage of about 5 V/cm. Electrophoresis was continued until the bromophenol dye travelled to about 1 cm from the edge of the gel. The positions of the DNA bands were revealed by exposing the gels to transilluminated UV light to produce the ethidium bromide fluorescence. While still exposing to UV light the gel was photographed using the Polaroid film 665 or 667.

2.10. Recovery of DNA fragment

DNA fragment were recovered from agarose gels by one of the following two methods.

2.10.1. Electroelution method

A gel slice containing the DNA band of interest was placed in a dialysis bag with 0.5 X Tris-borate-EDTA buffer (44.5 mM Tris and boric acid and 1 mM EDTA pH 8.3) sufficient to immerse the gel slice completely. The dialysis bag was then submerged in a shallow layer of 0.5 X Tris-borate-EDTA in the buffer reservoir of a small electrophoresis apparatus. The electric current (100 V) was then passed for 2-3 h so that the DNA was electroeluted from the gel slice and bound to the inner wall of the dialysis bag. The polarity of the current was reversed for 2 min to elute the DNA from the walls of the dialysis bag and the solution containing the DNA was then collected in a microfuge tube. The gel slice was restained and observed under UV light to ensure that all of the DNA was electroeluted from it before discarding. The electroeluted DNA solution was then subjected to phenol extraction (section 2.5) followed by ethanol precipitation (section 2.6).

2.10.2. Geneclean method

Purification of DNA fragments was also achieved by binding to Glassmilk (a suspension of silica matrix in water) and using other solutions provided in the Geneclean kit. The gel slice containing the DNA fragment of interest was placed in

a 1.5 ml microfuge tube. About two to three volumes of the saturated sodium iodide solution was added to cover the gel slice which was then incubated at 50°C for 5-15 min or until the agarose in the gel slice was dissolved completely. About 5 μl of Glassmilk suspension for each 5 μg or less DNA was added and the mixture maintained on ice for 5 min to allow for the adsorption of the DNA on the Glassmilk. The Glassmilk particles collected by a brief centrifugation (5 sec in an Eppendorf centrifuge), were resuspended in the washing solution provided in the kit and the pellet was washed twice more. The DNA was eluted from the Glassmilk in 10-20 μl water by incubating the tubes at 50°C for 5 min. The DNA solution thus obtained was separated from the Glassmilk by a brief centrifugation (1 min in the Eppendorf centrifuge), transferred to a clean tube and stored at -20°C until use.

2.11. Oligonucleotides probes for screening

The complete amino acid sequence of glycophorin A was known at the time when this work was initiated (Tomita and Marchesi, 1975). Two oligonucleotide mixtures GPA-N1 (a 21-mer mixture) and GPA-C (a 17-mer mixture) which are complementary to the coding sequence spanning the amino acid numbers 24-30 and 122-127, respectively, were synthesized and used initially to screen a K562 λgt10 cDNA LIbrary. While work described here was still in progress a partial cDNA sequence was published by Siebert and Fukuda (1986b). Therefore, an exact sequence oligonucleotide, GPA-N2 (33 nucleotide long) complementary to the coding nucleotide sequences spanning amino acid number 30-40 was designed and used

subsequently for screening. The selection of the amino acid number 30-40 was based upon the information that this region is unique for glycophorin A and is absent in the glycophorin B which is very similar to glycophorin A in many regions (Blanchard *et al.*, 1987; Siebert and Fukuda, 1987).

2.11.1. Purification of the oligonucleotides

For purification of the oligonucleotide mixtures, GPA-C and GPA-N1, and the exact sequence oligonucleotide GPA-N2, the crude oligonucleotides were released from the solid support by treatment with concentrated NH_4OH solution. These crude oligonucleotides were then purified through a 20% polyacrylamide gel containing 7 M urea as described below (section 2.11.1 and 2.11.1.2).

2.11.1.1. Separation of oligonucleotides from solid support

The fully protected oligonucleotides attached to the solid support were placed in 1.5 ml microfuge tubes and centrifuged for one min and the acetonitrile supernatant discarded, the remaining acetonitrile was evaporated by drying under vacuum. Enough concentrated NH_4OH solution (about 1.2 ml) was added making sure that the solid support with oligonucleotides attached to it was immersed completely in the solution. The top of the tube was closed and sealed tightly with parafilm and placed in a small vial (liquid scintillation counting vial) containing about 5 ml NH_4OH solution. The top of the scintillation vial was closed and incubated at 50°C in a water bath for about 18 h. The vial was removed from the water bath and

allowed to cool slowly to the room temperature. The pressure of the NH_3 vapours developed during the incubation at 50°C was released very slowly by opening the top of the scintillation vial and the NH_4OH solution was removed from it and discarded. The oligonucleotides now dissolved in NH_4OH solution were distributed into 10 clean microfuge tubes. To each of these tubes 0.1 volume of 3 M sodium acetate pH 7.5, was added and the oligonucleotides were precipitated by addition of five volumes of pre-chilled ethanol and storage at -20°C for about 18 h. The precipitated oligonucleotides were collected by centrifugation for 30 min in an Eppendorf centrifuge at 4°C. The oligonucleotide precipitates were then dried using a vacuum oven at room temperature, dissolved in sterile water and purified as described in the next section.

2.11.1.2. Polyacrylamide gel electrophoresis of oligonucleotides

Polyacrylamide gel (20%) containing 7 M urea was prepared as described in section 2.19.3.3. To prepare the samples for loading on the gel, 10 μl crude oligonucleotide solution (from section 2.11.1.1) was mixed with 20 μl gel loading dye containing formamide (0.1% bromophenol blue and xylene cyanol in 80% formamide) and heated at 90°C for 3 min followed by quick chilling on ice. An aliquot (10 μl) of this denatured oligonucleotide solution was loaded per well and electrophoresis performed at a constant voltage (1600 volts) until the bromophenol dye travelled up to about the 2/3rd of the length of the gel.

Following the electrophoresis, the gel was transferred to a sheet of plastic film (Saran Wrap) which was placed on a thin layer chromatography plate containing silica. When the short wave length UV light is illuminated from above, the TLC plate fluoresces, the oligonucleotides which absorb the UV light appear as dark band against a fluorescent background. While still exposed to the UV light the gel was photographed using a Polaroid camera. The area of the gel containing the major band of oligonucleotide of appropriate size was sliced out using a sharp scalpel. The gel slice was then placed in a microfuge tube and covered with 0.5 M ammonium acetate solution (pH 7.5). Following the incubation at 37°C for about 18 h the tubes were vortexed for 1 min and centrifuged for 5 min in an Eppendorf centrifuge. The supernatant containing the oligonucleotide was withdrawn and transferred to a clean microfuge tube making sure not to carry over any particulate material. The volume of the oligonucleotide solution was then reduced to about 100 μl by repeated extraction with n-butanol and centrifugation in an Eppendorf centrifuge to remove the upper organic layer between each extraction step. About 1 ml of absolute ethanol was added to the concentrated oligonucleotide solution and the precipitates were allowed to form by storage at -70°C for about 30 min. The precipitated oligonucleotides were then collected by centrifugation in an Eppendorf centrifuge for 5 min at 4°C. The supernatant was discarded and the precipitated oligonucleotide pellets were dried under vacuum, dissolved in sterile water and stored at -20°C till use.

2.11.2. 5′ End-labelling of oligonucleotides

Oligonucleotides to be used as the probes for screening the cDNA library and Northern and Southern blots were labelled at their 5′ ends using [γ-^{32}P]ATP (Specific activity, >3,000-5,000 Ci/mmol) and T4 polynucleotide kinase (Sgaramella and Khorana, 1972) using the 5′ DNA terminus labelling system kit as described below.

About 2 μl of a solution containing 5-10 pmol oligonucleotide in water was added to a microfuge tube which was heated at 65°C for 3 min to denature the oligonucleotide and chilled on ice. The denatured oligonucleotide was mixed with about 50-100 μCi [γ-^{32}P]ATP in 5-10 μl, 2.5 μl of a 10 X forward reaction buffer supplied in kit and 1 μl of T4 polynucleotide kinase (10 units) and spun for a few seconds in an Eppendorf centrifuge. Following the incubation for 30 min at 37°C the enzyme was inactivated by heating the tube for 3 min at 65°C. The unincorporated [γ-^{32}P]ATP was separated as described next.

2.11.3. Separation of the unincorporated [γ-^{32}P]ATP from the 5′ end-labelled oligonucleotide

The unincorporated [γ-^{32}P]ATP was separated from the 5′ end-labelled oligonucleotide by passage through a Sephadex G-50 column as described by Davis *et al.* (1986) as follows. About 10 ml Sephadex G-50 slurry previously autoclaved and equilibrated in TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) was poured in a sterile

disposable pipette (10 ml) containing a siliconized glass wool plug at the bottom. The column was washed with several volume of the TE and the sample was applied to the column in a total volume of about 100 μl. The solution was allowed to pass through the column, the sample tube was rinsed with a few portions (100 μl each) of TE which were added and allowed to pass through the column. These were then followed by addition of several 1 ml samples of TE and collection of 10-15 fractions (1 ml each) of column eluate. About 2 μl sample was withdrawn from each fraction into a separate scintillation vial containing 10 ml Aquasol, and the amount of the ^{32}P incorporated in the nucleic acids was determined by counting in a Beckman scintillation counter (model LS 8100 or LS 335). The 5´end labelled oligonucleotides present in the first peak of the radioactivity were pooled together, and the specific activity (counts per min/μg, cpm/μg) was calculated from the total cpm and the amount of oligonucleotide that was used for labelling (section 2.11.2). The unincorporated ATP in the second peak was discarded.

2.11.4. Hybridization with oligonucleotides

The hybridization as well as the washing temperatures for each of the oligonucleotide were determined experimentally using one of the following formulae as a guideline and the procedure of Wallace *et al.* (1979, 1981) with some modifications.

1) Temperature of dissociation (Td) = 4 (G + C) + 2 (A + T)

Where: G, C, A and T are the total number of each of these nucleotides in the oligonucleotide (Bolton and McCarthy, 1962).

2) Temperature of Incubation (Ti) = Melting temperature (Tm) -15°C

$$Tm = 16.6 \log[M] + 0.41 [P_{gc}] + 81.5 - P_m - B/L - 0.65[P_f]$$

(Bonner *et al.*, 1973)

Where:

M = Molar concentration of Na^+, to a maximum of 0.5 (1 X SSC contains 0.165 M Na^+)

P_{gc} = Percent of G or C bases in the oligonucleotide probe (between 30 and 70)

P_m = Percent of mismatched bases

P_f = Percent of formamide in the buffer

B = 675 for synthetic oligonucleotide up to 100 bases

L = Length of the oligonucleotide

The oligonucleotides were allowed to hybridize initially with Southern and Northern blots and plaque replicas at a temperature which was about 15-25°C lower than their Tm or Td as described below. Following hybridization the blots were washed successively at higher temperatures raising the temperature 3-5°C after each

successive wash and exposing the blots to X ray film for about 14-18 h. The washing temperature which gave a high signal to background ratio as determined by autoradiography was then used in subsequent experiments. The following solutions were prepared in advance.

Saline-sodium citrate (SSC) 20 X

NaCl	3 M
Trisodium citrate, disodium salt	0.3 M

The solution was adjusted to pH 7.0 with concentrated HCl then autoclaved and stored at room temperature.

Saline-sodium phosphate-EDTA (SSPE) 20 X

This was an alternate solution used instead of SSC for Northern blotting and consisted of the following reagents dissolved in water;

NaCl	3 M
$NaH_2PO_4.H_2O$	0.2 M
EDTA, disodium salt	0.02 M

The pH was adjusted to 7.0 by the addition of NaOH (10 M), the solution was then autoclaved and stored at room temperature.

Yeast tRNA

About 100 mg tRNA was dissolved in 1 ml water and stored at -20°C. Just

before use, the solution was thawed at room temperature and denatured by boiling for 5 min in a water bath followed by quick chilling on ice.

Denhardt's solution (50 X)

BSA	1% (w/v)
Ficoll	1% (w/v)
Polyvinylpyrrolidone	1% (w/v)

The solution was sterilized by filtration and stored at -20°C.

For hybridization of the oligonucleotides the Northern and Southern blots or plaque replicas were sealed in a plastic bag and prehybridized for about 2-6 h at a temperature which was about 15-25°C lower than the calculated Tm or Td (see above) in a prehybridization solution which consisted of the following reagents at the stated final concentrations;

SSC or SSPE	(6 X or 5X respectively)
Denhardt's solution	(5-10 X)
SDS	(0.1-0.5%, w/v)
Denatured salmon sperm DNA	(50-100 μg/ml, and (or)
Denatured yeast tRNA	(100 μg/ml)
Total volume	25-50 ml

Following prehybridization the fluid was drained from the bag and about 5-25 ml of hybridization solution was added to the bag. The hybridization solution was identical in composition to the prehybridization solution except that it contained about 5-10 pmol of oligonucleotides labelled at their 5´ end with ^{32}P (specific activity, about 10^9 cpm/μg). The hybridization was allowed to take place for about 18 h which was then followed by washing as described below.

2.11.5. Washing Northern and Southern blots or plaque replicas hybridized with ^{32}P-labelled oligonucleotides

The hybridization solution was collected in a tube, by cutting a corner of the bag. If desired this solution was reused after denaturation by heating at 65°C for 10 minutes, up to about two weeks after labelling. The filters now with the hybridized oligonucleotides were immersed in 2-6 X SSC (section 2.11.4) with or without 0.05-0.1% (w/v) SDS and maintained at a temperature which was based upon the calculated Tm or Td of oligonucleotide (section 2.11.4). The membranes were generally washed for 5 min, wrapped between two layers of Saran Wrap while still quite damp and exposed to the X ray film (Kodak X-Omat AR or RP) with two intensifying screens at -70°C for about 18 h. To determine the temperature that gave the highest signal with a minimal background, the blots were washed successively at increasing temperatures (raising the temperature by about 3-5°C after each successive wash) and then re-exposed to the X ray film as above.

2.12. Double-stranded DNA as a probe

The DNA to be used as a probe was either in the form of a plasmid or a cDNA fragment purified from the λ vector by one of the methods described previously (section 2.10).

2.12.1. DNA labelling

The DNA fragments or plasmid DNA were labelled with [α-^{32}P]dCTP using the technique of nick translation (Kelly *et al.*, 1970) or by random primer labelling techniques (Fienberg and Vogelstein, 1983, 1984) as described in the following sections.

2.12.1.1. Nick translation

DNA was labelled by DNA polymerase I and [α-^{32}P]dCTP (>3,000 Ci/mmol) using the Nick translation kit containing all solutions (except the [α-^{32}P]dCTP) and following the instruction of the supplier of the kit. The reaction was performed at 15°C for 2 h using the following conditions.

DNA	0.5 μg
[α-^{32}P]dCTP	100 μCi
Nucleotide buffer	30 μM
DNA polymerase I	2.5 units
DNase I	50 pg
Tris-HCl pH 7.5	10 mM

$MgCl_2$	10 mM
Final volume	25 μl

The reaction was terminated by the addition of 0.5 M EDTA to a final concentration of 0.2 M. The unincorporated [α-^{32}P]dCTP was then separated by a Spun-column procedure (section 2.12.2) or by gel filtration on Sephadex G-50 columns (section 2.11.3).

2.12.1.2. Random primer labelling

Double-stranded DNA was labelled by a random primer method using the Multiprime DNA labelling system kit obtained from Amersham. The reaction was performed following the manufacturer's instruction and using the reagents provided in the kit. The reaction tube contained the following reagents;

DNA	25 ng
Multiprime Buffer solution containing dATP, dGTP and dTTP	10 μl
Primer solution containing random hexanucleotides in aqueous solution with BSA	5 μl
[α-^{32}P]dCTP (>3,000 Ci/mmol)	5 μl
DNA polymerase I (Klenow fragment)	2 units
Water to a final volume	50 μl

The reaction was performed for 5 h at room temperature or overnight at 15°C. The radioactivity incorporated into the DNA was determined by subtracting the trichloroacetic acid (TCA) precipitable cpm measured by scintillation counting (section 2.12.3) from the total cpm. The labelled DNA was used without further purification since in this protocol more than 70-90% ^{32}P gets incorporated into the DNA and the remaining unincorporated ^{32}P does not cause a high background.

2.12.2. Spun-column procedure

Sephadex G-50 slurry equilibrated in STE (10 mM Tris-HCl pH 8.0, 1 mM EDTA and 100 mM NaCl) was added to a 1 ml syringe that was plugged with glass wool. This mini column was then placed in a 15 ml glass centrifuge tube (Corex) containing a capless 1.5 ml microfuge tube so that the column eluate could be collected in the microfuge tube. This was centrifuged at 3,000 rpm in a Beckman centrifuge, using the JS-13 rotor at 4°C for 5 min. More Sephadex was added as required and the centrifugation was repeated until the packed column volume was about 0.9 ml. This was followed by the addition of 0.1 ml STE to the column and centrifugation as before. The addition of STE and centrifugation was repeated once more. The nick translation reaction mixture (total volume about 0.1 ml) was applied to the packed column which was centrifuged exactly as before collecting the column effluent now containing the ^{32}P-labelled DNA in a clean microfuge tube. An aliquot of the purified DNA was then used to estimate the specific activity (cpm/μg) by scintillation counting as described in the next section.

2.12.3. TCA precipitation of nucleic acids for determination of incorporated cpm

The sample of nucleic acid whose cpm was to be determined was diluted in TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) and 1-2 μl of diluted sample was spotted in duplicate on 2.4 cm Whatman 540 filter discs. After air drying for a few minutes, the filter discs together with two blanks for background counting were immersed in a beaker containing about 10-20 ml 10% TCA (w/v) per disc and allowed to wash for 15 min with intermittent shaking. The washing was repeated twice in fresh TCA solution (15 min each time) and the filters were dropped subsequently in 95% ethanol and acetone followed by air drying and counting with 10 ml Aquasol. The cpm and the amount of the DNA used for labelling were then used to calculate specific activity (cpm/μg) of labelled probe.

2.12.4. Hybridization of ^{32}P-labelled double-stranded DNA probe

Northern or Southern blots were placed in a plastic bag which was then sealed on all sides. One of the corners of the bag was then excised and prehybridization solution of the following composition added.

SSC	6 X
Denhardt's solution	5-10 X
SDS	0.1-0.5% (w/v)
Denatured salmon sperm DNA	50-100 μg/ml
Deionized Formamide	50% (v/v)
Total volume	25-50 ml/bag

Northern and Southern blots were soaked in prehybridization buffer for 2-6 h in a sealed bag on a rotator placed in a oven maintained at 42°C. Following the prehybridization, the solution was drained out of the bag by cutting a corner of the plastic bag and replaced with a hybridization solution, identical in composition to the prehybridization solution except that it contained about 10^6-10^7 cpm of the ^{32}P labelled probe denatured by heating at 90°C for 3-5 min immediately before adding to prehybridization solution. The bag was resealed and hybridization was allowed to take place at 42°C for about 18 h. Following the hybridization, the fluid was collected in a tube by cutting the edge of the bag and stored at -20°C for reuse if desired. The membranes were washed and autoradiographed as described in the next section.

2.12.5. Washing of the blots hybridized with ^{32}P-labelled double-stranded DNA probes

Following hybridization the blots were washed in an SSC solution (0.1-6 X SSC) with or without SDS (0.05-0.1%, w/v) at various temperatures and exposed to X ray film (Kodak X-Omat AR or RP) in a cassette with two intensifying screens at -70°C for about 18 h.

2.13. cDNA cloning

K562 is a human erythroleukaemic cell line isolated from a patient with the diagnosis of chronic myeloid leukaemia (Lozzio and Lozzio, 1975). These cells express glycophorins on their surface (Gahmberg *et al.*, 1979). A cDNA library constructed from K562 cell poly(A)$^+$ RNA and cloned into the *Eco*RI site of the

λgt10 vector using standard methods (Huynh *et al.*, 1988) was purchased from Clontech Laboratories (USA). In this protocol, total poly(A)$^+$ RNA is reverse transcribed into single-stranded cDNA using oligo(dT) primers. This is followed by partial digestion of template RNA with RNAse H to generate RNA primers, that are utilized for DNA polymerase I catalyzed second strand synthesis using the first strand of cDNA as template. The double-stranded cDNA is then treated with *Eco*RI methylase to protect internal *Eco*RI sites if present in cDNAs from subsequent action of *Eco*RI in the next step. *Eco*RI linkers are ligated at both ends of the double-stranded cDNA, the excess *Eco*RI linkers possibly ligated in tandem to ends of the cDNAs are removed by digestion with *Eco*RI and the double-stranded cDNA fragments containing *Eco*RI linkers at their ends are ligated into λgt10 arms. These arms are generated by the treatment of the λgt10 DNA with *Eco*RI, which utilizes the unique *Eco*RI site present in the nucleotide sequence of λ coding for the λ repressor (Huynh *et al.*, 1988). The insertion of foreign DNA into this region inhibits the λ repressor production and hence the recombinant phage can not lysogenize in the subsequent steps when the ligated cDNA-vector are packaged using the λ packaging extract and amplified by growth in *E.coli* C600 Hfl (Hfl, high frequency lysogenization). Therefore, in this strain of *E.coli* only the recombinant phage can produce cell lysis and form plaques (Huynh *et al.*, 1988).

2.13.1. Estimation of phage titre

The titre of the phage in the cDNA library or in clear lysate was measured as

described here. A few 10 fold serial dilutions of the phage suspension were made by diluting 1-10 μl in SM medium (section 2.4.9). An aliquot (100-300 μl) of an overnight culture of bacterial host strain C600 Hfl (an *E.coli* derivative supplied with the K562 cell cDNA library) was mixed with top agarose (section 2.4.3, 2.5 ml for 100 x 15 mm plate or 7.5 ml for 150 x 15 mm plates) containing 0.2% maltose, poured on top of the preset plates containing LB-agar with 0.2% maltose (section 2.4.2) and allowed to set for a few minutes. An aliquot (5-10 μl) of each of the phage dilution was then spotted on the top agarose, spread to a small area and allowed to adsorb completely by keeping the culture plate at room temperature (with the top cover slightly open) for about 20 min. The plates were then incubated at 37°C to allow for the phage growth and appearance of the plaques (4-8 h). The number of the plaques was counted and used to calculate the plaque forming units/ml (PFU/ml) in the original phage suspension.

2.13.2. Preparation of λgt10 plaques replicas on nylon membrane

2.13.2.1. Plaque formation

The plaques of cDNA library were formed on LB-agar plates containing maltose (0.2%), by growing a sample of the cDNA library (phage suspension) in *E.coli* C600 Hfl by the method of Benton and Davis (1977) following the instructions provided by the supplier of the cDNA library as described here.

A sample of the phage suspension containing about 25,000 PFU was mixed

with 300 μl of an overnight bacterial culture of C600 Hfl that was grown in LB medium containing 0.2% (w/v) maltose (section 2.4.1). The mixture was incubated at 37°C for 10-20 min to allow for the adsorption of the phage to the cells. About 7.5 ml top agarose was liquefied by heating in a boiling water bath or in a microwave oven and cooled to a temperature of about 45°C. It was then mixed with the phage-cell mixture and poured on the top of the large LB-maltose plates (150 x 15 mm, at least two day old, prewarmed by incubation at 37°C). The top agarose was allowed to set by keeping the plates at room temperature for about 10 min. This was followed by incubation of the plates at 37°C for 4-8 h till the plaques appeared, making sure that they did not grow more than about 1.5 mm in size. After the plaque formation, the plates were sealed with parafilm and stored at 4°C for a few hours to harden the top agarose before making membrane replicas as described in the next section.

2.13.2.2. Plaque transfer and DNA blotting

The DNA from the plaques (2.13.2.1) was fixed following the instruction provided by Amersham manual, *Membrane transfer and detection methods*. A dry nylon membrane (Hybond-N, 132 mm diameter) was placed gently on the culture plate containing plaques using sterile flat tipped forceps avoiding any trapped air bubbles between the plate and the membrane. The membrane as well as the agarose underneath were marked at three or more asymmetric places for later alignment by puncturing holes with a 18-gauge needle attached to a syringe containing water proof ink. The membrane was peeled off carefully after a total of 2 min incubation to allow

for phage particles to adsorb on the membrane which was immersed immediately in the denaturing solution containing 0.5 M NaOH and 1.5 M NaCl. Following the denaturation for 1 min, the membrane was placed sequentially in the neutralization solution (0.5 M Tris-HCl pH 8.0, 1.5 M NaCl) and SSC (3 X) for eight and five minutes, respectively, after which the membrane was placed on Whatman 3 MM paper and allowed to air dry for a few minutes. A second replica was made in the same way from each master plate except that 3 min was allowed for adsorption of the plaques. When all replicas were made, the dried membranes were placed between layers of Whatman 3 MM paper, baked in a vacuum oven at 80°C for 2 h, then stored in sealed plastic bags. The master plates were sealed with parafilm and stored at 4°C.

2.13.2. Screening the cDNA library

The plaque replicas made in duplicate from a K562 λgt10 cDNA library were screened for the clone(s) encoding glycophorin A by hybridization with 5′ end-labelled oligonucleotide GPA-N2 as described in section 2.11.4. Prehybridization was performed for 2-6 h at 50°C in a solution containing the following reagents (prepared as described in section 2.11.4).

SSC	6 X
Denhardt's solution	10 X
SDS	0.5% (w/v)

Denatured yeast tRNA	100 μg/ml
Final volume	25 ml/bag

Hybridization was performed in about 10 ml prehybridization solution containing 5´ ^{32}P-labelled GPA-N2 (about 10^8 cpm, specific activity, 10^9 cpm/μg), at the appropriate temperature (section 2.11.4) for about 18 h. The duplicate membranes were next washed in 6 X SSC at 55°C for 5 min. While the washed membranes were still quite damp, radioactive ink (prepared by diluting an aliquot of ^{32}P in water proof ink) was spotted on the needle marks made previously during plaque transfer procedure. The membranes were then wrapped between two layers of Saran Wrap and exposed to Kodak X-Omat RP film for about 18 h with two intensifying screens at -70°C. Plaques showing hybridization and retaining the signal on both replicas were identified on their corresponding master culture plates and lifted as described in the following section.

2.13.4. Plaque purification

Membranes showing a positive hybridization signal were superimposed on their corresponding master plates, making use of the orientation marks on the plates, membranes and the films. The plaques or area showing positive hybridization signals were removed using the wide end of a sterile pasteur pipette. The agar-agarose plug containing the putative positive clone was then placed in a tube containing about 5 ml SM medium with a drop of chloroform and stored at 4°C for 6 h to overnight to

allow for the phage particles to elute into the medium. Following the incubation at 4°C, the tube was vortexed for a few seconds, centrifuged in an Eppendorf centrifuge to separate the agarose and cellular debris from the phage suspension and the clear supernatant containing the phage particles released from the plaques transferred to a clean tube. The titre of the phage suspension was determined as described in section 2.13.1, an appropriate dilution was replated as described in 2.13.2, plaque replicas were made and rescreened (secondary screening). Each sample was re-plated and screened with GPA-N2 several times until a membrane showing 100% positive plaques was obtained.

Once the λgt10 plaques containing cDNA encoding glycophorin A were biologically pure, they were grown on a large scale, and their DNA extracted and purified (sections 2.14). The DNA was then subjected to restriction digestion with the enzyme *Eco*RI as described in section 2.8, analyzed on an agarose gel (section 2.9) and blotted on the nylon membrane by the procedure of Southern (1975) as described in section 2.17. and probed with oligonucleotides and (or) cDNA fragment (sections 2.11.4 and 2.12.4).

2.14. Isolation of λ DNA

DNA from various λgt10 cDNA clones was isolated by a large-scale method of bacteriophage λ preparation (Maniatis *et al.*, 1982) which involved three major steps described next.

2.14.1. Preparation of phage lysates

A single isolated colony of *E.coli* C600 Hfl was introduced into 50 ml LB medium with 0.2% (w/v) maltose (section 2.4.1) and incubated overnight at 37°C on a rotator. Next day the OD_{600} of the culture was measured in a cuvette of 1 cm path length and an appropriate sample inoculated into a 2 L flask containing about 200 ml LB medium such that the final OD_{600} was 0.1 (or 8 x 10^7 cell/ml, assuming 1 OD_{600} = 8 x 10^8 cells/ml, Maniatis *et al.*, 1982). The culture was allowed to grow as before for about 2 h or till the OD_{600} of the culture reached about 0.4 (about 4 x 10^{10} cell/ 200 ml culture). This logarithmic phase culture was then split into four parts and each was mixed with a volume of λgt10 lysate to give about 5 x 10^5 PFU per 10^{10} cells and incubated at 37°C for 20 min, mixing intermittently to allow for the phage adsorption to take place. Each phage-cell mix was added to a 2 L flask containing 500 ml LB medium with 0.2% (w/v) maltose and incubated at 37°C with vigorous shaking for a few hours until complete lysis was apparent. About 10 ml chloroform was added and the lysate was agitated for a further 30 min at 37°C.

2.14.2. Purification of λ phage

The lysate was cooled on ice to room temperature and pancreatic DNase and RNase A were added to give final concentration of 1 μg/ml each and the flask kept at room temperature for 30 min. Solid NaCl was dissolved in the lysate to a final concentration of 1 M and allowed to stand on ice for 1 h. The cellular debris was removed by centrifugation at 10,000 rpm in a Beckman centrifuge using JA-14 rotor

for 10 min at 4°C. The cleared lysate was transferred to 500 ml Erlenmeyer flasks and the phage particles precipitated by dissolving solid polyethylene glycol (PEG 6,000) to a final concentration of 10% (w/v) and storage on ice for about 18 h. The precipitated phage particles were collected by centrifugation at 10,000 rpm in a Beckman centrifuge using JA-14 rotor for 10 min at 4°C. The supernatant was discarded and the centrifuge bottles containing the precipitated phage particles were allowed to drain completely by inverting them on paper towels. The pellets were then resuspended in SM medium (about 32 ml/2 L lysate) by shaking on a rotator for 12-24 h. The resuspended phage particles were then vortexed with an equal volume of chloroform for 30 sec and centrifuged at 3,000 rpm in a low speed centrifuge (IEC model PR6) for 15 min at 4°C. The upper aqueous layer containing the phage was removed and 0.5 g solid CsCl was added for each millilitre of phage suspension. This solution was then layered on a step gradient prepared in a Beckman SW 41 tube by layering 2.5 ml CsCl solutions prepared in SM medium with the final densities of ρ 1.45, ρ 1.5 and ρ 1.7. Following centrifugation in a Beckman ultracentrifuge (model L5-65) at 22,000 rpm for 2 h at 4°C, the band containing the bacteriophage particles visible at the interface of 1.45 and 1.5 layers was collected by puncturing the side of the centrifuge tube with a 21-gauge needle connected to a 2 ml syringe. Enough CsCl solution (1.5 g/ml) was added to the phage suspension to fill a SW 50.1 rotor tube which was then centrifuged at 35,000 rpm for 24 h at 4°C. The band containing the phage particles was collected with the help of a needle and syringe and the DNA isolated as described below.

2.14.3. Extraction of λ DNA

The CsCl was removed from the purified bacteriophage solution by dialysing for about 1-2 h against a 1000 fold volume of dialysis buffer (50 mM Tris-HCl pH 8.0, 10 mM NaCl and 10 mM $MgCl_2$). Dialysis was repeated once more with fresh buffer and the phage solution was transferred to 1.5 ml tubes (0.5 ml/tube). This was followed by addition of EDTA, proteinase K solutions and SDS to give final concentrations of 20 mM, 50 μg/ml, and 0.5%, respectively, and the tubes were incubated at 65°C for 1 h for protease digestion. The digest was extracted once with phenol equilibrated in TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) then with phenol-chloroform (1:1). Finally, the aqueous layer was extracted with chloroform only. The phage DNA was precipitated on ice by addition of 0.1 volume of 3 M sodium acetate and 2.0 volumes of ethanol. The precipitated DNA was collected either by centrifugation in an Eppendorf centrifuge or by spooling out the large DNA fibres on the closed end of a sterile pasteur pipette. The DNA was then washed in 70% ethanol dried under vacuum, dissolved in water and the concentration determined by measuring OD_{260} as described in section 2.7. The DNA solution was stored at 4°C.

2.15. Sub-cloning

2.15.1. Competent cell preparation

Competent cells were prepared by a modification of the Hannahan procedure (1988) and the following reagents were prepared as described next.

RF1

RbCl	100 mM
$MnCl_2.4H_2O$	50 mM
Potassium acetate	30 mM
$CaCl_2.2H_2O$	10 mM
Glycerol	15% (w/v)

The resulting solution was adjusted to pH 5.8 with 0.2 M acetic acid and sterilized by filtration through a 0.22 μm membrane and stored frozen at -20°C till just before use.

RF2

Morpholinopropanesulfonic acid pH 7.5	30 mM
RbCl	10 mM
$CaCl_2.2H_2O$	75 mM
Glycerol	15% (w/v)

Adjusted to pH 6.8 with 10 M NaOH and sterilized by filtration through a 0.22 μm membrane and stored frozen at -20°C till just before use.

S.O.C medium

Bacto-tryptone	2% (w/v)
Bacto-yeast extract	0.5% (w/v)
NaCl	10 mM
KCl	2.5 mM

$MgCl_2$-$MgSO_4$	20 mM (10 mM of each salt)
glucose	20 mM

All reagents except glucose and $MgCl_2$-$MgSO_4$ were dissolved in the appropriate amount of water and the solution was adjusted to pH 7.4 by addition of 10 M NaOH and autoclaved. If necessary 1.5% agar was added before autoclaving. Just before use glucose and $MgCl_2$-$MgSO_4$ solutions sterilized by filtration were added from 20%, w/v and 1M stock solution, respectively.

A freshly isolated single colony obtained by streaking a stock culture either of DH5α or XL1-Blue cell directly on S.O.C medium plates was inoculated into 10 ml S.O.C medium and incubated at 37°C. The bacteria were allowed to grow overnight at 37°C with shaking after which 5 ml culture was added to a 2 L flask containing 100 ml S.O.C medium and allowed to grow at 37°C with shaking until the OD_{600} of the culture reached about 0.5. The culture medium was then chilled on ice and centrifuged at 5,000 rpm at 4°C for 15 min in a Beckman JA-14 rotor. The cell pellet was resuspended in 40 ml pre-chilled RF1 and kept on ice for 15 min. The cell suspension was again centrifuged and the pellet obtained resuspended in 8 ml RF2. The resulting cell suspension was then added in 100-200 μl/1.5 ml microfuge tube for storage at -70°C. Just before use for transformation these cells were taken out of the freezer and placed on ice till just thawed. An aliquot (about 100 μl) was then added directly into the tubes containing the DNA sample.

2.15.2. Preparation of plasmid vectors for cloning

Purified plasmid DNA (section 2.16.1) was digested with an appropriate restriction endonuclease as described in section 2.8 except that 5 μg DNA was used for digestion. The completion of digestion was examined by analyzing the digest by agarose gel electrophoresis (section 2.9). The digested DNA was phenol extracted to destroy the restriction endonuclease and ethanol precipitated (section 2.5 and 2.6) before dissolving in water. The DNA was then treated to remove the 5´ terminal phosphate groups from the digested vector DNA with 11 units of calf intestinal alkaline phosphatase for 30 min at 37°C in 50 mM Tris-HCl buffer pH 9.0 containing 1 mM $MgCl_2$ and spermidine and 0.1 mM $ZnCl_2$. The reaction mixture was heated at 65°C for 15 min to inactivate the phosphatase and phenol extracted (section 2.5) before ethanol precipitation (section 2.6). The precipitates were dissolved in water (50 ng/μl) and stored at -20°C till use.

2.15.3. Ligation

Ligation buffer (5 X)

Tris-HCl pH 7.4	0.5 M
$MgCl_2$	0.1 M
DDT	0.1 M
Spermidine	10 mM
ATP	10 mM
BSA	1 mg/ml

The following reagents were mixed in a 0.8 ml microfuge tube

DNA restriction fragment	3 μl (about 0.5 μg)
Plasmid DNA (Bluescript or pUC 19)	1 μl (about 50 ng)
ligation buffer	4 μl
T4 DNA ligase	0.5 μl (4 units/μl)
Final volume	10-20 μl

The tubes were incubated at 15°C for about 18 h or at room temperature for 4 h. and the ligated DNA was diluted 1-5 fold with TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) depending upon the initial concentration of DNA in the ligation reaction and used either immediately for transformation or stored at -20°C.

2.15.4. Transformation

Transformation of competent bacterial cells was performed using the following protocol modified from that of Hannahan (1988). The DNA in 5-10 μl water or TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) were gently swirled for 5 sec with about 100 μl competent cells in RF2 (section 2.15.1) in a Falcon tube (2059). A control plasmid DNA (pUC 19 or Bluescript) was also included to monitor the efficiency of transformation. The mixture was kept on ice for a further 30 min then given a heat shock for exactly 45 sec by transferring the tubes into a 42°C water bath. The tubes were immediately placed on ice for 5 min, 0.9 ml S.O.C medium that was kept at

room temperature added and the tubes incubated at 37°C with gentle shaking for a further 1 h. The transformation mixture was plated on S.O.C agar plates (section 2.15.1) with appropriate antibiotic(s) either directly by spreading on the surface or by mixing with 2.5 or 7.5 ml top agarose containing Bluo-gal (50 μg/2.5 ml or 150 μg/ 7.5 ml top agarose/agar) and IPTG (10 mM) (section 2.4.4). The plates were incubated at 37°C for about 18 h to allow for the growth of the transformants. The number of transformants was determined by counting the colonies which were then screened for the presence of fragments of interest as described in the following section.

2.15.5. Screening the transformants

A few white colonies were transferred using a sterile tooth pick to the surface of a plate containing Bluo-gal and IPTG with appropriate antibiotic. A few of the white colonies which grew after about 18 h at 37°C were then transferred to about 5 ml LB medium with 0.2% (w/v) glucose containing appropriate antibiotic (section 2.4.1) and incubated at 37°C for overnight on a shaker. The plasmid DNA was isolated by a quick plasmid preparation method (section 2.16.2) and subjected to restriction followed by gel analysis (sections 2.8 and 2.9) to identify clones containing fragments of the desired size. Whenever identification of the appropriate clone by simple restriction analysis was difficult, the digested DNA was blotted on nylon membrane and the clones of interest then identified by hybridization with ^{32}P-labelled oligonucleotides (section 2.11.4) or ^{32}P-labelled cDNA (section 2.12.4). Following the

identification of a particular clone, a stock culture was made by growth in LB medium and stored at -70°C until required at which time a sample was grown as described in section 2.4.

2.16. Isolation of plasmid DNA

Plasmid DNA was isolated by an alkaline lysis method (section 2.16.1) (Birnboim and Doly, 1979) using the protocol given in *Molecular Cloning* (Maniatis *et al.*, 1982). For quick screening of transformants, following the subcloning or exonuclease III deletion, a mini plasmid DNA preparation (Budowle, 1985) (section 2.16.2) was also used.

2.16.1. Large scale isolation of plasmid DNA

The following reagent were prepared freshly from stock solutions:

Solution I

Glucose	50 mM
Tris-HCl pH 8.0	25 mM
EDTA pH 8.0	10 mM
Lysozyme	5 mg/ml

Solution II

NaOH	0.2 M
SDS	1% (w/v)

Potassium acetate solution pH 4.8

To 60 ml of 5 M potassium acetate stock solution 11.5 ml glacial acetic acid was added and the volume was made up to 100 ml. The resulting solution was 3 M with respect to potassium and 5 M with respect to acetate (Maniatis *et al.*, 1982).

About 10 ml LB medium (section 2.4.1) with appropriate antibiotic(s) was inoculated with a single isolated bacterial colony that was grown on a selective medium and the tube was incubated at 37°C with shaking for about 18 h. About 0.5 ml of the overnight culture was added to a 2 L flask containing 500 ml LB medium and incubated at 37°C with vigorous shaking until the OD_{600} of the culture reached to about 0.6. The culture medium was centrifuged at 4,000 rpm for 10 min at 4°C using a Beckman JA-14 rotor to collect the bacterial pellet which was then resuspended in about 50 ml STE (10 mM Tris-HCl pH 8.0, 1 mM EDTA and 100 mM NaCl) and transferred to a polypropylene tube and centrifuged at about 3,000 rpm in a bench top centrifuge for 5 min. The pellet was resuspended in 7.5 ml solution I by mixing gently and then transferred to a SW 28.1 tube (pollyallomer) and left at room temperature for 5 min. To this solution 15 ml solution II was added, the tube was covered with parafilm, mixed by inversion and left on ice for 10 min. About 11 ml of potassium acetate solution was added to the tube, covered with parafilm, mixed by inverting the tube sharply several times and then left on ice for 10 min. The precipitated high molecular weight DNA was then separated from the plasmid solution by centrifugation at 22,000 rpm at 4°C for 20 min using Beckman SW 28.1

rotor. The supernatant was collected in two Corex tubes (about 17 ml/tube), 10.2 ml isopropanol was added per tube, mixed well and incubated at room temperature for 15 min. The precipitated plasmid DNA was collected by centrifugation in a Beckman JA 14 rotor at 10,000 rpm, for 30 min at room temperature, resuspended in 70% ethanol and recentrifuged at 10,000 rpm for 10 min. The washed pellet was dried under vacuum for about 30 min then dissolved in 9 ml TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) and purified further to isolate closed circular plasmid DNA.

One gram of solid CsCl was added for every millilitre of DNA solution and mixed gently to completely dissolve the CsCl. To this solution, 0.8 ml ethidium bromide (10 mg/ml) for every 10 ml of the solution was added and mixed thoroughly. The density of the resulting solution (determined by weighing a sample) was adjusted to 1.55 g/ml by dissolving more CsCl if necessary. This solution together with the furry, purple aggregates due to the complexes formed between ethidium bromide and bacterial proteins was transferred to Beckman 75 Ti rotor cellulose nitrate tubes. The tubes were topped up with light parafilm oil and centrifuged at 45,000 rpm for 36 h at 20°C. After the centrifugation, two bands were visible, the upper band consisting of linear chromosomal DNA and nicked circular plasmid DNA, and the lower band containing closed circular plasmid DNA. The lower band was collected by puncturing the side of the centrifuge tube with a 21-gauge needle attached to a 2 ml syringe. The ethidium bromide was removed by vigorous shaking of equal volumes of the plasmid solution and n-butanol followed by brief centrifugation in a bench top centrifuge at

about 2,000 rpm for 3 min. The upper n-butanol layer was discarded and the lower aqueous layer transferred to a clean tube and the extraction was repeated until the colour due to ethidium bromide disappeared completely. The resulting aqueous solution was dialysed against several changes of TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) and the DNA collected by ethanol precipitation (section 2.6).

2.16.2. Mini preparation of plasmid DNA

The plasmid DNA was also isolated by the rapid procedure of Budowle (1985) which is based upon the alkaline lysis method of Birnboim and Doly (1979) described in section 2.16.1. An isolated colony with the appropriate plasmid was added into 5 ml of LB medium with relevant antibiotic and incubated at 37°C with vigorous shaking for about 18 h. About 1.5 ml overnight culture was centrifuged for 30 sec in a microfuge tube and the pellet obtained resuspended in 0.1 ml solution I (section 2.16.1) by vortexing for a few seconds. After keeping on ice for 30 min 0.2 ml alkaline-SDS solution (solution II, section 2.16.1) was added and the contents of the tube mixed thoroughly by vortexing the tube which was then left on ice for 5 min and 0.15 ml 3 M sodium acetate solution pH 4.8 was added. The contents of the tube were mixed as before and the tube was incubated on ice for 30 min with occasional shaking to allow for precipitation to occur. The solution was centrifuged in an Eppendorf centrifuge for 15 min at 4°C and the supernatant was collected in a clean microfuge tube. Two volumes of absolute ethanol was added, the contents of the tube mixed well by inverting several times and precipitation of plasmid DNA was allowed

to occur by incubation of the tube at -70°C for 10 min. The precipitated plasmid DNA was collected by centrifugation in an Eppendorf centrifuge for 15 min at 4°C. The pellet containing the plasmid DNA was washed twice in 70% ethanol, dried under vacuum, dissolved in 20-30 μl of water or TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) and stored at 4°C.

2.17. Southern blotting

2.17.1. DNA blotting

The DNAs with or without restriction endonuclease digestion were fractionated on 0.6-1.5% (w/v) agarose gels (section 2.9) and transferred to nylon membranes by the techniques of Southern (1975) and Amersham manual, *Membrane transfer and detection methods* with some modifications. Following electrophoresis and photography of the gel it was transferred to a glass dish and the DNA was denatured by shaking the gel in the denaturation solution (0.5 M NaOH, 1.5 M NaCl) for 15 min. Denaturation was repeated once more with a fresh solution and the gel neutralized by immersing in neutralizing solution (1.5 M Tris-HCl pH 8.0, 1.5 M NaCl) and shaking for 15 min. Neutralization was repeated once more followed by a brief rinsing in 20 X SSC.

The transfer of the DNA was allowed to take place for about 18 h. After the transfer, the position of the wells was marked on the nylon membrane, the gel was restained briefly in 1 μg/ml ethidium bromide to check for completion of the transfer.

The membrane, now with the transferred DNA was baked in a vacuum oven at 80°C for 2 h then stored in a sealed plastic bag till required for hybridization as described in the sections 2.11.4 or 2.12.4.

2.17.2. Probe stripping and reuse of the DNA blots

If the Southern blots were to be reused, the first probe was removed by immersing the membranes in 0.4 M NaOH that was maintained at 45°C. The incubation at 45°C was continued with agitation for 30-45 min. This was then followed by neutralization in a solution containing 0.2 M Tris-HCl pH 8.0, 0.1 X SSC and 0.1% SDS (w/v) at 45°C for 30-45 min. The excess solution was removed from the membranes by placing between two layers of Whatman 3 MM paper. The membranes were then exposed to Kodak X-Omat AR or RP X ray film for about 18 h to check for the complete removal of the previous hybridization signal.

2.18. RNA methods

2.18.1. Preparation of glassware and solutions for RNA work

Glassware to be used for RNA work was sterilized by autoclaving at 121°C at 15 lb/square inch, followed by baking for 30 min at 180°C, or for more than 6 h at 121°C. The glass centrifuge tubes used for RNA extraction were siliconized by immersing in dimethyldichlorosilane (2% in 1,1,1,trichloroethane). The siliconized tubes were then washed thoroughly with water followed by autoclaving, and baking as described above. The polyallomer centrifuged tubes were treated with 0.2% DEP

solution (in autoclaved water) followed by autoclaving. The reagents to be used for RNA work were kept separate from the rest of the reagents and dispensed directly from the original containers into the glassware specially treated for RNA work without the use of a spatula. The solutions were routinely made in autoclaved water, usually treated with 0.2% DEP for 12-18 h and autoclaved. The non-autoclavable solutions were made in autoclaved water and sterilized by passage through 0.22 μm cellulose nitrate membranes.

2.18.2. Isolation of RNA

Total cytoplasmic RNA was isolated by an SDS-phenol method (Hiatt, 1962, Taylor, 1979) as described by Gahmberg *et al.* (1979) followed by passage through an oligo(dT)-cellulose column (Aviv and Leder, 1972) to separate poly(A)$^+$ RNA. A quick method (Badley *et al.*, 1988) was also employed. These methods are described in the following sections.

2.18.2.1. SDS-phenol method of RNA extraction

K562 cells were grown in RPMI 1640 medium as described in section 2.3. The cells were collected by centrifugation at 2,000 rpm for 5 min in a bench top centrifuge and washed by resuspension in PBS (0.01 M phosphate buffer pH 7.4 containing 0.15 M NaCl and 0.1 mg/ml cycloheximide) and centrifuging as before. The cells were resuspended in a volume of lysis solution A (0.01 M Tris-HCl pH 7.4, 0.01 M NaCl, 15 mM $MgCl_2$ and 30 μg/ml Polyvinyl sulphate) that was about 10 times the volume

of cells and the tube was incubated on ice for 20-30 min to allow for cell swelling. The swollen cells were broken by 20-25 strokes in a Dounce homogenizer. The broken cell suspension was centrifuged at 3,000 rpm at 4°C for 10 min in a Beckman JS-13 rotor to collect the nuclei which were discarded. The supernatant containing the total cytoplasmic RNA was collected in a fresh tube and SDS was added to a final concentration of 2% and the resulting solution was heated in a water bath at 65°C for 5 min. An equal volume of phenol solution maintained at 65°C was added and the procedure of phenol extraction (section 2.5) followed by ethanol precipitation (section 2.6) were employed.

The precipitated RNA was collected by centrifugation at 10,000 rpm for 30 min at 4°C using a Beckman JS-13 rotor. The pellet was resuspended in 70% ethanol followed by centrifugation as before. Washing the RNA pellet with ethanol was repeated twice. The washed pellet was dried under vacuum, dissolved in DEP-treated water, subjected to spectrophotometric measurement and one more round of ethanol precipitation. The resulting RNA solution was then stored at -70°C till oligo(dT)-cellulose chromatography was performed as described in the next section.

2.18.2.2. Oligo(dT)-cellulose chromatography

About 0.2-0.5 g oligo(dT)-cellulose (type 3) was resuspended in DEP-treated water, poured into a small column (10 ml, Econo column, BioRad) and the oligo(dT) activated by passing about 10 column volumes of 0.1 N NaOH through the column.

This was followed by washing the column with 10 or more column volumes of water until the pH of the effluent dropped to neutrality. The column was then equilibrated by passing through about 10 column volumes of binding buffer (25 mM Tris-HCl pH 7.4, 0.5 M NaCl containing 0.5% SDS (w/v). The RNA in binding buffer was heated at 65°C for 5 min and cooled to the room temperature. This denatured RNA was then passed through the equilibrated oligo(dT)-cellulose column. Unbound material from the first passage of RNA through the column was applied once more to achieve maximum binding of the poly(A)$^+$ RNA. The column was then washed with binding buffer until the OD_{260} of the buffer reached back ground. The RNA was then eluted in 3 ml DEP-treated water. The eluted RNA solution was denatured, readjusted quickly to the binding conditions as before (now without SDS) and allowed to pass through the column which was now equilibrated in the binding buffer without SDS. The poly(A)$^+$ RNA was then eluted in 3 ml DEP-treated water and subject to ethanol precipitation (section 2.6). The precipitate was collected by centrifugation, dried under vacuum and dissolved in DEP-treated water and stored at -70°C until use.

2.18.2.3. Quick RNA extraction method

The SDS-phenol method is a very lengthy procedure requiring multiple steps including phenol extraction and two ethanol precipitations which may result in losses of the RNA. Therefore in later stages of this work a simpler method (Badley *et al.*,

1988) requiring no phenol extraction and a single ethanol precipitation was employed which is described now.

K562 cell were grown in RPMI 1640 medium as described in section 2.3 and collected by centrifugation at 2,000 rpm for 10 min in a bench top centrifuge. The cells were washed by resuspension in serum-free RPMI 1640 medium and centrifuged as before. The washing procedure was repeated twice more followed by suspension of the cells in about 10 ml of the lysis solution B [0.2 M NaCl, 0.2 M Tris-HCl pH 7.5, 1.5 mM $MgCl_2$, 2% SDS (w/v), 200 μg/ml Proteinase K] at a concentration of about 10^9 cell/ml. The cells were homogenized with a polytron homogenizer (Ultra Turrax, Janke Knukel) at medium speed for 10-15 sec and the resulting cell lysate was incubated for 1 h in a water bath maintained at 50°C.

About 0.4 g oligo(dT)-cellulose was prepared as follows. In a 50 ml polypropylene centrifuge tube the dried oligo(dT)-cellulose was resuspended in DEP-treated water. The resuspended cellulose was collected at the bottom of the tube by centrifugation at about 500-1,000 rpm in a bench top centrifuge for about 20 sec, the supernatant discarded and the pellet was resuspended in the binding buffer (10 mM Tris-HCl pH 7.5, 0.5 M NaCl). The cellulose was allowed to equilibrate in the buffer for a few minutes followed by centrifugation as described above. A few millilitres of the binding buffer was added to the pellet until incubation of the lysate at 50°C was complete. The salt concentration of the lysate was adjusted to 0.5 M by the addition

of 5 M NaCl. The cell lysate was then mixed with the equilibrated oligo(dT)-cellulose. The poly(A)$^+$ RNA was allowed to bind to the cellulose by agitating the tube on a rotator for about 15 min at room temperature followed by incubation for 5 min without agitation. This was followed by centrifugation as before to obtain the oligo(dT)-cellulose pellet now with the bound poly(A)$^+$ RNA. The supernatant was discarded and the pellet was washed in large volumes of binding buffer, centrifuging between each wash until the supernatant was no longer cloudy. The oligo(dT)-cellulose was then resuspended in about 10 ml fresh binding buffer and poured into a small column and washing continued until the OD_{260} of the effluent was less then 0.05. The column was then allowed to run dry and the poly(A)$^+$ RNA was eluted using several 0.5 ml fractions of DEP-treated water. The OD of an aliquot from each fraction was measured at 260 nm. The fractions containing poly(A)$^+$ RNA were pooled and subjected to ethanol precipitation (section 2.6) and the poly(A)$^+$ RNA was dissolved in DEP-treated water and stored at -70°C till use.

2.18.3. Northern transfer

Poly(A)$^+$ RNA was fractionated according to size on a 1% (w/v) agarose gel containing 2.2 M formaldehyde (Lehrich *et al.*, 1977, Goldberg, 1980) with some modifications (Maniatis 1982) and the instructions provided in the Amersham manual, *Membrane transfer and detection methods* as described next.

The agarose gel electrophoresis was performed as described in section 2.9 except that the agarose solution and the buffers were as described here. About 1 g agarose was suspended in about 50 ml of DEP-treated water, autoclaved and allowed to cool to a temperature of about 50°C. To this agarose solution reagents were added so that the final concentrations were as follows: 0.04 M morpholinopropanesulfonic acid (MOPS), 0.01 M sodium acetate pH 7.0, 1 mM EDTA pH 7.5, 2.2 M formaldehyde and 0.2 μg/ml ethidium bromide.

To prepare the RNA sample for electrophoresis, 10 μg RNA dissolved in 5 μl DEP-treated water was mixed with 2 μl 10 X MOPS buffer (0.4 M MOPS, 0.1 M sodium acetate pH 7.0, 10 mM EDTA pH 7.5), 3.5 μl formaldehyde, 10 μl formamide which was previously deionized by shaking with a mixed bed resin, AG 501-X8. The RNA was denatured by incubating at 65°C for 15 min and the tubes were chilled quickly by placing on ice and 2 μl gel loading dye, containing 50% glycerol, 0.25% bromophenol blue and xylene cyanol and 1mM EDTA pH 8.0, was added to each tube. The RNA was electrophoresed at a constant voltage (5 V/cm) in 1 X MOPS buffer for about 4 h. Following the electrophoresis, the RNA was visualized by UV transillumination and photographed (section 2.9). The gel was rinsed briefly in the transfer buffer (20 X SSPE, section 2.11.4) and blotted onto nylon membrane for about 18 h as described for Southern blotting (section 2.17.1) but without any further treatment of the gel. Transfer of the gel was monitored by restaining the gel in ethidium bromide followed by visualization under UV light. The membrane with the

transferred RNA was air dried and baked at 80°C for 2 h to fix the RNA on the membrane. These blots were then stored in a sealed plastic bag until required for hybridization.

2.18.4. Removal of the probes from Northern blots

Whenever a Northern blot was to be reused, the first probe was removed by following the instructions of the Amersham manual *Membrane Transfer and Detection Methods.* A solution containing 5 mM Tris-HCl pH 8.0, 2 mM EDTA and 0.1 X Denhardt's solution (section 2.11.4) was heated up to about 80°C and poured in a glass tray. The membrane was immersed in this solution and the tray incubated in a shaking water bath maintained at 65°C. Incubation was continued for 2 h. The blot was removed from the tray, blotted briefly on Whatman 3 MM paper, wrapped in plastic film and exposed to the Kodak X-Omat AR or RP X ray film for about 18 h to check for the complete removal of the probe.

2.18.5. Primer extension analysis

Either 50-100 μg of total RNA or 10 μg poly(A)$^+$ RNA from K562 cells was mixed with 0.5-1 p mole ^{32}P-labelled GPA-N2 in annealing buffer (10 mM Tris-HCl pH 8.3) and the volume was made up to 25 μl. The RNA was then denatured by heating at 80°C for 3 min and cooling on ice for 5 min, followed by annealing of the primer at 42°C for 1 h. This annealed RNA-primer mixture was diluted 1:1.5 with reverse transcriptase buffer and 7 U (1 μl) of reverse transcriptase were added. The

reaction was allowed to take place by incubation of the tube at 42°C for one h. The reaction mixture was then treated with 20 μg DNase-free RNase per tube at 37°C for 20 min. The primer extended products were then purified by extraction with phenol solution (section 2.5) followed by ethanol precipitation (section 2.6). The pellet was dissolved in water and analyzed by electrophoresis on a sequencing gel containing 5% polyacrylamide and 7 M urea as described in the section 2.19.3.3, followed by autoradiography using Kodak X-Omat AR film with two intensifying screens at -70°C.

2.19. Nucleotide sequencing

The nucleotide sequence of the cDNAs was determined by Sanger's dideoxy nucleotide sequencing method (Sanger *et al.*, 1972). *Eco*RI generated fragments were subcloned into the plasmid vector pUC19 and nucleotide sequenced using double-stranded plasmid DNA. The *Eco*RI generated fragments were also cloned into the plasmid vectors Bluescript-SK or Bluescript-KS and their nucleotide sequences were determined using single- or double-stranded plasmid DNA with or without exonuclease III-mung bean nuclease deletion. The various oligonucleotides used for nucleotide sequencing included the M13 universal primer, M13 reverse sequencing primer, SK and KS primers (for the Bluescript vectors only), and GPA-N2 and GPA-C2 an 18 mer of composition based upon the coding sequence of glycophorin A. The sequencing reactions were performed using the instructions provided in the manual for the Sequenase kit. These different methods will now be described in the following sections (2.19.1 to 2.19.3).

2.19.1. Exonuclease III-mung bean nuclease deletion

The cDNA fragments to be sequenced were subcloned into the *Eco*RI site of the Bluescript or pUC 19 by the procedures described in section 2.15. The cDNA fragments subcloned into the Bluescript vectors were also subjected to the exonuclease III-mung bean nuclease deletion procedure using the exonuclease III deletion kit to produce a nested set of deletions in the cDNA fragments. The technique takes the advantage of the requirement of exonuclease III for double-stranded DNA; exonuclease III does not digest 3´ single-stranded overhangs, but digests 3´ ends from blunt ends or 5´ overhangs. The polylinker in Bluescript vectors places unique 3´ restriction sites on the outside edge of the polylinker and 5´ and blunt restriction sites internally. To create deletions in the insert but not in the vector the plasmid DNA is digested with two restriction enzymes, one producing a 3´ and other a 5´ overhang or blunt end. The double digestion is then followed by exonuclease III-mung bean nuclease treatment. Taking advantage of the convenient location of restriction sites and predictable progression of the exonuclease III reaction a nested set of deletions can be produced easily (*Bluescript, Exo/Mung DNA Sequencing System, Instruction manual*). The details of the deletion procedure is described next (sections 2.19.1.1 to 2.19.1.3).

2.19.1.1. Digestion with two restriction enzymes

The plasmid DNA to be subjected to the deletion procedure was isolated by the large scale method (section 2.16.1). About 25 μg of this DNA was then digested

sequentially with appropriate restriction endonuclease producing a 3´ overhang and another producing a 5´ overhang. Completion of each digestion was monitored by electrophoresing an aliquot on a 0.8% agarose gel. After each restriction endonuclease digestion, the DNA was purified by phenol extraction (section 2.5) and ethanol precipitation (section 2.6). The double-digested DNA was then dissolved in water and treated with exonuclease III as described below using the reagents provided in the exonuclease deletion kit.

2.19.1.2. Exonuclease III treatment

To produce a nested set of deletions of varying length simultaneously, a single exonuclease III reaction was set up from which aliquots were withdrawn at different time points. In a microfuge tube the following reagents were placed. The volume of each reagent was multiplied by the number of samples to be withdrawn at different time points.

Digested DNA from section 2.19.1.1	5 μg
Exonuclease III buffer (2 X)	12 μl
2 Mercaptoethanol (100 mM)	2 μl
Exonuclease III (80 units)	1 μl
Water to a final volume	25 μl

Before initiating the exonuclease III reaction, a separate stop solution was prepared for each time point by diluting 40 μl mung bean nuclease buffer (5 X, provided in the

exonuclease deletion kit) in 135 μl of water, and the tubes were placed on ice. This buffer solution in addition to terminating the exonuclease III reaction at various time points, served as a buffer for mung bean reaction in the next step. To begin the exonuclease III treatment the tube was placed in a water bath maintained at 37°C and the reaction was initiated by addition of the exonuclease III. Samples of about 25 μl were then withdrawn at the time intervals of 30-60 sec (for about 5 min) and immediately mixed with the diluted mung bean buffer and the mixture flash frozen by submerging the sealed tube in liquid nitrogen. When all the samples were withdrawn, the tubes were heated at 65°C for 15 min, placed on ice and treated with mung bean nuclease.

2.19.1.3. Mung bean nuclease treatment

Each exonuclease III-digest withdrawn at different time intervals was mixed with 1 μl (15 units) of mung bean nuclease and incubated at 30°C for 30 min after which nuclease was inactivated by adding the following reagents.

SDS (20%, w/v)	4 μl
Tris-HCl pH 9.5 (1 M)	10 μl
LiCl (8 M)	20 μl
Phenol-chloroform solution (1:1)	250 μl

The reagents were mixed thoroughly and the aqueous and organic phases were separated by centrifugation in an Eppendorf centrifuge at room temperature for 5

min. The aqueous layer was transferred to a clean tube and extracted with an equal volume of chloroform followed by centrifugation to separate the organic and aqueous layers. The aqueous layer was again transferred to a clean tube, yeast tRNA was added to give a final concentration of 10 ng/μl, and the nucleic acids were precipitated by the adding 0.5 ml cold ethanol and storage at -20°C for about 18 h. The precipitates were collected by centrifugation in an Eppendorf centrifuge, and the pellet dried under vacuum then dissolved in water. An aliquot from each time point was analyzed on a 0.8% (w/v) agarose gel (section 2.9) to determine the average size of the exonuclease 3-mung bean nuclease treated DNA. The fragment of the most suitable size for nucleotide sequencing was re-ligated (section 2.15.3) and used to transform competent *E.coli* XL1-Blue cells (section 2.15.4). Clones with an appropriate deletion were then selected by the procedures described previously (section 2.15.5) and nucleotide sequenced using either the double-stranded plasmid DNA isolated by a quick method (section 2.16.2) or a large scale method (section 2.16.1). In some cases single-stranded DNA isolated by the procedure described below (section 2.19.2) was also used.

2.19.2. Isolation and purification of single-stranded plasmid DNA

The single-stranded DNA was isolated and purified by a polyethylene glycol method (Dente *et al.*, 1983) described in *Bluescript Exo/Mung DNA sequencing system instruction manual* as follows. An isolated colony from LB-agar plate containing ampicillin and tetracycline, was added into double strength YT medium (1.6% Bacto-

tryptone, 1% Bacto-yeast extract and NaCl) and incubated at 37°C for about 18 h. About 300 μl of the overnight culture (7.5 x 10^8 cells) was added into a 50 ml tube containing 3 ml of 2 X YT medium and allowed to grow for about 2 h at 37°C with shaking until the OD_{600} of the culture reached about 0.3/ml (about 2.5 x 10^8 cells/ml). At this stage the helper phage, an M13 derivative, R408 or VCSM13 (provided in the exonuclease deletion kit) was added at a multiplicity of infection of about 20:1 (phage:cells) and the mixture was once again shaken at 37°C for about 8-12 h. The bacterial cells and debris were removed by centrifugation in an Eppendorf centrifuge for about 30 min and the supernatant was transferred to a clean tube. To the clear supernatant 0.25 volume of precipitating solution (20%, w/v PEG 6000, 3.5 M ammonium acetate pH 7.5) was added and the precipitate was allowed to form at 0°C for 15 min. The precipitate was then collected by centrifugation for 15 minutes in an Eppendorf centrifuge. The tubes were drained thoroughly and the pellet was resuspended in about 200 μl water. The single-stranded plasmid DNA solution thus obtained was vortexed for 1 min with 200 μl phenol-chloroform (1:1). The organic and the aqueous phases were separated by centrifugation for 5 min in an Eppendorf centrifuge at the room temperature. Phenol-chloroform extraction was repeated a few more times (usually 3-5) until a clear interface was obtained. The resultant solution was extracted once with chloroform only. The single-stranded DNA was then precipitated by the adding 150 μl of 7.5 M ammonium acetate pH 7.5 and 600 μl chilled absolute ethanol and keeping on ice for 15 min. The precipitated single-stranded DNA was collected by centrifugation in an Eppendorf centrifuge at 4°C for 20 min. The resultant pellet was dried under vacuum for about 15 min and dissolved

in water (25 μl) and stored at 4°C. The DNA was examined by agarose gel electrophoresis (section 2.9).

2.19.3. Nucleotide sequencing of double-stranded or single-stranded plasmid DNA

Single-stranded plasmid DNA was used without any further treatment for nucleotide sequencing, whereas the double-stranded plasmid DNA was subjected to denaturation by treatment with NaOH (Hattori and Sakaki, 1986) as describe below.

2.19.3.1. Denaturation of the double-stranded plasmid DNA

About 2-3 μg of double-stranded plasmid DNA (10 μl) was mixed with an equal volume of 0.4 M NaOH and incubated at room temperature for 5 min. The denatured DNA was precipitated by the addition of 0.8 volume 5 M ammonium acetate pH 7.5 and 4 volumes pre-chilled absolute ethanol and transferring to a -70°C freezer for 10 min. The precipitate was collected by centrifugation in an Eppendorf centrifuge for 15 min at 4°C, dried under vacuum and dissolved in 7 μl water.

2.19.3.2. Sequencing reaction

The sequencing reactions using either single-stranded or double-stranded plasmid DNA were performed using the Sequenase kit containing the modified T7 polymerase following the instructions and with the reagents provided by the manufacturer. The reaction was divided in three major steps, annealing, labelling and termination.

2.19.3.2.1. Annealing

About 7 μl template plasmid DNA, 2 μl sequenase buffer and 1 μl primer solution (0.5-1 pmol) were mixed together, heated at 65°C for 3 min and allowed to cool slowly (about 30 min) till the temperature of the water bath (a 250 ml beaker) decreased to about 30-35°C.

2.19.3.2.2. Labelling reaction

The annealed DNA solution was centrifuged briefly in an Eppendorf centrifuge to bring the evaporated material to the bottom of the tube and the following reagents were added in the in the given order.

Dithiothreitol (DTT, 0.1 M)	1 μl
Labelling mix	2 μl
[^{35}S] dATP	1 μl (10 μ Ci)
Sequenase (diluted 1:7)	2 μl

After each addition, the content of the tubes were mixed by pipetting back and forth with the help of a Pipette Man. The reaction was allowed to take place by incubation at room temperature for 5 min.

2.19.3.2.3. Termination reaction

Four tubes labelled C, T, A, and G were prepared containing 2.5 μl of the appropriate termination mixture and warmed to 37°C while the labelling reaction was taking place. Following the 5 min incubation of the labelling reaction, 3.5 μl was transferred to each of the four termination reaction tubes (labelled C, T, A, and G),

mixed quickly and the incubation continued for a further 5 min at 37°C to allow for the termination reaction to take place. The reactions were stopped by adding 4 μl stop solution provided by the supplier of the kit. The tubes were heated at 75°C for 3 min, chilled on ice and analyzed on a sequencing gel immediately or stored at -20°C till used later.

2.19.3.3. Sequencing gel

Sequencing gels 0.35 mm x 380 mm x 430 mm containing 6% polyacrylamide and 7 M urea prepared in Tris-borate-EDTA buffer (1M Tris and boric Acid, 0.4 M EDTA pH 8.3) were pre-run for about 30 min at a constant power of about 90 watts in a Hoefer Scientific Instrument SE 1500 sequencing apparatus with Tris-borate-EDTA as the running buffer. Samples for analysis were denatured by heating at 75°C then chilled on ice immediately before transferring about 4 μl of each to wells formed using a shark's tooth comb. Electrophoresis was usually performed until the bromophenol blue dye had reached the bottom of the gel (about 2.5 hours). To read more than 150 nucleotides, multiple loading of the same sequencing reaction were performed. After the electrophoresis was complete each gel was fixed in 1 L of a 10% acetic acid-10% methanol solution for about 30 min to leach out the urea from the gel, dried at 90°C under vacuum using a gel dryer (Biorad-model 583) for about 1 h then exposed to Kodak X-Omat AR film with two intensifying screen for 12-18 h at -70°C. The nucleotide sequences were read manually and analyzed using Microgenie from Beckman or PC/GENE from IntelliGenetics sequencing programmes or both.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Isolation of cDNA clones encoding glycophorin A

The primary structure of glycophorin A is very similar to that of glycophorin B in many regions (Blanchard *et al.*, 1987; Siebert and Fukuda, 1987). The NH_2-terminus of glycophorin A^N (positions 1 to 26 from the NH_2-terminus of glycophorin A) has an amino acid sequence which is identical to the NH_2-terminal sequence of glycophorin B (glycophorin A numbering is used in the present report). In addition, both have highly homologous regions between amino acids numbered 59 to 100. A segment comprising amino acids 27 to about 56 and about 30 amino acids of the carboxyl terminus of glycophorin A are missing from glycophorin B (Siebert and Fukuda, 1987). To select cDNA clones encoding glycophorin A, oligonucleotide mixtures GPA-N1 and GPA-C which were derived from the glycophorin A amino acid sequence spanning residues 24 to 30 (21-mer mixture) and 122 to 127 (a 17-mer mixture), respectively, were synthesized and used for hybridization to screen a λgt10 cDNA library constructed from K562 cell RNA. While this work was still in progress a partial cDNA sequence was reported (Siebert and Fukuda, 1986b). Therefore, the exact sequence oligonucleotide GPA-N2 was synthesized based upon this published sequence, complementary to the nucleotides encoding residues 30 to 40 of glycophorin A, and used to screen the K562 cell cDNA library.

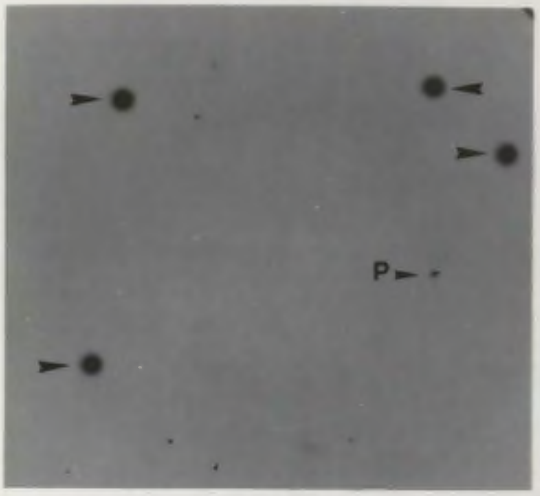
Northern blotting experiments performed previously (section 4.3.2.1) had shown that hybridization of GPA-N2 to K562 cell RNA at 50°C followed by washing in 6 X SSC for 5 min at 60°C gave the best signal of mRNA bands over a very low background. Therefore, oligonucleotide GPA-N2 was hybridized with a λgt10 K562 cell cDNA library that was blotted on nylon membranes under the conditions established by Northern blotting. About 20 membranes containing a total of about 5×10^5 plaques (about 2.5×10^4 plaques/membrane) were screened using these conditions and several showed spots of GPA-N2 hybridization. Only those spots which were present on duplicate membrane replicas were considered to represent potentially positive clones, for example, the spot P in Fig. 3.1(A) and (B). Any other spot which was present only on one of the two duplicate membranes, for example, the spot B in Fig. 3.1(B), was ignored. In this way, the primary screening of the K562 cell cDNA library resulted in the detection of seven individual spots. The plaques and (or) area surrounding them, corresponding to the spots on the membranes, were identified on their respective master plates. These potentially positive plaques and (or) the area surrounding them were then lifted off the culture plates by removing an agar plug and suspending it in SM medium. The phage particles were allowed to elute out of the agar into the SM medium.

Each potentially positive plaque identified by primary screening was subjected to secondary screening. The number of PFU/ml in the phage suspension was determined and a sample of an appropriate dilution was mixed with *E.coli* C600 Hfl

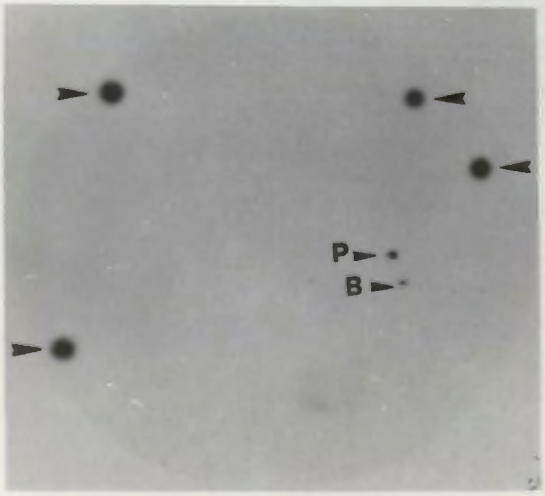
Fig. 3.1. Screening a K562 library to the isolate cDNA clones encoding glycophorin A: Primary Screening.

A sample containing about 2.5 x 10^4 PFU of the K562 cDNA library (λgt10 phage) was mixed with *E.coli* C600 Hfl, spread on a 150 x 10 mm plate containing LB agar with maltose and allowed to grow until plaques were visible. Membrane replicas were prepared in duplicate (A and B) by transferring and fixing the DNA on a nylon membrane (Hybond-N, 132 mm), hybridized with GPA-N2 at 50°C and washed at 60°C. The large, dark, circular spots (arrow heads) represent the marks obtained by spotting diluted radioactive ink on membranes for orientation with the master plate. P and B with an arrow head represent a potentially positive plaque present on both of the membrane replicas and a background spot, respectively.

(A)



(B)



and plated to produce a small number of plaques (100-1000/plate). Duplicate membrane replicas were made from these plates and hybridized with oligonucleotide GPA-N2 as described for primary screening. This secondary screening resulted in identification of several plaques from one primary spot on each plate (for example, see in Fig. 3.2). Out of about 100 plaques present, about nine showed hybridization with GPA-N2. This indicated that the phage suspension obtained after first round of screening (i.e. primary screening) contained extraneous plaques in addition to the ones showing specific hybridization. These unrelated plaques must have been picked up along with the truly positive clone; this is quite normal for high density screening in which the presence of a large number of plaques over a small area makes it very difficult to accurately identify the plaque corresponding to a hybridization spot on a membrane. Therefore, the process of identifying the positive clones followed by its isolation was repeated at least twice more or until a plaque was biologically purified and was completely free of any other contaminating plaques. For example, out of the several plaques showing hybridization in the second round of screening, a well-separated plaque (P in Fig. 3.2) was lifted from its corresponding master culture plate and subjected to two more rounds of screening. Fig. 3.3 shows the results of the tertiary screening of the plaque identified in Fig. 3.2. The master culture plate corresponding to this membrane contained about 300 plaques. A large proportion of these plaques hybridized with GPA-N2 indicating further enrichment but not a complete biological purification of the primary clone. One of these tertiary plaques was lifted from the master culture plate corresponding to this membrane and

Fig. 3.2. Screening a K562 library to the isolate cDNA clones encoding glycophorin A: Secondary screening.

The plaques obtained by growing a sample containing about 100 PFU of the phage suspension after primary screening were transferred in duplicate (A and B) to 90 mm diameter nylon membranes and hybridized with GPA-N2 as for primary screening. The large, dark, circular spots indicated by arrow heads represent marks produced by spotting diluted radioactive ink on membranes for orientation with the master plate. P represents the plaque picked and used for the next round of screening (tertiary).

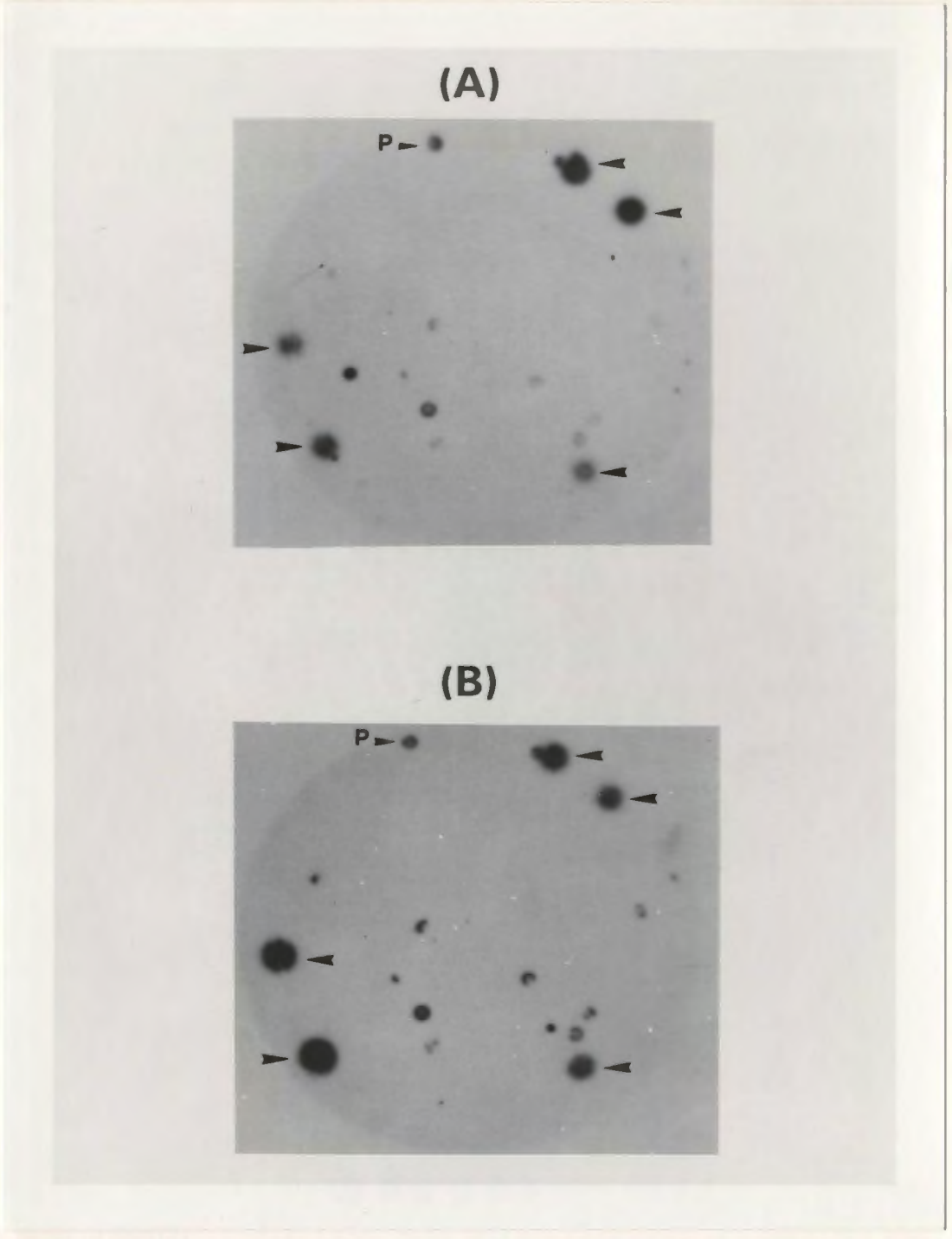
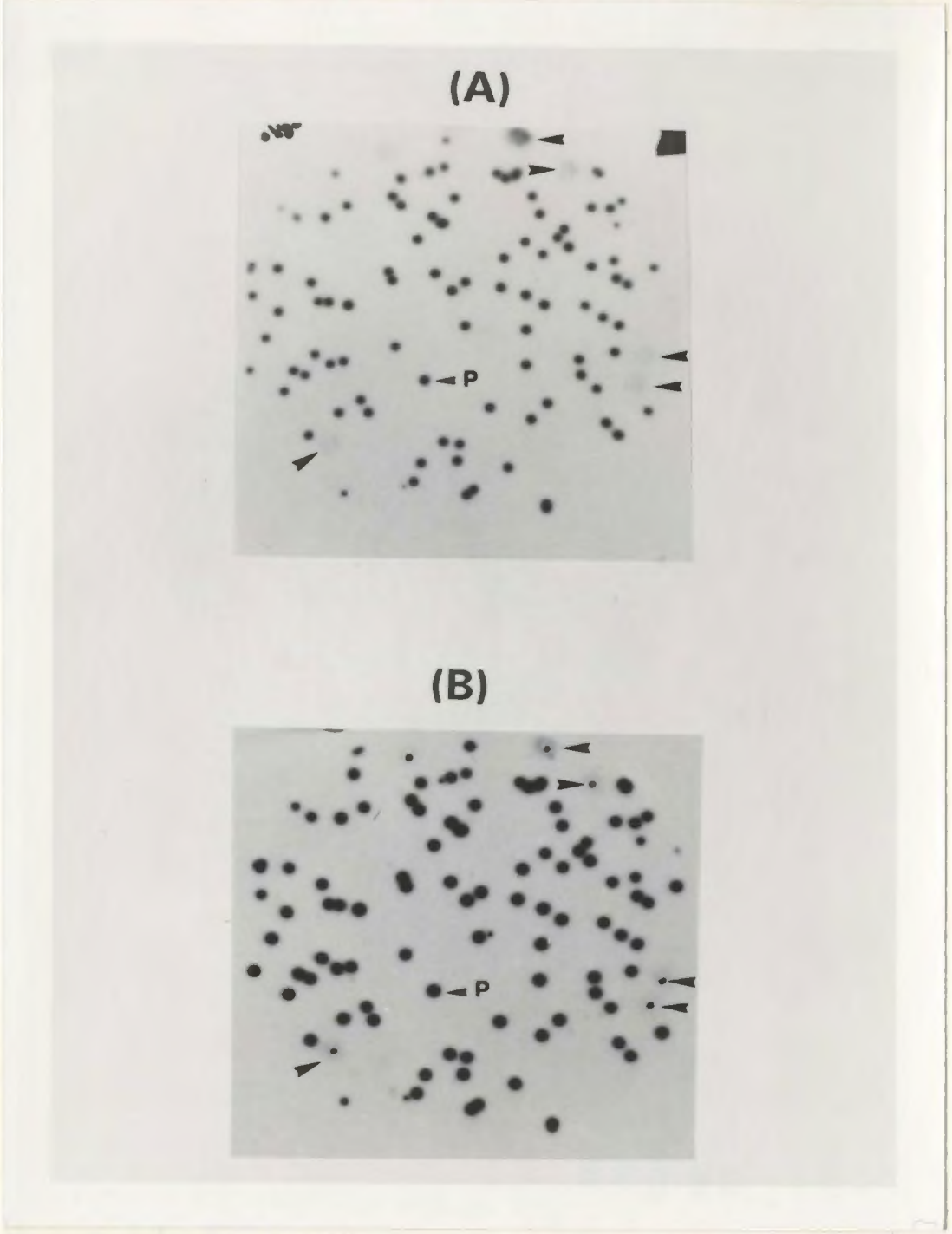

(A)
P
(B)
P

Fig. 3.3. Screening a K562 library to the isolate cDNA clones encoding glycophorin A: Tertiary screening.

A sample containing about 300 PFU of the phage suspension obtained after secondary screening was grown and transferred in duplicate (A and B) to 90 mm diameter nylon membranes. The membranes were hybridized with GPA-N2 as for primary screening. The arrow heads represent marks obtained by spotting diluted radioactive ink on membranes for orientation with the master plate. P represents the plaque picked and used for biological purification of the clone in the final round of screening (quaternary).


(A)
P
(B)
P

subjected to another round of screening. At this stage, the membrane contained about 200 plaques and every one of them showed hybridization with GPA-N2 indicating the biological purification of the clone (Fig. 3.4). Therefore, one plaque was picked from the plate corresponding to this membrane, resuspended in SM medium and the resulting phage suspension used as a source of inoculum for growth in subsequent steps.

The process of repeated screening was continued for each primary spot until a plaque was considered to be biologically purified. In this way, out of seven primary clones, six were biologically purified. These six individual clones denoted by the numbers λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, λ-gpa6 and λ-gpa7 and were grown up on a large scale. DNA was extracted from each of these clones and after restriction endonuclease digestion analyzed on an agarose gel as described in next section.

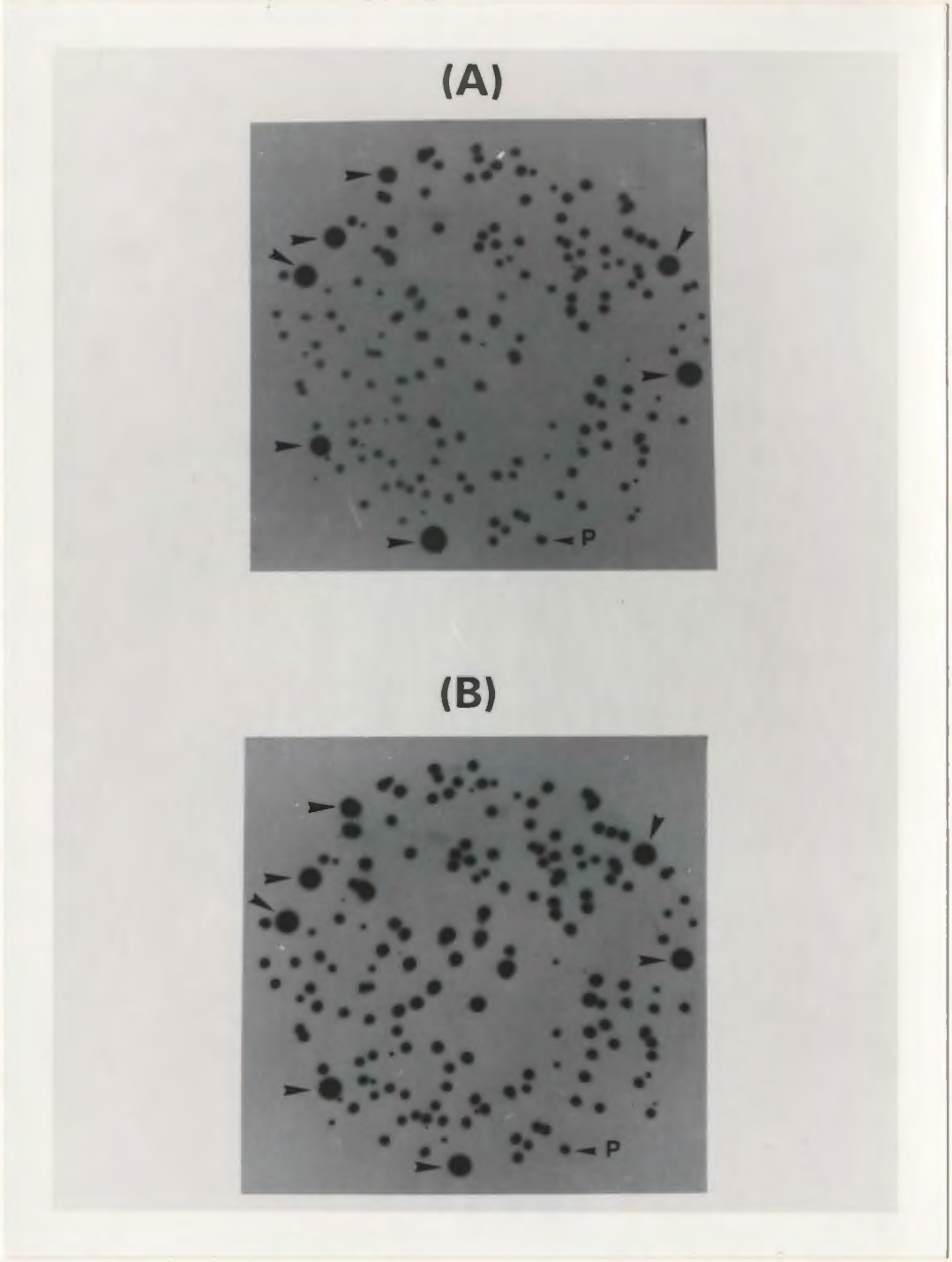
3.2. Confirming the identity of cDNA clones

3.2.1. Agarose gel electrophoresis

λgt10 is a 43 kbp, linear, double-stranded DNA λ bacteriophage-derived cloning vehicle into which cDNAs that have *Eco*RI linkers ligated to their ends are cloned at the unique *Eco*RI site (Huynh et. al., 1984, also section 2.13). Therefore, as a first step to characterize the cDNAs, λ DNA was isolated from each of the six clones. About 2 μg of the λgt10 DNA was digested with *Eco*RI to release the cDNA fragment(s) from the cloning site in the vector. The digestion mixture was divided

Fig. 3.4. Screening a K562 library to the isolate cDNA clones encoding glycophorin A: Quaternary screening.

A sample containing about 100-300 PFU of the phage suspension obtained after tertiary screening was grown and transferred in duplicate (A and B) on a nylon membranes of 90 mm. The membranes were then hybridized with GPA-N2 as for primary screening. The arrow heads represent marks obtained by spotting diluted radioactive ink on membranes for orientation with the master plate. P represents the plaque picked up and used as a source of inoculum for phage growth.


(A)
P
(B)
P

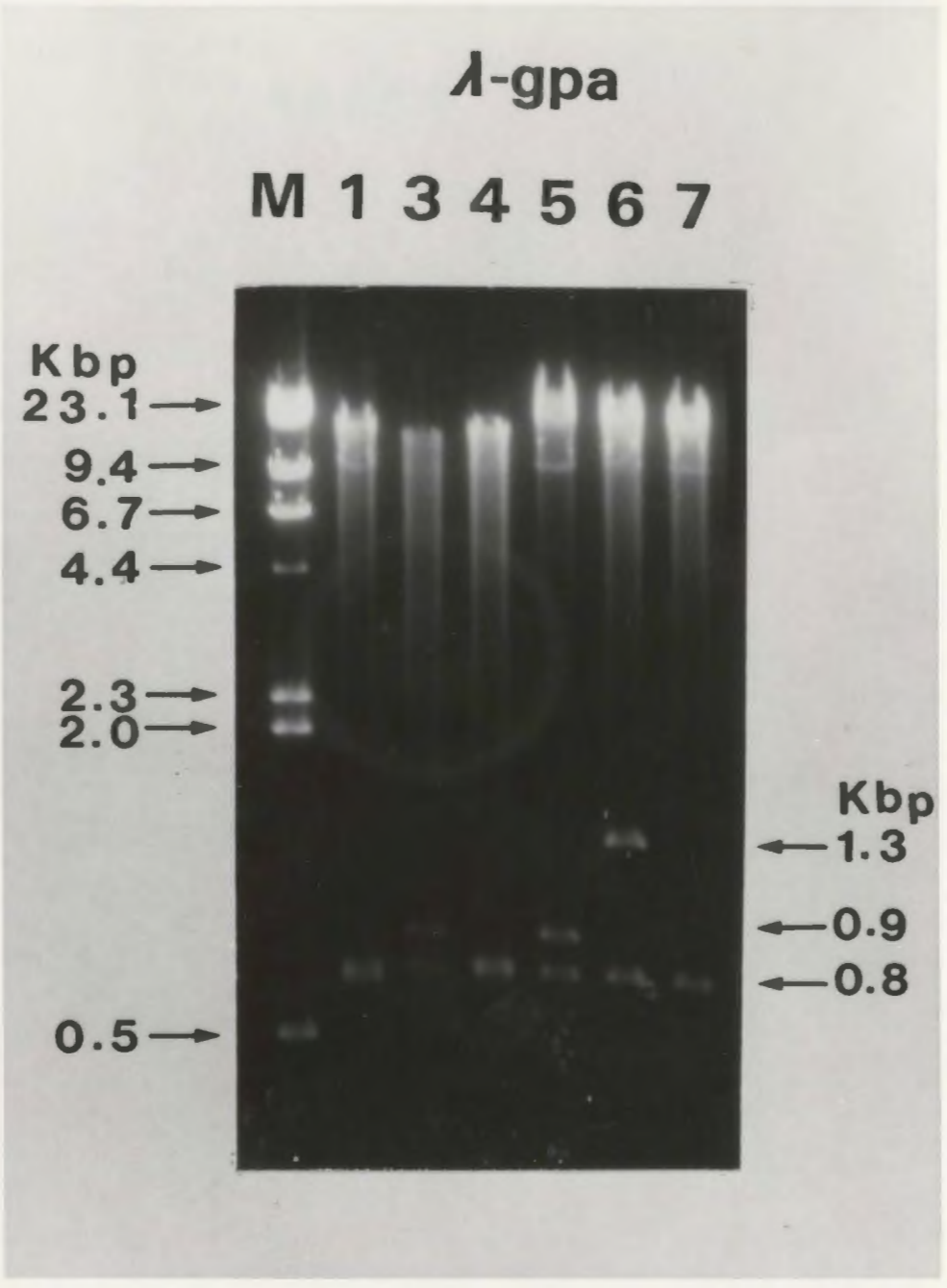
into two equal portions and both were analyzed on a 0.8% agarose gel. DNA bands were visualized by ethidium bromide fluorescence (Fig. 3.5); note that only one half of the gel containing one set of digested DNA is shown. The sizes of cDNA fragments were calculated by comparing with the mobilities of molecular size markers run on a parallel lane (Fig. 3.5, lane M). *Eco*RI digestion of λgt10 DNA containing cDNA produced a fragment of about 0.8 kbp from each of the six clones (Fig. 3.5, lanes 1 and 3 to 7). In addition to this common 0.8 kbp fragment, two of the six clones, λ-gpa3 and λ-gpa5 also contained a fragment of about 0.9 kbp (Fig. 3.5, lanes 3 and 5) and still another, λ-gpa6, contained an additional fragment of about 1.3 kbp (Fig. 3.5, lane 6). Therefore, the combined size of the cDNA fragments in various clones were 0.8 (λ-gpa1, λ-gpa4 and λ-gpa7), 1.7 (λ-gpa3 and λ-gpa5) and 2.1 kbp (λ-gpa6).

3.2.2. Southern blotting

To further characterize the cDNA clones, the DNA from the gel shown in Fig. 3.5 together with the remaining half of the gel referred to above was blotted on a nylon membrane. The membrane was divided into two halves, each containing one set of the six cDNAs. One half of the membrane was used for the hybridization with oligonucleotide GPA-N2. The same membrane was also used to hybridize with the 0.8 kbp fragment of λ-gpa6 after the removal of the first probe (GPA-N2). Similarly the other half of the membrane was used for hybridization with oligonucleotide GPA-N1 and subsequently with GPA-C. These oligonucleotides are described

Fig. 3.5. Agarose gel electrophoresis of DNA isolated from glycophorin A cDNA clones and digested with restriction endonuclease *Eco*RI.

DNA fragments were detected by staining the gel with 0.5 μg/ml ethidium bromide. Numbers above the lanes indicate clones λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, λ-gpa6 and λ-gpa7. Lane M contains *Hin*d III-generated phage λ DNA fragments as molecular size markers run on the same gel. The position and sizes in kbp of the markers and various cDNA fragments is also indicated.


λ-gpa
M 1 3 4 5 6 7
Kbp
23.1
9.4
6.7
4.4
2.3
2.0
0.5
Kbp
1.3
0.9
0.8

in section 3.1. The results of all these hybridizations are described in the next section.

3.2.2.1. Oligonucleotide hybridization

When oligonucleotide GPA-N2 was incubated with the Southern blot of cDNAs at 50°C and washed for 5 min at 60°C (the same conditions were used for screening the cDNA library using this oligonucleotide), only the 0.8 kbp bands of the six clones hybridized (Fig. 3.6, lanes 1 and 3 to 7). The same 0.8 kbp band from each of the six clones hybridized with oligonucleotides GPA-C [Fig. 3.7 (A)] as well as with GPA-N1 [Fig. 3.8 (A)]. Since GPA-N1 and GPA-C were mixtures containing 508 and 98 different oligonucleotides, respectively, it was necessary to make sure that they were hybridizing specifically.

The Td for exact sequence oligonucleotide GPA-C was calculated to be 44°C (section 2.11.4). Therefore, GPA-C was hybridized at 37°C with the Southern blot, washed for 5 min in a solution containing 6 X SSC at 39°C and exposed to X ray film [Fig. 3.7 (A)]. The blot was washed again in a similar solution for 5 min but at higher temperatures [45°C and 50°C, Fig. 3.7 (B) and 3.7 (C) respectively]. The 0.8 kbp fragments of six clones retained the signal of hybridization even after washing at 45°C and 50°C, the temperatures which were 1°C and 6°C higher than the calculated Td, respectively. The retention of signal of hybridization after washing at a temperatures close to the Td of GPA-C indicated that it was hybridizing specifically.

Fig. 3.6. Hybridization of ^{32}P-labelled oligonucleotide GPA-N2 with a Southern blot of DNA isolated from glycophorin A cDNA clones and digested with restriction endonuclease *Eco*RI.

The blot was hybridized at 50°C and washed at 60°C as described in methods. Numbers above the lanes indicate clones λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, λ-gpa6 and λ-gpa7. The position of the 0.8 kbp fragment is indicated by the arrow.

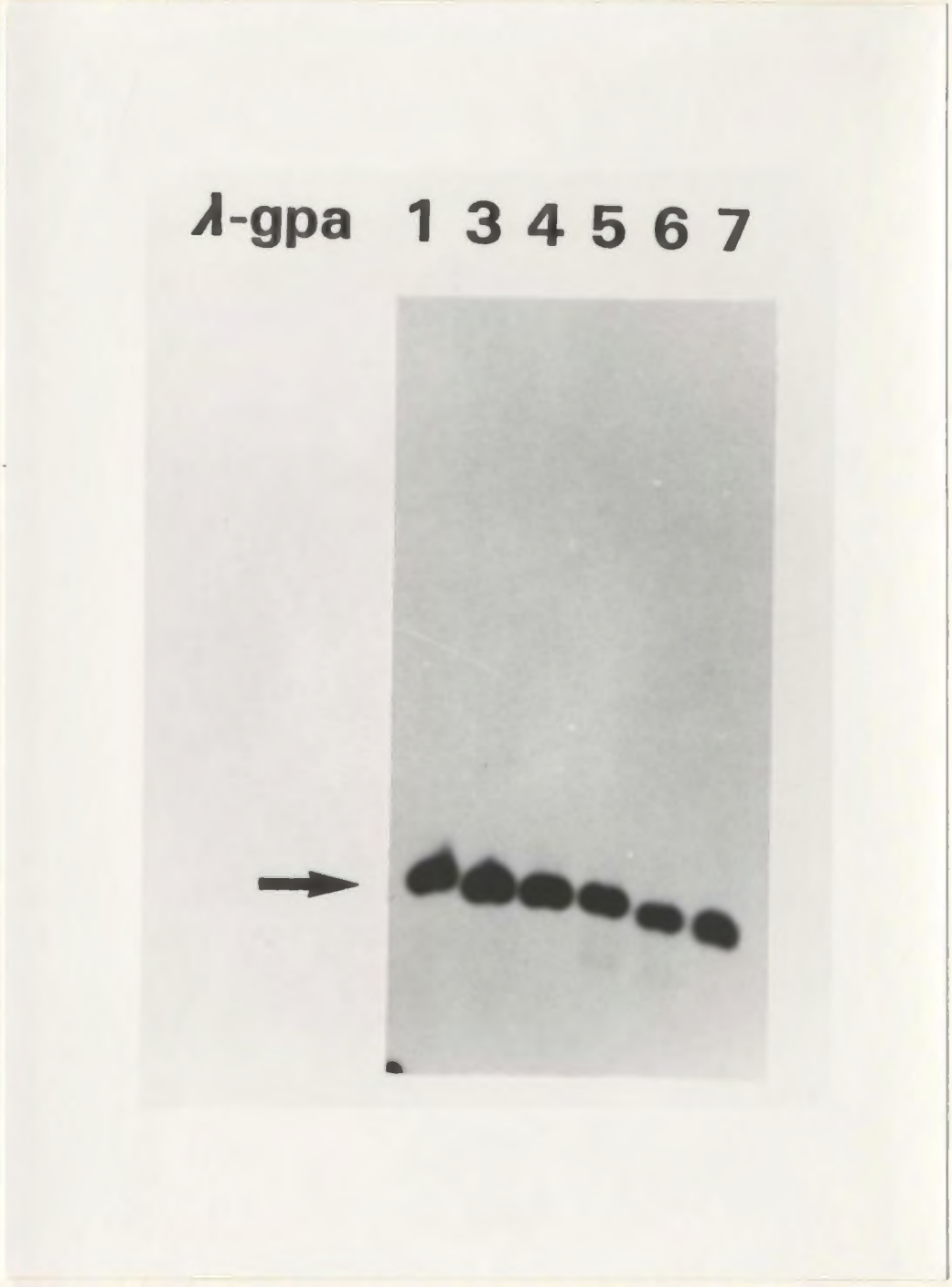
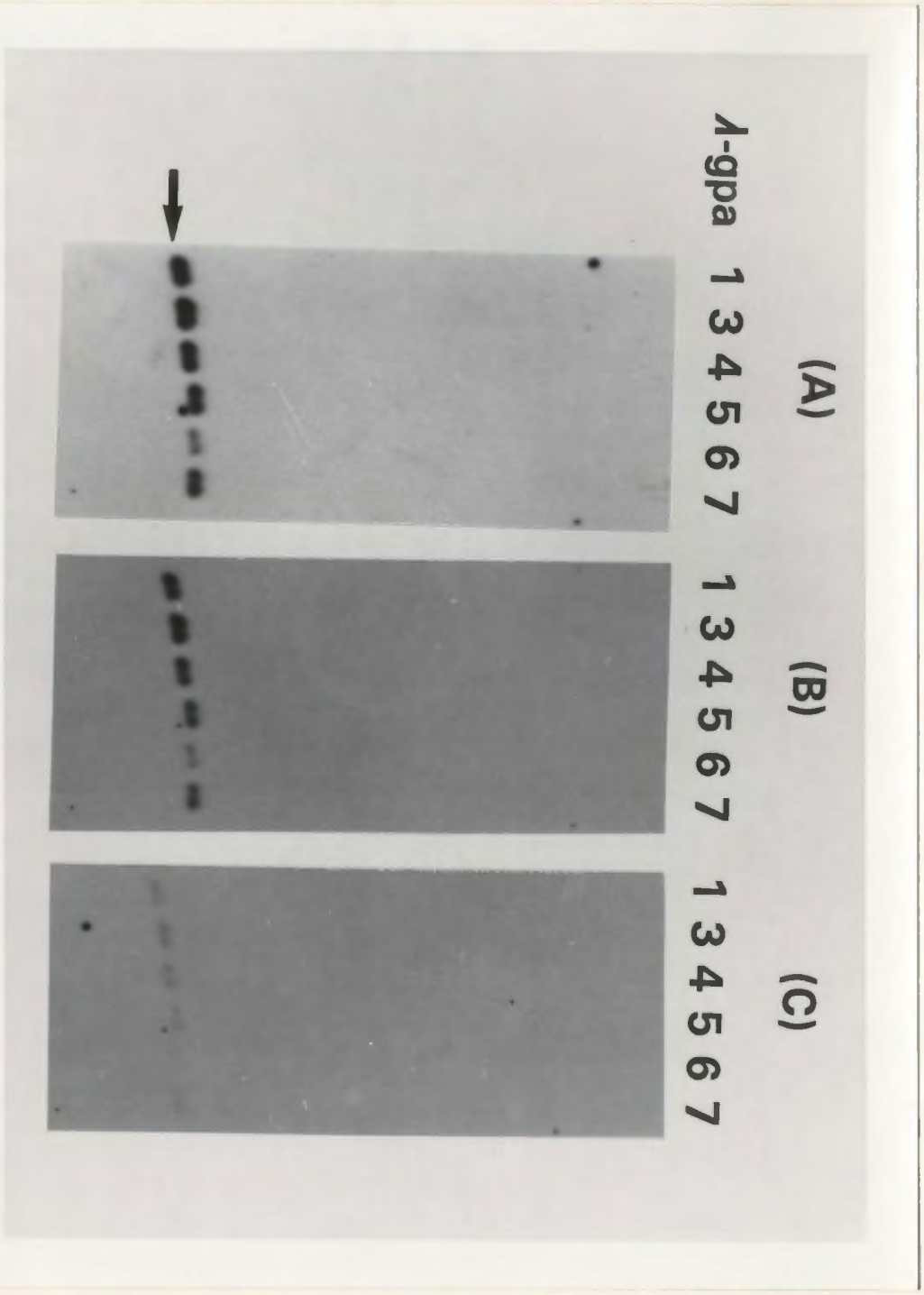

λ-gpa 1 3 4 5 6 7

Fig. 3.7. Hybridization of ^{32}P-labelled oligonucleotide GPA-C with a Southern blot of DNA isolated from glycophorin A cDNA clones and digested with restriction endonuclease *Eco*RI.

The blot was hybridized at 37°C and washed subsequently at: 39°C (A), 45°C (B) and 50°C (C) and exposed to X ray film between each wash as described in methods. Numbers above the lanes indicate clones λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, λ-gpa6 and λ-gpa7. The position of the 0.8 kbp band is indicated by the arrow.


(A)
(B)
(C)
A-gpa 1 3 4 5 6 7 1 3 4 5 6 7 1 3 4 5 6 7

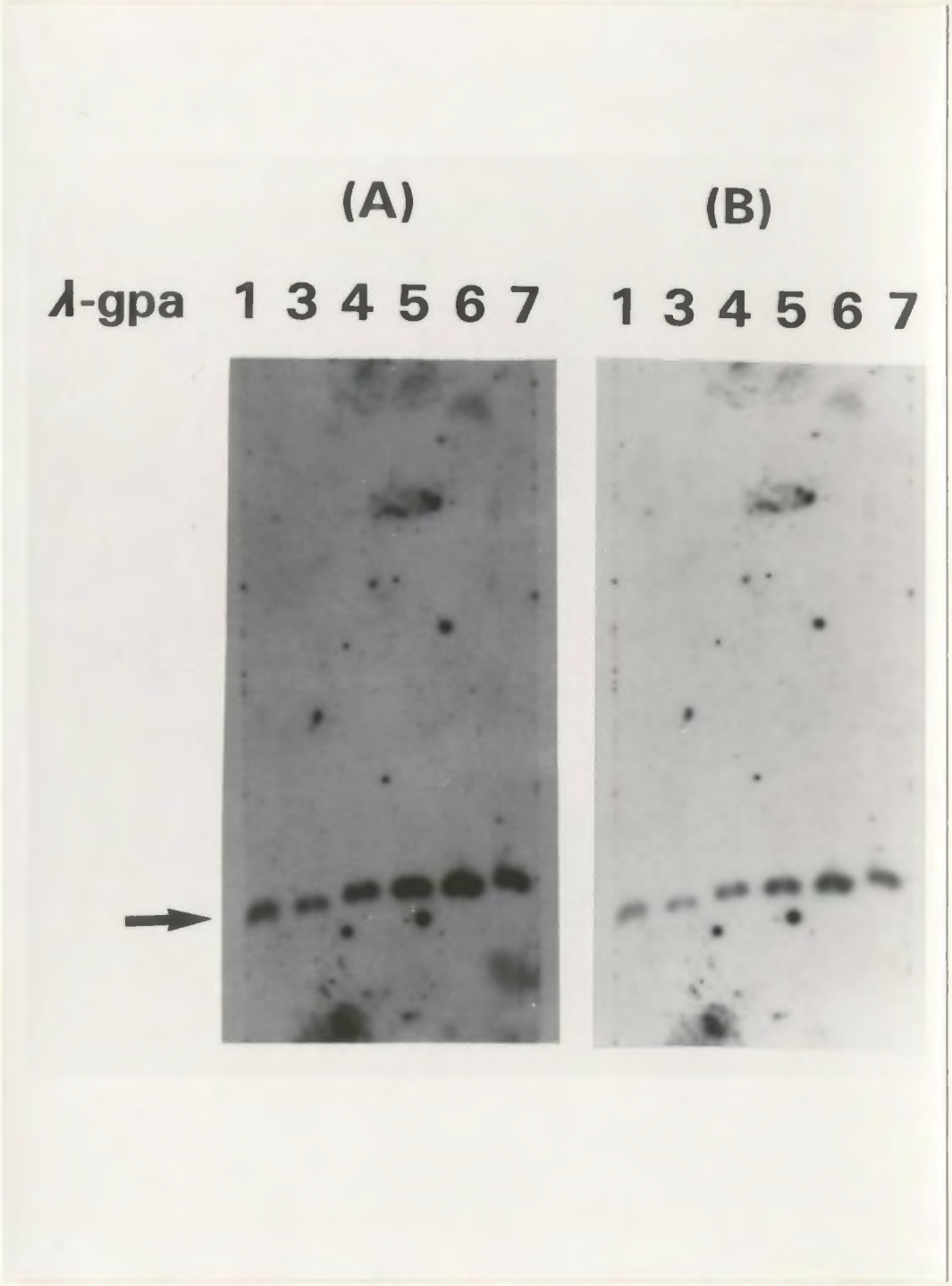
The Td for exact sequence oligonucleotide GPA-N1 was calculated to be 58°C (section 2.11.4) Therefore, hybridization with oligonucleotide GPA-N1 at 52°C was followed by washing in a solution containing 6 X SSC for 5 min at 52°C and exposure to X ray film [Fig. 3.8 (A)]; the blot was washed once again in a similar solution for 5 min at a 57°C [Fig. 3.8 (B)]. The 0.8 kbp fragments of the six clones retained the signal of hybridization of this oligonucleotide after washing at both temperatures indicating specific hybridization.

Hybridization of Southern blots containing cDNAs with GPA-N2, GPA-C and GPA-N1 clearly demonstrated that the 0.8 kbp fragment from each of the six clones contained sequences complementary to these oligonucleotides (Fig. 3.6 to 3.8). In contrast, neither the 1.3 nor the 0.9 kbp fragments showed hybridization with any of the three oligonucleotides (Fig. 3.5). Since oligonucleotides synthesized to recognize glycophorin A hybridized with the 0.8 kbp fragments from the six clones, it was clear that all six cDNAs were derived from glycophorin A specific mRNA(s).

The oligonucleotides represented three different segments of the glycophorin A amino acid sequence spanning regions containing amino acid 24 to 30 (GPA-N1), 30 to 40 (GPA-N2) and 122 to 127 (GPA-C). All these oligonucleotides hybridized to the 0.8 kbp fragments suggesting that the cDNAs contained the region spanning amino acids 22 to 127 and probably the full coding sequence for glycophorin A since this contains only 131 amino acids. These oligonucleotide hybridization experiments

Fig. 3.8. Hybridization of ^{32}P-labelled oligonucleotide GPA-N1 with a Southern blot of DNA isolated from glycophorin A cDNA clones and digested with restriction endonuclease *Eco*RI.

The blot was hybridized at 50°C and washed subsequently at: 52°C (A) and 57°C (B) and exposed to the film between each wash as described in methods. Numbers above the lanes indicate clones λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, λ-gpa6 and λ-gpa7. The position of the 0.8 kbp band is indicated by the arrow.


(A)
(B)
λ-gpa 1 3 4 5 6 7
1 3 4 5 6 7

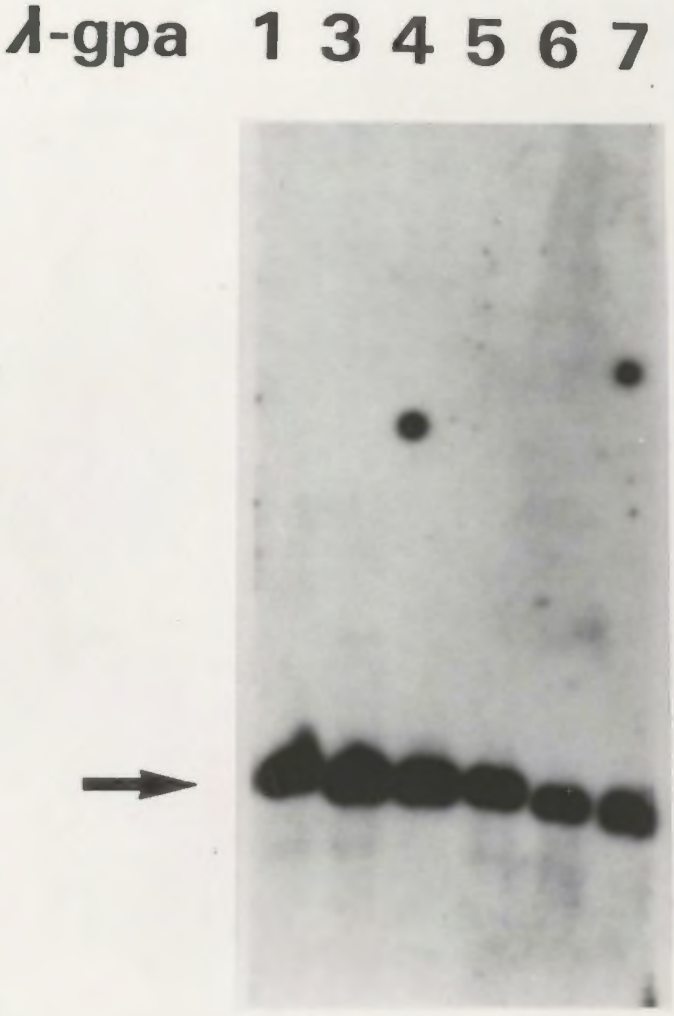
also suggested that the 0.8 kbp fragments, in addition to probably containing the full coding sequence, also contained some nucleotides representing the 5′, 3′ or both untranslated regions since only 393 nucleotides are required to code for the glycophorin A amino acid sequence. The lack of hybridization of any of the oligonucleotides to the 1.3 and 0.9 kbp fragments implied that these fragments must represent either the 3′ or 5′ untranslated region of glycophorin A mRNA. Another possibility that these fragments represented some unrelated DNA ligated to authentic glycophorin A cDNA during the cloning procedure although very unlikely was not ruled out at this stage.

3.2.2.2. Hybridization with 0.8 kbp fragment

The 0.8 kbp *Eco*RI fragment of λ-gpa6 was isolated by electroelution, labelled with ^{32}P by the random primers labelling method and used as a probe for hybridization with the Southern blot that had previously been used for hybridization with GPA-N2 (section 3.2.2.1). The 0.8 kbp fragment of each of the six clones showed hybridization after a very stringent wash in 0.1 X SSC at 65°C (Fig. 3.9, lanes 1 and 3 to 7). This indicated that the 0.8 kbp fragments from the six cDNA clones have similar sequences. In contrast neither the 0.9 kbp fragments from λ-gpa3 or λ-gpa5, nor the 1.3 kbp fragment from λ-gpa6 showed hybridization, strengthening the notion that these fragments did not share sequences with any of the 0.8 kbp fragment and did not code for the glycophorin A amino acid sequence.

Fig. 3.9. Hybridization of the ^{32}P-labelled 0.8 kbp fragment of λ-gpa6 with a Southern blot of DNA isolated from cDNA clones and digested with restriction endonuclease *Eco*RI.

The blot was hybridized at 42°C with 50% formamide, the final stringent wash was performed at 65°C and the blot was exposed to X ray film as described in methods. Numbers above the lanes indicate clones λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, λ-gpa6 and λ-gpa7. The position of the 0.8 kbp band is indicated by the arrow.


λ-gpa 1 3 4 5 6 7

The conclusions drawn from the Southern blot analysis of the cDNAs with either the 0.8 kbp fragment of λ-gpa6 or with oligonucleotides as probes can be summarized as follows; (i) the six cDNA clones isolated by screening the K562 cell cDNA library with oligonucleotide GPA-N1 encode glycophorin A; (ii) the 0.8 kbp fragment of the six cDNA clones had similar sequences which probably spanned the full coding region of the of glycophorin A in addition to some nucleotides of 3′ or 5′ untranslated regions; and (iii) the 1.3 and 0.9 kbp fragments probably represented the 3′ or 5′ untranslated regions of glycophorin A mRNA.

3.3. Restriction mapping

λ DNA containing a cDNA insert isolated from each of the six clones was subjected to restriction analysis using endonucleases *Kpn*I, *Xho*I, *Sac*I, *Xba*I, *Hind*III, *Stu*I and *Dra*I to characterize the cDNA clones and to establish the relationship among them. However, the restriction analysis did not unveil any dissimilarities apart from the differences of size of various cDNA fragments.

The identical size of the 0.9 kbp fragment of λ-gpa3 and λ-gpa5 (section 3.2.1) suggested that perhaps they had similar nucleotide sequences. It was also possible, however, that they differed internally which might or might not result in creation or removal of some of the restriction sites. Restriction analysis was performed to investigate the relationships between the two 0.9 kbp fragments of λ-gpa3 and λ-gpa5 and the 1.3 kbp fragment of λ-gpa6. Bluescript plasmids which contained these cDNA

fragments were digested with a selection of restriction enzymes (section 3.4.2). The enzymes used were chosen based upon the prediction of restriction sites, obtained by examination of the complete nucleotide sequence of λ-gpa6 with MicroGenie and (or) PC/GENE computer programmes. The restriction analysis using the enzymes *Hind*III, *Hpa*II, *Xba*I, *Alu*I, *Xmn*I and *Rsa*I revealed no internal differences except that both of the 0.9 kbp fragments lacked some of the sequences which were present in the 3′ end region of the 1.3 kbp fragment.

The restriction analysis of λ DNA from the six clones performed directly or after subcloning into plasmid vectors provided further support to the view that the six cDNA clones had similar sequences in their 0.8 kbp fragments. Further, the two 0.9 fragments shared some sequences with the 5′ region of the 1.3 kbp fragment.

3.4. Subcloning

3.4.1. Subcloning in pUC 19 plasmids

DNA isolated from λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, λ-gpa6 and λ-gpa7 was separately digested with *Eco*RI and a sample of each was analyzed by agarose gel electrophoresis to confirm completion of restriction digestion. The remaining digestion mixture from each of the cDNA clones was then purified by phenol-chloroform extraction and the DNA fragments ligated into the plasmid vector pUC19 which had been linearized at its unique *Eco*RI site. A sample of this ligated DNA was then used to transform competent *E.coli* DH5α cells. A large number of

transformants were obtained a few of which were picked, grown up and subjected to plasmid DNA isolation by the quick method (section 2.16.2). The plasmid DNA, with or without *Eco*RI digestion was then subjected to agarose gel analysis which revealed that many of the clones contained DNA inserts of size other than 0.8, 0.9 or 1.3 kbp. This indicated that perhaps they represented some unrelated DNAs accidentally cloned into the plasmid vector probably because the entire restriction digestion mixture of λgt10 DNA containing cDNA fragments was used for ligation with the plasmid vector without prior purification of the cDNA inserts. Hence, it became evident that identifying the clones containing the cDNA fragment would be very difficult if digestion with *Eco*RI followed by gel analysis was used for identification of subclones. Therefore, to identify the subclones, plasmid DNA was blotted on nylon membrane and subjected to Southern analysis using the following as probes: ^{32}P-labelled oligonucleotide GPA-N2 or 0.8 or 1.3 kbp fragments of λ-gpa6. Many subclones containing the 0.8 kbp fragment of λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5 and λ-gpa6 were identified using this approach. One clone was selected for each of the fragments to be subcloned, grown up and subjected to the large scale plasmid DNA isolation procedure (section 2.16.1). At this stage, no subclones representing the 0.8 kbp fragment of λ-gpa7, the 0.9 kbp fragments of λ-gpa3 and λ-gpa5, or the 1.3 fragment of λ-gpa6 were obtained. However, these fragments were later successfully sub-cloned into the Bluescript vectors.

3.4.2. Subcloning in Bluescript plasmids

The experiment described in the previous section showed that if the entire digestion mixture was used for ligation with vector, identification of subclones was difficult. Therefore, in subsequent experiments the 0.9 kbp cDNA fragments from λ-gpa3 and λ-gpa5, the 0.8 kbp from λ-gpa6 and λ-gpa7, and the 1.3 kbp from λ-gpa6 were purified by Geneclean prior to ligation with the Bluescript plasmids (KS and SK) in their unique *Eco*RI site. The ligated DNAs were then used to transform competent *E.coli* XL1-Blue cells. A large number of transformants were obtained. About 30 were picked, grown up and subjected to plasmid DNA isolation using the quick method (section 2.16.2). The digestion of the resultant plasmid DNA with *Eco*RI followed by agarose gel analysis revealed that about 80% contained cDNA inserts of appropriate size. One clone was selected for each of the cDNA fragments to be subcloned, grown up and subjected to the large scale plasmid DNA isolation procedure (section 2.16.1).

3.5. Nucleotide sequencing

cDNAs cloned in plasmid vectors were sequenced by the dideoxy sequencing method (Sanger *et al.*, 1972) using the Sequenase kit. Either double-stranded or single-stranded plasmid DNA were sequenced using various sequencing primers including the M13 universal sequencing primer, the M13 reverse sequencing primer, oligonucleotides GPA-N2 and GPA-C2 and two more primers (SK and KS) which were specific for the Bluescript plasmids only. Initially partial sequences of all cDNA

fragments were determined to find out the relationships among them (sections 3.5.1 and 3.5.2). The complete nucleotide sequence was then determined for the cDNA clone λ-gpa6 which is described in section 3.5.3.

3.5.1. Nucleotide sequence of the 0.8 kbp *Eco*RI fragment of various cDNAs

Partial nucleotide sequences of subclones representing the 0.8 kbp fragments of λ-gpa1, λ-gpa3, λ-gpa4, λ-gpa5, and λ-gpa7 were determined using various primers. The primers used for sequencing included the M13 universal, the M13 reverse, GPA-N2 and GPA-C2 oligonucleotides. Comparison of these partial nucleotide sequences with that of the λ-gpa6 revealed that although the cDNAs differed slightly from each other in size, they contained the complete nucleotide sequence of glycophorin A. This information is summarized in diagrammatic form in Fig. 3.10 (A).

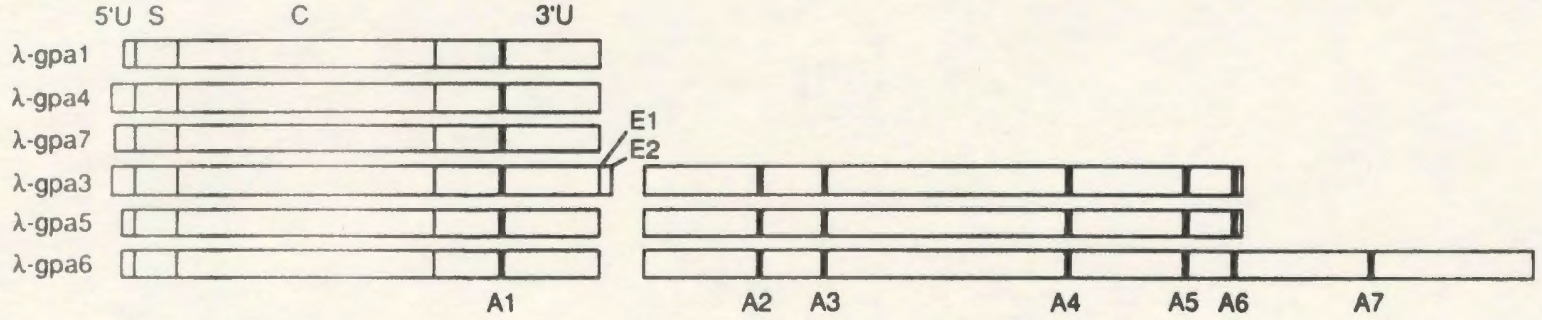
Glycophorin A is an integral membrane protein, like all other integral membrane proteins (Kreil, 1981; Sabatini *et al.*, 1982) it is also synthesized with a cleavable signal peptide (Jokinen *et al.*, 1981). It was found that all of the 0.8 kbp fragments of the six clones contained nucleotides representing a 19 amino acid-long signal peptide as reported earlier for a single clone encoding glycophorin A isolated by Siebert and Fukuda (1986b).

Examination of the nucleotide sequence of the 5 ´ untranslated regions of the various cDNAs showed that they vary slightly, probably a reflection of cDNA cloning

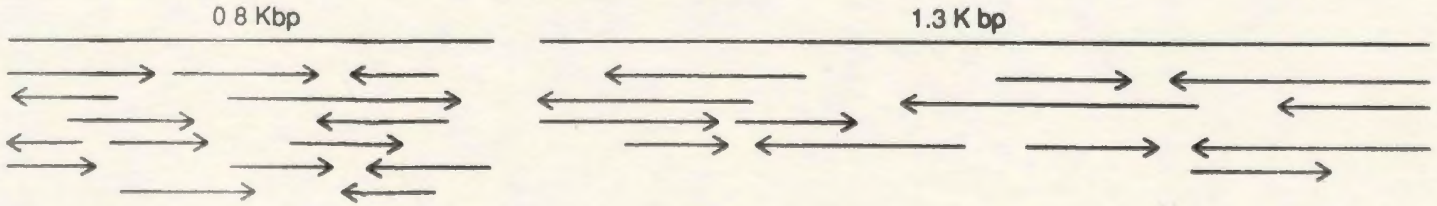
Fig. 3.10. Nucleotide sequence of glycophorin A cDNA derived from clones λ-gpa3 and λ-gpa6, together with the predicted amino acid sequence represented by the single letter code.

(A) 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. (B) Horizontal arrows indicate the direction and approximate position of the regions sequenced. (C) 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.

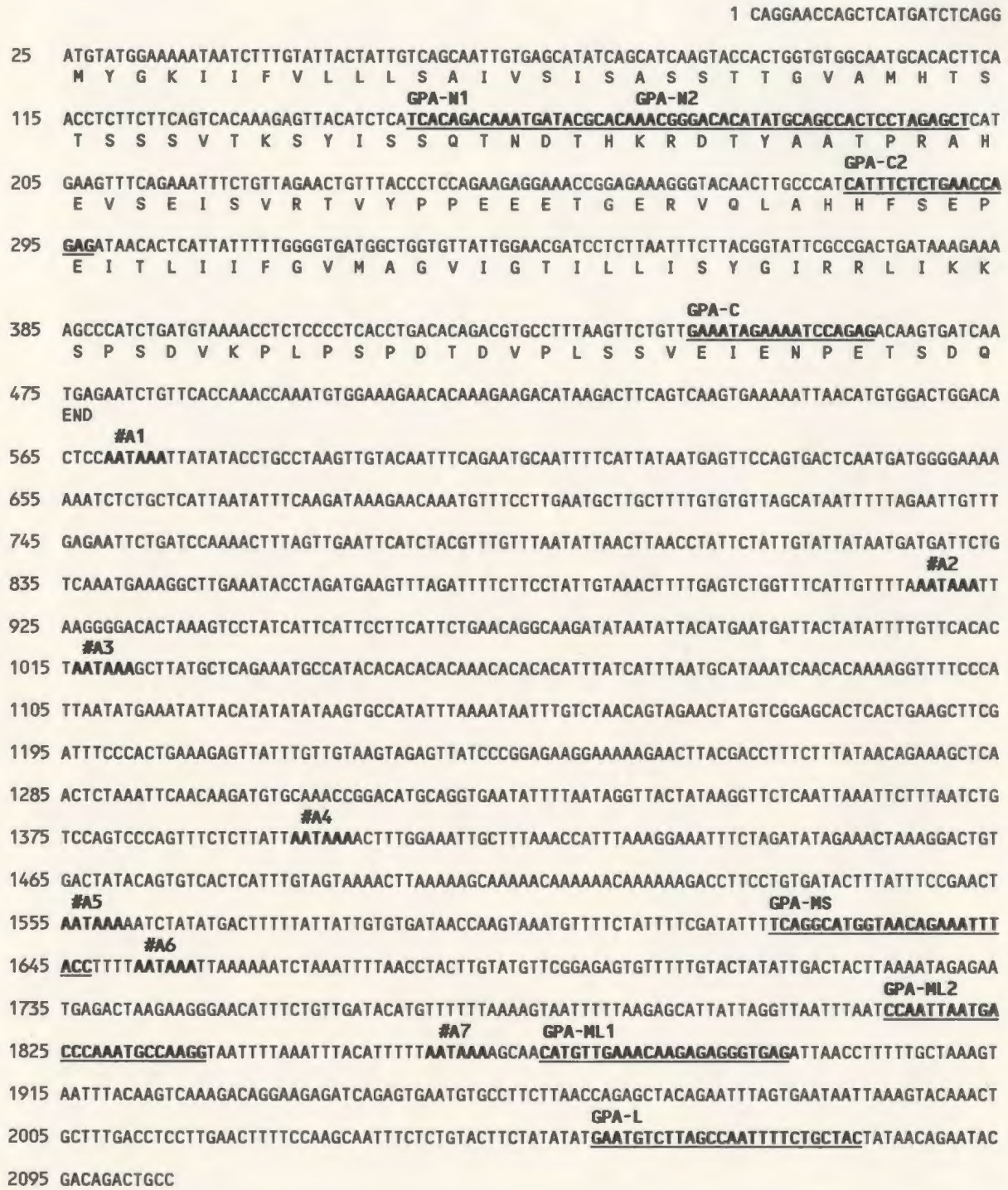
(A)


5'U
S
C
3'U
λ-gpa1
λ-gpa4
λ-gpa7
λ-gpa3
λ-gpa5
λ-gpa6
E1
E2
A1
A2
A3
A4
A5
A6
A7

(B)


0 8 Kbp
1.3 K bp

(C)


1 CAGGAACCAGCTCATGATCTCAGG
25 ATGTATGGAAAAATAATCTTTGTATTACTATTGTCAGCAATTGTGAGCATATCAGCATCAAGTACCACTGGTGTGGCAATGCACACTTCA
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
GPA-N1
GPA-N2
115 ACCTCTTCTTCAGTCACAAAGAGTTACATCTCATTCACAGACAAATGATACGCACAAACGGGACACATATGCAGCCACTCCTAGAGCTCAT
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
GPA-C2
205 GAAGTTTCAGAAATTTCTGTTAGAACTGTTTACCCTCCAGAAGAGGAAACCGGAGAAAGGGTACAACTTGCCCATCATTTCTCTGAACCA
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 GAGATAACACTCATTATTTTTGGGGTGATGGCTGGTGTTATTGGAACGATCCTCTTAATTTCTTACGGTATTCGCCGACTGATAAAGAAA
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
GPA-C
385 AGCCCATCTGATGTAAAACCTCTCCCCTCACCTGACACAGACGTGCCTTTAAGTTCTGTTGAAATAGAAAATCCAGAGACAAGTGATCAA
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 TGAGAATCTGTTCACCAAACCAAATGTGGAAAGAACACAAAGAAGACATAAGACTTCAGTCAAGTGAAAAATTAACATGTGGACTGGACA
END
#A1
565 CTCCAATAAATTATATACCTGCCTAAGTTGTACAATTTCAGAATGCAATTTTCATTATAATGAGTTCCAGTGACTCAATGATGGGGAAAA
655 AAATCTCTGCTCATTAATATTTCAAGATAAAGAACAAATGTTTCCTTGAATGCTTGCTTTTGTGTGTTAGCATAATTTTTAGAATTGTTT
745 GAGAATTCTGATCCAAAACTTTAGTTGAATTCATCTACGTTTGTTTAATATTAACTTAACCTATTCTATTGTATTATAATGATGATTCTG
#A2
835 TCAAATGAAAGGCTTGAAATACCTAGATGAAGTTTAGATTTTCTTCCTATTGTAAACTTTTGAGTCTGGTTTCATTGTTTTAAATAAATT
925 AAGGGGACACTAAAGTCCTATCATTCATTCCTTCATTCTGAACAGGCAAGATATAATATTACATGAATGATTACTATATTTTGTTCACAC
#A3
1015 TAATAAAGCTTATGCTCAGAAATGCCATACACACACACAAACACACACATTTATCATTTAATGCATAAATCAACACAAAAGGTTTTCCCA
1105 TTAATATGAAATATTACATATATATAAGTGCCATATTTAAAATAATTTGTCTAACAGTAGAACTATGTCGGAGCACTCACTGAAGCTTCG
1195 ATTTCCCACTGAAAGAGTTATTTGTTGTAAGTAGAGTTATCCCGGAGAAGGAAAAAGAACTTACGACCTTTCTTTATAACAGAAAGCTCA
1285 ACTCTAAATTCAACAAGATGTGCAAACCGGACATGCAGGTGAATATTTTAATAGGTTACTATAAGGTTCTCAATTAAATTCTTTAATCTG
#A4
1375 TCCAGTCCCAGTTTCTCTTATTAATAAAACTTTGGAAATTGCTTTAAACCATTTAAAGGAAATTTCTAGATATAGAAACTAAAGGACTGT
1465 GACTATACAGTGTCACTCATTTGTAGTAAAACTTAAAAAGCAAAAACAAAAAACAAAAAAGACCTTCCTGTGATACTTTATTTCCGAACT
#A5
GPA-MS
1555 AATAAAAATCTATATGACTTTTTATTATTGTGTGATAACCAAGTAAATGTTTTCTATTTTCGATATTTTCAGGCATGGTAACAGAAATTT
#A6
1645 ACCTTTTAATAAATTAAAAAATCTAAATTTTAACCTACTTGTATGTTCGGAGAGTGTTTTTGTACTATATTGACTACTTAAAATAGAGAA
GPA-ML2
1735 TGAGACTAAGAAGGGAACATTTCTGTTGATACATGTTTTTTAAAAGTAATTTTTAAGAGCATTATTAGGTTAATTTAATCCAATTAATGA
#A7
GPA-ML1
1825 CCCAAATGCCAAGGTAATTTTAAATTTACATTTTTAATAAAAGCAACATGTTGAAACAAGAGAGGGTGAGATTAACCTTTTTGCTAAAGT
1915 AATTTACAAGTCAAAGACAGGAAGAGATCAGAGTGAATGTGCCTTCTTAACCAGAGCTACAGAATTTAGTGAATAATTAAAGTACAAACT
GPA-L
2005 GCTTTGACCTCCTTGAACTTTTCCAAGCAATTTCTCTGTACTTCTATATATGAATGTCTTAGCCAATTTTCTGCTACTATAACAGAATAC
2095 GACAGACTGCC

artifacts, i.e. incomplete synthesis of cDNAs in this region. The 5´ untranslated region for λ-gpa1, λ-gpa5 and λ-gpa6 had 24 bases, that for λ-gpa7 had 36 bases and that for λ-gpa3 and λ-gpa4 had 42 bases. In addition as stated before, λ-gpa3 had 24 bases more at the 3´ end than the other 0.8 kbp fragments corresponding to nucleotides 747 to 771 in the nucleotide sequence of λ-gpa6 [section 3.5.3, Fig. 3.10 (C)]. Further, the 0.8 kbp fragment from λ-gpa5 had the complete coding sequence for the A^N polymorphic form of glycophorin A in which leucine and glycine are found in positions one and five (numbered from the NH_2-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 (Dahr *et al.*, 1977; Wasniowska *et al.*, 1977; Blumenfeld and Adamany, 1978; Furthmayr, 1978b; Lisowska and Wasniowska, 1978). Apart from these differences, the six 0.8 kbp fragments were identical.

3.5.2. Nucleotide sequences of the 1.3 and 0.9 kbp *Eco*R1 fragments

Partial nucleotide sequence of the 0.9 kbp fragments from λ-gpa3 and λ-gpa5 and a comparison with that of the 1.3 kbp fragment of λ-gpa6 revealed that both of the 0.9 kbp fragments were similar in sequence to one end of the 1.3 kbp fragment indicating that the 1.3 kbp fragment was an extension of the 0.9 kbp fragment. The complete nucleotide sequence of λ-gpa6 determined later was compared with that of the 0.9 kbp fragments; it revealed that both of the 0.9 kbp fragments had nucleotide sequences identical to those of λ-gpa6 extending from nucleotide 771 to

nucleotide 1667 [(Fig. 3.10 (C)].

Since neither the 1.3 kbp nor the 0.9 kbp fragments contained any overlapping sequences to 0.8 kbp fragments of the six clones, it was concluded that these fragments lay outside the 0.8 kbp fragments, i.e. in the 5′ or 3′ untranslated regions. This view was also supported by the lack of oligonucleotide hybridization to the 0.9 and 1.3 kbp fragments in the Southern blots of the six cDNA clones (Fig. 3.6 to 3.9). However, the placement of these fragments either at 5′ or 3′ untranslated region remained enigmatic. While this work was still in progress the nucleotide sequence of a cDNA encoding glycophorin A was published by Rahuel *et al.* (1988) comparison of which with that of the 1.3 kbp fragment of λ-gpa6 revealed some homology; more than 160 nucleotides of the 1.3 kbp fragment of λ-gpa6 [nucleotides 771 to 935, Fig. 3.10 (C)] were found to be similar to those in the 3′ untranslated region of the cDNA (nucleotides 803 to 967) isolated by Rahuel *et al.* (1988). Therefore, the 1.3 kbp fragment of λ-gpa6 as well as the 0.9 kbp fragments of λgpa3 and λgpa5 must have arisen from the 3′ untranslated region of glycophorin A mRNA (Fig. 3.10 A and B). The gaps indicate that no continuous fragment containing these sequences had been isolated in the present study. Once these fragments were localized to the 3′ untranslated region it became clear that beside having minor differences in 5′ untranslated regions, the cDNAs varied greatly in the length of their 3′ untranslated regions. There were about 270 nucleotides of the 3′ untranslated region in λ-gpa1, λ-gpa4, and λ-gpa7, 1190 in λ-gpa3 and λ-gpa5, and 1628 in λ-gpa6.

3.5.3. Complete nucleotide sequence of λ-gpa6

Since partial sequencing of the six cDNA clones did not reveal any major difference in the coding or 5′ untranslated region but revealed that they differed in the length of their 3′ untranslated regions, further efforts were concentrated on determining the complete nucleotide sequence of the largest clone, λ-gpa6. With the exception of a few regions the nucleotide sequence analysis of most of the λ-gpa6 cDNA was performed in both orientations to determine the complete nucleotide sequence without any ambiguities (Fig. 3.10 B). To achieve this goal, the 0.8 kbp fragment of λ-gpa6 was subcloned into the Bluescript vector. The 1.3 kbp fragment was also subcloned separately in two opposing orientations in the Bluescript vectors. This was followed by subjecting the Bluescript subclone containing the 0.8 kbp fragment of λ-gpa6 to exonuclease III-mung bean nuclease deletion procedure to remove nucleotides from the 3′ end (section 2.19.1). The two Bluescript subclones containing the 1.3 kbp fragment of λ-gpa6 in opposite orientation were also subjected to the exonuclease III-mung bean nuclease deletion procedure.

The nucleotide sequence of various subclones with different extents of deletions was determined using various primers, including the M13 universal, M13 reverse, KS, SK, GPA-N2 and GPA-C2. The nucleotide sequence determined for the λ-gpa6 insert is shown in Fig. 3.10 (C). This nucleotide sequence of λ-gpa6 also contains the sequence of the small *Eco*RI fragment (24 nucleotides) of λ-gpa3 located between two internal *Eco*RI sites [Fig. 3.10 (A)] corresponding to nucleotides 747 and

771 [Fig. 3.10 (C)]. This fragment was not detected in any of the subclones except in a subclone of λ-gpa3. It was possible that due to its small size this fragment remained undetected and was lost at some stage of cloning.

The sequence depicted in Fig. 3.10 (C) in addition to containing 393 nucleotides encoding 131 amino acids residues of glycophorin A, also contained 24 nucleotides of 5′ untranslated region, 57 nucleotides encoding a signal peptide 19 amino acids long and 1628 nucleotides of the 3′ untranslated region. As mentioned before (chapter 1) glycophorin A is encoded by a single gene and produces three different sized mRNAs (2.8, 1.7 and 1.0 kb, Siebert and Fukuda, 1986b). The examination of the nucleotide sequence in the 3′ untranslated region revealed the presence of seven polyadenylation signals (AAUAAA) giving rise to the notion that they may have a role in the production of multiple mRNAs from a single glycophorin A gene.

It was interesting to note that even though the coding region was only 450 nucleotides long and the 5′ untranslated region contained only 24 nucleotides, the cDNA contained a very large 3′ untranslated region more than three times the length of the coding region. It should be noted however, that this cDNA clone λ-gpa6 of size about 2.1 kbp did not fully represent the largest glycophorin A mRNA whose size is 2.8 kb. Therefore, the actual size of the 3′ untranslated region of the largest mRNA must be more than 1628 nucleotides long.

3.6. Discussion

3.6.1. Comparison with sequences for other glycophorin A cDNAs

Nucleotide sequences similar to that presented in Fig. 3.10 (C) have been reported by other groups (Rahuel *et al.*, 1988; Siebert and Fukuda, 1986a, b; Tate and Tanner, 1988). Comparison of the sequence I had found for nucleotides 1 to 488 in λ-gpa6 [Fig. 3.10 (C)], revealed that it was identical to that reported by Kudo and Fukuda (1989) for a glycophorin A genomic clone isolated from a K562 cell genomic DNA library. Another genomic clone termed λ6 isolated from a human leucocyte genomic DNA library also had nucleotide sequence similar to nucleotides 1 to 488 of λ-gpa6 (Vignal. *et al.*, 1990). Further, nucleotides 90 to 935 of λ-gpa6 were identical to nucleotides 10 to 854 for a clone isolated from a human foetal liver cDNA library containing almost the full coding region, starting from amino acid four of glycophorin A, and more than 400 nucleotides of 3′ untranslated region in addition to a poly (A) tail (Rahuel *et al.*, 1988). The sequence of the 5′ untranslated region and the coding region of the cDNA clones I isolated was similar to that of a composite cDNA reported by Rahuel *et al.* (1989) from a human foetal liver cDNA library, but varied slightly in the 3′ untranslated region. These differences in sequence of the 3′ untranslated region could be attributed to sequence polymorphism for the clones isolated from two different tissues, or to sequencing artifacts.

The λ-gpa6 sequence reported here also differed from the first partial glycophorin A cDNA sequence from a K562 cell cDNA library reported by Siebert

and Fukuda (1986b) but agreed, except for one base, with the revised sequence published by the same authors (Siebert and Fukuda, 1986a). The difference was that T not A was present in λ-gpa6 at position 228 [Fig. 3.10 (C)]; however, this difference would result in a silent mutation.

Tate and Tanner (1988) 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 5′ end and terminated at nucleotide 935 [Fig. 3.10 (C)] but also contained 16 nucleotides of a poly (A) tail. The size of approximately 970 nucleotides for this clone suggested that it was derived from the smallest glycophorin A mRNA of about 1.0 kb size. The following nucleotide positions in ALP1 which contained the bases given in parentheses differed from those in λ-gpa6 [Fig. 3.10 (C)]: 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 one and five, instead of serine and glycine which the sequence in the present report predicts [Fig. 3.10 (C)]. This demonstrates that ALP1 and λ-gpa6 code for the polymorphic N and M forms of glycophorin A, respectively. The difference at nucleotide 62 would result in glutamic acid at amino acid position minus seven in the signal sequence in ALP1 and alanine in λ-gpa6. It was suggested that the amino acid at this position represents yet another polymorphic site in glycophorins A^M and A^N with a glutamic acid in A^N and an alanine in A^M (Tate and Tanner, 1988). The sequence I 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 had the same base as λ-gpa6 in positions 62 which would result in alanine and not glutamic acid as predicted by Tate and Tanner (1988). Hence, the data presented in this thesis do not support the conclusion drawn by Tate and Tanner (1988) regarding the minus seven position. Finally the differences at positions 117 and 552 represent silent mutations.

3.6.2. Relationships between cDNAs and mRNAs

The cDNA clones I isolated can be grouped into three size classes; 0.8 kbp (λ-gpa1, λ-gpa4 and λ-gpa7), 1.7 kbp (λ-gpa3 and λ-gpa5) and 2.1 kbp (λ-gpa6). Mere comparison of these three size classes with those of the three glycophorin A mRNAs gave rise to the notion that these cDNAs were representative of the three glycophorin A mRNAs. The three smaller cDNAs (λ-gpa-1, λ-gpa4 and λ-gpa7) of size 0.8 kbp could arise from the smallest (1.0 kb) glycophorin A mRNA. The two medium sized-cDNA clones (λ-gpa3 and λ-gpa5) could represent the medium size (1.7 kb) mRNA. The cDNA clone (λ-gpa6) could be an incomplete representation of the largest (2.8 kb) glycophorin A mRNA.

However, a closer examination of the nucleotide sequence data revealed that the three smallest cDNAs of 0.8 kbp could be generated from any of the large cDNAs, since there were two sites [beginning at nucleotide 747 and 771, Fig. 3.10 (C)] containing the recognition sequence for the restriction endonuclease, *Eco*RI. The

significance of the two sites is that *Eco*RI was used during cloning to remove excess *Eco*RI linkers from cDNA fragments before insertion into the *Eco*RI site of the λgt10 vector (section 2.13). Hence, the 0.8 kbp fragment could have been generated by *Eco*RI digestion of cDNA derived from any one of the glycophorin A mRNAs due to incomplete protection of the two sites by *Eco*RI methylase during the cloning procedure. 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, since the latter was found to be the second most abundant of the three glycophorin A mRNAs (section 4.3.2.1).

The two medium sized-cDNAs (λ-gpa3 and λ-gpa5) of sizes 1.7 kbp seemed to be derived from the medium length (1.7 kb) glycophorin A mRNA. Both of these cDNA clones had similar coding regions except that λ-gpa5 encoded glycophorin A^N while clone λ-gpa3 encoded glycophorin A^M. Both clones had similar 5′ untranslated regions except λ-gpa3 had 17 nucleotides more than λ-gpa5 in this region. When the sequences at the 3′ end of these two clones were compared, they showed a complete identity up to nucleotide 1667 with λ-gpa6 [10 nucleotides downstream from poly(A) addition signal # 6, Fig. 3.10 (C)]. Since these two clones (i.e. λ-gpa3 and λ-gpa5) apparently derived from two distinct glycophorin A mRNAs (encoding glycophorin A^M and A^N) had identical 3′ end sequences it was reasonable to suggest that this identity could not arise by chance but represented the true 3′ end of the mRNAs of this class (i.e. 1.7 kb). Therefore, these cDNAs (λ-gpa3 and λ-gpa5) most probably represented the structure of medium size glycophorin A mRNA of similar size (1.7

kb) encoding glycophorins A^M and A^N. The presence of poly(A) sequence in either of these clones would have confirmed that they were derived from 1.7 kb mRNA. Unfortunately a poly(A) sequence was absent from the subclones for both of the λ-gpa3 and λ-gpa5. The sequence was probably lost at some stage of cloning and hence could not provide the evidence supporting the conclusion that the two medium size cDNAs (λ-gpa3 and λ-gpa5) were derived from 1.7 kb mRNA.

Clones λ-gpa3 and λ-gpa5 of size 1.7 kb could not have been derived from mRNA of 1.0 kb. However, another possibility was that the cDNAs were reverse transcribed from the largest glycophorin A mRNA by internal priming. The nucleotide sequence of the largest cDNA clone [λ-gpa6, Fig. 3.10 (C)] showed that in a position close to the 3´ end of the two clones, between nucleotides 1660 and 1671 in the λ-gpa6 sequence, there is an A-rich stretch which could serve as a template for oligo (dT) primed cDNA synthesis. Hence, it was also possible that the two medium size cDNA clones were derived from the large mRNA by priming in this region. However, the possible origin of λ-gpa3 and λ-gpa5 from 2.8 kb mRNA by any mechanism seemed very unlikely since it would imply that at least three of the six cDNA clones which I isolated were derived from the least abundant mRNA (section 4.3.2.1). Therefore, the most reasonable conclusion is that λ-gpa3 and λ-gpa5 were derived from the medium sized-mRNA.

The cDNA clone of size 2.1 kbp (λ-gpa6) certainly could not have been derived from the two smaller mRNAs of sizes 1.7 or 1.0 kb. Therefore, it seemed to represent, although incompletely, the largest glycophorin A mRNA of size 2.8 kb.

Regardless of whether or not the three size classes of the cDNA were representative of three mRNAs for glycophorin A, differences in the primary structure of the three glycophorin A mRNAs could be predicted on the basis of the cDNA sequences. No major differences were found in the 5′ untranslated or coding regions. Indeed the only differences were found in the length of the 3′ untranslated region. Furthermore, the presence of seven poly(A) addition signals in λ-gpa6 suggested that they may have a role in generating multiple glycophorin A mRNAs from a single gene. It is proposed that multiple glycophorin A mRNAs are produced by differential processing at the 3′ end of pre-mRNA, via mechanisms discussed in chapter 4.

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Investigation of mRNA structure

Nucleotide sequences of the six cDNA clones did not reveal any differences in the coding regions. If the three glycophorin A mRNAs are all represented by the six cDNAs it would suggest that these mRNAs also do not vary in the coding region. It was, however, possible that the three glycophorin A mRNAs differed in their 5′ or 3′ untranslated regions or in both. I performed primer extension analysis and Northern blotting experiments using K562 cell RNA to investigate these possibilities.

4.2. Primer extension analysis

K562 cell RNA [total and poly(A)$^+$] was subjected to primer extension analysis using oligonucleotide GPA-N2 as a primer which codes for amino acids 30 to 40 in the glycophorin A sequence [Fig. 3.10 (C)]. This analysis should establish the length of the 5′ untranslated region of the three glycophorin A mRNAs. Oligonucleotide GPA-N2 was selected as a primer since it would only anneal with the mRNAs for glycophorin A but not with that of glycophorin B (also see section 3.1). The size of the expected product was calculated as follows: the signal peptide was 19 amino acids long (57 nucleotides); this was followed by 29 amino acids (87 nucleotides) of glycophorin A sequence to the start of the region coded by GPA-N2; and the

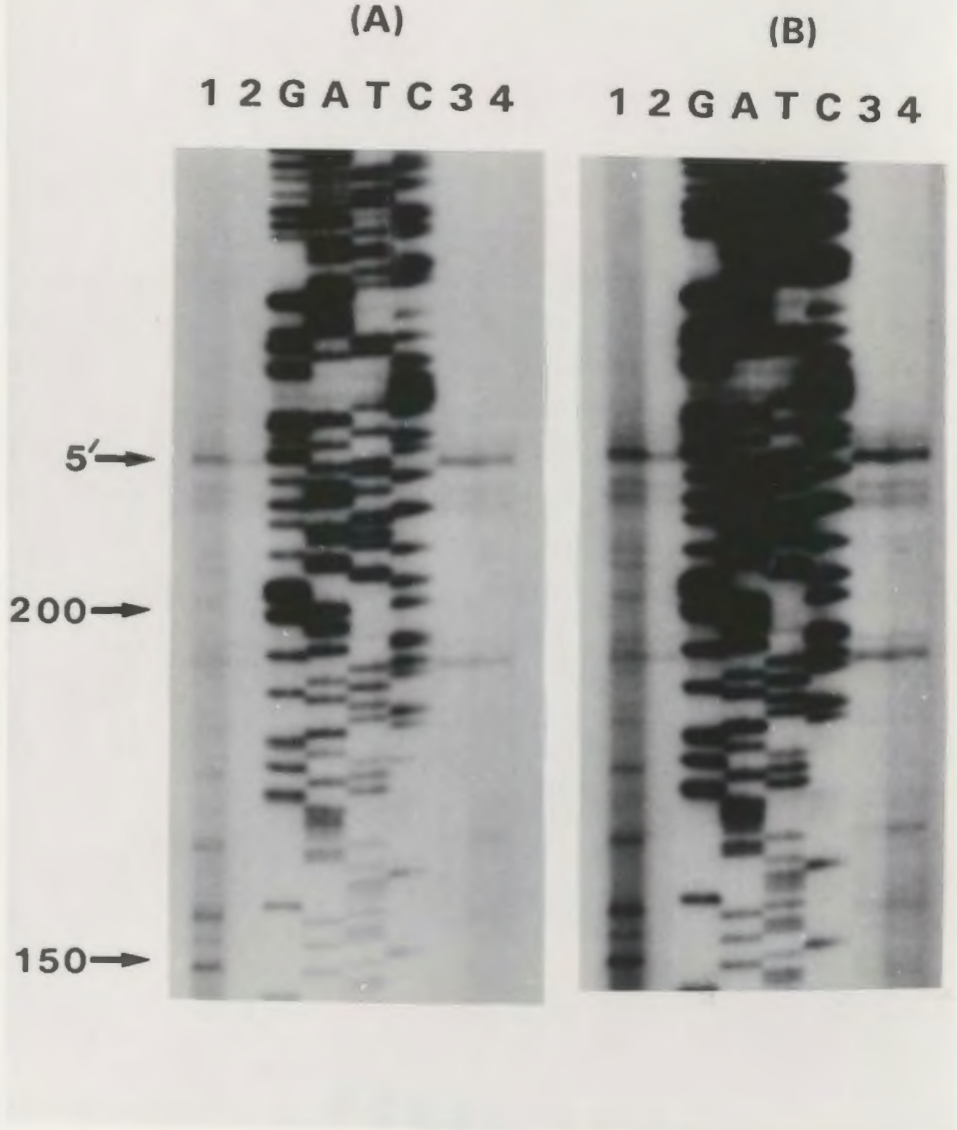
oligonucleotide GPA-N2 was itself 33 nucleotides long. Adding these nucleotides gave 177 nucleotides [Fig. 3.10 (C)]. Furthermore, nucleotide sequencing of the six cDNA clones had shown that they contained 24 (λ-gpa1, λ-gpa5, λ-gpa6), 36 (λ-gpa3, λ-gpa4) and 42 (λ-gpa7) nucleotides of 5′ untranslated region. Therefore, a fragment, at least 201 nucleotides long (including at least 24 nucleotides of the 5′ untranslated region) was expected to be produced using GPA-N2 as a primer in the reverse transcriptase reaction. Furthermore, if there was a single initiation site for all three glycophorin A mRNAs, only one major product was expected whereas if there was more than one initiation site giving rise to mRNAs of different sizes additional products of more than 201 nucleotides were also expected.

To determine the size of primer extended product the 0.8 kbp cDNA insert from λ-gpa6 subcloned in Bluescript plasmid was nucleotide sequenced using GPA-N2. This nucleotide sequencing reaction was run side by side with the primer extension reaction (lanes G, A, T, C in Fig. 4.1). By reading the nucleotide sequence of the cDNA fragment the size of various bands could be determined; the size of the primer extended product could then be deduced accurately by comparing its position with those of bands in sequencing lanes.

Primer extension analysis of K562 cell poly(A)$^+$ RNA revealed a single, major product greater than 201 nucleotides long [Fig. 4.1 (A), lanes 1 and 3]. A fragment

Fig. 4.1 Primer extension analysis of K562 cell RNA.

Oligonucleotide GPA-N2 was used as a primer to extend 10 and 5 μg K562 cell poly(A)$^+$ RNA (lanes 1 and 3, respectively), or 50 and 100 μg total RNA (lanes 2 and 4, respectively). The reaction products were analyzed on a 5% polyacrylamide gel containing 7 M urea and the dried gel was exposed to X ray film for: (A) 4 days, or (B) 14 days. 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 a Bluescript plasmid containing 0.8 kbp *Eco*RI fragment of λ-gpa6 and oligonucleotide GPA-N2 as a primer. The suggested position of the major, full-length, primer-extended product is indicated by 5´.


(A)
1 2 G A T C 3 4
(B)
1 2 G A T C 3 4
5′
200
150

of similar size was also obtained when total K562 cell RNA was used as template [Fig 4.1 (A), lanes 2 and 4]. The size of the major product was found to be about 230 nucleotides when its position was compared with the band in the nucleotide sequencing reactions of the Bluescript plasmid as described above.

With the exception of some faint bands which might represent the 5' untranslated region most of the bands in the lower part of the gel that were of smaller sizes than 201 nucleotides in length [Fig. 4.1 (A), lanes 1 and 3) were presumably due to premature termination during reverse transcription. None of these additional bands could be functional 5´ end as they were too small to even approach the 5´ end of the coding region. The above conclusion was consistent with the nucleotide sequence analysis I performed, as well as that of others (Siebert and Fukuda, 1986a, b; Tate and Tanner, 1988; Rahuel *et al.*, 1988). These nucleotide sequence analyses had shown that various clones did not have any major variation in the coding regions. Therefore, it was safe to assume that the three glycophorin A mRNAs did not differ in their coding sequence (also see section 3.5). Hence, the bands in the lower part of the gel were presumably due to premature termination.

It was also possible that minor components of higher molecular size remained undetected. Therefore, the gel was exposed for longer periods [Fig. 4.1 (B)]. No additional higher molecular size bands were visible even after a long exposure of 14 days [Fig. 4.1 (B), lanes 1 to 4).

Another point worthy of consideration was that the three mRNAs were present in approximate proportions of 30%, 60% and 10%, for the large, medium and small mRNAs, respectively (discussed in section 4.3.2.1). Thus, if each mRNA was transcribed from its own unique 5´ terminus, three bands in approximately these proportions were expected to be produced by reverse transcription of glycophorin A mRNAs. This was not the case for the multiple bands produced by primer extension [Fig. 4.1 (A)].

The primer extension experiments showed that all three glycophorin A mRNAs had the same 5´ end which in turn represented a unique initiation site for transcription of the glycophorin A gene. This analysis also revealed that the length of the 5´ untranslated region for the three glycophorin A mRNAs was about 53 nucleotides long. Similar lengths of 5´ untranslated region have been reported by other investigators (Tate and Tanner, 1988; Kudo and Fukuda, 1989, Rahuel *et al.*, 1989, Vignal *et al.*, 1990).

4.3. Northern blotting

Nucleotide sequencing of various cDNA clones isolated in the present study showed that they varied in the length of their 3´ untranslated region which suggested that the three glycophorin A mRNAs also differed in this region. Further, primer extension analysis performed around the same time when Northern blotting experiments were performed, revealed that the three glycophorin A mRNA do not

vary in their 5′ untranslated regions. Hence, to confirm these findings of primer extension analysis and nucleotide sequence of the cDNAs I performed Northern blotting experiments using various oligonucleotide probes. Furthermore, since the nucleotide sequence analysis of various cDNAs suggested that the three glycophorin A mRNAs vary in the length of their 3′ untranslated regions, the Northern blotting experiments were also performed to set up the limits of the 3′ untranslated regions in the three glycophorin A mRNAs as described in following sections.

4.3.1. Design of oligonucleotides

The sequences of oligonucleotides GPA-MS, GPA-ML1, GPA-ML2 and GPA-L were complementary to the sequences of 3′ untranslated regions in the largest cDNA clone (λ-gpa6). Additionally, oligonucleotide GPA-N2 that was complementary to the coding sequence and was used for screening the cDNA library was also used for Northern blotting experiments. The selection of these particular sequences was based upon the following rationale.

4.3.1.1. Oligonucleotide GPA-N2

This oligonucleotide (33 nucleotides long) contained sequences complementary to a region of glycophorin A not present in glycophorin B (amino acids 30 to 40, section 1.8.1 and section 3.1). This oligonucleotide was expected to hybridize only with the glycophorin A mRNAs and since the three glycophorin A mRNAs were not expected to vary in their coding regions it was predicted that GPA-N2 would hybridize to all three glycophorin A mRNAs.

4.3.1.2. Oligonucleotide GPA-MS

Nucleotide sequencing of two middle-sized cDNA clones (namely λ-gpa3 and λ-gpa5 of sizes 1.7 kbp) revealed that they had similar sequences with a minor difference in that λ-gpa3 encoded glycophorin A^M while λ-gpa5 encoded glycophorin A^N (section 3.5). Hence they must represent two distinct glycophorin A mRNAs. The similarity in size of clones λ-gpa3 and λ-gpa5 to that of the medium-sized mRNA suggested they were reverse transcribed from this mRNA. Furthermore, the nucleotide sequence of the two clones was identical to that of λ-gpa6 in the 3′ untranslated region up to nucleotide 1667 which was 10 nucleotides downstream from the poly(A) addition signal # 6. If the notion was correct that the two cDNA clones of 1.7 kbp in length represented the middle size mRNA species, then any oligonucleotide complementary to the region beyond poly(A) addition signal #2 extending up to nucleotide 1667 would hybridize with the mRNA of 1.7 and 2.8 kb and not with that of 1.0 kb in length. Therefore, oligonucleotide GPA-MS was synthesized which was complementary to nucleotides 1623 to 1647, four nucleotides 5′ to the poly(A) addition signal #6 in the sequence of λ-gpa6 [Fig. 3.10 (C)].

4.3.1.3. Oligonucleotide GPA-ML1 and GPA-ML2

There were seven poly(A) addition signals in λ-gpa6 and six in λ-gpa3 and λ-gpa5. There were about 200 nucleotides between poly(A) addition signal # 6 and # 7 [Fig. 3.10 (C)]. Since no poly(A) tail was detected in either subclones of λ-gpa3 and

λ-gpa5 it was not possible from nucleotide sequence data to determine if these clones completely represented the 1.7 kb mRNA species or if they lacked some sequences of the 3′ untranslated region including as far as poly(A) addition signal # 7. Therefore, to further define the 3′ boundary of the 1.7 kb mRNA, oligonucleotides GPA-ML1 and GPA-ML2 were synthesized. Oligonucleotide GPA-ML1 was complementary to the λ-gpa6 sequence from nucleotide 1871 to 1894 (five nucleotides 3′ to poly(A) addition signal # 7 and oligonucleotide GPA-ML2 was complementary to the nucleotides 1814 to 1838, 21 nucleotides 5′ to the poly(A) addition signal # 7 [Fig. 3.10 (C)]. If the poly(A) addition signal # 6 was used for producing the medium size (1.7 kb) glycophorin A mRNA then this mRNA would not show hybridization with both oligonucleotides GPA-ML1 and GPA-ML2. Since the small (1.0 kb) mRNA could not have sequences complementary to nucleotides 1814 to 1838 or to 1871 to 1894 it was predicted that it would not show hybridization with GPA-ML1 or GPA-ML2. Obviously the large mRNA (2.8) should show hybridization with the above two oligonucleotides (GPA-ML1 and GPA-ML2). If the 1.7 kb mRNA species was generated by using poly(A) addition signal # 7 both of the larger glycophorin A mRNAs (1.7 and 2.8 kb) would be expected to hybridize with oligonucleotides GPA-ML2 and only the large glycophorin A mRNA would hybridize with GPA-ML1.

4.3.1.4. Oligonucleotide GPA-L

cDNA clone λ-gpa6 was concluded to have been derived from the largest glycophorin A mRNA (section 3.6.2). If the prediction that the medium-sized glycophorin A mRNA was derived by 3´ end processing utilizing poly(A) addition signal # 6 was correct this mRNA (1.7 kb) would not contain sequences down stream from poly(A) addition signal #6 or #7. Hence, oligonucleotide GPA-L recognizing sequences [nucleotides 2056 to 2081, Fig. 3.10 (C)] downstream from the poly(A) addition signal # 7 (nucleotides 1661 to 1666) was not expected to hybridize with the medium size (1.7 kb) glycophorin A mRNA. Therefore, it was predicted that GPA-L would hybridize only with the large glycophorin A mRNA (2.8 kb) and not with the smaller mRNAs (medium, 1.7 kb or small, 1.0 kb).

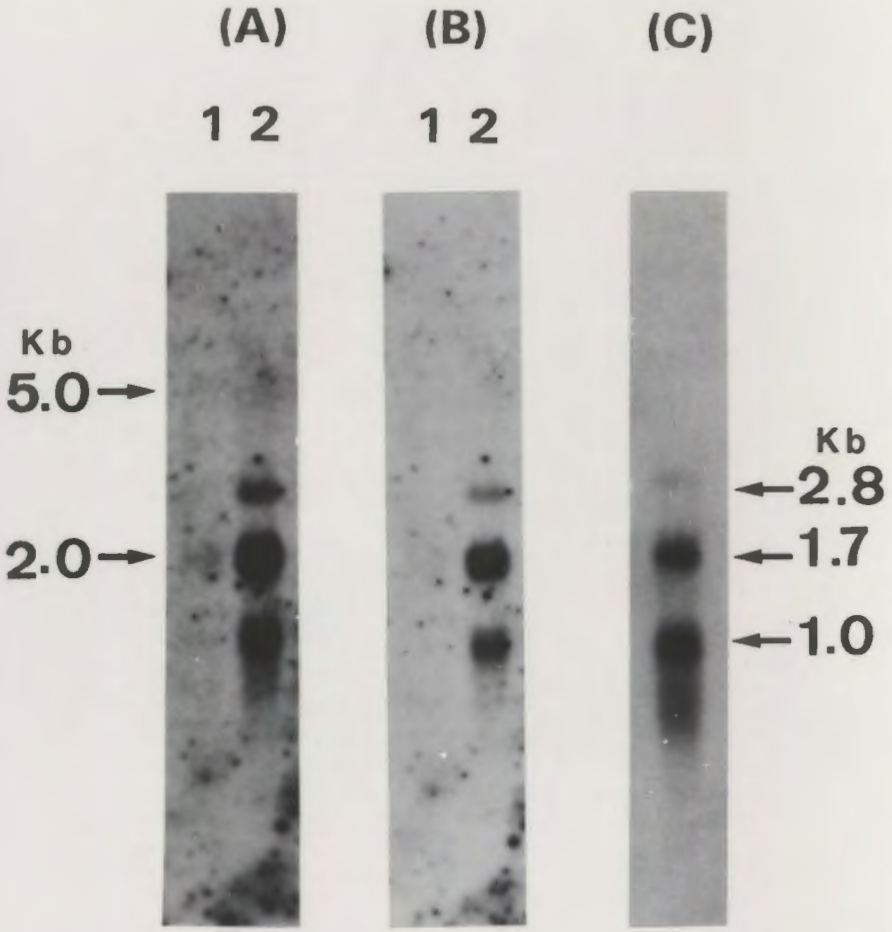
4.3.2. Oligonucleotide/ds cDNA fragment hybridization

4.3.2.1. RNA hybridization with GPA-N2/0.8 kbp *Eco*RI fragment of λ-gpa6

Oligonucleotide GPA-N2 was hybridized with K562 cell RNA at 50°C and washed subsequently for 5 min in 6 X SSC containing 0.1% SDS at 60°C and exposed to X ray film (Kodak X-Omat AR) with two intensifying screens at -70°C for about 18 h. After washing at 60°C and exposure to the film K562 cell poly(A)$^+$ RNA revealed a sharp band and two very broad bands labelled 2.8, 1.7 and 1.0 kb, respectively [Fig. 4.2 (A), lane 2]. In contrast, the poly(A)$^-$ RNA lanes showed a very low level of hybridization of GPA-N2 with 18S and 28S ribosomal RNAs [Fig. 4.2 (A) lane 1]. The membrane was washed once again at 65°C [Fig. 4.2 (B)]. Under these

Fig. 4.2. Northern blotting of K562 cell RNA using as probes oligonucleotide GPA-N2 and the 0.8 kbp *Eco*RI fragment of λ-gpa6.

K562 cell RNA was separated on a 1% agarose gel and blotted on nylon membrane which was then hybridized with ^{32}P-labelled GPA-N2 (A, B), or with ^{32}P-labelled 0.8 kbp *Eco*RI fragment of λ-gpa6 (C). The lanes contained: 1) Poly(A)$^-$ RNA, 2 μg, or 2) poly(A)$^+$ RNA, 10 μg. The single lane in (C) contained poly(A)$^+$ RNA. The final washing temperatures were: A) 60°C, B) and C) 65°C. The numbers 5.0 and 2.0 kb indicate the positions of the 28S and 18S ribosomal RNA, respectively. Positions of the three glycophorin A mRNAs large (2.8 kb), medium (1.7 kb) and small (1.0 kb) are also indicated.


(A)
(B)
(C)
1 2
1 2
Kb
5.0
2.0
Kb
2.8
1.7
1.0

conditions essentially the same result was obtained [Fig. 4.2 (B)]. The strongest signal of the three glycophorin A mRNAs bands was obtained after washing at 60°C but the background hybridization was slightly higher at this temperature than after washing at 65°C. This was clear in the lanes containing poly(A)$^-$ RNA which showed very little hybridization signal with 28S and 18S ribosomal RNAs only after washing at 65°C [Fig. 4.2 (B), lane 1). The washing at 65°C resulted in very little background hybridization but the intensity of specific hybridization signal was also reduced. Nevertheless the background level of hybridization was low enough at 60°C so as not to interfere with the detection of specific signals. Therefore, for subsequent experiments hybridization with GPA-N2 at 50°C was followed by washing at 60°C.

The size of the three RNAs detected by GPA-N2 was calculated by comparing their mobilities with those of the molecular size markers (0.24 to 9.5 kb RNA ladder, Bethesda Research Laboratories, Canada) run in parallel lanes (data not shown). The RNA band largest in size was found to be about 2.8 kb, whereas the size of two smaller RNAs proved to be difficult to measure accurately due to the broadness of the bands representing them. The larger of these two broad bands (labelled 1.7 kb in Fig. 4.2) had a range in size between 1.7 to 2.2 kb depending upon whether the leading or the lagging edge of the band was used to calculate the size. Similarly estimates for the smaller band ranged between the size of 1.0 to 1.2 kb. mRNAs of similar sizes have been reported previously by Siebert and Fukuda (1986b) (i.e 2.8, 1.7 and 1.0 kb) isolated from K562 and human reticulocyte cells. Tate and Tanner

(1988) characterised three mRNAs also from K562 cells with the sizes of 3.2, 2.0 and 1.2 kb. The three mRNAs from human spleen erythroblast were also described to have sizes of 2.2, 1.7 and 1.0 kb (Rahuel *et al.*, 1989). It is clear that although these various groups were dealing with the same three mRNAs encoding glycophorin A, their estimation of size was not very accurate, perhaps due to inherent difficulty in calculation of the accurate size of the broad RNA bands. I calculated the size of the large mRNA to be about 2.8 kb and clearly not 3.2 or 2.2 kb. Furthermore, the two smaller species, as mentioned above had at least some component of 1.7 and 1.0 kb as presented by Siebert and Fukuda (1986b). Therefore, for these reasons and also because the three mRNAs were first reported by Siebert and Fukuda (1986b) their nomenclature will be used through out this dissertation, i.e. the large mRNA of 2.8 kb, medium of 1.7 kb and small of 1.0 kb.

Densitometric scanning of the Northern blot shown in Fig. 4.2 revealed that the medium size glycophorin A mRNA was the most abundant (60%) followed by small (30%) and the large mRNAs (10%). Similar proportions for the three glycophorin A mRNAs was also observed when another blot containing poly(A)$^+$ RNA also from K562 cells was probed with the 0.8 kbp *Eco*RI fragment of λ-gpa6 [Fig. 4.2 (C)]. The smear below the 1.0 kb mRNA [Fig. 4.2 (C)] was most probably due to the glycophorin B mRNA which was reported to be of size about 0.6 kb (Siebert and Fukuda, 1986a, b) or 0.7 kb (Tate and Tanner, 1988). Glycophorin A and B share a considerable sequence homology both at amino acid as well as at the

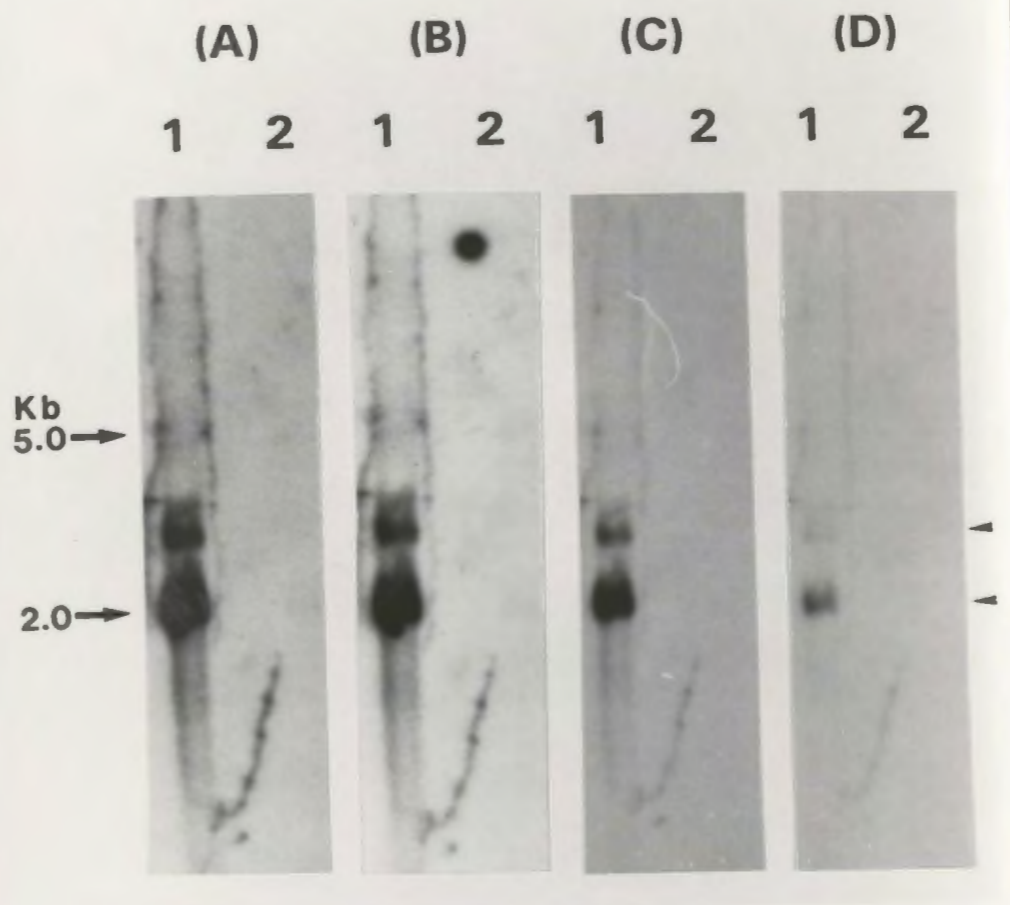
nucleotide sequence levels (Blanchard *et al.*, 1987; Siebert and Fukuda, 1987, also see section 3.1). The 0.8 kbp *Eco*RI fragment of λ-gpa6 which included the 24 nucleotides of the 5′ untranslated region as well as full coding region for glycophorin A might be expected to hybridize with the glycophorin B mRNA. Further, the smear below the 1.0 kb band could also be due to hybridization of the 0.8 kbp fragment with the transcript of a recently isolated novel glycophorin gene called E or *inv* which shares some sequence homology with glycophorin A, and which was reported to be of 0.7 kb in size (Kudo and Fukuda, 1989). Alternatively, the smear below the small mRNA band might represent the degradation products of glycophorin A mRNAs.

4.3.2.2. Hybridization with GPA-MS

Oligonucleotide GPA-MS which was complementary to the nucleotides 1623 to 1647 in the cDNA (λ-gpa6), terminating just four nucleotides 5′ to the poly(A) addition signal # 6 [Fig. 3.10 (C)] was hybridized with K562 cell RNA at 50°C and the final stringent washes were performed at 50, 53, 57, and 62°C in 6 X SSC containing 0.05% SDS [Fig. 4.3 (A), (B), (C) and (D), respectively]. Under these conditions, when GPA-MS was hybridized with the K562 poly(A)$^+$ RNA it revealed a broad band in the size range of about 1.7 kb and a less abundant sharp band of about 2.8 kb but no 1.0 kb mRNA [Fig. 4.3 (A), lane 1]. These bands retained the signal after washing at higher temperatures [Fig. 4.3 (B), (C) and (D), lanes "1"]. This oligonucleotide did not show hybridization with the poly(A)$^-$ RNA to any great extent [Fig. 4.3 (A), (B), (C) and (D), lanes "2"] even at the lowest washing temperature

Fig. 4.3. Northern blotting of K562 cell RNA using as probe ^{32}P-labelled oligonucleotide GPA-MS.

K562 cell RNA was separated on a 1% agarose gel and blotted on a nylon membrane which was then probed with ^{32}P-labelled GPA-MS. The lanes contained: 1) Poly(A)$^+$ RNA, 10 μg, or 2) poly(A)$^-$ RNA, 2 μg. The final washing temperatures were: A) 50°C, B) 53°C, C) 57°C and D) 62°C. The numbers 5.0 and 2.0 kb indicate the positions of the 28S and 18S ribosomal RNA. The positions of the two glycophorin A mRNAs of 2.8, 1.7 kb showing specific hybridization are also indicated by arrow heads on the right hand side of the figure.


(A)
(B)
(C)
(D)
1 2 1 2 1 2 1 2
Kb
5.0
2.0

(50°C, Fig. 4.3 lane 2). Therefore, it was concluded that the hybridization of oligonucleotide GPA-MS was specific and the bands observed in poly(A)$^+$ RNA lanes were due to specific hybridization with the glycophorin A mRNAs of 2.8 and 1.7 kb.

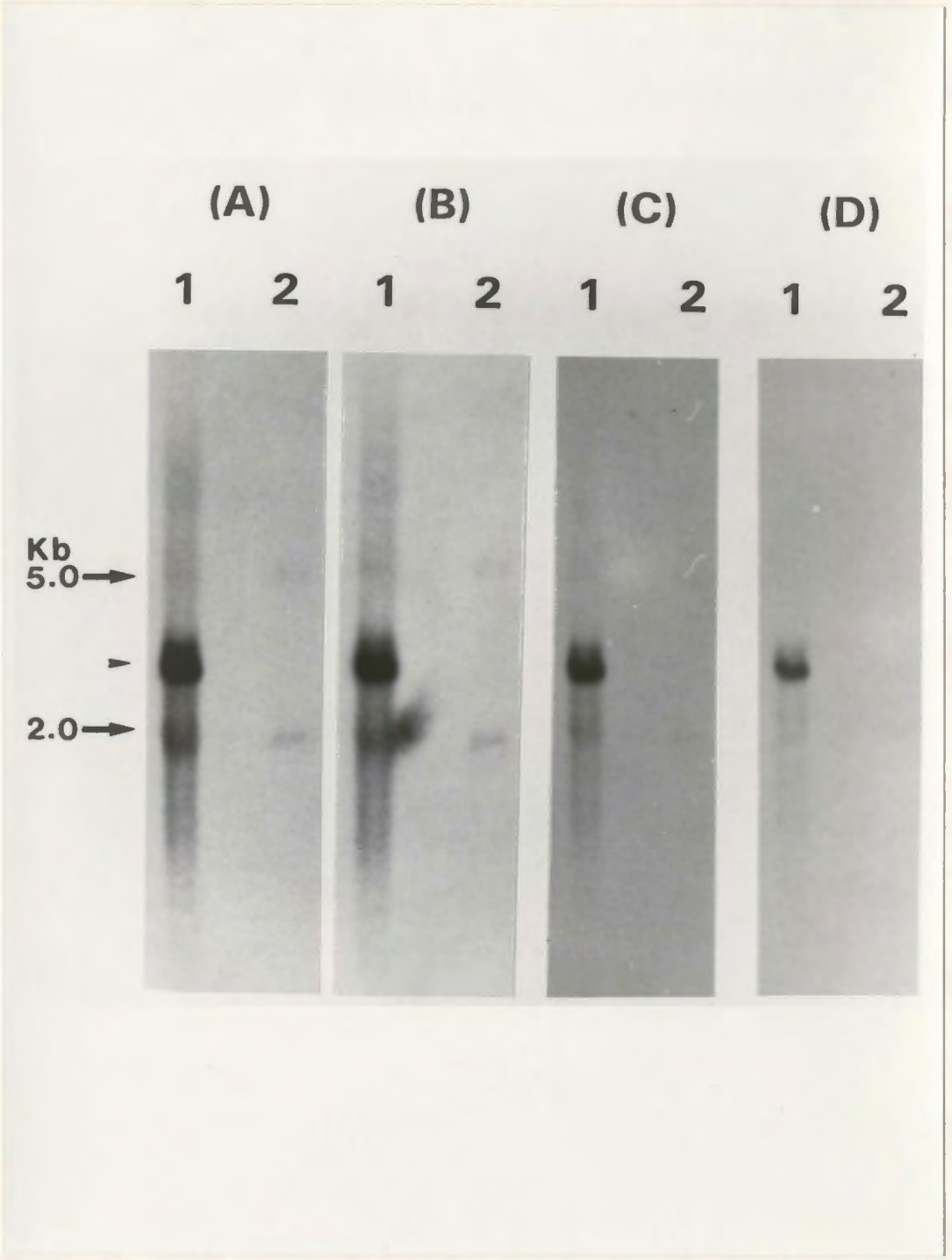
The hybridization of GPA-MS with the medium-sized glycophorin A mRNA suggested that this mRNA contained the sequences at least up to the poly(A) addition signal # 6, close to nucleotide 1647 (GPA-MS). The lack of hybridization to the smallest glycophorin A mRNA confirmed the prediction that this RNA did not contain the sequences complementary to oligonucleotide GPA-MS.

4.3.2.3. Hybridization with GPA-ML1 and GPA-ML2

Oligonucleotides GPA-ML1 and GPA-ML2 were hybridized with K562 cell RNA at 50°C and the final stringent washes were performed at 50, 53, 57, and 62°C in 6 X SSC containing 0.05% SDS. The oligonucleotide GPA-ML1 which was complementary to nucleotide 1871 to 1894 [beginning 5 nucleotides 3´ to poly(A) addition signal # 7] in the cDNA λ-gpa6 sequence, hybridized with the 2.8 kb mRNA and, in addition, with a much less abundant component of about 2.0 kb [Fig. 4.4 (A), (B), (C) and (D), lanes "1"]. This oligonucleotide showed very little hybridization with the poly(A)$^-$ RNA [Fig. 4.4 (A), (B), (C) and (D), lanes "2"] indicating that this oligonucleotide hybridized specifically with the poly(A)$^+$ RNA. The successive washes at increasing temperature showed that the 2.8 kb component retained the signal of hybridization to a great extent, while the 2.0 kb component showed a very low level

Fig. 4.4. Northern blotting of K562 cell RNA using as probe oligonucleotide GPA-ML1.

K562 cell RNA was separated on a 1% agarose gel and blotted on a nylon membrane which was then hybridized with ^{32}P-labelled GPA-ML1. The lanes contained: 1) poly(A)$^+$ RNA, 10 μg, or 2) poly(A)$^-$ RNA, 2 μg. The final washing temperatures were: A) 50°C, B) 53°C, C) 57°C and D) 62°C. The numbers 5.0 and 2.0 kb indicate the positions of the 28S and 18S ribosomal RNA. The position of the 2.8 kb mRNA is also indicated by the arrow head on the left hand side of the figure.

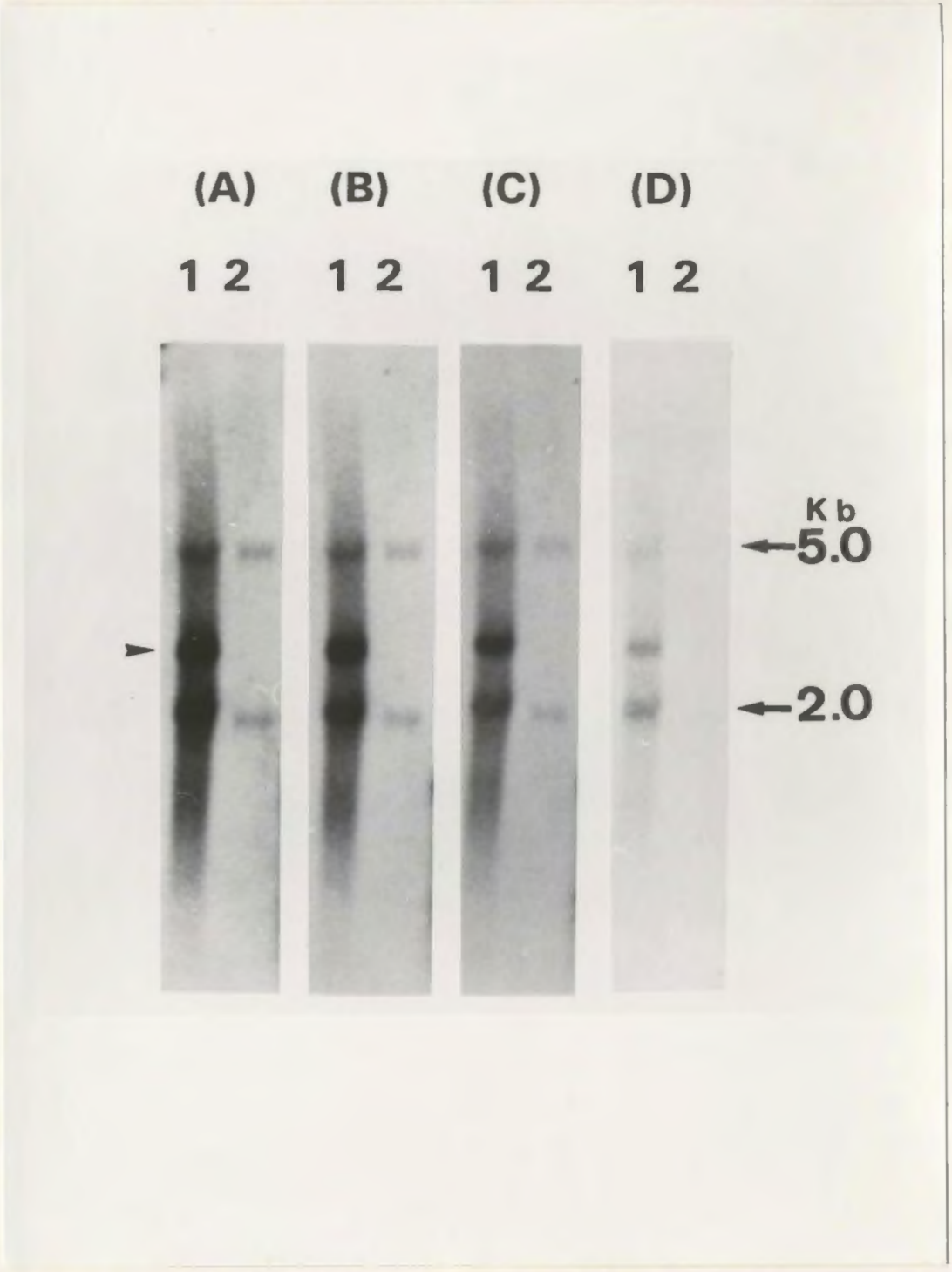

(A)
(B)
(C)
(D)
1
2
1
2
1
2
1
2
Kb
5.0
2.0

of hybridization signal after washing at 57°C [Fig. 4.4 (C), lane 1]. Therefore, this hybridization experiment was taken to conclude that the 2.8 kb mRNA included sequences up to at least nucleotide 1894 but the bulk of the medium-sized mRNA lacked these sequences.

Oligonucleotide GPA-ML2 which was complementary to nucleotides 1814 to 1838 [terminating 21 nucleotides 5´ to the poly(A) addition signal # 7] hybridized with the 2.8 kb mRNA and, in addition, to less abundant 2.0 kb and 5.0 kb components but not with the 1.0 kb mRNA [Fig. 4.5 (A), (B), (C) and (D), lanes "1"]. The hybridization of this oligonucleotide with the 2.0 kb component was much stronger than that observed with GPA-ML1 (compare Fig. 4.4 and 4.5, lanes "1"). The poly(A)$^-$ RNA also showed a detectible level of signal [Fig. 4.5 (A), (B) and (C), lanes "2"]. The same hybridization pattern was retained after each successive higher temperature wash except that the 5.0 kb component lost signal to a greater extent than the 2.0 kb component [Fig. 4.5 (B), (C) and (D), lanes "1", also see section 4.3.2.5] . In contrast, the poly(A)$^-$ lanes (Fig 4.5, lanes "2") showed that the 2.0 and 5.0 kb components retained the signal in a equal proportion which suggested that both 18S and 28S ribosomal RNAs hybridized with GPA-ML2 at a low level. As concluded above for GPA-ML1, hybridization with GPA-ML2 was taken to indicate that the 2.8 kb RNA contained sequences at least as far as nucleotide 1838 but that most of the medium size mRNA did not contain sequences up to nucleotide 1814 to 1838.

Fig. 4.5. Northern blotting of K562 cell RNA using as probe oligonucleotide GPA-ML2.

K562 cell RNA was separated on a 1% agarose gel and blotted on a nylon membrane which was then hybridized with ^{32}P-labelled GPA-ML2. The lanes contained: 1) poly(A)$^+$ RNA, 10 μg or 2) poly(A)$^-$ RNA, 2 μg. The final washing temperatures were: A) 50°C, B) 53°C, C) 57°C and D) 62°C. The numbers 5.0 and 2.0 kb indicate the positions of the 28S and 18S ribosomal RNA. The position of the glycophorin A mRNAs of 2.8 kb is also indicated by the arrow head on the left hand side of the figure.

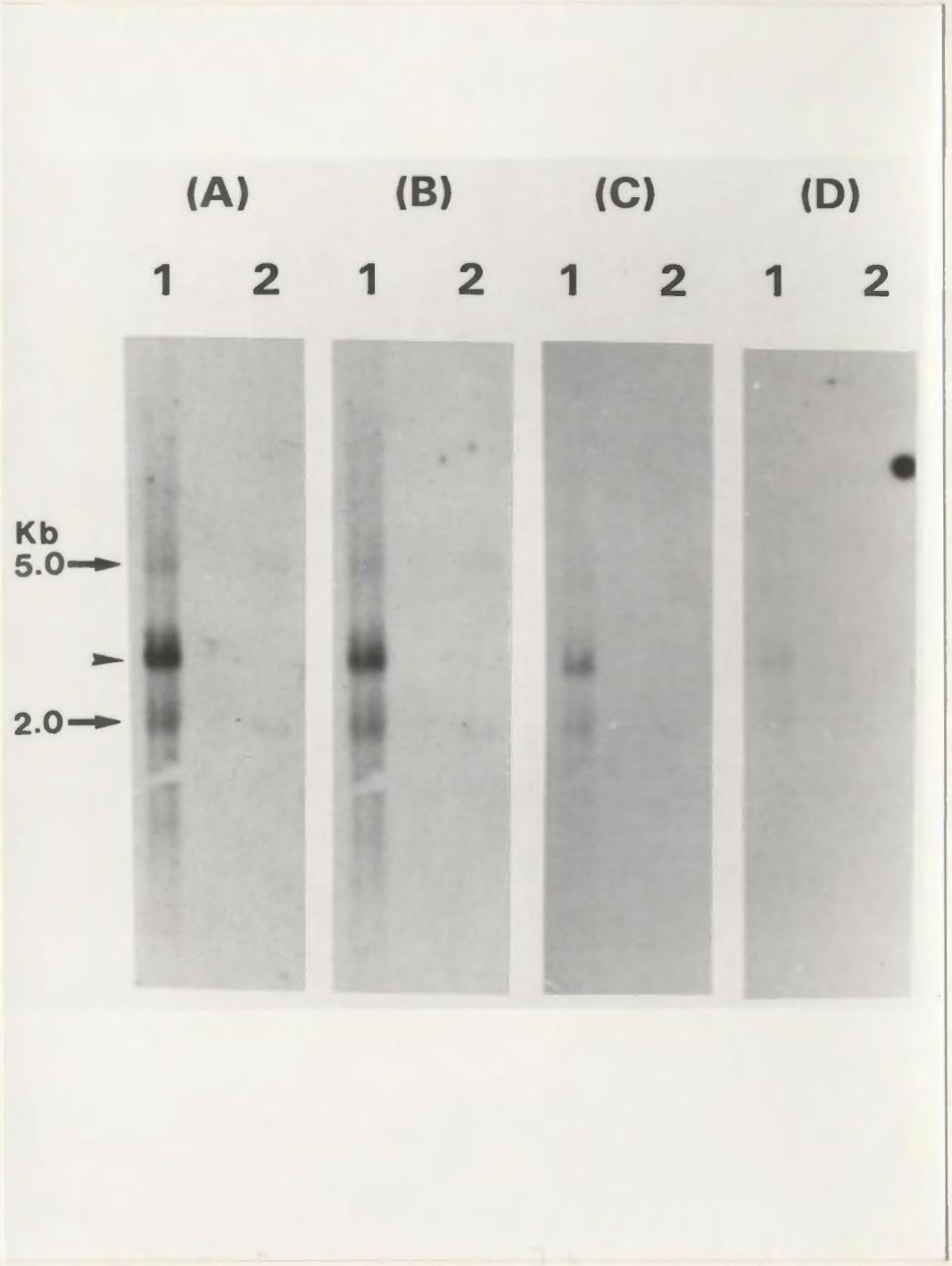

(A)
(B)
(C)
(D)
1 2
1 2
1 2
1 2
Kb
5.0
2.0

4.3.2.4. Hybridization with GPA-L

Oligonucleotide GPA-L which was complementary to nucleotides 2056 to 2081 in the cDNA (λ-gpa6) was hybridized with K562 cell RNA at 50°C and the final stringent washes were performed at 50, 53, 57, and 62°C in 6 X SSC containing 0.05% SDS. Under these conditions, GPA-L hybridized with the 2.8 kb mRNA species and with two much less abundant components of about 2.0 and 5.0 kb [Fig. 4.6 (A), (B), (C) and (D), lanes "1"]. Similar 2.0 and 5.0 kb components showed very low levels of hybridization in poly(A)$^-$ K562 cell RNA lanes when the membranes were washed at 50°C; this indicated that GPA-L did not hybridize with 18S and 28S ribosomal RNAs to a great extent [Fig. 4.6 (A), (B), (C) and (D), lanes, "2"]. The hybridization pattern in the lanes containing poly(A)$^+$ RNA remained the same after washing at higher temperatures except that the 5.0 kb component lost most of the signal after washing at 57°C [Fig. 4.6 (C), lane 1]. Furthermore, the intensity of the hybridization signal of GPA-L with the 2.0 kb component in poly(A)$^+$ lanes was less than with oligonucleotides GPA-ML1 and GPA-ML2 (compare Fig. 4.4 to 4.6, lanes"1"). The results of hybridization of oligonucleotide GPA-L were taken to indicate that the large glycophorin A mRNA contained sequences including and beyond the nucleotides 2056 to 2081. Moreover, it was also concluded that the bulk of medium size mRNA did not contain the sequence complementary to this oligonucleotide.

Fig. 4.6. Northern blotting of K562 cell poly(A)$^+$ RNA using as probe oligonucleotide GPA-L.

K562 cell RNA was separated on a 1% agarose gel and blotted on a nylon membrane which was then hybridized with ^{32}P-labelled GPA-L. The lanes contained: 1) poly(A)$^+$ RNA, 10 μg or 2) poly(A)$^-$ RNA, 2 μg. The final washing temperatures were: A) 50°C, B) 53°C, C) 57°C and D) 62°C. The numbers 5.0 and 2.0 kb indicate the positions of the 28S and 18S ribosomal RNA. The position of 2.8 kb mRNA is also indicated by the arrow head on the left hand side of the figure.


(A)
(B)
(C)
(D)
1
2
1
2
1
2
1
2
Kb
5.0
2.0

The following conclusions can be drawn from the results of Northern blotting: (i) the hybridization of the various oligonucleotides suggested that the structures of these cDNAs accurately reflected the structures of the mRNAs; (ii) the 1.0 kb mRNA did not extend as far as nucleotides 1623 to 1647 (GPA-MS, Fig. 4.3); (iii) most of medium size (1.7 kb) mRNA extended beyond nucleotides 1623 to 1647 (GPA-MS, Fig. 4.3) but not as far as far as 1814 to 1838 (GPA-ML2, Fig. 4.5) or 1871 to 1894 (GPA-ML1, Fig. 4.4); (iv) A minor component of about 2.0 kb recognized by GPA-ML1, GPA-ML2 and GPA-L (Fig. 4.4, 4.5 and 4.6) might be present; and (v) the 3´ region of the 2.8 kb mRNA extended beyond nucleotide 2105 (i.e. beyond the end of clone λ-gpa6).

4.3.2.5. The nature of the 2.0 kb component

Oligonucleotides GPA-ML1, GPA-ML2 and GPA-L detected minor components of sizes 2.0 and 5.0 kb which suggested non-specific hybridization of these oligonucleotides with 18S and 28S ribosomal RNAs since similar 2.0 and 5.0 kb component were also detected by Northern blot analysis of poly(A)$^-$ RNA. (Fig. 4.4 to 4.6, lanes "2"). However, the hybridization with the 2.0 kb component could not be attributed only to the hybridization with 18S ribosomal RNA as the level of hybridization with this component which varied for each oligonucleotide used was found to be different from that observed with the 5.0 kb component. Therefore, it was possible that the 2.0 kb component was in fact a fraction of a mixture of glycophorin A mRNA species designated medium-sized (1.7 kb). The highest level

of hybridization with this 2.0 kb component was observed with oligonucleotide GPA-ML2 (Fig. 4.5) followed by GPA-ML1 (Fig. 4.4) and GPA-L (Fig. 4.6), decreasing in that order. This indicated that a much larger fraction of the medium size mRNA species contained the sequences complementary to oligonucleotide GPA-ML2 than to GPA-ML1 or GPA-L. This implied that although the most of the medium size RNA did not contain sequences complementary to all three oligonucleotides some fraction of the mRNA did contain sequences complementary to these oligonucleotides, i.e. the 1.7 kb mRNA species was found to be heterogenous. Consistent with this idea was the observation that 1.7 kb mRNA was present as a broad band (Fig. 4.2). Furthermore, the different level of hybridization of the three oligonucleotides with the 2.0 kb component also suggested that the fractions of the medium size RNA containing the sequences complementary to these oligonucleotides was not equal, i.e. the fraction containing sequences complementary to GPA-ML2 was significantly higher than to GPA-ML1 and GPA-L.

4.4. Discussion

4.4.1. Investigations on differences among the three glycophorin A mRNAs

In eucaryotic cells differential processing of the primary transcript can give rise to multiple mRNAs that differ from each other in their coding regions or in their 5´- or 3´-untranslated regions (Leff *et al.*, 1986). Nucleotide sequencing of various cDNA clones, primer extension analysis of K562 cell mRNA and Northern blotting experiments indicated that although the three glycophorin A mRNAs had similar

nucleotide sequences they differed in the length of their 3′ untranslated region. The following is a description of the possible differences in primary structures of the three glycophorin A mRNAs.

4.4.1.1. Differences in the 5′ untranslated region

Nucleotide sequencing showed that the six cDNA clones encoding glycophorin A isolated in the present study did not differ significantly in their 5′ untranslated region. (section 3.5). The small differences in the number of nucleotides in the 5′ untranslated regions could not contribute to a significant degree to the size differences observed in three classes of cDNA. Assuming that the all three glycophorin A mRNAs were represented by the six cDNAs it was concluded that the three glycophorin A mRNAs also did not vary in the length of their 5′ untranslated region, a conclusion supported by primer extension analysis of K562 cell RNA as described next.

Primer extension analysis (section 4.1) revealed that the three glycophorin A mRNAs had similar 5′ untranslated region of about 53 nucleotides suggesting that the 5′ untranslated region of glycophorin A mRNAs did not contribute to the large size differences seen among the three glycophorin A mRNAs. These findings were consistent with a single initiation site for the synthesis of the three glycophorin A mRNAs, a conclusion that was also supported by other published reports. For instance, S1 nuclease mapping showed the presence of a single initiation site for the

glycophorin A gene isolated from a K562 cell genomic library (Kudu and Fukuda, 1989). The same report concluded that the 5´ untranslated region was 55 nucleotides long. An identical length (55 nucleotides) was reported for the 5´ untranslated region of a glycophorin A cDNA isolated from a human reticulocyte library (Tate and Tanner, 1988). Rahuel *et al.* (1989) also obtained an identical length (55 nucleotides) for the 5´ untranslated regions by primer extension analysis of the three glycophorin A mRNAs isolated from spleen erythroblasts. The difference of two nucleotides in the length of the 5´ untranslated region obtained in the present work and that reported by others could be attributed to sequencing gel artifacts. For instance, overexposure of the autoradiographs might have caused an error in determining the accurate length of the diffused band representing the primer extended product. Therefore, it was concluded that a similar 5´ untranslated region of 55 nucleotides was present in the three glycophorin A mRNAs. It is interesting to note that a similar length was also reported for the 5´ untranslated region of another sialoglycoprotein, glycophorin B, which is presumably derived from an ancestral gene in common with glycophorin A (Kudo and Fukuda, 1989).

4.4.1.2. Differences in the coding region

Nucleotide sequence analysis of the 0.8 kbp EcoRI fragments of the six cDNA clones did not reveal any differences in the segment corresponding to the coding region for glycophorin A except those representing the polymorphism of glycophorin A^M and A^N (section 3.5). If the assumption that all three mRNAs were represented

cDNA clones isolated in this work was correct, then the natural conclusion would be that the three glycophorin A mRNAs also had similar coding sequences. This assumption was supported by hybridization of the 0.8 kbp EcoRI fragment of cDNA clone λ-gpa6, and of oligonucleotide GPA-N2 to K562 cell RNA on Northern blots (section 4.2.2.1). Hybridization of other oligonucleotides also with the K562 cell RNA (Siebert and Fukuda, 1986b) supported this notion. Moreover, the nucleotide sequence I found and that reported by others for cDNAs isolated from a K562 cell cDNA library (Siebert and Fukuda, 1986a, b), a human reticulocyte cDNA library (Tate and Tanner, 1988) and a human fetal liver cDNA library (Rahuel *et al.*, 1988) were found to be virtually identical in the coding region and in the 5′ and 3′ untranslated regions (also see section 3.6.1) supporting the assumption that all three glycophorin A mRNAs shared common sequences. Finally, the amino acid sequence of glycophorin A has been known for a long time (Tomita and Marchesi, 1975; Tomita *et al.*, 1978) but apart from the polymorphism of the MN blood group locus (Dahr *et al.*, 1977; Wasniowska *et al.*, 1977; Blumenfeld and Adamany, 1978; Furthmayr *et al.*, 1978; Lisowska and Wasniowska, 1978) no other major distinct forms of glycophorin A have been reported. Therefore, it was concluded that all three mRNAs have identical coding regions thereby excluding this region as the basis for the observed size differences.

4.4.1.3. Differences in the 3′ untranslated region

The nucleotide sequence analysis of the six cDNA clones isolated in the present study showed that they varied in the length of their 3′ untranslated region (section 3.5). Furthermore, seven poly(A) addition signals were detected in the nucleotide sequence of the largest cDNA clone, λ-gpa6. The presence of multiple poly(A) addition signals and various lengths for the 3′ untranslated regions in the cDNA clones suggested that these poly(A) addition signals could be used differentially to produce the three glycophorin A mRNAs; this was possible only if the mRNAs also differed in the length of their 3′ untranslated regions. A close examination of poly(A) addition signals revealed that they were located in positions suitable to generate mRNAs of 1.0 and 1.7 kb. Obviously, the largest size cDNA (λ-gpa6) of only 2.1 kb could not encode the complete nucleotide sequence of the large, 2.8 kb mRNA. The Northern blotting experiments with oligonucleotides GPA-MS, GPA-ML1, GPA-ML2 and GPA-L (section 4.2.2.2, 4.2.2.3 and 4.2.2.4) supported the assumption that the three glycophorin A mRNAs differed in the length of their 3′ untranslated regions. Recently, Rahuel *et al.* (1989) reported similar results by investigation of cDNA clones isolated from a human fetal liver cDNA library.

4.4.2. Proposed primary structures of three glycophorin A mRNAs

Oligonucleotide hybridization with K562 cell RNA helped in defining the 3′ boundary of the three glycophorin A mRNAs. Based upon these results, the primary structures for the three glycophorin A mRNAs are proposed (Fig. 4.7). These

Fig.4.7. Proposed primary structures of the three glycophorin A mRNAs.

This figure depicts the composite nucleotide sequence of the three glycophorin A mRNAs generated from the sequence of λ-gpa6 (capital letters). The stretch of "N" (see text) and nucleotides presented in small letters are from Kudo and Fukuda (1989). The proposed boundaries of the 3´ untranslated region of the two smaller mRNAs (1.0 and 1.7 kb) for glycophorin A are marked with arrows. The poly(A) addition signals (bold) are denoted by #A1 to #A7 and #An. Other sequence elements: instability sequence, AUUUA; and poxvirus termination signals UUUUUNU, marked T1, T2 and T3 are underlined. The sequence corresponding to the conserved motif CAYTG is marked with double underlines.

5' 1 aguugucuuugguaguuuuuuugcacuaacuuCAGGAACCAGCUCAUGAUCUCAGG

56 AUGUAUGGAAAAAUAAUCUUUGUAUUACUAUUGUCAGCAAUUGUGAGCAUAUCAGCAUCAAGUACCACUGGUGUGGCAAUGCACACUUCA

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

147 ACCUCUUCUUCAGUCACAAAGAGUUACAUCUCAUCACAGACAAAUGAUACGCACAAACGGGACACAUAUGCAGCCACUCCUAGAGCUCAU

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

237 GAAGUUUCAGAAAUUUCUGUUAGAACUGUUUACCCUCCAGAAGAGGAAACCGGAGAAAGGGUACAACUUGCCCAUCAUUUCUCUGAACCA

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

327 GAGAUAACACUCAUUAUUUUUGGGGUGAUGGCUGGUGUUAUUGGAACGAUCCUCUUAAUUUCUUACGGUAUUCGCCGACUGAUAAAGAAA

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

417 AGCCCAUCUGAUGUAAAACCUCUCCCCUCACCUGACACAGACGUGCCUUUAAGUUCUGUUGAAAUAGAAAAUCCAGAGACAAGUGAUCAA

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

507 UGAGAAUCUGUUCACCAAACCAAAUGUGGAAAGAACACAAAGAAGACAUAAGACUUCAGUCAAGUGAAAAAUUAACAUGUGGACUGGACA

END

#A1

597 CUCC**AAUAAA**UUAUAUACCUGCCUAAGUUGUACAAUUUCAGAAUGCAAUUUUCAUUAUAAUGAGUUCCAGUGACUCAAUGAUGGGGAAAA

687 AAAUCUCUGCUCAUUAAUAUUUCAAGAUAAAGAACAAAUGUUUCCUUGAAUGCUUGCUUUUGUGUGUUAGCAUAAUUUUUAGAAUUGUUU

777 GAGAAUUCUGAUCCAAAACUUUAGUUGAAUUCAUCUACGUUUGUUUAAUAUUAACUUAACCUAUUCUAUUGUAUUAUAAUGAUGAUUCUG

#A2

867 UCAAAUGAAAGGCUUGAAAUACCUAGAUGAAGUUUAGAUUUUCUUCCUAUUGUAAACUUUUGAGUCUGGUUUCAUUGUUUUA**AAUAAA**UU

957 AAGGGGACACUAAAGUCCUAUCAUUCAUUCCUUCAUUCUGAACAGGCAAGAUAUAAUAUUACAUGAAUGAUUACUAUAUUUUGUUCACAC

#A3

1047 U**AAUAAA**GCUUAUGCUCAGAAAUGCCAUACACACACACAAACACACACAUUUAUCAUUUAAUGCAUAAAUCAACACAAAAGGUUUUCCCA

1137 UUAAUAUGAAAUAUUACAUAUAUAUAAGUGCCAUAUUUAAAAUAAUUUGUCUAACAGUAGAACUAUGUCGGAGCACUCACUGAAGCUUCG

1227 AUUUCCCACUGAAAGAGUUAUUUGUUGUAAGUAGAGUUAUCCCGGAGAAGGAAAAAGAACUUACGACCUUUCUUUAUAACAGAAAGCUCA

1317 ACUCUAAAUUCAACAAGAUGUGCAAACCGGACAUGCAGGUGAAUAUUUUAAUAGGUUACUAUAAGGUUCUCAAUUAAAUUCUUUAAUCUG

#A4

1407 UCCAGUCCCAGUUUCUCUUAUU**AAUAAA**ACUUGGAAAUUGCUUUAAACCAUUUAAAGGAAAUUUCUAGAUAUAGAAACUAAAGGACUGU

1497 GACUAUACAGUGUCACUCAUUUGUAGUAAAACUUAAAAAGCAAAAACAAAAAACAAAAAAGACCUUCCUGUGAUACUUUAUUUCCGAACU

#A5 T1

1587 **AAUAAA**AAUCUAUAUGACUUUUUAUUAUUGUGUGAUAACCAAGUAAAUGUUUUCUAUUUUCGAUAUUUUCAGGCAUGGUAACAGAAAUUU

#A6 T2

1677 ACCUUUU**AAUAAA**UUAAAAAAUCUAAAUUUUAACCUACUUGUAUGUUCGGAGAGUGUUUUUGUACUAUAUUGACUACUUAAAAUAGAGAA

1767 UGAGACUAAGAAGGGAACAUUUCUGUUGAUACAUGUUUUUUAAAAGUAAUUUUUAAGAGCAUUAUUAGGUUAAUUUAAUCCAAUUAAUGA

#A7

1857 CCCAAAUGCCAAGGUAAUUUUAAAUUUACAUUUUU**AAUAAA**AGCAACAUGUUGAAACAAGAGAGGGUGAGAUUAACCUUUUUGCUAAAGU

1947 AAUUUACAAGUCAAAGACAGGAAGAGAUCAGAGUGAAUGUGCCUUCUUAACCAGAGCUACAGAAUUUAGUGAAUAAUUAAAGUACAAACU

2037 GCUUUGACCUCCUUGAACUUUUCCAAGCAAUUUCUCUGUACUUCUAUAUAUGAAUGUCUUAGCCAAUUUUCUGCUACUAUAACAGAAUAC

2127 GACAGACUGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNuuauuuucuuccuaguucuggag

2217 gcugggaaggcgaagggcauggcacugacaucugccuuguaacugaugagaaccuucuuacugcaugauaacaaagcagcaaggcaagca

2307 aaagcguaagaugaagagagaggaaaugaagccaaacacauccuuucaucagaagcccauucccucuauaaggcguuacuacauuuauga

2397 gaauggaguccucaugaccuaaucgugaccuuaaaggccccucccaacacuguuacaauggcaauuaaauuucaacaaagguuccagagg

T3

2487 ugacauucgaaucagcaaugaaauuuucauaguuaaauugguauucgugggggaagaaaugaccauuucccuuguauuuuuauaauuaa

#An

2577 aucagcaaaauauugu**aauaaa**gaaaucuuucc $(A)_n$ 3'

proposed structures incorporate the oligonucleotide hybridization results as well as the cDNA sequences reported in the present work and by others (Siebert and Fukuda, 1986a, b; Rahuel *et al.*, 1988; Tate and Tanner, 1988) as well as genomic DNA sequences presented by others (Kudo and Fukuda, 1989; Vignal *et al.*, 1990). The evidence for these structures will be discussed individually for each of the three glycophorin A mRNAs in the following sections.

4.4.2.1. Small (1.0 kb) mRNA

Only oligonucleotide GPA-N2 hybridized to the smallest (1.0 kb) glycophorin A mRNA (Fig. 4.2-4.6). This was not surprising since GPA-N2 recognized nucleotides 168 to 201 in the coding region of glycophorin A whereas the other oligonucleotides recognized different parts of the 3´ untranslated region of the glycophorin A cDNA (Fig. 3.10, and section 3.1). Furthermore, an mRNA of size 1.0 kb would not be expected to contain sequences complementary to nucleotides 1623 to 1647 (GPA-MS) and beyond (complementary to oligonucleotides GPA-ML1, GPA-ML2 and GPA-L) unless the three glycophorin A mRNAs were generated by alternative splicing. However, alternative splicing was ruled out as a mechanism for the generation of the three glycophorin mRNAs, based upon the organization of the glycophorin A gene (Kudo and Fukuda, 1989; Vignal *et al.*, 1990) and the absence of internal differences in the three glycophorin A mRNAs (section 4.4.1).

Hybridization of the 0.8 kbp fragment of λ-gpa6 with all three glycophorin A

mRNAs suggested all three contained sequences similar to the 0.8 kbp fragment of λ-gpa6 [i.e. from nucleotide 1 to nucleotide 747 in Fig. 3. 10 (C), the 3′ boundary of the 0.8 kbp fragment of λ-gpa6]. The idea that all three glycophorin A mRNAs contained similar sequences up to nucleotide 747, about 200 nucleotides downstream beyond the poly(A) addition signal # 1 [at nucleotides 569 to 574, Fig. 3.10 (C)] was consistent with the experimental results reported by Rahuel *et al.* (1988) who used an oligonucleotide complementary to nucleotides 901 to 931 [Fig. 3.10 (C)], 17 bases upstream from the poly(A) addition signal # 2 [at nucleotides 917 to 922, Fig. 3.10 (C)] to probe a Northern blot containing erythroblast RNA and found that the oligonucleotide hybridized efficiently with all three glycophorin A mRNAs and hence provided evidence that the sequences up to nucleotide 931 were present in all three glycophorin A mRNAs.

The boundary of the 3′ untranslated region of the small mRNA can be deduced from cDNA sequences reported by other investigators (Rahuel *et al.*, 1988; Tate and Tanner, 1988). Rahuel *et al.* (1988) published a partial glycophorin A cDNA sequence comprising 854 nucleotides with a 70 nucleotide long poly(A) tail attached. This cDNA sequence lacked the 5′ untranslated region and the nucleotides encoding the three NH_2-terminal amino acids of glycophorin A. Tate and Tanner (1988) reported a cDNA sequence comprising 983 nucleotides also containing a poly(A) tail of 16 nucleotides. The 3′ end of both of these clones corresponded to nucleotide 935 in Fig. 3.10. (C), 13 nucleotides beyond the end of the poly(A) addition signal # 2.

The similarity in size of these clones to that of the smallest glycophorin mRNA and the presence of a poly(A) tail suggested that they represented the complete nucleotide sequence of the small glycophorin A mRNA (1.0 kb). Therefore, it is concluded that the 3′ boundary of the smallest glycophorin A mRNA corresponds to nucleotide 935, in the sequence of λ-gpa6 [Fig. 3.10. (C)] which is equivalent to nucleotide 967 in Fig. 4.7 (marked by arrow).

4.4.2.2. Medium (1.7 kb) mRNA

Hybridization of the medium-sized (1.7 kb) glycophorin A mRNA with the oligonucleotide GPA-MS provided evidence that these mRNAs contained sequences corresponding to nucleotides 1623 to 1647 [GPA-MS sequence, Fig. 3.10 (C)]. The lack of hybridization of oligonucleotides GPA-ML1, GPA-ML2 and GPA-L with the major fraction of the 1.7 kb mRNA species indicated that it did not contain the sequences complementary to nucleotides 1814 to 1838 (GPA-ML2) and beyond [GPA-ML1, GPA-L, Fig. 3.10 (C)]. Therefore, the bulk of medium size glycophorin A mRNA contains sequences up to and including the region covered by GPA-MS [nucleotide 1647, Fig. 3.10 (C)], close to poly(A) addition signal #6. The exact location of the 3′ boundary of the medium-sized glycophorin A mRNA species could not be predicted accurately. The reason is that so far no poly(A)-containing cDNA that falls into the size range of medium-sized glycophorin A mRNA has been isolated thus making it difficult to define the exact 3′ boundary of this mRNA. However, as discussed before (section 3.6.2 and 4.2.1), two of the cDNAs, λ-gpa3 and λ-gpa5

ended exactly at nucleotide 1667, 10 nucleotides downstream from poly(A) addition signal # 6 [Fig. 3.10 (C)]. Therefore, it is tentatively proposed that nucleotide 1667 corresponding to nucleotide 1700 in Fig. 4.7 (arrow), or a few nucleotides downstream from this defines the 3′ boundary of the bulk of medium-sized mRNA. The conclusion that the medium size mRNA does not contain a significant number of nucleotides beyond poly(A) addition signal #6 was also supported by the findings of Rahuel *et al.* (1989) who concluded that poly(A) addition signal # 6 was used for 3′ end processing to produce the medium-sized glycophorin A mRNA, based also on oligonucleotide hybridization with Northern blots of human spleen erythroblast RNA.

4.4.2.3. Large (2.8 kb) mRNA

As predicted on the basis of the nucleotide sequence of λ-gpa6, oligonucleotides GPA-MS, GPA-ML1, GPA-ML2 and GPA-L hybridized with the large-sized (2.8 kb) glycophorin A mRNA. This showed that the large mRNA contained sequences up to nucleotides 2056 to 2081 recognized by GPA-L, [Fig. 3.10 (C)] and beyond. It is obvious that the complete nucleotide sequence of a 2.8 kb mRNA was not represented in any of the cDNAs, the largest, λ-gpa6, being of about 2.1 kb. Therefore, the large-sized (2.8 kb) glycophorin A mRNA must extend in the 3′ direction beyond the region covered by any of the cDNAs reported so far including the present work.

The gene for glycophorin A was reported to be organized in seven exons

(Kudo and Fukuda, 1989; Vignal *et al.*, 1990), the first six exons include the 5′ untranslated region, the signal peptide and all of the coding region except for about four amino acids near the COOH-terminus. Nucleotides for these remaining four amino acids and for all of the 3′ untranslated region are found in a single exon of about 2.1 kb (termed 3′ UT, Kudo and Fukuda, 1989). Therefore, this exon contains sufficient nucleotides to generate the large (2.8 kb) glycophorin A mRNA when spliced with the first six exons. A partial sequence of this exon was deposited in the EMBL data bank (# M24133) by Kudo and Fukuda(1989). The sequence contained 416 nucleotides at the 3′ end of the 3′ UT exon but did not contain all of the 5′ end. My sequence for λ-gpa6 contained the 5′ end of the 3 UT exon but none of the sequence I determined overlapped with that reported by Kudo and Fukuda (1989). However, in the partial sequence of the 3′ UT exon of Kudo and Fukuda (1989), I noted a poly(A) addition signal (termed # An, Fig. 4.7). Further, the study on glycophorin A gene organization also reported the 3′ boundary of the glycophorin A gene (Kudo and Fukuda, 1989; Vignal *et al.*, 1990).

Therefore, based upon the information on the 3′ end of the glycophorin A genes in these two reports and the presence of a single exon of about 2.1 kb for all of the 3′ untranslated region (Kudo and Fukuda, 1989; Vignal *et al.*, 1990), I predict that the 3′ end of the large mRNA would be in close proximity to polyadenylation signal, #An. Combining the 1628 nucleotides in the 3′ untranslated region in λ-gpa6 with the 416 nucleotides for 3′ UT present in the EMBL data bank (# M24133)

gives a length of 2044 nucleotides out of the 2100 reported to be present in the 3′ UT exon (Kudo and Fukuda, 1989; Vignal *et al.*, 1990). This leaves a gap of about 66 nucleotides in the middle of the 3′ untranslated region of the large glycophorin A mRNA whose sequence has not been reported; this sequence is represented as a stretch of "N"s (Fig. 4.7). The actual number of these nucleotides would depend on the total number of nucleotides in the 3′ UT exon. I have included 66 N residues to make the 3′ UT exon 2.1 kb in size as reported (Kudo and Fukuda, 1989; Vignal *et al.*, 1990). Therefore, based upon the structure presented here (Fig. 4.7), the largest glycophorin A mRNA would contain about 2600 nucleotides, together with a poly(A) tail of about 200 to 250 nucleotides it would be close to the size of large glycophorin A mRNA of size 2.8 kb.

In summary, the three glycophorin A mRNAs have similar 5′ untranslated regions and coding sequences. However, the three mRNAs differ in the length of their 3′ untranslated regions as depicted in Fig. 4.7. The smallest mRNA (1.0 kb) includes sequences about 13 nucleotides downstream from poly(A) addition signal # 2. The medium size mRNA (1.7 kb) has sequences up to a few nucleotides 3′ beyond poly(A) addition signal # 6. The largest glycophorin A mRNA (2.8 kb) contains all the nucleotides represented by λ-gpa6 reported here plus the remaining part of the 3′ UT exon which, however, is not represented in any of the cDNA clones.

The above discussion on the primary structure of glycophorin A mRNAs

revealed two important features of these mRNAs. First, these investigations revealed the presence of long 3′ untranslated regions of more than 458, about 1200 and 2100 nucleotides in the small, medium and large size mRNA, respectively. The length of the 3′ untranslated region in the largest glycophorin A mRNA was more than 4.5 times the combined length of the 5′ untranslated region and coding region of glycophorin A. The significance of the 3′ untranslated region is discussed in the following section (4.4.3).

The second important finding was that the three glycophorin A mRNAs contained multiple poly(A) addition signals. Since no differences were found in any region of the three mRNAs except in the 3′ untranslated regions, it was natural to assume that these multiple poly(A) addition signals may have a significant role in the production of multiple glycophorin A mRNAs. A close examination of these poly(A) addition signals revealed that they were conveniently located so as to produce mRNAs which varied in the length of their 3′ untranslated region as discussed in section 4.4.5.

4.4.3. Functions of 3′ untranslated region

The presence of 3′ untranslated regions is a common feature of eucaryotic mRNAs. Many genes such as the rabbit and human β-globin genes produce mRNAs with a very short 3′ untranslated region, containing only 95 and 135 nucleotides of 3′ untranslated region, respectively (Littauer and Soreq, 1982). Other genes produce

mRNAs with very long 3´ untranslated regions. For example, the mRNA for human estrogen receptor (Keaveney *et al.*, 1989) and for rat liver Golgi mannosidase II contain more than 4000 nucleotides of 3´ untranslated sequence. The 3´ untranslated region of mRNAs is believed to serve one or more important roles. Evidence accumulated in recent years implicates this region in post-transcriptional regulatory mechanisms, such as the control of mRNA half life, and in translation. These possible roles will be discussed in the following sections.

4.4.3.1. Regulation of mRNA half life

Many eucaryotic mRNAs contain sequences in their 3´ untranslated region that are involved in controlling their half lives. Human β- and δ-globin have very different half lives, the latter turns over four to six times faster than the former (Ross, 1988). Comparison of the nucleotide sequences of these mRNAs revealed that they differed by more than 50% in their 3´ untranslated region but by only about 8% in other regions. It was shown that these differences in the 3´ untranslated region sequences accounted for their different half lives (Ross, 1988).

The mRNAs for proto-oncogenes, c-myc, c-fos, some growth factors and cytokines have very short half lives. In contrast, as mentioned above, the β-globin gene produces a very stable mRNA. Transfection with chimeric genes containing sequences from stable and unstable mRNAs demonstrated experimentally the importance of 3´ untranslated regions in the regulation of mRNA stability. A

chimeric mRNA with 5′ c-fos sequences linked to 3′ β-globin sequences was stable but unstable when 5′ β-globin sequences were linked to the 3′ untranslated region of c-fos mRNA. This observation indicated that the 3′ untranslated region of c-fos contained sequences responsible for its instability (Triesman, 1985). A 75 nucleotide long AU-rich element was identified in the 3′ untranslated region of c-fos mRNA, that was responsible for conferring instability to normally very stable β-globin mRNA (Shyu *et al.*, 1989).

Caput *et al.* (1986) had first detected the presence of a consensus sequence (TTATTAT) present in 3′ untranslated regions of genes for human and mouse tumour necrosis factors, human lymphotoxin and colony stimulating factor, human and mouse interleukin 1, human and rat fibronectin and most of the human and mouse interferons (reviewed in Reeves and Magnuson, 1990). A conserved AU-rich sequence that conferred instability on transcribed mRNAs has been identified and experimentally verified in mRNAs of many transiently expressed genes (Reeves *et al.*, 1985; Buetler *et al.*, 1986; Shaw and Kamen, 1986). This highly conserved stretch of DNA frequently consisted of tandem repeats of the tetranucleotide, TATT [i.e.$(TATT)_n$] transcribed as the cognate $(UAUU)_n$ sequence in the 3′ untranslated regions of mRNAs of most of the known cytokine genes, for example, interferons and interleukins; colony stimulating factors, for example, granulocyte-macrophage colony stimulating factor; and the proto-oncogenes, c-fos and c-myc (Reeves and Magnuson, 1990). The AT-rich motifs in 3′ untranslated regions of some of these genes were

more highly conserved than the protein coding regions. This tendency of the conservation of 3′ untranslated regions over the protein coding regions of homologous genes emphasizes the importance of this region (Caput *et al.*, 1986; Reeves and Magnuson, 1990). Further investigations with various chimeric constructs containing instability sequences from the aforementioned genes and various other genes demonstrated that the poly (UAUU) sequence but not an AU-rich sequence with a different base order, $(UUAU)_n$, confers instability on mRNAs (Reeves and Magnuson, 1990). In many of the unstable mRNAs, the instability sequence $(UUAU)_n$ was present in multiple copies, resulting in formation of the AUUUA motif (i.e. UUAUUUAU--, Shyu *et al.*, 1991).

Further studies on the mechanism responsible for conferring instability by AU-rich sequences revealed that for degradation of c-fos proto-oncogene mRNA, continuing translation was also required (Wilson and Treisman, 1990). It was shown that soon after mRNA synthesis, the poly(A) tail was removed leading to eventual degradation of the transcribed sequences of the mRNA. Therefore, it was concluded that the AU-rich sequences acted to destabilize the mRNA by directing rapid removal of the mRNA poly(A) tail (Wilson and Treisman, 1990). Another study on the role of the AU-rich sequence in destabilization of c-fos mRNA showed that the AU-rich sequence controlled two steps: removal of the poly(A) tail which did not require an intact pentanucleotide, AUUUA within the AU-rich sequence; and

subsequent degradation of the mRNA that seemed to be dependent on AUUUA (Shyu *et al.*, 1991).

Like c-fos, c-myc mRNA is very unstable and its instability determinant was also localized to a region of about 140 nucleotides in its 3´ untranslated region (Jones and Cole, 1987). The region consisted of two AU-rich elements containing AUUUA sequences (Laird-Offringa *et al.*, 1991). Studies on intermediates of the degradative pathway of this mRNA showed that the AU-rich sequences were directly involved in the degradation of c-myc mRNA (Brewer and Ross, 1988). It was suggested that for degradation of this mRNA, the poly(A) binding protein migrated from the poly(A) possibly to AU-rich sequences making the naked poly(A) tail vulnerable to nuclease attack (Brewer and Ross, 1988). The observation that the degradation of the c-myc mRNA was initiated with poly(A) shortening which generated a pool of oligoadenylated mRNAs [i.e mRNAs with larger variation in length of poly(A) tail than is observed when the mRNA is newly synthesized] supported the suggestion that poly(A) binding protein migrated from the poly(A). The poly(A) shortening was followed by degradation of the oligoadenylated mRNA pool generating decay products with 3´ termini located within the AU-rich sequences of the 3´ untranslated region; this was claimed to provide evidence that the AU-rich sequences were directly involved in mRNA degradation (Brewer and Ross, 1988). The direct involvement of AU-rich sequences in the degradation of c-myc mRNA was further supported by the observation that an endonuclease, endoribonuclease V digested various cytokine

transcripts at AUUUA sequences (Jochum *et al.*, 1990).

Although the mechanism responsible for the AUUUA-dependent instability of mRNAs containing AUUUA sequence is not fully understood, the recent isolation of specific factors that bind stably to the AUUUA elements in cytokines and proto-oncogenes may provide some clues to this phenomenon (Malter, 1989; Bohjanen *et al.*, 1991; Brewer, 1991; Vakalopoulou *et al.*, 1991). Malter (1989) isolated an AUUUA-specific mRNA-binding protein from the cytoplasm of a lymphocyte cell line, Jurkat cells. Stable binding of the protein factor to mRNAs required four reiterated AUUUA elements (Malter, 1989). In contrast, a 32 kilodalton factor isolated from HeLa cells required only one AUUUA sequence in an AU-rich context for efficient binding to the 3´ untranslated region of granulocyte-macrophage colony stimulating factor, c-fos and c-myc mRNAs as well as to a similar domain downstream of the poly(A) addition site of adenovirus IVa2 mRNA (Vakalopoulou *et al.*, 1991). Binding of the 32 kilodalton factor to AUUUA-containing mRNAs reduced their accumulation suggesting that binding of the factor marked the mRNAs for degradation (Vakalopoulou *et al.*, 1991). It is possible that there is some sort of interaction between the poly(A) binding protein and the AUUUA-binding factor, resulting in the degradation of mRNAs containing the instability sequence, AUUUA. As mentioned above, the poly(A) binding protein migrates to the c-fos AU-rich element and this in turn results in the degradation of that mRNA.

The 3′ untranslated regions in the three glycophorin A mRNAs are AU-rich and contain eight AUUUA elements at various locations in the sequence of λ-gpa6 and two more in the 3′ UT exon deposited in the EMBL data bank (# M24133, Kudo and Fukuda, 1989, underlined in Fig. 4.7). Most probably the AUUUA sequence is not functional in the regulation of glycophorin A mRNA levels, because this sequence is present in the two larger glycophorin A mRNAs (i.e. of 1.7 and 2.8 kb) and not in the smaller mRNA (i.e. of 1.0 kb) yet the 1.7 kb mRNA is the most abundant of the three glycophorin A mRNAs.

Secondary structure in the 3′ untranslated region of mRNAs also play an important role in regulation of the expression of some genes, particularly at the level of mRNA stability. For example, the regulation of transferrin receptor by iron (Owen and Kuhn, 1987) and the cell cycle-dependent histone gene expression (Schumperli, 1988) both involve secondary structures in the 3′ untranslated regions of their respective mRNAs. Transferrin receptor level is increased by treatment of cells with iron and decreased when the cells expressing transferrin receptor are treated with iron chelators. The control of expression of transferrin receptor gene was found to be post-transcriptional. The regulatory region was localized in a 2.3 kb segment within the 3′ untranslated region of transferrin receptor mRNA (Owen and Kuhn, 1987). Further investigations revealed the existence of two distinct domains in the 3′ untranslated region both of which were essential for regulation of the transferrin receptor level in mouse L cells. One of the two regulatory domains consisted of a

stem and loop structure of about 60 nucleotides and the other contained five repeats of a palindromic sequence (Mulner and Kuhn, 1988). A specific protein factor was found to bind to the stem and loop structure in the 3′ untranslated region and was involved in the regulation of transferrin receptor mRNA levels (Mulner, 1989).

A different secondary structure is found in the cell cycle-dependent histone mRNAs. These mRNAs are very unstable and are not polyadenylated (Schumperli, 1988) but end in a stem-loop structure. The half lives of histone mRNAs in S-phase is about one hour which decreases to about 10 to 15 minutes at the end of S-phase (Sittman *et al.*, 1983). The instability determinant of the histone H4 mRNA was localized within the 3′ untranslated region, was composed of a stem and loop structure and was a target of a 3′ to 5′ exonuclease activity (Ross *et al.*, 1986). There is no evidence in the literature suggesting that the secondary structures in transferrin receptor and histone mRNAs have any similarities. However, both sequence motifs achieve the same end result i.e. regulate their respective levels post-transcriptionally by decreasing their half lives.

4.4.3.2. Regulation of mRNA translation

The 3′ untranslated region of mRNA in addition to providing the translation-enhancing affects of poly(A) tails (section 1.14.6.1.3) also has been implicated in the regulation of translation through transcribed sequences [i.e. sequences other than poly(A), reviewed in Jackson and Standart, 1990; Munroe and Jacobson, 1990].

Littauer and Soreq (1982) studied translation in *Xenopus* oocytes of human fibroblast β_1 and β_2 interferon mRNAs lacking a substantial portion of the 3´ untranslated region and poly(A) tails. They concluded that the poly(A) tail and the 3´ untranslated region did not affect translational efficiency or stability of these mRNAs in *Xenopus* oocytes (Littauer and Soreq, 1982). However, they found that unfractionated mRNAs lacking a substantial portion of the 3´ untranslated region and poly(A) tails also from human fibroblasts showed decreased translational efficiency and stability *in vitro*. Therefore, it was concluded that the discrepancy of results came from using two different translation systems (i.e. in *Xenopus* oocytes and *in vitro*).

Recently, Kruys *et al.* (1989) showed that the AU-rich sequence (section 4.4.3.1) in addition to destabilizing some mRNAs such as interferon, granulocyte-macrophage colony stimulating factor and c-fos mRNAs also decreased their translation. It was demonstrated that a 62 nucleotide-long segment in the β-interferon mRNA 3´ untranslated region, that was 85% AU-rich and contained several UUAUUUAU sequences was responsible for the decrease in its translational efficiency. It was further shown by studies in *Xenopus* oocytes, that the decrease in the translational efficiency was unrelated to the role of AU-rich sequences in decreasing the stability of the β-interferon mRNA. Only a single UUAUUUAU element was needed to decrease the translational efficiency (Kruys *et al.*, 1989).

Translational regulation by the 3´ untranslated region also plays an important

role in the control of the expression of many genes during spermatogenesis, including mouse phosphoglycerate kinase-2 gene, mouse and trout protamine genes and rat nuclear transition protein 1 gene (Braun *et al.*, 1989). Mouse protamine 1 gene is transcribed exclusively in haploid round spermatids (Hecht *et al.*, 1986) and the resulting protamine 1 mRNA is stored for about one week before it is translated in elongating spermatids (Balhorn *et al.*, 1984 ; Kleene *et al.*, 1984). Braun *et al.* (1989) demonstrated that 156 nucleotides of the 3´ untranslated region of mouse protamine 1 mRNA were sufficient to confer mouse protamine 1-like translational regulation to human growth hormone gene in a chimeric transgene containing the promoter and transcriptional regulatory elements of the mouse protamine 1 gene, the human growth hormone structural gene and the mouse protamine 1 3´ untranslated region.

Conservation of the 3´ untranslated region of creatine kinase-B gene across many species also suggested its importance which was shown to be in translational control (Ch'ng *et al.*, 1990). However, in this case the mechanism seems to be different from that observed in developmentally regulated genes discussed above (Ch'ng *et al.*, 1990). A cell line, U937 had undetectable creatine kinase-B activity yet had the creatine kinase-B mRNA associated with polysomes, suggesting that most of the creatine kinase-B mRNA was blocked in this cell line at a step subsequent to the ribosome binding. It was further shown that binding of a protein factor to the 3´ untranslated region of creatine kinase-B mRNA was responsible for the repression of translation (Ch'ng *et al.*, 1990).

Biosynthesis of ferritin, the iron storage protein, provides another example of translational control. When the iron supply is low, the ferritin mRNA is stored in cells and is translated at a very low rate (Dickey *et al.*, 1988). When the cellular iron levels increase, the ferritin mRNA moves from the nonpolyribosomal RNA fraction to the polyribosomal RNA fraction and its rate of translation increases 40 to 50 fold. The translational repression of ferritin mRNA required a 70 nucleotides long fragment in the 3′ untranslated region (Dickey *et al.*, 1988). Comparison of nucleotide sequences of eight ferritin mRNAs from humans, rats, chicken and frogs revealed that the 3′ untranslated region was conserved among these species and each mRNA had sequences that were involved in base pairing to form secondary structures (Dickey *et al.*, 1988).

Recently Jackson and Standart (1990) reviewing the roles of poly(A) tails and 3′ untranslated regions of various mRNAs suggested that they influence translation to different extents and perhaps serve as modulators of translational efficiency rather than as indispensable elements. It was further suggested that events at the 3′ end of the mRNA, such as the polyadenylation state or protein-3′ untranslated region interaction, could influence upstream events such as translation (re)initiation. This interaction between the two ends of mRNA could take place in the form of a rearrangement between alternative secondary structural forms affecting whole mRNA. Alternatively, there might be a direct interaction simply between 5′-proximal and 3′-proximal sequences (Jackson and Standart, 1990).

4.4.4. Polyadenylation site choice and 3´ end formation

Many genes produce multiple forms of proteins as a result of alternative polyadenylation in combination with alternative splicing and (or) different initiation site usage. Examples of such genes include immunoglobulin heavy chain, adenovirus late transcription unit (reviewed in Moore and Sharp, 1984), human fibrinogen and interleukin-2 receptor, rat calcitonin and prolactin, chicken myosin light chain, and *Drosophila* myosin alkali light chain and glycinamide ribotide transformylase (reviewed in Leff *et al.*, 1986). There are examples of genes that use alternate polyadenylation sites to produce multiple mRNAs which differ in the length of their 3´ untranslated region without affecting the final product. Examples of this type of gene include human insulin receptor (Goldstein and Kahn, 1989) and N-ras (Hall and Brown, 1985), rat insulin-like growth factor-I (Hepler *et al.*, 1990), $\alpha 2\mu$-globulin (Unterman *et al.*, 1985) and malic enzyme (Morioka *et al.*, 1989), mouse proto-oncogene, c-rel, (Grumont and Gerondakis, 1990), chicken vimentin (Capetanaki *et al.*, 1983), and X gene (Heilig *et al.*, 1980), mouse α amylase and dihydro folate reductase (Moore and Sharp, 1984) and yeast Cox 6 (Wright *et al.*, 1989). Evidence presented in the present work and by others suggested that glycophorin A gene also falls in this category.

In addition to the regulation at the transcription level, differential processing can provide further control for the maintenance of a particular level of a protein product. There is evidence that the selection of a poly(A) addition site can be a regulated event controlling gene expression. In many cells, differential processing of

pre-mRNA is related to the developmental stage, for example, the selection of adenovirus late poly(A) addition sites and the immunoglobulin μ and δ poly(A) addition sites. Differentially processed mRNAs can also be expressed in different cell types, for example, the calcitonin and calcitonin gene-related peptide (Darnell *et al.*, 1990).

It has been suggested that either the presence of tissue specific factors (in the case of calcitonin/calcitonin gene-related peptide) or factors appearing during a developmental pathway (adenovirus and immunoglobulin) are responsible for selective poly(A) addition site usage (Hart *et al.*, 1985). Calcitonin and calcitonin gene-related peptide are formed preferentially in thyroid and brain, respectively (Leff *et al.*, 1987). Complete primary transcripts are produced in both cell types; formation of final mRNA is cell-specific and involves poly(A) addition site choice and splice site choice. However, in this case it is directed more by splice site commitment than by poly(A) addition site choice (Leff *et al.*, 1987).

Rat insulin like growth factor-I gene produces two mRNAs (7.0-7.5 kb and 0.9-1.2 kb) differing in the length of their 3´ untranslated region (Hepler *et al.*, 1990). These two mRNAs were found to be very AU-rich in their 3´ untranslated regions. The longer transcript had a shorter half-life suggesting that the two mRNAs produced from a single insulin-like growth factor-I gene of rat were regulated differentially. Yeast Cox 6 produces three classes of mRNAs which vary in the length of their 3´

untranslated region, all three were shown to be functionally active as demonstrated by their presence in the polysomal fraction (Wright *et al.*, 1989). It was further shown by glucose repression/derepression that the largest Cox6 mRNA and the two smaller RNAs were regulated differentially (Wright *et al.*, 1989).

In switching between mRNAs encoding the membrane-bound (m) and secreted (s) forms of IgM heavy chain (μ), the developmentally regulated production of μm and μs mRNAs is dependent on the relative efficiency of utilization of a poly(A) addition site and splice site. From mutation studies of μs and μm poly(A) addition sites, Peterson and Perry (1989) concluded that the splice site choice was the major determining factor in the selection of the μs and μm poly(A) addition sites.

In adenovirus, a temporal switch in processing affects the mRNA population (Akusjarvi and Persson, 1980; Shaw and Ziff, 1980; Nevins and Wilson, 1981). The early transcription unit 3 (E3) from adenovirus 2 produces a series of mRNAs which are transcribed from a unique initiation site but which differ in polyadenylation and splice sites. The 5′ region of E3 contains a polyadenylation site which is very rarely used in producing early mRNA transcripts. However, this same polyadenylation site is utilized efficiently to produce late mRNA transcripts. The switching of early to late transcription was shown to be accomplished by cleavage in the 3′ region of E3 transcription unit inactivating a splice site. This allowed the less active polyadenylation site to be utilized efficiently for the production of late mRNAs and

hence induced early to late mRNA switching (Gillinaro *et al.*, 1988).

The late transcription unit of adenovirus contains five mRNA families distinguished by the polyadenylation elements (L1 to L5). All of the polyadenylation signals (L1 to L5) are used in late viral infection. Falck-Pedersen and Logan (1989) characterized the regulation of polyadenylation site choice using the reconstituted E1A gene as a site for insertion of major late transcription unit poly(A) addition sites (L1 and L3) and reported that different poly(A) addition sites present in a single mRNA precursor were used preferentially on the first come first served rule basis early in infection. It was suggested in mRNAs with many functional polyadenylation elements regardless of their numbers, the first poly(A) addition site (i.e. 5′ proximal) was selected; this restriction to the use of the first poly(A) addition site was not due to a transcription termination or to a sequence-specific factor binding, rather it was mediated by cis interaction (Falck-Pedersen and Logan, 1989).

Iwasaki and Temin (1990) showed that the efficiency of RNA 3′ end formation in spleen necrosis virus which contains two different poly(A) addition sites was dependent upon the distance between the cap site and the poly(A) addition site. When this distance was shorter than 500 nucleotides, only 3 to 9% of the RNA was polyadenylated at the second, 3′ poly(A) addition site. However, when the distance between the cap site and the poly(A) addition site was 1400 nucleotides or more, 70% of total RNA was polyadenylated at the 3′ poly(A) addition site. In comparison

to spleen necrosis virus, the poly(A) addition signals of thymidine kinase and simian virus 40 late genes functioned at high efficiency, even when the distance between the cap site and the poly(A) addition site was small enough to inactivate the 3′ poly(A) addition site of the spleen necrosis virus (Iwasaki and Temin, 1990). The differences in the efficiency of utilization of the poly(A) addition sites in these different mRNAs indicated that many alternative mechanisms are responsible for poly(A) addition site selection.

To investigate the mechanism involved in the selection of poly(A) addition site in transcription units containing multiple poly(A) addition signals, Denome and Cole (1988) constructed a series of plasmids containing multiple poly(A) addition signals downstream from the herpes simplex type 1 thymidine kinase coding region. The poly(A) addition signals used for construction of these plasmids were from simian virus 40 late region and the herpes simplex virus thymidine kinase gene in addition to a very poor signal from the simian virus 40 early region. These plasmids were transfected in Cos-1 cells and assays on the cytoplasmic RNA showed that: a) all poly(A) addition signals were used in all constructs; b) increasing the distance between two signals caused an increase in the use of the 5′ signal and a decrease in the use of the 3′ signal. Four possible models were described: i) The polyadenylation signal is recognized by a soluble factor associating randomly with pre-mRNA; ii) the polyadenylation site is recognized by some factor associated with RNA polymerase II; iii) the polyadenylation site is recognized by a factor scanning 5′ to 3′ along the

RNA precursor independent of the RNA polymerase II, and iv) polyadenylation sites are recognized by a factor scanning 3′ to 5′ along the nascent transcript (Denome and Cole, 1988).

The previous discussion illustrates the complexities of mechanisms involved in differential use of poly(A) addition sites to generate multiple mRNAs. It can be a simple first-come first-served choice as observed in the case of adenovirus late transcription unit (Falck-Pedersen and Logan, 1989), or could also involve the complexity of splice site choice as well as poly(A) addition site choice as observed in the case of calcitonin/calcitonin gene related peptide (Leff *et al.*, 1987).

4.4.5. The mechanism of generation of three glycophorin A mRNAs

The data presented in this report show that the three glycophorin A mRNAs differ in the length of their 3′ untranslated regions (section 4.4.1, Fig. 4.7). In this section I will discuss the possible mechanism of production of these three glycophorin A mRNAs transcribed from a single gene. Additionally, I will also discuss possible mechanisms responsible for differential abundance of the three glycophorin A mRNAs.

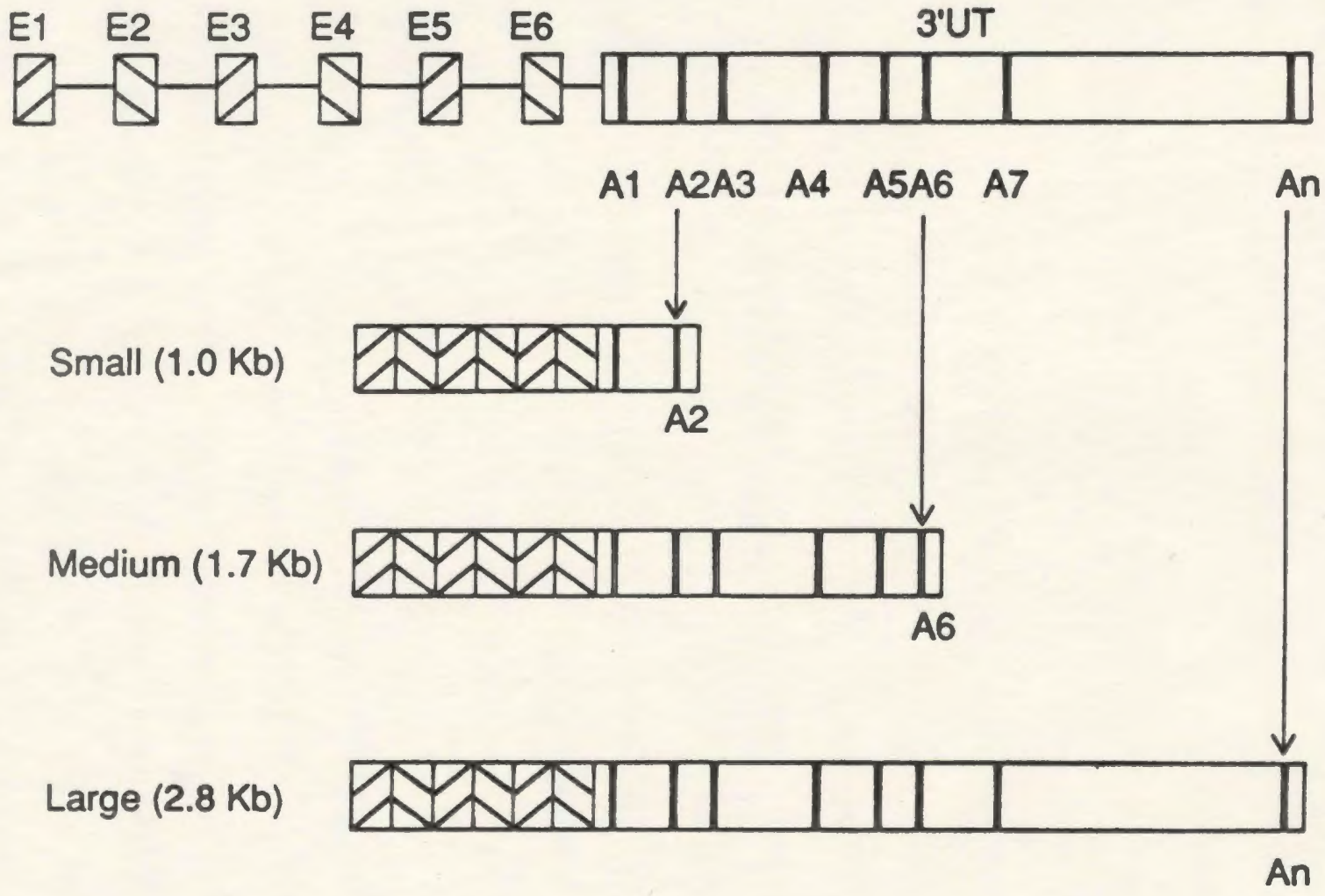
The presence of a single exon for all of the 3′ untranslated region (section 4.4.2.3) and the presence of an identical sequence in the three glycophorin A mRNAs (section 4.4.2) ruled out the possibility of alternative splicing which would generate

multiple mRNAs having internal differences. Furthermore, the nucleotide sequence of the cDNAs in the present work and that by others revealed multiple poly(A) addition signals. Multiple mRNAs that differ in the length of their 3′ untranslated region, without any internal sequence difference can be produced from a single gene utilizing multiple poly(A) addition signals or by alternate termination. I propose that the three glycophorin A mRNAs are produced by differential utilization of these poly(A) addition signals during the processing of the 3′ end of glycophorin A pre-mRNA.

The mechanism proposed for the generation of three glycophorin A mRNAs is depicted in Fig. 4.8. According to the mechanism, a single primary transcript is produced by initiation of transcription of the glycophorin A gene at a unique site followed by termination at the end of the glycophorin A gene at an appropriate site. The primary transcript of the glycophorin A gene in addition to intron sequences must contain the complete sequence of the seven glycophorin A exons. The primary transcript of the glycophorin A gene can undergo differential processing at the 3′ end to generate the three glycophorin A mRNAs which differ only in the length of their 3′ untranslated region. For instance, if poly(A) addition signals 2, 6 and An are selected for 3′ end processing of the primary transcript, the small- (1.0 kb), medium- (1.7 kb) and large-sized (2.8 kb) glycophorin A mRNAs, respectively, would be produced (Fig. 4.8).

Fig. 4.8. Proposed mechanism of production of three 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 three amino acids of glycophorin A sequence, the termination signal and the complete 3′ untranslated region (Kudo and Fukuda, 1989). 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 (1989). Lines between the exon boxes represents the introns.


E1
E2
E3
E4
E5
E6
3'UT
A1
A2A3
A4
A5A6
A7
An
Small (1.0 Kb)
A2
Medium (1.7 Kb)
A6
Large (2.8 Kb)
An

It is conceivable that three different sized primary transcripts are synthesized rather than one. The poly(A) addition signal choice could still be involved in the production of the three glycophorin A mRNAs. In this case, one can envision termination of the primary transcript synthesis close to the appropriate poly(A) addition signal rather than differential cleavage by the endonuclease to form 3´ ends of the three glycophorin A mRNAs.

Despite the general belief that 3´ ends of mRNAs are produced post-transcriptionally by endonucleolytic cleavage and that transcription termination has little, if any role in the 3´ end formation, investigations of a variety of genes are beginning to uncover distinct termination sites (see section 1.14.5). In poxviruses, a sequence element TTTTTNT has been detected and shown to be involved in transcription termination (Earl *et al.*, 1990, section 1.14.5). This sequence is also present in close proximity to the 3´ end of a number of genes including the human immunodeficiency virus type 1 envelope gene (Earl *et al.*, 1990). I have detected equivalent sequences occurring three times in the glycophorin A sequence. One of these termination sequences is located 132 nucleotides downstream from poly(A) addition signal #5 (T1 in Fig. 4.7) and another 43 nucleotide downstream from poly(A) addition signal #6 (T2 in Fig. 4.7). Yet another poxvirus termination element (T3 in Fig. 4.7) is present in the 3´ UT exon sequence deposited in the EMBL data bank (# M24133) by Kudo and Fukuda (1989), about 22 nucleotide upstream from the last poly(A) addition signal, # An. It is possible that these poxvirus termination

signals may also be involved in the generation of multiple mRNAs from a single glycophorin A gene. The presence of two such signals in the vicinity of poly(A) addition signals that would produce medium sized mRNA rather than large RNA might also explain the abundance of the medium-sized glycophorin A mRNA. Hence, termination in the vicinity of termination signal 1 and (or) 2 (T1 and T2 in Fig. 4.7) would eliminate the production of large mRNA and only relatively few transcripts that somehow escape termination and processing at these earlier sites (T1 and T2) would be able to give rise to large-sized mRNA which would perhaps be produced due to termination at T3 (Fig. 4.7).

To produce the small glycophorin A mRNA a sequence other than poxvirus termination sequence in close proximity to the poly(A) addition signal #2 might be responsible for termination. For instance, the presence of a sequence equivalent to CAYTG consensus sequence in close proximity to poly(A) addition signal # 2 (Fig. 4.7, section 1.14.6.2.2.) only, might suggest that the sequence has a role in 3´ end processing of the precursor RNA to produce the small mRNA (1.0 kb). Alternatively, termination at poly(A) addition signals 1, 2 or 3 followed by differential polyadenylation of the precursor RNA may also generate the small mRNA (1.0 kb).

4.4.6. Selection of poly(A) addition site and the glycophorin A mRNA abundance

As mentioned before there are multiple poly(A) addition signals in the glycophorin A gene and work reported in the present study, as well by others, has

shown that at least three of the eight poly(A) addition signals detected so far are used preferentially. Resolution of the features of pre-mRNA encoding glycophorin A responsible for the selection of these signals over the others is not apparent from these studies. There is considerable evidence in the literature that sequences located downstream from poly(A) addition signals are important for the 3′ end processing of mRNAs. One such sequence is a GU cluster reported to be present in the region about 30 bases downstream of many poly(A) addition signals (Birnstiel *et al.*, 1985, also section 1.14.6.2.2). A computer as well a manual search for GU-rich sequences in the regions as far as 50 nucleotides downstream from the various poly(A) addition signals producing the three glycophorin A mRNAs did not uncover any particularly GU-rich regions except following poly(A) addition signal #6. The absence of the GU-rich sequences in the vicinity of various poly(A) addition signals suggested that perhaps the GU-rich sequences in glycophorin A mRNAs do not account for preferential use of poly(A) addition signals 2, 6 and An in the 3′ end processing of the glycophorin A mRNA precursor.

The absence of a well defined GU cluster may also account for the apparent heterogeneity of the medium-sized glycophorin A mRNA (sections 4.3.2.3 to 4.3.2.4). Due to lack of a defined GU cluster or any other sequence motif important in the 3′ end processing in close proximity of various poly(A) addition signals, perhaps more than one poly(A) addition site is used for the generation of small- and

medium-sized glycophorin A mRNAs, i.e. perhaps the small- and medium-sized glycophorin A mRNAs are heterogeneous.

It is also possible that perhaps all poly(A) addition signals are used in glycophorin A pre-mRNA processing. If poly(A) addition signal #s 1, 2, 3, 4, 5, 6, 7, and An were used for 3´ end processing, mRNAs of sizes about 900, 1200, 1400, 1700, 1900, 2200 and 2800 nucleotides, respectively, would be produced. These mRNAs would still fall in the range of 1.0 to 2.8 kb, the mRNA length range observed experimentally in the present work (section 4.3.2.1). This would also be consistent with the notion that at least the small- and medium-sized glycophorin A mRNAs are heterogeneous (section 4.3.2.1 to 4.3.2.5). Assuming there are no more poly(A) addition signals in the yet unknown sequence of the large mRNA (i.e. in the stretch of "N" in Fig. 4.7), the 2.8 kb mRNA could not be heterogeneous as there are no other poly(A) addition signals for a distance of about 700 nucleotides upstream from the poly(A) addition signal # An [up to poly(A) addition signal #7]. Therefore, the poly(A) addition signal #An must be used for 3´ end processing producing the production of the large (2.8 kb) glycophorin A mRNA (Fig. 4.7) with a distinct 3´ end (i.e. non-heterogeneous).

Like glycophorin A, the human dopamine β-hydroxylase gene produces multiple mRNAs having different 3´ terminal regions generated by the utilization of different poly(A) addition sites (Kobayashi *et al.*, 1989). The dopamine β-hydroxylase

gene consists of 12 exons, the last one encoding the entire 3′ untranslated region, again a situation similar to that found in the glycophorin A gene. However, unlike glycophorin A, the most abundant mRNA of the dopamine β-hydroxylase gene is the longer of the two transcripts (Kobayashi *et al.*, 1989). The other difference is that in case of the dopamine β-hydroxylase gene, the less abundant transcript is presumably produced by processing at the 3′ end utilizing a variant of the AAUAAA (AGUAAA). The presence of a variant of AAUAAA in the pre-mRNA of dopamine β-hydroxylase gene could explain its rarity (Kobayashi *et al.*, 1989). In contrast, both glycophorin A mRNAs (large and medium) are apparently produced by utilization of the normal poly(A) addition signals, (AAUAAA), for the 3′ end processing.

The distance between the cap site and the poly(A) addition site may also regulate selection of a particular poly(A) addition site over other site present in the same gene (section 4.4.4). In the case of glycophorin A mRNAs, the 1.7 kb-sized mRNA produced apparently utilizing the poly(A) addition signal # 6 [nucleotides 1652 to 1657, Fig. 3.10 (C)] with a distance of at least 1652 nucleotides between the cap and poly(A) addition site was the most abundant (about 70%, section 4.3.2.1), while the poly(A) addition signal #1 [nucleotides 569 to 574, Fig. 3.10 (C)] apparently did not seem to be used perhaps consistent with the observation made in the spleen necrosis virus poly(A) addition sites (section 4.4.4, Iwasaki and Temin, 1990). The assumption that more nucleotides between the cap site and poly(A) addition sites a mRNA has the more abundant this mRNA would be, seems to hold if the level (30%,

section 4.3.2.1) of small mRNA (1.0 kb) produced using poly(A) addition signal #2 is compared with that of the medium-sized glycophorin A mRNAs (70%, section 4.3.2.1) which is produced using the poly(A) addition signal # 6. However, for the largest mRNA (2.8 kb) which is the least abundant of the three glycophorin A mRNA the same mechanism does not seem to apply since according to the cap-poly(A) addition site distance model the large glycophorin A mRNA should be the most abundant, which it is not. Possibly, besides a minimum length between cap and poly(A) addition site, the maximum allowed distance also governs the efficient utilization of a poly(A) addition signal/site or some other mechanism is involved in the selection of poly(A) addition signal # An. Alternatively it is just a coincidence that the small- and medium-sized glycophorin A mRNAs seem to follow the cap-poly(A) addition site distance model as observed in spleen necrosis virus and this mechanism is not involved in the selection of a poly(A) addition site in the glycophorin A pre-mRNA.

The paucity of the 2.8 kb mRNA could also be explained in terms of the model presented by Denom and Cole (1988) that involves 5′ to 3′ scanning by a factor (section 4.4.4). A consequence of this scanning model is that if there is an efficient poly(A) addition site upstream, there would be very little unprocessed transcript leftover. The abundance of the medium-sized mRNA compared with that of the large (2.8 kb) mRNA is consistent with this postulate. If the poly(A) addition site following poly(A) addition signal # 6 is efficiently recognized by the factor

compared with poly A site following poly(A) addition signal # 2, it would explain the greater abundance of medium (1.7 kb) mRNA over small (1.0 kb) mRNA.

The identification of the features in glycophorin A pre-mRNA responsible for the selection of one poly(A) addition site over the others is not resolved. As discussed above and in section 1.14.6.2, various sequence motifs have been proposed to be involved in the selection of poly(A) addition sites in different eucaryotic systems. It would be interesting to investigate this aspect of glycophorin A mRNA production, for example, by linking a reporter gene sequences and the downstream sequences from the various poly(A) addition signals of glycophorin A pre-mRNA and assessing the efficiency with which each signal is used in an *in vitro* processing reaction in HeLa nuclear extracts for instance.

References

Adamany, A.M. and Blumenfeld, O.O. (1983). A carbohydrate structural variant of MM glycoprotein (Glycophorin A). J. Biol. Chem. *258*, 11537-11545.

Adamany, A.M. and Kathan, R.H. (1969). Isolation of a tetrasaccharide common to MM, NN and MN antigens. Biochem. Biophys. Res. Comm. *37*, 171-178.

Akusjarvi, G. and Persson, M. (1980). Controls of splicing and termination in the major late adenovirus transcription unit. Nature *292*, 420-426.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. (1989). In Molecular Biology of the Cell. (New York: Garland Publishing, Inc.), pp. 406-456.

Allaway, G.P., Pardoe, I.U., Tavakkol, A., and Burness, A.T.H. (1986). Encephalomyocarditis virus attachment. In Virus Attachment and Entry into Cells. R.L. Crowell and K. Lonberg-Holm, ed. (American Society for Microbilology), pp. 116-125.

Allaway, G.P. and Burness, A.T.H. (1986). Site of attachment of encephalomyocarditis virus on human erythrocytes. J. Virol. *59*, 768-770.

Allaway, G.P. and Burness, A.T.H. (1987). Analysis of the bond between encephalomyocarditis virus and erythrocyte receptor by affinity chromatography on Sepharose columns. J. Gen. Virol. *68*, 1849-1856.

Alloisio, N., Morle, L., Bachir, D., Guetarni, D., Colonna, P., and Delaunay, J. (1985). Red cell membrane sialoglycoprotein β in homozygous and heterozygous 4.1(-) heriditary elliptocytosis. Biochim. Biophys. Acta *816*, 57-62.

Aminoff, D. (1988). The role of sialoglycoconjugates in the aging and sequestration of red cells from circulation. Blood Cells *14*, 229-257.

Anderson, K. and Bond, C.W. (1987). Biological properties of mengovirus: Characterization of avirulent, hemagglutination defective mutants. Arch. Virol. *93*, 31-49.

Anderson, R.A. and Lovrien, R.E. (1984). Glycophorin is linked by band 4.1 protein to the human erythrocyte membrane skeleton. Nature *307*, 655-658.

Angel, M.A. and Burness, A.T.H. (1977). The attachment of encephalomyocarditis virus to erythrocytes from several animal species. Virology *83*, 428-432.

Anstee, D.J., Barker, D.M., Judson, P.A., and Tanner, M.J. (1977). Inherited sialoglycoprotein deficiencies in human erythrocytes of type En(a-). Br. J. Haematol. *35*, 309-320.

Anstee, D.J., Mawby, W.J., and Tanner, M.J.A. (1979). Abnormal Blood group Ss active sialoglycoproteins in the membrane of Miltenberger class III, IV and V human erythrocytes. Biochem. J. *183*, 193-203.

Anstee, D.J. (1981). The blood group MNSs active sialoglycoproteins. Sem. Hematol. *18*, 13-31.

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., and Fraser, I.D. (1984). Two individuals with elliptocytic red cells apparently lack three minor erythrocyte membrane sialoglycoproteins. Biochem. J. *218*, 615-619.

Anstee, D.J. (1990). Blood group active surface molecules of the human red blood cells. Vox Sang. *58*, 1-20.

Anstee, D.J. and Tanner, M.J.A. (1974). The distribution of Blood group antigens on butanol extraction of human erythrocyte 'ghosts'. Biochem. J. *138*, 381-386.

Anstee, D.J. and Tanner, M.J.A. (1978). Genetic variants involving the major membrane sialoglycoprotein of human erythrocytes; studies on erythrocytes of type Mk, Mitenberger class V and Mg. Biochem. J. *175*, 149-157.

Anstee, D.J. and Tanner, M.J.A. (1986). Structure and function of the red cell membrane sialoglycoproteins. Br. J. Haematol. *64*, 211-215.

Aviv, H. and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA *69*, 1408-1412.

Badley, J.E., Bishop, G.A., St.John, T., and Frelinger, J.A. (1988). A simple, rapid method for the purification of poly A^+ RNA. Biotechniques *6*, 114-116.

Baer, B. and Kornberg, R.D. (1983). The protein responsible for the repeating structure of cytoplasmic poly(A)$^-$ ribonucleoprotein. J. Cell Biol. *96*, 717-721.

Balhorn, R., Weston, S., Thomas, T., and Wyrobek, A.J. (1984). DNA packaging in mouse spermatids. Synthesis of protamine variants and four transition proteins. Exp. Cell Res. *150*, 298-308.

Baranowski, T., Lisowska, E., Morawiecki, A., Romanowska, E., and Strozecka, K. (1959). Studies on blood group antigens M and N. III. Chemical composition of purified antigens. Arch. Immunol. Ther. Exp. *7*, 15-27.

Baranowski, T. and Lisowska, E. (1963). Studies on blood group antigens M and N. VIII. Fractionation of the products of proteolytic digestion. Arch. Immunol. Ther. Exp. *11*, 631-640.

Bard, E., Efron, D., Marcus, A., and Perry, R.P. (1974). Translational capacity of deadenylated messenger RNA. Cell *1*, 101-106.

Bardwell, V.J., Zarkower, D., Edmonds, M., and Wickens, M. (1990). The enzyme that adds poly(A) to mRNAs is a classical poly(A) polymerase. Mol. Cell Biol. *10*, 846-849.

Bardwell, V.J. and Wickens, M. (1990). Polyadenylation specific complexes undergo a transition early in the polymerization of a poly(A) tail. Mol. Cell Biol. *10*, 295-302.

Bennet, V. (1985). The membrane skeleton of human erythrocytes and its implications for more complex cells. Ann. Rev. Biochem. *54*, 273-304.

Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980). The ovalbumin gene sequence of putative control regions. Nucl. Acids Res. *8*, 127-142.

Bentley, D.L. and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature *321*, 702-706.

Bentley, D.L. and Groudine, M. (1988). Sequence requirements for premature termination of transcription in the human c-myc gene. Cell *53*, 245-256.

Benton, W.D. and Davis, R.W. (1977). Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. Science *196*, 180-181.

Berget, S.M. (1984). Are U4 small nuclear ribonucleoproteins involved in polyadenylation. Nature *309*, 179-182.

Bernstein, P., Peltz, S.W., and Ross, J. (1989). The poly(A) binding protein complex is a major determinant of mRNA stability *in vitro*. Mol. Cell Biol. *9*, 659-670.

Bernstein, P. and Ross, J. (1989). Poly(A), poly(A) binding protein and the regulation of mRNA stability. Trends Biochem. Sci. *14*, 373-377.

Birnboim, H.C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol. *100*, 243-255.

Birnstiel, M.L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3´ processing: The end is in site. Cell *41*, 349-359.

Blackburn, G.M. and Gait, M.J. (1990). RNA saequence information and transmission. In Nucleic Acids in Chemistry and Biology. G.M. Blackburn and M.J. Gait, eds. (Oxford: IRL Press), pp. 217-257.

Blanchard, D., Dahr, W., Hummel, M., Latron, F., Beyreuther, K., and Cartron, J. (1987). Glycophorins B and C from human erythrocyte membranes. J. Biol. Chem. *262*, 5808-5811.

Bloy, C., Blanchard, D., Lambin, P., Goosens, D., Rouger, P., Salmon, C., and Cartron, J.P. (1987). Human monoclonal antibody against Rh(D) antigen: Partial characterization of the Rh(D) polypeptide from human eythrocytes. Blood *69*, 1491-1497.

Blumenfeld, O.O., Callup, P.M., Howe, C., and Lee, L.T. (1970). Erythrocyte membrane proteins: Their study using aqueous pyridine solutions. Biochim. Biophys. Acta *211*, 109-123.

Blumenfeld, O.O. and Adamany, A.M. (1978). Structural polymorphism within the amino-terminal region of MM, NN, and MN glycoproteins (glycophorins) of the human erythrocyte membrane. Proc. Natl. Acad. Sci. USA *75*, 2727-2731.

Bohjanen, P.R., Petryniak, B., June, C.H., Thompson, C.B., and Lindsten, T. (1991). An inducible cytoplasmic factor (AU-B) binds selectively to AUUUA multimers in the 3′ untranslated region of lymphokine mRNA. Mol. Cell Biol. *11*, 3288-3295.

Bolton, E.T. and McCarty, B.J. (1962). A general method for the isolation of RNA complementary to DNA. Proc. Natl. Acad. Sci. USA *48*, 1390-1397.

Bonner, T.I., Brenner, D.J., Neufeld, B.R., and Britten, R.J. (1973). Reduction in the rate of DNA reassociation by sequence divergence. J. Mol. Biol. *81*, 123-135.

Bormann, B.J., Knowles, W.J., and Marchesi, V.T. (1989). Synthetic peptides mimic the assembly of transmembrane glycoproteins. J. Biol. Chem. *264*, 4033-4037.

Braun, R.E., Peschon, J.J., Behringer, R.R., Brinster, R.L., and Palmiter, R.D. (1989). Protamine 3′ untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. Genes Dev. *3*, 793-802.

Brawerman, G. (1976). Characteristics and significance of the polyadenylate sequence in mammalian messenger RNA. Prog. Nucl. Acids Res. Mol. Biol. *17*, 117-148.

Bretscher, M.S. (1971). A major protein which spans the human erythrocyte membrane. J. Mol. Biol. *59*, 351-357.

Bretscher, M.S. (1975). C-terminal region of the major erythrocyte sialoglycoprotein is on the cytoplasmic side of the membrane. J. Mol. Biol. *98*, 831-833.

Brewer, G. (1991). An A+U-rich element RNA binding factor regulates c-myc mRNA stability *in vitro*. Mol. Cell Biol. *11*, 2460-2466.

Brewer, G. and Ross, J. (1988). Poly(A) shortening and degradation of the 3´ A+U-rich sequences of human c-myc mRNA in a cell-free system. Mol. Cell Biol. *8*, 1697-1708.

Brown, P.H., Tiley, L.S., and Cullen, B.R. (1991). Effect of RNA secondary structure on polyadenylation site selection. Genes Dev. *5*, 1277-1284.

Budowle, B. (1985). A rapid, small scale procedure for the preparation of plasmid DNA for screening mutants by the deoxyribonucleotide chain terminator sequencing method. Biotechniques *93*, 98-101.

Buetler, B., Kochiin, N., Milsark, I., Leudke, C., and Cerami, A. (1986). Control of cachectin (tumor necrosis factor) synthesis: Mechanism of endotoxin resistance. Science *232*, 977-980.

Burness, A.T.H. and Pardoe, I.U. (1981). Effect of enzymes on the attachment of influenza and encephalomyocarditis viruses to erythrocytes. J. Gen. Virol. *55*, 275-288.

Burness, A.T.H. and Pardoe, I.U. (1983). A sialoglycopeptide from human erythrocytes with receptor like properties for encephalomyocarditis and influenza viruses. J. Gen. Virol. *64*, 1137-1148.

Burness, A.T.H. (1981). Glycophorin and Sialylated Components as Receptors for Viruses. In Receptors and Recognition, Series B. K.Lonberg-Holm ed. (London: Chapman Hall), *8*, 65-84.

Butterfield, A., Farmer II, B.T., and Feix, J.B. (1983). Induced alterations in the physical state of sialic acid and membrane proteins in human erythrocyte ghosts: Implications for the topology of the major sialoglycoprotein. Ann. N. Y. Acad. Sci. *414*, 169-179.

Capetanaki, Y.G., Nagai, J., Flytzanis, C.N., and Lazarides, E. (1983). Tissue specific expression of two mRNA species transcribed from a single vimentin gene. Cell *35*, 411-420.

Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S., and Cerami, A. (1986). Identification of a common nucleotide sequence in the 3´ untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl. Acad. Sci. USA *83*, 1670-1674.

Carlsson, S.R., Sasaki, H., and Fukuda, M. (1986). Structural variations of O-linked oligosaccharides present in Leukosialin isolated from erythroid, Myeloid, and T-lymphoid cell Lines. J. Biol. Chem. *261*, 12787-12795.

Cartron, J.-P., Colin, Y., Kudo, S., and Fukuda, M. (1990). Molecular Genetics of human erythrocyte sialoglycoproteins glycophorin A, B, C, and D. In Blood Cell Biochemistry. R.H. Harris, ed. (New York: Plenum Publishing Corporation), pp. 299-335.

Ch'ng, J.L., Shoemaker, D.L., Schimmel, P., and Holmes, E.W. (1990). Reversal of creatine kinase translational repression by 3′ untranslated sequences. Science *248*, 1003-1006.

Chasis, J.A., Mohandas, N., and Shohet, S.B. (1985). Erythrocyte membrane rigidity induced by glycophorin A ligand interaction: Evidence for a ligand induced association between glycophorin A and skeletal proteins. J. Clin. Invest. *75*, 1919-1926.

Chasis, J.A. and Schrier, S.L. (1989). Membrane deformability and the capacity for shape change in the erythrocyte. Blood *74*, 2562-2568.

Chasis, J.A. and Shohet, S.B. (1987). Red cell biochemical anatomy and membrane properties. Ann. Rev. Physiol. *49*, 237-48.

Christofori, G. and Keller, W. (1988). 3′ cleavage and polyadenylation of mRNA precursors *in vitro* requires a poly(A) polymerase, a cleavage factor, and a snRNP. Cell *54*, 875-889.

Christofori, G. and Keller, W. (1989). Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA *in vitro*. Mol. Cell Biol. *9*, 193-203.

Cohen, C.M. (1983). The molecular organization of the red cell membrane skeleton. Sem. Hematol. *20*, 141-158.

Colin, Y., Rahuel, C., London, J., Romeo, P.H., Auriol, L., Galibert, F., and Cartron, J. (1986). Isolation of cDNA clones and complete amino acid sequence of human erythrocyte glycophorin C. J. Biol. Chem. *261*, 229-233.

Connelly, S. and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. *2*, 440-452.

Conway, L. and Wickens, M. (1987). Analysis of mRNA 3′ end formation by modification interference: The only modifications which prevent processing lie in

AAUAAA and the poly(A) site. EMBO J. *6*, 4177-4184.

Cook, G.M.W., Heard, D.H., and Seaman, G.V.F. (1960). A sialomucopeptide liberated by trypsin from the human erythrocyte. Nature *188*, 1011-1012.

Corradini, D., El Rassi, Z., Horváth, C., Guerra, G., and Horne, W. (1988). Combined lectin affinity and metal interaction chromatography for the separation of glycophorins by high performance liquid chromatography. J. Chromatogr. *458*, 1-11.

Cotmore, S.F., Furthmayr, H., and Marchesi, V.T. (1977). Immunochemical evidence for the transmembrane orientation of glycophorin A. Localization of ferritin antibody conjugates in intact cells. J. Mol. Biol. *113*, 539-553.

Dahr, W., Uhlenbruck, G., and Knott, H. (1975a). Immunochemical aspects of the MNSs blood group system. J. Immunogenet. *2*, 87-100.

Dahr, W., Uhlenbruk, G., Issit, P.D., and Allen, F.H.,Jr. (1975b). SDS polyacrylamide gel electrophoretic anallysis of the membrane glycoproteins from S-s-U- erythrocytes. J. Immunogenet. *2*, 249-251.

Dahr, W., Uhlenbruck, G., Janssen, E., and Schmalisch, R. (1977). Different N-terminal amino acids in the MN glycoprotein from MM and NN erythrocytes. Hum. Genet. *35*, 335-343.

Dahr, W., Uhlenbruck, G., Leikola, J., and Wagstaff, W. (1978). Studies on the membrane glycoprotein defect of En(a-) erythrocytes: III. N-terminal amino acids of sialoglycoproteins from normal and En(a-) red cells. J. Immunogenet. *5*, 117-127.

Dahr, W., Beyreuther, K., Steinbach, H., Gielen, W., and Kruger, J. (1980a). Structure of the Ss blood group antigens, II: A methionine/threonine polymorphism within the N-terminal sequence of the Ss glycoprotein. Z. Physiol. Chem. *361*, 895-906.

Dahr, W., Gielen, W., Beyreuther, K., and Kruger, J. (1980b). Structutre of the Ss blood group antigens.I. Isolation of Ss-active glycopeptides and differentiation of the antigens by modification of methionine. Z. Physiol. Chem. *361*, 145-152.

Dahr, W., Beyreuther, K., Kordowicz, M., and Kruger, J. (1982). N-terminal amino acid sequence of sialoglycoprotein D (glycophorin C) from human erythrocyte membranes. Eur. J. Biochem. *125*, 57-62.

Dahr, W. (1986). Immunochemistry of sialoglycoproteins in human red blood cell membranes. In Recent Advances in Blood Group Biochemistry. V. Vengelen-Tyler and W.J. Judd, eds. (Arlington: American Association of Blood Banks), pp. 23-65.

Dahr, W., Kiedrowski, S., Blanchard, D., Hermand, P., Moulds, J.J., and Cartron, J.P. (1987a). High frequency antigens of human membrane sialoglycoproteins, V: Characterization of the Gerbich blood group antigens: Ge2 and Ge3. Z. Biol. Chem. *368*, 1375-1383.

Dahr, W., Kordowicz, M., Moulds, J., Gielen, W., Lebeck, L., and Kruger, J. (1987b). Characterization of the Ss sialoglycoprotein and its antigens in Rh(null) erythrocytes. Blut *54*, 13-24.

Dahr, W. and Beyreuther, K. (1985). A revision of the N-terminal structure of sialoglycoprotein D (glycophorin C) from human erythrocyte membranes. Biol. Chem. *366*, 1067-1070.

Dahr, W. and Kruger, J. (1983). Solubilization of various group antigens by detergent Triton X-100. In The 10th International Congress of the Society for Forensic and Haemogenetics. (Munich: Schmitt, Meyer and Wurzburg), pp. 141-155.

Dahr, W. and Longster, G. (1984). Studies on Mv red cells. II Immunochemical investigations. Blut *49*, 299-306.

Darnell, J., Lodish, H., and Baltimore, D. (1990). In Molecular Cell Biology (New York: W.H. Freeman Company), pp. 305-464.

Davis, L.G., Dibner, M.D., and Battey, J.F. (1986). In Basic Methods in Molecular Biology (New York: Elsevier), pp. 68-78.

Denome, R.M. and Cole, C.N. (1988). Patterns of polyadenylation site selection in gene constructs containing multiple polyadenylation signals. Mol. Cell Biol. *8*, 4829-4839.

Dente, L., Cesareni, G., and Cortese, R. (1983). pEMBL: A new family of single stranded plasmids. Nucl. Acids Res. *11*, 1645-1655.

Deshpande, A., Chatterjee, B., and Roy, A. (1979). Translation and stability of rat liver messenger RNA for $\alpha 2\mu$-globulin in *Xenopus* oocyte. J. Biol. Chem. *254*, 8937-8942.

Dickey, L.F., Wang, Y-H., Shull, G.E., Wortman III, I.A., and Theil, E.C. (1988). The importance of the 3′ untranslated region in the translational control of ferritin mRNA. J. Biol. Chem. *263*, 3071-3074.

Doel, M.T. and Carey, N.H. (1976). The translational capacity of deadenylated ovalbumin messenger RNA. Cell *8*, 51-58.

Dolci, E.D. and Palade, G.E. (1989). Ontogenic expression of the murine erythrocyte glycophorins. J. Cell. Sci. *93*, 191-197.

Drummond, D.R., Armstrong, J., and Colman, A. (1985). The effect of capping and polyadenylation on the stability, movement, and translation of synthetic messenger RNAs in *Xenopus* oocytes. Nucl. Acids Res. *13*, 7375-7395.

Duval, C., Bouvet, P., Omilli, F., Roghi, C., Dorel, C., LeGuellec, R., Paris, J., and Osborne, H.B. (1990). Stability of maternal mRNA in *Xenopus* embryos: Role of transcription and translation. Mol. Cell Biol. *10*, 4123-4129.

Earl, P.L., Hugin, A.W., and Moss, B. (1990). Removal of cryptic poxvirus transcription termination signals from the human immunodeficiency virus type 1 envelope gene enhances expression and immunogenicity of a recombinant vaccinia virus. J. Virol. *64*, 2448-2451.

Eaton, B.T. and Crameri, G.S. (1989). The site of bluetongue virus attachment to glycophorins from a number of animal erythrocytes. J. Gen. Virol. *70*, 3347-3353.

Ebert, W., Metz, J., and Roelcke, D. (1972). Modifications of n-acetylneuraminic acid and their influence on the antigen activity of erythrocyte glycoproteins. Eur. J. Biochem. *27*, 470-472.

Edmonds, M. and Winters, M.A. (1976). Polyadenylate polymerases. Prog. Nucl. Acids Res. Mol. Biol. *17*, 149-179.

Ekblom, M., Gahmberg, C.G., and Andersson, L.C. (1985). Late expression of M and N antigens on glycophorin A during erythroid differentiation. Blood *66*, 233-236.

El-Maliki, B., Blanchard, D., Dahr, W., Beyreuther, K., and Cartron, J.P. (1989). Structural homology between glycophorin C and D of human erythrocyte. Eur. J. Biochem. *183*, 639-643.

Enegren, B.J. and Burness, A.T.H. (1977). Chemical structure of attachment sites for viruses on human erythrocytes. Nature *268*, 536-537.

Fairbanks, G., Steck, T.L., and Wallach, D.F.H. (1971). Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry *10*, 2606-2617.

Falck-Pedersen, E. and Logan, J. (1989). Regulation of poly(A) site selection in adenovirus. J. Virol. *63*, 532-541.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *132*, 6-13.

Feinberg, A.P. and Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *137*, 266-267.

Fitzgerald, M. and Shenk, T. (1981). The sequence 5′ AAUAAA 3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell *42*, 251-260.

Fletcher, M.A., Lo, T.M., and Graves, W.R. (1976). Immunochemical studies of infectious mononucleosis. V. Isolation and characterization of a glycoprotein from goat erythrocyte membranes. J. Immunol. *117*, 722-729.

Fox, C.A., Sheets, M.D., and Wickens, M.P. (1989). Poly(A) addition during maturation of frog oocytes: Distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. Genes Dev. *3*, 2151-2162.

Fox, C.A. and Wickens, M. (1990). Poly(A) removal during oocyte maturation: A default reaction selectively prevented by specific sequences in the 3′ UTR of certain maternal mRNAs. Genes Dev. *4*, 2287-2298.

Friedman, M.J., Fukuda, M., and Laine, R.A. (1985). Evidence for a malarial parasite interaction site on the major transmembrane protein of the human erythrocyte. Science *228*, 75-77.

Fukuda, K., Tomita, M., and Hamada, A. (1980). Isolation and characterization of alkali-labile oligosaccharide units from horse glycophorin. J. Biochem. *87*, 687-693.

Fukuda, M., Lauffenburger, M., Sasaki, H., Rogers, M.E., and Dell, A. (1987). Structures of novel sialylated O-linked oligosaccharides isolated from human erythrocyte glycophorins. J. Biol. Chem. *262*, 11952-11957.

Fukuda, M. and Fukuda, M.N. (1984). Cell surface glycoproteins and carbohydrate antigens in development and differentiation of human erythroid cells. In The Biology of Glycoproteins. R.J. Ivatt, ed. (New York: Plenum Press), pp. 183-234.

Furthmayr, H., Tomita, M., and Marchesi, V.T. (1975). Fractionation of the major sialoglycopeptides of the human red blood cell membrane. Biochem. Biophys. Res. Comm. *65*, 113-121.

Furthmayr, H. (1978a). Structural comparison of glycophorins and immunochemical analysis of genetic variants. Nature *271*, 519-524.

Furthmayr, H. (1978b). Glycophorins A, B, and C: A family of sialoglycoproteins. Isolation and preliminary characterization of trypsin derived peptides. J. Supramol. Struct. *9*, 79-95.

Furthmayr, H. and Marchesi, V.T. (1976). Subunit structure of human erythrocyte glycophorin A. Biochemistry *15*, 1137-1144.

Furthmayr, H. and Marchesi, V.T. (1983). Glycophorins: Isolation, orientation, and localization of specific domains. Methods Enzymol. *96*, 268-279.

Gahmberg, C.G., Jokinen, M., and Andersson, L.C. (1978). Expression of the major sialoglycoprotein (glycophorin) on erythroid cells in human bone marrow. Blood *52*, 379-387.

Gahmberg, C.G., Jokinen, M., and Andersson, L.C. (1979). Expression of the major red cell sialoglycoprotein, glycophorin A, in the human leukemic cell line K562. J. Biol. Chem. *254*, 7442-7448.

Gahmberg, C.G., Jokinen, M., Karhi, K.K., and Andersson, L.C. (1980). Effect of tunicamycin on the biosynthesis of the major human red cell sialoglycoprotein, glycophorin A, in the leukemia cell line K562. J. Biol. Chem. *255*, 2169-2175.

Gahmberg, C.G., Jokinen, M., Karhi, K.K., Kampe, O., Peterson, Per A., and Andersson, L.C. (1983). Glycophorin A: *In vitro* biogenesis and processing. Methods Enzymol. *96*, 281-298.

Gahmberg, C.G., Ekblom, M., and Andersson, L.C. (1984). Differentiation of human erythroid cells is associated with increased O-glycosylation of the major sialoglycoprotein, glycophorin A. Proc. Natl. Acad. Sci. USA *81*, 6752-6756.

Galili, G., Kawata, E., Smith, L.D., and Larkins, B.A. (1988). Role of the 3´ poly(A) sequence in translational regulation of mRNAs in *Xenopus laevis* oocytes. J. Biol. Chem. *263*, 5764-5770.

Gallinaro, H., Vincendon, P., Sittler, A., and Jacob, M. (1988). Alternative use of a polyadenylation signal and of a downstream 3´ splice site. Effect of 5, 6-dichloro-1-β-D- ribofuranosyl benzimidazole. J. Mol. Biol. *204*, 1031-1040.

Gil, A. and Proudfoot, N.J. (1984). A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3´ end formation. Nature *312*, 473-474.

Gil, A. and Proudfoot, N.J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3´ end formation. Cell *49*, 399-406.

Gilmartin, G.M., McDevitt, M.A., and Nevins, J.R. (1988). Multiple factors are required for specific RNA cleavage at a poly(A) addition site. Genes Dev. *2*, 578-587.

Gilmartin, G.M. and Nevins, J.R. (1989). An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev. *3*, 2180-2190.

Goldstein, B.J. and Kahn, C.R. (1989). Analysis of mRNA heterogeneity by ribonuclease H mapping: Application to the insulin receptor. Biochem. Biophys. Res. Comm. *159*, 664-669.

Gottschalk, A. (1960). In The Chemistry and Biology of Sialic Acids and Related Substances (Cambridge: University Press).

Graham, J.M. and Sandall, J.K. (1979). Tissue culture cell fractionation; fractionation of membranes from tissue culture cells homogenized by glycerol induced lysis. Biochem. J. *182*, 157-164.

Graves, R.A., Pandey, N.B., Chodchoy, N., and Marzluff, W.F. (1987). Translation is required for regulation of histone mRNA degradation. Cell *48*, 615-626.

Greenwalt, T.J., Sasaki, T., Sanger, R., Sneath, J., and Race, R.R. (1954). Proc. Natl. Acad. Sci. USA *40*, 1126, as cited in Issitt, 1981.

Grefrath, S.P. and Reynolds, J.A. (1974). The molecular weight of the major glycoprotein from the human erythrocyte membrane. Proc. Natl. Acad. Sci. USA *71*, 3913-3916.

Grumont, R.J. and Gerondakis, S. (1990). The murine c-rel proto-oncogene encodes two mRNAs the expression of which is modulated by lymphoid stimuli. Oncogene Res. *5*, 245-254.

Hadjvassiliou, H. and Brawerman, G. (1967). Template and ribosomal ribonucleic acid components in the nucleus and the cytoplasm of rat liver. Biochemistry *6*, 1934-1941.

Hadley, T.J., Klotz, F.W., and Miller, L.H. (1986). Invasion of erythrocytes by malaria parasites: A cellular and molecular overview. Ann. Rev. Microbiol *40*, 451-477.

Hadley, T.J., Klotz, F.W., Pasvol, G., Haynes, J.D., McGinniss, M.H., Okubo, Y., and Miller, L.H. (1987). Falciparum malaria parasites invade erythrocytes that lack glycophorin A and B (MkBk); strain differences indicate receptor heterogeneity and two pathways for invasion. J. Clin. Investig. *80*, 1190-1193.

Hall, A. and Brown, R. (1985). Human N-ras: cDNA cloning and gene structure. Nucl. Acids Res. *13*, 5255-5268.

Hall, J.L. and Baker, D.A. (1977). Structure and properties of cell membranes. In Cell Membrane and Ion Transport. J.L. Hall and D.A. Baker, eds. (London: Longman Group Limited), pp. 1-30.

Hamaguchi, H. and Cleve, H. (1972). Solubulization of human erythrocyte membrane glycoproteins and separation of the MN glycoprotein from a glycoprotein with I, S, and A activity. Biochim. Biophys. Acta. *278*, 271-280.

Hanahan, D. (1988). Techniques for transformation of E.coli. In DNA Cloning, A Practical Approach. D.M. Glover, ed. (Oxford: IRL Press), pp. 109-135.

Harrison, R. and Lunt, G.G. (1980). In Biological Membranes: Their Structure and Function (Glasgow and London: Blackie and Son Limited).

Hart, R.P., McDevitt, M.A., and Nevins, J.R. (1985). Poly(A) site cleavage in HeLa nuclear extract is dependent on downstream sequences. Cell *43*, 677-683.

Hattori, M. and Sakaki, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. *152*, 232-238.

Hecht, N.B, Bower, P.A., Waters, S.H., Yelick, P.C., and Distel, R.J. (1986). Evidence for haploid expression of mouse testicular genes. Exp. Cell Res. *164*, 183-190.

Heilig, R., Perrin, F., Gannon, F., Mandel, J.L., and Chambon, P. (1980). The ovalbumin gene family: Structure of the X gene and evolution of duplicated split genes. Cell *20*, 625-637.

Hepler, J.E., Van Wyk, J.J., and Lund, P.K. (1990). Different half lives of insulin-like growth factor I mRNAs that differ in length of 3′ untranslated sequence. Endocrinology. *127*, 1550-1552.

Hermentin, P. (1987). Malaria invasion of human erythrocytes. Parasitol. Today *3*, 52-55.

Hiatt, H.H. (1962). A rapidly labeled RNA in rat liver nuclei. J. Mol. Biol. *5*, 217-229.

Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J., and Proudfoot, N.J. (1983). α-thalassaemia caused by a polyadenylation signal mutation. Nature *306*, 398-400.

Hirst, G.K. (1942). Adsorption of influenza hemagglutinins and virus by red blood cells. J. Exp. Med. *76*, 195-209.

Honma, K., Tomita, M., and Hamada, A. (1980). Amino acid sequence and attachment sites of oligosaccharide of porcine erythrocyte glycophorin. J. Biochem. *88*, 1679-1691.

Honma, K., Manabe, H., Tomita, M., and Hamada, A. (1982). Isolation and preliminary characterization of two glycophorins from rabbit erythrocyte membranes. Chem. Pharm. Bull. *30*, 966-972.

Howe, C., Rose, H.M., and Schneider, L. (1957). Enzymic action of influenza virus on human erythrocyte stroma components. Proc. Soc. Exp. Biol. Med. *96*, 88-94.

Hoyer, B.C., McCarty, B.J., and Bolton, E.T. (1963). Complementary RNA in nucleus and cytoplasm of mouse liver cells. Science *140*, 1408-1411.

Huez, G., Marbaix, G., Gallwitz, D., Weinberg, E., Devos, R., Hubert, E., and Cleuter, Y. (1978). Functional stabilization of HeLa cell histone messenger RNAs injected into *Xenopus* oocytes by 3´ -OH polyadenylation. Nature *271*, 572-573.

Humphrey, T. and Proudfoot, N.J. (1988). A beginning to the biochemistry of polyadenylation. Trends. Genet. *4*, 243-255.

Huynh, T., Young, R., and Davis, R. (1984). Construction and screening of cDNA libraries in lambda gt10 and lambda gt11. In DNA Cloning: A Practical Approach. D. Glover, ed. (Oxford: IRL Press) pp. 17-43.

Irimura, T., Tsuji, T., Tagami, S., Yamamoto, K., and Osawa, T. (1981). Structure of a complex type sugar chain of human glycophorin A. Biochemistry *20*, 560-566.

Issit, P.D. (1981). The MN Blood Group System (Cincinnati: Montgomery Scientific Publications).

Iwasaki, K. and Temin, H.M. (1990). The efficiency of RNA 3´-end formation is determined by the distance between the cap site and the poly(A) site in spleen necrosis virus. Genes Dev. *4*, 2299-2307.

Jackson, R.J. and Standart, N. (1990). Do the poly(A) tail and 3´ untranslated region control mRNA translation? Cell *62*, 15-24.

Jackson, R.L., Segrest, J.P., Kahane, I., and Marchesi, V.T. (1973). Studies on the major sialoglycoprotein of the human red cell membrane. Isolation and characterization of tryptic glycopeptides. Biochemistry *12*, 3131-3138.

Jacobson, A. and Favreau, M. (1983). Possible involvemnt of poly(A) in protein synthesis. Nucl. Acids Res. *11*, 6353-6368.

Jefferey, W.R. and Brawerman, G. (1974). Characterization of the steady state population of messenger RNA and its poly (adenylic acid) segment in mammalian cells. Biochemistry *13*, 4633-4637.

Jochum, C., Voth, R., Rossol, S., Meyer zum Buschenfelde, K.H., Hess, G., Will, H., Schroder, H.C., Steffen, R., and Muller, W.E. (1990). Immunosuppressive function of hepatitis B antigens *in vitro*: Role of endoribonuclease V as one potential trans inactivator for cytokines in macrophages and human hepatoma cells. J. Virol. *64*, 1956-1963.

Jokinen, M., Gahmberg, C.G., and Andersson, L.C. (1979). Biosynthesis of the major human red cell sialoglycoprotein, glycophorin A, in a continuous cell line. Nature *279*, 604-607.

Jokinen, M., Ulmanen, I., Andersson, L.C., Kaariainen, L., and Gahmberg, C.G. (1981). Cell free synthesis and glycosylation of the major human red cell sialoglycoprotein, glycophorin A. Eur. J. Biochem. *114*, 393-397.

Jokinen, M., Andersson, L.C., and Gahmberg, C.G. (1985). Biosynthesis of the major human red cell sialoglycoprotein, glycophorin A: O-glycosylation. J. Biol. Chem. *260*, 11314-11321.

Jones, T.R. and Cole, M.D. (1987). Rapid cytoplasmic turnover of c-myc mRNA: Requirement of the 3′ untranslated sequences. Mol. Cell Biol. *7*, 4513-4521.

Karp, G. (1984). The cell surface. In Cell Biology. (New York: McGraw Hill Inc.), pp. 221-251.

Kathan, R.H., Winzler, R.J., and Johnson, C.A. (1961). Preparation of an inhibitor of viral hemagglutination from human erythrocytes. J. Exp. Med. *113*, 37-45.

Kathan, R.H. and Winzler, R.J. (1963). Structure studies on the myxovirus hemagglutination inhibitor of human erythrocytes. J. Biol. Chem. *238*, 21-25.

Kawashima, I., Fukuda, K., Tomita, M., and Hamada, A. (1982). Isolation and characterization of alkali-labile oligosaccharides from porcine erythrocyte glycophorin. J. Biochem. *91*, 865-872.

Keaveney, M., Neilan, J., and Gannon, F. (1989). An analysis by restriction enzymes of the genomic structure of the 3′ untranslated region of the human estrogen receptor gene. Biochim. Biophys. Acta *1007*, 289-294.

Kleene, K.C., Distel, R.J., and Hecht, N.B. (1983). cDNA clones encoding cytoplasmic poly(A)$^+$ RNAs which first appear at detectable levels in haploid phases of spermatogenesis in the mouse. Dev. Biol. *98*, 455-464.

Kobayashi, K., Kurosawa, Y., Fujita, K., and Nagatsu, T. (1989). Human dopamine β-hydroxylase gene: Two mRNA types having different 3´ terminal regions are produced through alternative polyadenylation. Nucl. Acids Res. *17*, 1089-1102.

Kornfeld, R. and Kornfeld, S. (1970). The structure of a phytohaemagglutinin receptor site from human erythrocytes. J. Biol. Chem. *245*, 2536-2545.

Krotkiewski, H. (1988). The structure of glycophorins of animal erythrocytes. Glycoconjugate J. *5*, 35-48.

Kruys, V., Marinx, O., Shaw, G., Deschamps, J., and Huez, G. (1989). Translational blockade imposed by cytokine derived UA-rich sequences. Science *245*, 852-855.

Kuczmarski, C.A., Orlina, A.R., Delahanty, L.K., and Kuczmarski, E.R. (1987). Instability of red cell shape associated with the absence of membrane glycophorin C. Vox Sang. *52*, 36-42.

Kudo, S. and Fukuda, M. (1989). Structural organization of glycophorin A and B genes: Evolved by homologous recombination at Alu repeat sequences. Proc. Natl. Acad. Sci. USA *86*, 4619-4623.

Kudo, S. and Fukuda, M. (1990). Identification of a novel human glycophorin, glycophorin E, by isolation of genomic clones and complementary DNA clones utilizing polymerase chain reaction. J. Biol. Chem. *265*, 1102-1110.

Lai, E.C., Stein, J.P., Catterall, J.F., Woo, S.L.C., Mace, M.L., Means, A.R., and O'Malley, B.W. (1979). Molecular structure and flanking nucleotide sequences of the natural chicken ovomucoid gene. Cell *18*, 829-842.

Laing, P., Appleby, L.A., Culbert, E.J., and Elson, C.J. (1988). Antigenicity of rat erythrocyte glycophorins. Immunol. Lett. *18*, 125-128.

Laird-Fryer, B., Moulds, J.J., Dahr, W., Min, Y.O., and Chandanayingyong, D. (1986). Anti En(a) Ab detected in the serum of an Mi(VII) homozygote. Transfusion *26*, 51-56.

Laird-Offringa, I.A., Elfferich, P., and Van der Eb, A.J. (1991). Rapid c-myc mRNA degradation does not require (A+U)-rich sequences or complete translation of the mRNA. Nucl. Acids Res. *19*, 2387-2394.

Landsteiner, K. and Levine, P. (1927). Proc. Soc. Exp. Biol. *24*, 600. As cited in Issitt, 1981.

Lee, S.Y., Mendecki, J., and Brawerman, G. (1971). A polynucleotide segment rich in adenylic acid in the rapidly labelled polysomal RNA component of mouse sarcoma 180 ascites cells. Proc. Natl. Acad. Sci. USA *68*, 1331-1335.

Leff, S.E., Rosenfeld, M.G., and Evans, R.M. (1986). Complex transcriptional units: Diversity in gene expression by alternative RNA processing. Ann. Rev. Biochem. *55*, 1091-1117.

Leff, S.E., Evans, R.M., and Rosenfeld, M.G. (1987). Splice commitment dictates neuron specific alternative RNA processing in calcitonin/CGRP gene expression. Cell *48*, 517-524.

Lehrach, H., Diamond, D., Wozney, J.M., and Boedtker, H. (1977). RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry *16*, 4743-4752.

Lenard, J. (1970). Protein and glycolipid components of human erythrocyte membranes. Biochemistry *9*, 1129-1132.

Levine, P., Kuhmichel, A.B., Wigod, M., and Koch, E. (1951). Proc. Soc. Exp. Biol. *87*, 218. As cited in Issit, 1981.

Levitt, N., Briggs, D., Gil, A., and Proudfoot, N.J. (1989). Definition of an efficient synthetic poly(A) site. Genes Dev. *3*, 1019-1025.

Lewin, B. (1987). Genes III (New York: John Wiley & Sons) pp. 458-482.

Lim, L. and Canellakis, K.S. (1970). Adenine rich polymer associated with rabbit reticulocyte messenger RNA. Nature *227*, 710-715.

Lisowska, E. (1968). Fatty acids in M and N blood group substances. Nature *218*, 569-70.

Lisowska, E. (1988). Antigenic properties of human erythrocyte glycophorins. Adv. Exp. Med. Biol. *228*, 265-315.

Lisowska, E. and Duk, M. (1972). The reaction of products of sequential periodate oxidation of human erythrocyte glycoproteins with hemagglutinin from *Helix pomatia*. Arch. Immunol. Ther. Exp. *20*, 869-875.

Lisowska, E. and Duk, M. (1975). Effect of modification of amino groups of human erythrocytes on M, N and Nvg blood group specificities. Vox Sang. *28*, 392-397.

Lisowska, E. and Jeanloz, R.W. (1973). Composition and distribution of carbohydrate chains in glycoproteins of human erythrocyte membrane. Carbohydr. Res. *29*, 181-191.

Lisowska, E. and Wasniowska, K. (1978). Immunochemical characterization of cyanogen bromide degradation products of M and N blood group glycopeptides. Eur. J. Biochem. *88*, 247-252.

Littauer, U.Z. and Soreq, H. (1982). The regulatory function of poly(A) and adjascent 3′ sequences in translated RNA. Prog. Nucl. Acids Res. Mol. Biol. *27*, 53-83.

Logan, J., Falck-Pedersen, F., Darnell, J.E.,Jr., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse β maj-globin gene. Proc. Natl. Acad. Sci. USA *84*, 8306-8310.

Loken, M.R., Civin, C.I., Bigbee, W.L., Langlois, R.G., and Jensen, R.H. (1987). Coordinate glycosylation and cell surface expression of glycophorin A during normal human erythropoiesis. Blood *70*, 1959-1961.

Lozzio, C.B. and Lozzio, B.B. (1975). Human chronic myelogenous leukemia cell line with positive Philadelphia chromosome. Blood *45*, 321-334.

Lutz, H.U., Barber, R., and McGuire, R.F. (1976). Glycoprotein enriched vesicles from sheep erythrocyte ghosts obtained by spontaneous vesiculation. J. Biol. Chem. *251*, 3500-3510.

Makela, O., Miettinen, T., and Pesola, R. (1960). Release of sialic acid and carbohydrates from human red cells by trypsin treatment. Vox Sang. *5*, 492-496.

Malhotra, S.K. (1980). Organization, composition, and biogenesis of animal cell membranes. In Membrane Structure and Function. E.E. Bittar, ed. (New York: John Wiley & Sons), pp. 1-72.

Malter, J.S. (1989). Identification of an AUUUA specific messenger RNA binding protein. Science *246*, 664-666.

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

Manley, J.L. (1983). Accurate and specific polyadenylation of mRNA preursor in soluble whole cell lysate. Cell *33*, 595-605.

Manley, J.L., Yu, H., and Ryner, L. (1985). RNA containing hexanucleotide AAUAAA directs efficient mRNA polyadenylation *in vitro*. Mol. Cell Biol. *5*, 373-379.

Manley, J.L. (1988). Polyadenylation of mRNA precursors. Biochim. Biophys. Acta *950*, 1-12.

Manley, J.L., Proudfoot, N.J., and Platt, T. (1989). RNA 3′ end formation. Genes Dev. *3*, 2218-2222.

Marchesi, V.T., Tillack, T.W., Jackson, R.L., Segrest, J.P., and Scott, R.E. (1972). Chemical characterization and surface orientation of the major glycoprotein of the human erythrocyte membrane. Proc. Natl. Acad. Sci. USA *69*, 1445-1449.

Marchesi, V.T., Furthmayr, H., and Tomita, M. (1976). The red cell membrane. Ann. Rev. Biochem. *45*, 667-698.

Marchesi, V.T. (1979). Functional proteins of the human red blood cell membrane. Semin. Hematol. *16*, 3-20.

Marchesi, V.T. and Andrews, E.P. (1971). Glycoproteins: Isolation from cell membranes with lithium diiodosalicylate. Science *174*, 1247-1248.

Marshall, L.M., Thureson-Klein, A., and Hunt, R.C. (1984). Exclusion of erythrocyte specific membrane proteins from clathrin coated pits during differnetiation of human erythroleukemic cells. J. Cell Biol. *98*, 2055-2063.

Mather, E.L., Nelson, K.J., Haimovich, J., and Perry, R.P. (1984). Mode of regulation of immunoglobulin μ- and δ-chain expression varies during B lymphocyte maturation. Cell *36*, 329-338.

Matsui, Y., Natori, S., and Obinata, M. (1989). Isolation of the cDNA clone for mouse glycophorin, erythroid specific membrane protein. Gene *77*, 325-332.

McDevitt, M.A., Imperiale, M.J., Ali, H., and Nevins, J.R. (1984). Requirement of a downstream sequence for generation of a poly(A) addition site. Cell *37*, 993-999.

McDevitt, M.A., Hart, R.P., Wong, W.W., and Nevins, J.R. (1986). Sequences capable of restoring poly(A) site function define two distinct downstream elements. EMBO. J. *5*, 2907-2913.

McDevitt, M.A., Gilmartin, G.M., Reeves, W.H., and Nevins, J.R. (1988). Multiple factors are required for poly(A) addition to a mRNA 3′ end. Genes Dev. *2*, 588-597.

McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985). The consensus YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucl. Acids Res. *13*, 1347-1368.

Meinkoth, J. and Wahl, G. (1984). Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. *138*, 267-284.

Mizumoto, K. and Kaziro, Y. (1987). Messenger RNA capping enzymes from eukaryotic cells. Prog. Nucl. Acids Res. Mol. Biol. *34*, 1-28.

Montell, C., Fisher, E.F., Caruthers, M.H., and Berk, A.J. (1983). Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3′ consensus sequence AAUAAA. Nature *305*, 600-605.

Moore, C.L. and Sharp, P.A. (1984). Site specific polyadenylation in a cell free reaction. Cell *36*, 581-591.

Moore, C.L. and Sharp, P.A. (1985). Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell *41*, 845-855.

Morawieki, A. (1964). Dissociation of M and N group mucoproteins into subunits in detergent solutions. Biochim. Biophys. Acta. *83*, 339-347.

Morioka, H., Magnuson, M.A., Mitsuhashi, T., Song, M.K., Rall, J.E., and Nikodem, V.M. (1989). Structural characterization of the rat malic enzyme gene. Proc. Natl. Acad. Sci. USA *86*, 4912-4916.

Morrison, M., Mueller, T.J., and Huber, C.T. (1974). Transmembrane orientation of the glycoproteins in normal human erythrocytes. J. Biol. Chem. *249*, 2658-2660.

Morrow, B. and Rubin, C.S. (1987). Biogenesis of glycophorin A in K562 human erythroleukemia cells. J. Biol. Chem. *262*, 13812-13820.

Moulds, J.J. and Dahr, W. (1989). MNSs and Gerbich blood systems. In Human Immunogenetics: Basic Principles and Clinical Relevance. S.D. Litwin, ed. (New York: Marcell Dekker Inc.), pp. 713-741.

Mueller, T.J. and Morrison, M. (1974). The transmembrane proteins in the plasma membrane of normal human erythrocytes. J. Biol. Chem. *249*, 7568-7573.

Mueller, T.J. and Morrison, M. (1981). Glycoconnectin (PAS 2), a membrane attachment site for the human erythrocyte cytoskeleton. Prog. Clin. Biol. Res. *56*, 95-116.

Mullner, E.W., Neupert, B., and Kuhn, L.C. (1989). A specific mRNA binding factor regulates the iron dependent stability of cytoplasmic transferrin receptor mRNA. Cell *58*, 373-382.

Mullner, E.W. and Kuhn, L.C. (1988). A stem-loops in the 3′ untranslated region mediates iron dependent regulation of transferrin receptor mRNA stability in the cytoplasm. Cell *53*, 815-825.

Munoz, R.F. and Darnell, J. (1974). Poly(A) in mRNA does not contribute to secondary structure necessary necessary for protein sunthesis. Cell *2*, 247-252.

Munroe, D. and Jacobson, A. (1990). Tales of poly(A): A review. Gene *91*, 151-158.

Murayama, J.I, Takeshita, K., Tomita, M., Hamada, A. (1981). Isolation and characterization of two glycophorins from horse erythrocyte membranes. J. Biochem. *89*, 1593-1598.

Murayama, J.I., Utsusmi, H., and Hamada, A. (1989). Amino acid sequence of monkey erythrocyte glycophorin MK. Its amino acid sequence has a striking homology with that of human glycophorin A. Biochim. Biophys. Acta. *999*, 273-280.

Murayama, J.I., Tomita, M., and Hamada, A. (1982). Glycophorins of bovine erythrocyte membranes. Isolation and preliminary characterization of the major component. J. Biochem. *91*, 1829-1836.

Murayama, J.I., Tomita, M., and Hamada, A. (1982). Primary structure of horse erythrocyte glycophorin HA. Its amino acid sequence has a unique homology with those of human and porcine erythrocyte glycophorins. J. Membrane Biol. *64*, 205-215.

Murayama, J.I., Yamashita, T., Tomita, M., and Hamada, A. (1983). Amino acid sequence and oligosaccharide attachment sites of the glycosylated domain of dog erythrocyte glycophorin. Biochim. Biophys. Acta. *742*, 477-483.

Nevins, J.R. and Darnell, J.E.Jr. (1978). Steps in the processing of Ad2 mRNA: Poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell *15*, 1477-1493.

Nevins, J.R. and Wilson, M.J. (1981). Regulation of adenovirus 2 gene expression at the level of the transcriptional termination and RNA processing. Nature *290*, 113-118.

Nudel, U., Soreq, H., Littauer, U.Z., Mabaix, G., Huez, G., Leclercq, M., Hubert, E., and Chantrenne, H. (1976). Globin mRNA species containing poly(A) segments of different lengths. Their functional stability in *Xenopus* oocytes. Eur. J. Biochem. *64*, 115-121.

Ohkuma, S. and Ikemoto, S. (1966). A sialoglycopeptide liberated from human red cells by treatment with trypsin. Nature *212*, 198-199.

Ohkuma, S. and Shinohara, T. (1967). MN Blood group active sialoglycopeptides released from human erythrocytes by pronase treatment. Biochim. Biophys. Acta *147*, 169-171.

Okoye, V.C.N. and Bennet, V. (1985). *Plasmodium falciparum* malaria: Band 3 as a possible receptor during invasion of human erythrocytes. Science *227*, 167-171.

Orkin, S.H., Cheng, T.C., Autonarakis, S.E., and Kazazian, H.H.Jr. (1985). Thalassemia due to a mutation in the cleavage-polyadenylation signal of the huuman β-globin gene. EMBO. J. *4*, 453-456.

Owen, D. and Kuhn, C. (1987). Noncoding 3´ sequences of the transferrin receptor gene are required for mRNA regulation by iron. EMBO J. *6*, 1287-1293.

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P.A. (1986). Splicing of messenger RNA precursors. Ann. Rev. Biochem. *55*, 1119-1150.

Papayannopoulou, T., Nakamoto, B., Kurachi, S., and Nelson, R. (1986). Analysis of the erythroid phenotype of HEL cells: Clonal variation and the effect of inducers. Blood *70*, 1764-1772.

Papayannopoulou, T., Nakamoto, B., Kurachi, S., Tweeddale, M., and Messner, H. (1988). Surface antigenic profile and globin phenotype of two new human erythroleukemia lines: Characterization and interpretations. Blood *72*, 1029-1038.

Pasvol, G., Jungery, M., Weatherall, D.J., Parsons, S.F., Anstee, D.J., and Tanner, M.J.A. (1982a). Glycophorin as a possible receptor for *Plasmodium falciparum*. Lancet 947-950.

Pasvol, G., Wainscoat, J.S., and Weatherall, D.J. (1982b). Erythrocytes deficient in glycophorin resist invasion by the malarial parasite *plasmodium falciparum*. Nature *297*, 64-66.

Paul, R.W. and Lee, P.W.K. (1987). Glycophorin is the reovirus receptor on human erythrocytes. Virology *159*, 94-101.

Pedersen, J.T., Kaplan, H., Wedeck, L., Murphy, R.B., LoBue, J., and Pincus, M.R. (1990). Octapeptide segments from the amino terminus of glycophorin A contain the antigenic determinants of the M and N blood groups systems. J. Lab. Clin. Med. *116*, 527-34.

Perry, R.P., Kelley, D.E., and LaTorre, J. (1974). Synthesis and turnover of nuclear and cytoplasmic polyadenylic acid in mouse L cells. J. Mol. Biol. *82*, 315-331.

Peterson, M.L. and Perry, R.P. (1989). The regulated production of μm and μs mRNA is dependent on the relative efficiencies of μs poly(A) site usage and the cμ4 to μ1 splice. Mol. Cell Biol. *9*, 726-738.

Pinnaduwage, P. and Huang, L. (1989). The role of protein-linked oligosaccharide in the bilayer stabilization activity of glycophorin A for dioleoyl-phosphatidylethanolamine liposomes. Biochim. Biophys. Acta *986*, 106-114.

Platt, T. (1986). Transcription termination and the regulation of gene expression. Ann. Rev. Biochem. *55*, 339-372.

Ponder, E. (1951). Blood *6*, 350. As cited in Cook et al., 1960.

Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends. Biochem. Sci. *14*, 105-110.

Proudfoot, N.J. and Brownlee, G.G. (1976). 3´ non coding region sequences in eukaryotic messenger RNA. Nature *263*, 211-214.

Rahuel, C., London, J., D'Auriol, L., Mattei, M.G., Tournamille, C., Skrzynia, C., Lebouc, Y., Galibert, F., and Cartron, J.P. (1988). Characterization of cDNA clones for human glycophorin A. Use for gene localization and for analysis of normal of glycophorin A deficient (Finnish type) genomic DNA. Eur. J. Biochem. *172*, 147-153.

Rahuel, C., Vignal, A., London, J., Hamel, S., Roméro, P-H., Colin, Y., and Cartron, J.P. (1989). Structure of the 5´ flanking region of the gene encoding human glycophorin A and analysis of its multiple transcripts. Gene *85*, 471-477.

Rearden, A. (1986). Evolution of glycophorin A in the homonoid primates studied with monoclonal antibodies, and description of a sialoglycoprotein analogous to human glycophorin B in chimpanzee. J. Immunol. *136*, 2504-2509.

Reeves, M., Spies, A.G., Nissen, M.S., Buck, C.D., Weinberg, A.D., Barr, P.J., Magnuson, N.S., and Magnuson, J.A. (1985). Molecular cloning of a functional bovine interleukin 2 cDNA. Proc. Natl. Acad. Sci. USA *83*, 3228-3232.

Reeves, R. and Magnuson, N. (1990). Mechanisms regulating transient expression of mammalian cytokine genes and cellular oncogenes. Prog. Nucl. Acids Res. Mol. Biol. *38*, 241-282.

Reid, M.E., Anstee, D.J., Jensen, R.H., and Mohandas, N. (1987a). Normal

membrane function of abnormal β-related erythrocyte sialoglycoproteins. Br. J. Haematol. *67*, 467-472.

Reid, M.E., Chasis, J.A., and Mohandas, N. (1987b). Identification of a functional role for erythrocyte sialoglycoproteins. Blood *69*, 1068-1072.

Renan, M.J. (1987). Conserved 12 bp element downstream from mRNA polyadenylation sites. Gene *60*, 245-254.

Ridgwell, K., Tanner, M.J.A., and Anstee, D.J. (1983). The Wr(b) antigen, a receptor for *Plasmodium falciparum* malaria, is located on a helical region of the major membrane sialoglycoprotein of human red blood cells. Biochem. J. *209*, 273-276.

Robinson, P.J., Bull, F.G., Anderton, B.H., and Riott, I. (1975). Direct autoradiographic visualization in SDS gel of lectin-binding components of human erythrocyte membrane. FEBS Lett. *58*, 330-333.

Ross, A.H., Radhakrishnan, R., Robson, R.J., and Khorana, H.G. (1982). The transmembrane domain of glycophorin A as studied by crosslinking using photoactivatable phospholipids. J. Biol. Chem. *257*, 4152-4161.

Ross, J., Peltz, S.W., Kobs, G., and Brewer, G. (1986). Histone mRNA degradation *in vivo*: The first detectable step occurs at or near the 3´ terminus. Mol. Cell Biol. *6*, 4362-4371.

Ross, J., Kobs, G., Brewer, G., and Peltz, S.W. (1987). Properties of the exonuclease activity that degrades H4 histone mRNA. J. Biol. Chem. *262*, 9374-9381.

Ross, J. (1988). Messenger RNA turnover in eukaryotic cells. Mol. Biol. Med. *5*, 1-14.

Ross, J. and Kobs, G. (1986). H4 histone messenger RNA decay in cell free extracts initiates at or near 3´ terminus and proceds 3´ to 5´. J. Mol. Biol. *188*, 579-593.

Rubin, R.W. and Milikowski, C. (1978). Over two hundred polypeptides resolved from the human erythrocyte membrane. Biochim. Biophys. Acta *509*, 100-110.

Ryner, L.C., Takagaki, Y., and Manley, J.L. (1989). Sequences downstream of AAUAAA signals affect pre-mRNA cleavage and polyadenylation *in vitro* both directly and indirectly. Mol. Cell Biol. *9*, 1759-1771.

Sachs, A.B., Davis, R.W., and Kornberg, R.D. (1987). A single domain of yeast poly(A) binding protein is necessary and sufficient for RNA binding and cell viability. Mol. Cell Biol. *7*, 3268-3276.

Sadler, J.E., Paulson, J.C., and Hill, R.L. (1979). The role of sialic acid in the expression of human MN blood group antigens. J. Biol. Chem. *254*, 2112-2119.

Salzman, N.P., Shatkin, A.J., and Sebring, E.D. (1964). The synthesis of a DNA like RNA in the cytoplasm of HeLa cells infected with vaccinia virus. J. Mol. Biol. *8*, 405-416.

Sanger, F., Nicholson, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463.

Sanger, R. and Race, R.R. (1947). Nature *160*, 505. As cited in Issitt, 1981.

Sarris A.H., and Palade G.E. (1982). Isolation and partial characterization of the sialoglycoprotein fraction of murine erythrocyte ghosts. J. Cell Biol. *93*, 583-590.

Schumperli, D. (1988). Multilevel regulation of replication dependent histone genes. Trends Genet. *4*, 187-191.

Segrest, J.P., Jackson, R.L., Marchesi, V.T., Guyer, R.B., and Terry, W. (1972). Red cell membrane glycoprotein: Amino acid sequence of an intramembranous region. Biochem. Biophys. Res. Comm. *49*, 964-969.

Segrest, J.P., Kahane, I., Jackson, R.L., and Marchesi, V.T. (1973). Major glycoprotein of the human erythrocyte membrane: Evidence for an amphipathic molecular structure. Arch. Biochem. Biophys. *155*, 167-183.

Sgaramella, V. and Khorana, H.G. (1972). Total synthesis of the structural gene for an alanine transfer RNA from yeast. Enzymic joining of the chemically synthesized polydeoxynucleotides to form the DNA duplex representing nucleotide sequence. J. Mol. Biol. *72*, 427-444.

Sharp, P.A. (1987). Splicing of messenger RNA precursors. Science *235*, 766-771.

Sharp, P.A. (1988). RNA splicing and genes. JAMA *260*, 3035-3041.

Shaw, A. and Ziff, E.B. (1980). Transcripts from the major late promoter yield a single early family of 3´ coterminal mRNAs and five late families. Cell *22*, 905-916.

Shaw, G. and Kamen, R. (1986). A conserved AU sequence from the 3´ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell *46*, 659-667.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: Effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. Nucl. Acids Res. *18*, 5799-5805.

Sheets, M.D. and Wickens, M. (1989). Two phases in the addition of a poly(A) tail. Genes Dev. *3*, 1401-1412.

Shyu, A-B., Greenberg, M.E., and Belasco, J.G. (1989). the c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes Dev. *3*, 60-72.

Shyu, A-B., Belasco, J.G., and Greenberg, M.E. (1991). Two distinct destabilizing elements in the c-fos message trigger deadenylation as a first step in rapid mRNA decay. Genes Dev. *5*, 221-231.

Siebert, P.D. and Fukuda, M. (1985). Regulation of glycophorin gene expression by a tumor promoting phorbol ester in human leukemic K562 cells. J. Biol. Chem. *260*, 640-645.

Siebert, P.D. and Fukuda, M. (1986a). Human glycophorin A and B are encoded by separate, single copy genes coordinately regulated by a tumor promoting phorbol ester. J. Biol. Chem. *261*, 12433-12436.

Siebert, P.D. and Fukuda, M. (1986b). Isolation and characterization of human glycophorin A cDNA clones by a synthetic oligonucleotide approach: Nucleotide sequence and mRNA structure. Proc. Natl. Acad. Sci. USA *83*, 1665-1669.

Siebert, P.D. and Fukuda, M. (1987). Molecular cloning of a human glycophorin B cDNA: Nucleotide sequence and genomic relationship to glycophorin A. Proc. Natl. Acad. Sci. USA *84*, 6735-6739.

Silver, R.E., Adamany, A.M., and Blumenfeld, O.O. (1987). Glycophorins of human erythroleukemic K562 cells. Arch. Biochem. Biophys. *256*, 285-294.

Singer, S.J. (1971). The molecular organization of biological membranes. In Structure and Function of Biological Membrane. L.I. Rothfield Ed. (New York: Academic press), pp. 145-222.

Singer, S.J. and Nicolson, G.L. (1972). The fluid mosaic model of the cell membrane. Science *175*, 720-731.

Sippel, A.E., Stavrianopoulos, J.G., Schutz, G., and Fiegelson, P. (1974). Translational properties of rabbit globin mRNA after specific removal of poly(A) with ribonuclease H. Proc. Natl. Acad. Sci. USA *71*, 4635-4639.

Sittman, D.B., Graves, R.A., and Marzluff, W.F. (1983). Histone mRNA concentrations are regulated at the level of transcription and mRNA degradation. Proc. Natl. Acad. Sci. USA *80*, 1849-1853.

Skolnik-David, H., Moore, C.L., and Sharp, P.A. (1987). Electrophoretic separation of polyadenylation specific complexes. Genes Dev. *1*, 672-682.

Skolnik-David, H., Moore, C.L., Sharp, P.A., Patel, D.D., Pickup, D.J., Law, R., Kuwabara, M.D., Briskin, M., Fasel, N., Hermanson, G., Sigman, D.S., Wall, R., Logan, J., Falck-Pedersen, E., Darnell, J.E.,Jr., and Shenk, T. (1987b). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse β maj-globin gene. Proc. Natl. Acad. Sci. USA *84*, 8306-8310.

Soreq, H., Nudel, U., Salomon, R., Revel, M., and Littauer, U.Z. (1974). *In vitro* translation of polyadenylic acid free rabbit globin messenger RNA. J. Mol. Biol. *88*, 233-245.

Southern, E. (1977). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503-517.

Spector, D.H., Villa-Komaroff, L., and Baltimore, D. (1975). Studies on the function of polyadenylic acid on poliovirus RNA. Cell *6*, 41-44.

Springer, G.F., Nagai, Y., and Tegtmeyer, H. (1966). Isolation and properties of human blood group MN and meconium Vg antigens. Biochemistry *5*, 3254-3272.

Springer, G.F., Desai, P.R., and Yang, H.J. (1976). Interrelations of blood group M and precursor specificities and their significance in human carcinoma, In Human Blood Groups. 5th International Convocation on Immunology, Buffalo. pp. 179-187.

Springer, G.F. and Desai, P.R. (1975). Human blood group MN and precursor specificities. Carbohydr. Res. *40*, 183-192.

Steck, T.L. (1974). The organization of proteins in the human red blood cell membrane. J. Cell. Biol. *62*, 1-19.

Steck, T.L. and Hainfield, J.F. (1977). Protein ensembles in the human red cell membrane. Int. J. Cell Biol. 6-14.

Swimmer, C. and Shenk, T. (1981). Selection of sequence elements that substitute for the standard AATAAA motif which signals 3´ processing and polyadenylation of late simian virus 40 mRNA. Nucl. Acids Res. *13*, 8053-8063.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell *52*, 731-742.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1989). Four factors are required for 3′ end cleavage of pre-mRNAs. Genes Dev. *3*, 1711-1724.

Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J., and Shenk, T. (1990). A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. Genes Dev. *4*, 2112-2120.

Tanford, C. and Reynolds, J.A. (1976). Characterization of membrane proteins in detergent solution. Biochim. Biophys. Acta *457*, 133-170.

Tanner, M.J.A., Anstee, D.J., and Judson, P.A. (1977). A carbohydrate deficient membrane glycoprotein in human erythrocyte of S-s-. Biochem. J. *165*, 157-161.

Tanner, M.J.A. (1978). Erythrocyte glycoproteins. Cur. Top. Mem. Transp. *11*, 279-325.

Tanner, M.J.A., High, S., Martin, P.G., and et, a1. (1988). Genetic variants of human red cells membrane sialoglycoprotein β: Study of the alteration occuring in the sialoglycoprotein β gene. Biochem. J. *250*, 407-414.

Tanner, M.J.A. and Anstee, D.J. (1976a). The membrane change in En(a-) human erythrocytes; absence of the major erythrocyte sialoglycoprotein. Biochem. J. *153*, 271-277.

Tanner, M.J.A. and Anstee, D.J. (1976b). A method for the direct demonstration of the lectin-binding components of the human erythrocyte membrane. Biochem. J. *153*, 265-270.

Tanner, M.J.A. and Boxer, D.H. (1972). Separation and some properties of the major proteins of the human erythrocyte membrane. Biochem. J. *129*, 333-347.

Tate, C.G. and Tanner, M.J.A. (1988). Isolation of cDNA clones for human erythrocyte membrane sialoglycoproteins α and δ. Biochem. J. *254*, 743-750.

Tavakkol, A. and Burness, A.T.H. (1990). Evidence for a direct role for sialic acid in the attachment of encephalomyocarditis virus to human erythrocytes. Biochemistry *29*, 10684-10690.

Taya, Y., Devos, R., Tavernier, J., Cheroutre, H., Engler, G., and Fiers, W. (1982). Cloning and structure of the human immune interferon γ chromosomal gene. EMBO J. *1*, 953-958.

Taylor, J.M. (1979). The isolation of eukaryotic messenger RNA. Ann. Rev. Biochem. *48*, 681-717.

Terns, M.P. and Jacob, S.T. (1989). Role of poly(A) polymerase in the cleavage and polyadenylation of mRNA precursor. Mol. Cell Biol. *9*, 1435-1444.

Thomas, D.B. and Winzler, R.J. (1969a). Structural studies on human erythrocyte glycoproteins. J. Biol. Chem. *244*, 5943-5946.

Thomas, D.B. and Winzler, R.J. (1969b). Oligosaccharides from M-active sialoglycopeptides of human erythrocytes. Biochem. Biophys. Res. Comm. *35*, 811-818.

Thomas, D.B. and Winzler, R.J. (1971). Structure of glycoproteins of human erythrocytes: Alkali-stable oligosaccharides. Biochem. J. *124*, 55-59.

Tomita, M., Furthmayr, H., and Marchesi, V.T. (1978). Primary structure of human erythrocyte glycophorin A. Isolation and characterization of peptides and complete amino acid sequence. Biochemistry *17*, 4756-4770.

Tomita, M. and Marchesi, V.T. (1975). Amino acid sequence and oligosaccharide attachment sites of human erythrocyte glycoprotein. Proc. Natl. Acad. Sci. USA *72*, 2964-2968.

Tonkonow, B.L., Hoffman, R., Burger, D., Elder, J.T., Mazur, E.M., Jo Murnane, M., and Benz Jr., E.J. (1982). Differing responses of globin and glycophorin gene expression to hemin in the human leukemia cell line K562. Blood *59*, 738-746.

Treisman, R. (1985). Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5´ element and c-fos 3´ sequence. Cell *42*, 889-902.

Tuech, J.K. and Morrison, M. (1974). Human erythrocyte membrane sialoglycoproteins: A study of interconversion. Biochem. Biophys. Res. Comm. *59*, 352-360.

Unterman, R.D., Lynch, K.R., Nakahashi, H.L., Dolan, K.P., and Hamilton, J.W. (1985). Cloning and sequence of several $\alpha 2\mu$-globulin cDNAs. Proc. Natl. Acad. Sci. USA *78*, 3478-3482.

Vakalopoulou, E., Schaack, J., and Shenk, T. (1991). A 32-kilodalton protein binds to AU-rich domains in the 3´ untranslated regions of rapidly degraded mRNAs. Mol. Cell Biol. *11*, 3355-3364.

Van Hoogvest, P., Du Maine, A.P.M., de Kruijff, B., and De Gier, J. (1984). The influence of lipid composition on glycophorin induced bilayer permeability. Biochim. Biophys. Acta *771*, 119-126.

Varnum, S.M. and Wormington, W.M. (1990). Deadenylation of maternal mRNAs during *Xenopus* oocyte maturation does not require specific cis sequences: A default mechanism for translational control. Genes Dev. *4*, 2278-2286.

Vassalli, J-D., Huarte, J., Belin, D., Gubler, P., Vassalli, A., O, Connell,M.L., Parton, L.A., Rickles, R.J., and Strickland, S. (1989). Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. Genes Dev. *3*, 2163-2171.

Vignal, A., Rahuel, C., London, J., Cherif Zahar, B., Schaff, S., Hataab, C., Okubo, Y., and Cartron, J.P. (1990). A novel gene member of the human glycophorin A and B gene family. Molecular cloning and expression. Eur. J. Biochem. *191*, 619-625.

Villeval, J.L., Pelicci, P.G., Tabilio, A., Titeux, M., Henri, A., Houesche, F., Thomopoulos, P., Vainchenker, W., Garbaz, M., Rochant, H., Breton-Gorius, J., Edwards, P.A., and Testa, U. (1983). Erythroid properties of K562 cells. Effect of hemin, butyrate and TPA induction. Exp. Cell Res. *146*, 428-435.

Vitala, J. and Jarnefelt, J. (1985). The red cell surface revisited. Trends Biochem. Sci. *10*, 112-117.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T., and Itakura, K. (1979). Hybridization of synthetic oligodeoxyribonucleotides to ϕ X 174 DNA: The effect of single base pair mismatch. Nucl. Acids Res. *6*, 3543-3557.

Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E., and Itakura, K. (1981). The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit β-globin DNA. Nucl. Acids Res. *9*, 879-894.

Wasniowska, K., Drzenieck, Z., and Lisowska, E. (1977). The amino acids of M and N blood group glycoproteins are different. Biochem. Biophys. Res. Comm. *76*, 385-390.

Weise, M.J. and Ingram, V.M. (1976). Proteins and glycoproteins of membranes from developing chick red cells. J. Biol. Chem. *251*, 6667-6673.

Wells, E. and Findlay, J.B.C. (1979). Labelling of the intramembranous region of the major sialoglycoprotein of human erythrocytes with a photosensitive hydrophobic probe. Biochem. J. *179*, 265-272.

Welsh, E.J., Thom, D., Morris, E.R., and Rees, D.A. (1985). Molecular organization of glycophorin A: Implications for membranae interactions. Biopolymers *24*, 2301-2332.

Whitelaw, E. and Proudfoot, N. (1986). α-Thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3´ end processing in the human α2 globin gene. EMBO J. *5*, 2915-2922.

Wickens, M. (1990a). In the beginning is the end: Regulation of poly(A) addition and removal during early development. Trends Biochem. Sci. *15*, 320-324.

Wickens, M. (1990b). How the messenger got its tail: Addition of poly(A) in the nucleus. Trends Biochem. Sci. *15*, 277-281.

Wiener, A.S., Unger, L.J., and Cohen, E.B. (1953). JAMA *153*, 1444. As cited in Issitt, 1981.

Williamson, R., Crossley, J., and Humphries, S. (1974). Translation of mouse globin messenger ribonucleic acid from which the poly(adenylic acid) sequence has been removed. Biochemistry *13*, 703-707.

Wilson, T. and Triesman, R. (1988). Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3´ AU-rich sequences. Nature *336*, 396-399.

Wilusz, J., Shenk, T., Takagaki, Y., and Manley, J.L. (1990). A multicomponent complex is required for the AAUAAA-dependent cross linking of a 64 killodalton protein to polyadenylation substrates. Mol. Cell Biol. *10*, 1244-1248.

Wilusz, J. and Shenk, T. (1988). A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell *52*, 221-228.

Winzler, R.J., Harris, E.D., Pekas, D.J., Johnson, C.A., and Weber, P. (1967). Studies on glycopeptides released by trypsin from intact human erythrocytes. Biochemistry *6*, 2195-2202.

Winzler, R.J. (1969). Red cell membrane. G.A. Jamieson and T.J. Greenwalt, eds. (Philadelphia: Lippincott), pp. 157-171.

Winzler, R.J. (1970). Carbohydrates in cell surfaces. Int. Rev. Cytol. *29,* 77-125.

Winzler, R.J. (1972). Glycoproteins of plasma membranes. Chemistry and function. In Glycoproteins: Their Compositions, Structure and Function. Part B. A. Gottschalk, ed. (New York: Elsevier Publishing Company), pp. 1268-1293.

Wright, R.M., Rosenzweig, B., and Poyton, R.O. (1989). Organization and expression of the COX6 genetic locus in *Saccharomyces cerevisiae*: Multiple mRNAs with different 3´ termini are transcribed from COX6 and regulated differentially. Nucl. Acids Res. *17*, 1103-1120.

Wyse, J.W. and Butterfield, A. (1989). Interaction of hemin with erythrocyte membranes: Alterations in the physical state of the major sialoglycoprotein. Biochim. Biophys. Acta *979*, 121-126.

Yokoyama, M. and Trams, E.G. (1962). Effect of enzymes on blood group antigens. Nature *194*, 1048-1059.

Yoshima, H., Furthmayr, H., and Kobata, A. (1980). Structures of the asparagine-linked sugar chains of glycophorin A. J. Biol. Chem. *255*, 9713-9718.

Yurchenco, P.D. and Furthmayr, H. (1980). Expression of red cell membrane proteins in erythroid precursor cells. J. Supramol. Struct. *13*, 255-269.

Zarkower, D., Stephenson, P., Sheets, M., and Wickens, M. (1986). The AAUAAA sequence is required both for cleavage and for polyadenyalation of simian virus 40 pre-mRNA *in vitro*. Mol. Cell Biol. *6*, 2317-2323.

Zarkower, D. and Wickens, M. (1987a). Specific pre-cleavage and post cleavage complexes involved in the formation of SV40 late mRNA 3´ termini *in vitro*. EMBO J. *6*, 4185-4192.

Zarkower, D. and Wickens, M. (1987b). Formation of mRNA 3´ termini: Stability and dissociation of a complex involving the AAUAAA sequence. EMBO J. *6*, 177-186.

Zeevi, M., Nevins, J.R., and Darnell, J.E.J. (1981). Nuclear RNA is spliced in the absence of poly(A) addition. Cell *26*, 39-46.