A FUNCTIONAL ANALYSIS OF CD33 AND CD34

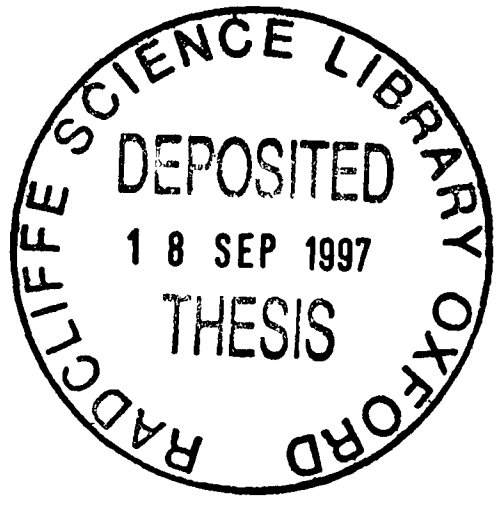
A Thesis Submitted for the degree of

Doctor in Philosophy

to the Board of the Faculty of Biological Sciences
Department of Biochemistry

Elizabeth Kathryn Barber

Brasenose College
University of Oxford
Michaelmas Term
1996



This thesis is dedicated to the patients who may benefit from the work.

Acknowledgements

I particularly thank Dr. Paul Crocker and Professor Andrew McMichael at the Institute of Molecular Medicine, Oxford for providing lab space and valuable supervision to elucidate the final stages of this project. I cannot thank Professor Sir David Weatherall adequately for his support and encouragement in completing this work and am indebted to the related scientific example set by his lab's work on Thalassemia. David Simmons provided lab space, helpful advice on gene cloning and protein structure. I thank the Imperial Cancer Research Fund for providing my scholarship to study at the Institute of Molecular Medicine, John Radcliffe Hospital, Oxford University and the British Council for granting an Overseas Research Scholarship.

I was particularly fortunate to collaborate with Drs. Jamel Chelly and Tony Monaco and I am very thankful for their help and experience. I also thank my previous supervisor Dr. Christopher Rudd at Harvard for his outstanding scientific example, continuing to provide helpful discussions and send antibodies on request. I thank Professor Alan Williams, a Fellow of Brasenose College and the Chairman of the Sir William Dunn School of Pathology for his example and work on the Immunoglobulin Superfamily and on the field of signal transduction. I had the privilege of meeting him shortly before his passing and was inspired by his view that science should always be fun. I especially thank Professor Nick Proudfoot and Brasenose College for their fellowship and support throughout this project. For her genius with computer visuals, I thank Helene Mellor and Geoff Richardson for their help in preparing the colour photography for the manuscript and everyone in Medical Illustration, Tony Brett for computer assistance and Reg Boone for his assistance with the Flow Cytometer and related programs.

Among my coworkers, I would like to thank Justin Newton for his genius, Dr. David Bossy and Dr. Adele Hartnell for their good humor and wisdom, and Paul Bradfield and Mary Vinson from the Crocker lab. Also Sean McCarthy, Amir Mohagghadam, Emmanuel Huguet, Rangana Choudhury, Lyna Zhang and Dr. Roy Bicknell from the Bicknell lab along with Dr. David Ussery, Pascale Romano, Stephen Tucker for their friendship and Professor

Chris Higgins from the Higgins lab for setting a valuable professional example and for his scientific support. Amanda Stafford and Ann Walker were good friends from the Monaco lab.

Lisa Walker, Nick Wells, Richard Issacs and Dr. Ian Hickson from the Hickson Lab made the transcription knowledge a bit easier to absorb. In the Cell Adhesion Lab, Jon Fawcett provided a legendary example of the might of man over machine while Dr. Stuart Bloom and Dr. Claire Holness had inimitable wit. Amanda Littler will be missed. The ICRF and IMM staff were always cheerful, particularly Carol, Inez, Nicky and especially John Moffatt and Gerry Pocklington. I particularly thank Karen Douglass for her intuition and aplomb.

I cannot adequately thank my mother and father both for their support and for the superb example of family that they have set. I thank James especially for his friendship, optimism, unfailing support and for letting me borrow the car!

A FUNCTIONAL ANALYSIS OF CD33 AND CD34

Elizabeth K. Barber
D.Phil.
Brasenose College
University of Oxford

Michaelmas Term
November 1996

Abstract

In the bone marrow, CD34 is the best currently available marker of human multipotential stem cells and is downregulated upon the commitment of cells to the myeloid pathway. CD33 is the earliest marker of stem cell commitment to the myeloid lineage but is down-regulated as myeloid cells mature to granulocytes while expression is retained on monocytes, dendritic cells and macrophages. CD34 is a mucosialin and CD33 a member of the immunoglobulin superfamily. To determine the functions of CD34 and CD33 in early haematopoiesis, soluble forms of CD34 and CD33 have been constructed by PCR based construction of extracellular domain-IgG1Fc (ECD-Fc) fusion plasmids. These chimaeric proteins have been used to define the ligand binding of CD33 and CD34 in a range of assay systems including: screening cell lines by immunofluorescence and iodination analysis; immunohistochemical staining; screening of cDNA libraries transiently expressed in COS cells by "panning" for ligands; and phosphorylation studies to assess the potential of phosphoproteins to regulate these molecules and their ligands in vitro. Iodination demonstrated that CD33 binds increasing numbers of ligands heterophilically on erythroleukemic (K562) and promonocytic (U937) cell lines at sizes ranging from 54-69Kd and 97-110Kd in a differentiation-dependent manner. CD33 was established to associate with several *src*-like kinases while Fc-CD33 precipitates phosphoproteins in the same region in KG-1a, K562 and U937. Novel divalent-cation-dependent CD34 ligands of 66Kd and 110Kd were isolated on HUVEC and both CD34 and Fc-CD34 precipitated phosphoproteins from HUVEC. From panning studies, Fc-CD33 demonstrated low levels of binding to ICAM-1 and a cDNA product sharing homology to the 3' end of dystrophin, termed here as apo-dystrophin-4. Apo-4 appears to give rise to two major proteins, at least one of which may provide an in vitro ligand for CD33 and contain a 3′ enhancer. Models for both CD33 phosphoprotein and ligand binding behaviour are presented.

Table of Contents

Abbreviations ..... **viii**

Chapter 1: Introduction **1**

1.0 CD33 and CD34 as differentiation markers 2

1.1 The role of CD33 and CD34 in myeloid leukemia 4

1.2 Characteristics of CD33 5

1.3 Characteristics of CD34 8

1.4 History of immunoadhesins 9

1.5 CD33 and CD34 as immunoadhesins 12

1.6 CDM8-based transient expression cloning 13

1.6.1 A description of frequently used DNA viral vectors 13

1.6.2 The advantage of SV40-based plasmids 14

1.6.3 The advantage of COS cells 16

1.7 The phosphorylation of CD33, CD34 and their ligands 18

1.8 Ligand studies on CD34 20

1.9 A dystrophin homologue is isolated from Fc-CD33 panning 21

1.10 Aims of the thesis 25

Chapter 2 : Materials and Methods **26**

2.0 Materials 27

2.1 Monoclonal antibodies, antisera and cells 27

2.1.1 Description of leukemic cell lines used 28

2.1.2 HUVEC endothelial cell preparation and culture 29

2.1.3 YAC and related reagents 31

2.1.4 Plasmid vectors 31

2.2 Methods.....32
2.2.1 Molecular biology methods: introductory remarks.....32
2.2.2 Plasmid preparation.....33
2.2.3 Large scale DNA preparation ("Maxiprep").....33
2.2.4 Small scale plasmid DNA preparation ("miniprep").....35
2.2.5 Oligonucleotide synthesis.....36
2.2.6 Oligonucleotide deprotection for PCR.....37
2.2.7 Construction of Fc-adhesins.....37
2.2.8 Polymerase Chain Reaction (PCR).....38
2.2.9 Purification of DNA fragments using Geneclean™.....40
2.2.10 End filling of DNA for blunt end ligation.....40
2.2.11 TA cloning system.....44
2.2.12 Transformation and preparation of competent cells.....44
2.2.13 DNA probe preparation.....46
2.2.14 DNA analysis.....48
2.3 RNA analysis.....54
2.3.1 RNA isolation.....54
2.3.2 PolyA+ mRNA isolation using DynabeadsR.....56
2.3.3 Poly A+ mRNA purification with cellulose columns.....58
2.3.4 RNA electrophoresis.....59
2.3.5 Northern hybridisation.....61
2.3.6 Preparation of radioactive RNA probes.....61
2.3.7 RNAase protection analysis.....64
2.3.8 Reverse transcriptase-polymerase chain reaction.....64
2.3.9 Primer extension.....65
2.3.10 *In vitro* transcription and translation of cDNA..... 67
2.4 Protein expression and adhesion Work.....70
2.4.1 Production of recombinant Fc-adhesins.....70

2.4.2 Large scale manufacture of Fc-adhesins.....71
2.4.3 Immunoselection and Episomal Rescue.....72
2.4.4 Adhesion Assays to ECM proteins.....81
2.4.5 Adhesion assay to HUVEC.....82
2.4.6 Immunoprecipitation and In Vitro Kinase Assay.....82
2.4.7 Sequential Depletion of Lysates.....83
2.4.8 Immunoprecipitation - stringent wash method.....84
2.4.9 Immunoprecipitation with Peptide Blocking.....85
2.4.10 Phosphoamino acid analysis.....85
2.4.11 Cell surface labelling.....86
2.4.12 Metabolic labelling of Fc-chimeras or cell lines.....89
2.4.13 SDS-PAGE analysis of proteins.....89
2.4.14 Western blotting.....90
2.4.15 N-glycanase digestion.....92
2.4.16 Immunofluorescent cell staining.....93
2.4.17 ELISA to test anti-peptide antibody reactivity.....94
2.4.18 Affinity purification of polyclonal antibodies.....95
2.4.19 Immunohistochemistry.....98
2.5 Standard reagents and solutions.....99
2.5.1 Media and standard buffers.....99
2.5.2 RNAse protection reagents.....99
2.5.3 Primer extension reagents.....101
2.5.4 Immunoprecipitation - stringent wash method.....102
2.5.5 Standard stringency immunoprecipitation buffer.....102
2.5.6 Cell culture.....103
2.5.7 Bacterial culture solutions.....103
2.5.8 Molecular biology solutions and standard buffers.....104

Chapter 3: Fc-CD33 and Fc-CD34 Cloning and Binding Studies 112

3.0 Specific aims 113

3.1 Vectors used in Fc construction 113

3.2 CD33 and CD34 as immunoadhesins 118

3.2.1 Fc-CD33 and Fc-CD34 constructs in the Rg vector 119

3.2.2 N-Glycanase digestion of CD33 125

3.2.3 CD33 and CD34 extracellular domain PCR products 126

3.3 CD33 ligand binding studies using Fc-adhesins 131

3.3.1 Characteristics of myeloid cell lines 131

3.3.2 Fc-CD33 binding increases with differentiation 133

3.3.3 Fc-CD33 binds specifically among Fc-adhesins 136

3.3.4 Fc-CD33 binds ligand in a proportional manner 138

3.3.5 Surface iodination of cell lines 141

3.3.6 Fc-CD33 and Fc-CD34 show no specific ECM binding 143

3.3.7 Immunohistochemical analysis of CD33 and CD34 147

3.4 Discussion 150

Chapter 4: Phosphorylation Studies on CD33 and its Ligands 153

4.0 Specific aims 154

4.1 Phosphorylation in cell signalling 154

4.1.1 CD33 possesses homologous YXXL/V motifs 155

4.1.2 CD33 shares MAG homology at a Major Phosphorylation Site 156

4.1.3 Importance of spacing between YXXL motifs 161

4.1.4 Other motifs identified 161

4.1.5 PXXP motifs and kinase associations 162

4.2 ITAMs and ITIMS in kinase signalling 162

4.3 Phosphorylation studies on CD33 and potential ligands 163

4.3.1 CD33 phosphoproteins increase with differentiation 164
4.3.2 *Src*-related kinases associate with CD33 166
4.4 Immunoprecipitation of phosphoproteins with Fc-CD33 179
4.5 Fc-CD33 ligands associate with *lck* in K562 171
4.6 Discussion 176

Chapter 5: CD34 Binding and Ligand Studies 179
5.0 Specific aims 180
5.1 Binding studies on Fc-chimeras to HUVEC 180
5.2 Fc-CD34 binds to a 110 Kd protein on iodinated HUVEC 183
5.3 Fc-CD34 binding to HUVEC is divalent cation-dependent 185
5.4 Fc-CD34 shows no homophilic binding to myeloid cell lines 186
5.5 Fc-CD34 binds at a similar level to all candidate cDNA products 188
5.6 Fc-CD34 precipitates a 70Kd HUVEC phosphoprotein 192
5.7 CD34 and ligand phosphoproteins are labelled on T and Y 195
5.8 Discussion 198

Chapter 6: Characterisation of a Potential CD33 Ligand 200
6.0 Specific aims 201
6.1 Fc-CD33 as a probe with the panning method 201
6.2 FACS staining of "candidate" cDNAs with Fc-CD33 204
6.3 An Fc-CD33 candidate ligand is termed apo-dystrophin-4 209
6.3.1 Apo-4 shares a similar inversion with globin Gγ(γβ)o 214
6.3.2 Yeast Artificial Chromosome (YAC) hybridization 214
6.3.3 Homologous regions in cDNA, genomic and mouse DNA 219
6.4 GRAIL identifies an exon upstream of apo-dystrophin-4 222

6.11.4 Transcription of apo-4 does not require a TATA box 289

6.11.5 Proposed explanations for stop-codon "readthrough" 289

Chapter 7: Concluding Discussion 293

7.0 The limits and advantages of Fc-adhesins 295

7.1 Chapter 3: Fc-CD33 Binding Increases with Differentiation 295

7.2 Chapter 4: CD33 and ligand kinase associations 296

7.3 Chapter 5: Fc-CD34 binds unique ligands on HUVEC 299

7.4 Chapter 6: Characterisation of an apo-dystrophin cDNA 300

Chapter 8: Bibliography 308

Appendix: Apo-4 Primers and Structural Information 329

Appendix I 330

Appendix II 331

Appendix III 333

Appendix IV 334

6.4.1 A possible upstream transcription start site is found 223
6.4.2 The start site contains a potential Pol II binding region 226
6.5 Northern blotting shows a major apo-4 transcript of 1.62Kb 227
6.5.1 RNAse protection reveals no protected transcript 229
6.5.2 RT-PCR analysis produces evidence for apo-4 230
6.5.3 Evidence for splicing and a reproduced inversion 232
6.5.4 Subcloning reveals close homology with apo-4 242
6.6 Inversion breakpoints are identified on genomic DNA 243
6.6.1 Downstream RSS's flank an inversion breakpoint 246
6.6.2 Potential mechanisms to produce the inversion 248
6.7. The inversion is necessary to produce proteins *in vitro* 250
6.8 Peptide antisera designed against putative antigenic regions 253
6.8.1 Antisera precipitate specific protein products 254
6.8.2 Surface labelling demonstrates at least four species 255
6.8.3 Apo-4 antibody and Fc-CD33 precipitate a 50Kd band 258
6.8.4 Fc-CD33 precipitates a 50Kd band from K562 and TFs 260
6.8.5 p50 can be specifically blocked by peptide 262
6.8.6 Apo-4 appears to exist as a cell surface heterodimer 264
6.9 Predicted STADEN and GeneID splice sites 265
6.9.1 Predicted structure of full-length apo-4 protein 268
6.9.2 TopPred analysis predicts five TM domains 271
6.10 Apo-4 P1 antibodies suggest specific staining in tissues 275
6.10.1 An extracellular N-terminus is stained 277
6.10.2 Fc-CD33 shows low-level apo-4 staining 278
6.11 Discussion 285
6.11.1 Apo-4 expression and Fc-CD33 binding 286
6.11.2 The inversion could result from several factors 287
6.11.3 A potential role for RAG-1 and RAG-2 in the inversion 289

Abbreviations

anti-HulgFc	anti-human Fc region immunoglobulin
BcR	B cell receptor
β-ME	β-mercaptoethanol
BSA	bovine serum albumin
CD	cluster of differentiation
CDM8	the pCDM8 vector
C domain	constant domain
C terminus	carboxyl termins
cDNA	complementary deoxyribonucleic acid
CEA	carcinoembryonic antigen
CHO	Chinese hamster ovary
CNBr	cyanogen bromide
Con A	Concanvalin A
cpm	counts per minute
Dig	Digitonin
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECD	extracellular domain
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EGFR	epidermal growth factor receptor
EGTA	ethyleneglycol-*bis* (b-aminoethyl ether)-N,N,N',N'-tetra acetic acid
FACS	fluorescence-activated cell sorter
FCS	foetal calf serum
FITC	fluorescein isothiocyanate
FPLC	fast protein liquid chromatography
GAR	goat anti rat-immunoglobulin antibodies
glg	goat anti-mouse immunoglobulin antibodies
gp	glycoprotein
GPI	glycosyl-phosphatidylinositol
GTPase	guanosine triphosphatase
HBS	HEPES-buffered saline

HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HIV	human immunodeficiency virus
HRF	homologous restriction factor
IAA	iodoacetic acid
ICAM	intercellular adhesion molecule
Ig	immunoglobulin
IgSF	immunoglobulin superfamily
Kd	kilodaltons
LFA	lymphocyte function-associated antigen
LPS	lipopolysaccharide
M	Molar
mIg	membrane immunoglobulin
mRNA	messenger ribonucleic acid
NP-40	Nonidet® P 40
NK cell	natural killer cell
N-terminus	N amino terminus
PAGE	polyacrylamide gel electrophoresis
PAS	protein A sepharose
pBS	the pBluescript SK+ vector
PBS	phosphate-buffered saline
PBSA	phosphate-buffered saline A (without Ca^{++} or Mg^{++})
PCR	polymerase chain reaction
PI	phosphatidylinositide
pIG1	plasmid immunoglobulin 1 vector
PLC-γ	phospholipase C gamma
PMA	phorbol myristate acetate
PTP	protein tyrosine phosphatase
RNA	ribonucleic acid
RT	room temperature (25°C)
SAC	*Staphylococcus aureus* cells
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SF	superfamily
SH	*src* homology
sIg	surface immunoglobulin

SV40	simian virus 40
TBS	Tris-buffered saline
TcR	T cell receptor
TEMED	N,N,N',N'-tetramethylethylenediamide
TNF	tumour necrosis factor
TX-100	Triton® X-100
V domain	variable domain

Chapter 1

Introduction

1.0 CD33 and CD34 as differentiation markers

CD33 is an important diagnostic cell surface marker used to differentiate between Acute Lymphocytic and Acute Myelocytic Leukaemias while CD34 is an important marker for the identification and purification of lymphohematopoetic stem cells for research and potentially for clinical stem cell transplantation to reconstitute bone marrow (Civin 1990). To identify their cognate ligands will give us important information as to their precise role in normal and aberrant cell differentiation. It is important to look at their function in tandem as they appear to be closely linked in defining the differentiation pathway. CD34 is downregulated upon progenitor cell commitment to the lymphoid and myeloid pathways while CD33 is the earliest marker to indicate stem cell commitment to the myeloid pathway.

CD33 (gp67) is a differentiation antigen mainly restricted to cells of the monocytic/myeloid lineage (Andrews et al. 1983; Griffin et al. 1984; Matutes et al. 1985). In bone marrow, the earliest colony forming units (CFU) appear to lack CD33 surface expression in contrast to CD34 (Andrews 1989). $CD33^+$ cells comprise virtually all the colony forming units for progenitors of granulocytes mononuclear phagocytes (CFU-GM) (Matutes et al. 1985) and a variable proportion of the colony forming units for granulocytes, erythrocytes, monocytes and megakaryocytes (CFU-GEMM), early erythroid progenitors (BFU-E), mast cells and leukemic myeloblasts from nearly all Acute Myelocytic Leukemia (AML) cells (Andrews 1989; Griffin et al. 1984; Sabbath 1986). CD33 monoclonal antibodies (Mabs) (MY9, L1B2 and L4F3) react with approximately 90% of AML cells but virtually no ALL or Chronic Lymphocytic Leukemia (CLL) or erythroid leukaemic cells making them a useful tool both for a differential diagnosis of the leukemias (Andrews 1989;

Griffin et al. 1984; Sabbath 1986). The MY9 and L4F3 antibodies have also been used to purge minimal residual leukemic myeloblasts from autologous bone marrow grafts given to patients with AML in remission, as CD33 is expressed by leukemic myeloblasts but not the early progenitors necessary for bone marrow recovery (Ritz 1987).

In the myeloid lineage, CD33 is strongly expressed on promonocytes and monocytes and on myeloblasts, promyeloblasts, myelocytes and metamyelocytes (Andrews 1989; Griffin et al. 1984). CD33 is weakly expressed on band forms and mature granulocytes as well as from primitive pluripotent hematopoietic stem cells (which express CD34 (MY10)) and is undetectable on early lymphoid precursors, normal and malignant lymphoid cells, nonhematopoietic cells and terminally differentiated granulocytes (Andrews 1989). It is occasionally found in some patients with poorly differentiated ALL (Bain 1990).

CD33 has also been found on T lymphocytes activated with CD3 (Nakamura 1994) as well as on a subset of human umbilical cord blood-derived natural killer (NK) cells which coexpress CD56 (Handgretinger 1993). The new subset of NK cells of the $CD56^{+}/CD33^{+}$ resembled large granular lymphocytes (LGLs) and exhibited decreased cytolytic activity relative to $CD56^{+}/CD33^{-}$ NK cells (Handgretinger 1993). CD33 expression increases as cells mature along the monocytic and granulocytic pathways, with a decline upon terminal differentiation. CD34 is most abundant on stem cells and begins to decline upon the appearance of CD33.

1.1 The role of CD33 and CD34 in myeloid leukemia

In clinical diagnosis, CD33 (gp67) antibodies are used to differentiate between Acute Lymphocytic (ALL) and Acute Myelocytic Leukemias (AML) (Griffin et al. 1984). CD33 is expressed on clonogenic leukemic cells from about 90% of patients with AML (Griffin et al. 1984; Neame 1986; Schwarzinger 1990) as well as being expressed on normal myeloid progenitors. Early hematopoietic $CD33^-$ progenitors are able to grow normally in long-term bone-marrow culture suggesting that their primary role is expressed later in the differentiation pathway (Andrews et al. 1986). In vitro studies suggest that the majority of leukemic blasts are not capable of long-term replication but are generated by a small percentage of rapidly proliferating stem cells (Griffin 1986). The end-stage phenotype of these cells often differs from the clonogenic stem cell (Wouters and Lowenberg 1984). Thus the clonogenic stem cell was used as a target in bone-marrow purging with anti-CD33 antibodies prior to transplantation (Robertson 1992). Following transplantation, $CD33^-$ populations required a significantly longer time to achieve neutrophil engraftment than those purged with other surface markers. $CD33^+$ committed myeloid progenitor cells thus appear crucial to the early stages of bone marrow engraftment after transplant, but are not essential for the eventual recovery or hemopoiesis (Robertson 1992). From this study it could be inferred that ligands for CD33 may be on neutrophils, as their recovery was greatly inhibited by its absence. Neutrophils are difficult to work with experimentally, only living for a few hours after isolation (K. Patel, personal communication) unless they are maintained in growth factors like GM-CSF but would be interesting to explore for CD33 ligand in future work.

Radiolabelled anti-CD33 antibodies have also been used to augment bone marrow transplants; as with soluble recombinant ECDs, however, a relatively short half-life limited their usefulness in therapy (Appelbaum 1992). Recently, the M195 anti-CD33 antibody has been "humanised" with the addition of a human Fc portion, thus improving the ability of the radiolabelled antibody to be internalised and to retain the radioisotope for improved effector functions such as ADCC (Caron 1994). Human CD33-Fc will be used in a similar way to target ligand on leukemic cells in an effort to eliminate them, ideally, without compromising healthy cells.

1.2 Characteristics of CD33

Peiper et al. sublocalised the CD33 gene to the q13.3 region of chromosome 19 by *in situ* hybridisation. RNA transcripts of ~1.6 Kb (predominant in the acute promyelocytic leukemia cell line HL-60) and ~1.4 Kb (predominant in the erythroleukemic cell line K562) were isolated (Peiper 1988). Dots next to chromosomal regions show relative regions of hybridisation

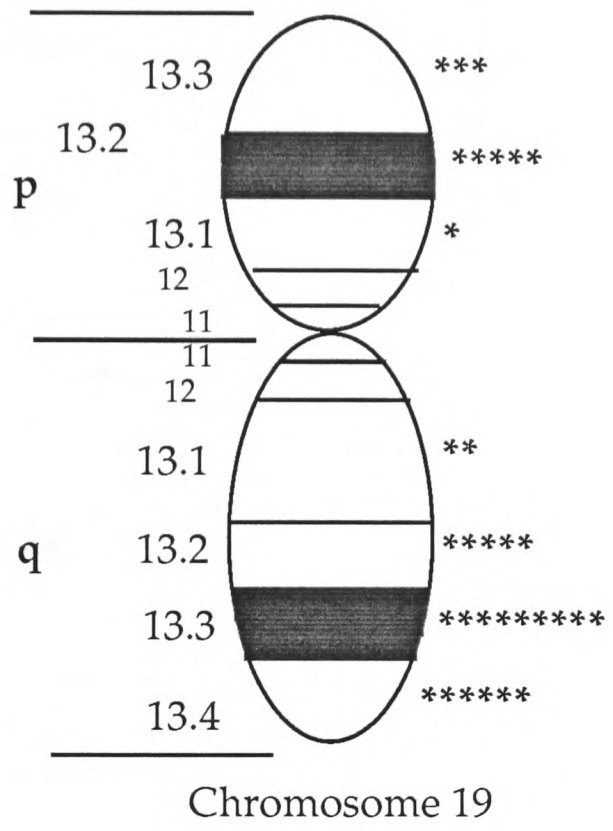


Figure 1.1 The chromosomal location of the CD33 gene. CD33 was localised to the q13.3 region of chromosome 19 and RNA transcripts of 1.6 Kb from HL-60 and 1.4 Kb from K562 were isolated (Peiper 1989).

Subsequently, the CD33 transcripts were also reported as 1.8 Kb and 1.4 Kb in a panel of myeloid leukemic cell lines which also included the myeloid leukemic cell lines KG-1 (stem cell-like), HEL (erythroleukemic), and U937 (promonocytic lymphoma) which predominantly expressed CD33 and from which a cDNA library was constructed. The CD33 cDNA clone was isolated from this library after 3 rounds of transient expression in COS cells and clonal enrichment through panning with the anti-CD33 Mabs MY9, L1B2 and L4F3. The CD33 insert from transfected COS cells was shown to be 1.5Kb, possessing a mass slightly smaller than the mass of CD33 expressed on myeloid cells (Simmons 1988b). The CD33 cDNA sequence predicts a 40Kd polypeptide glycosylated to 67 Kd (Simmons 1988b). Mouse CD33 has recently been isolated and predicts a similar structure for two isoforms (Tchilian 1994). The cDNA sequence of CD33 codes for two immunoglobulin superfamily (IgSF)

domains which share the greatest sequence homology (in the V domain) with myelin-associated glycoprotein (MAG), CD22, (Simmons 1988b) and sialoadhesin (Crocker 1994) and it has recently been included in a newly defined family of sialic acid binding proteins, the sialoadhesins (Kelm 1994a).

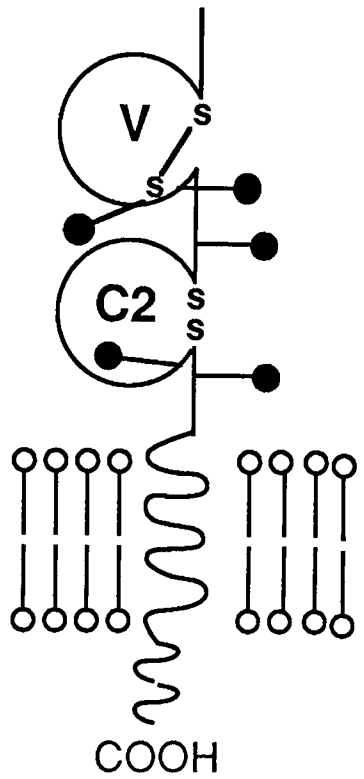


Figure 1.2 The structure of CD33. The predicted protein structure of CD33 with putative N-glycosylation sites shown (reproduced with permission from (Barclay 1993)).

CD33 is a member of the immunoglobulin superfamily (IgSF) in that it possesses a variable (V) and a C2-set domain (commonly seen in non-lymphoid cells) (Williams 1989). The position of the cysteines in the V domain are unusual however, and, like its homologue in MAG, its structure predicts an intra-β-pleated sheet disulfide bond. Thus the unusual position of the disulfide bonds between the V and C2-set Ig domains. This unusual bond position would result in the masking of the potential R-G-D binding sequence (to which integrins often bind) in MAG due to its location within one rather than between two "Ig folds". CD33 shares an identical positioning of the R-G-

D residues and is thus not predicted to bind via an R-G-D motif as do many matrix proteins (Williams 1989).

1.3 Characteristics of CD34

CD34 is unique in its narrow specificity of expression on lymphohaematopoietic progenitor (stem) cells. It stains positive in approximately 70% of B-lineage acute lymphoblastic leukemia (ALL) cases, less than 5% of T-lineage ALL cases, and approximately 40% of acute myeloid leukemia (AML) cases. CD34 is negative in chronic lymphoblastic leukemia (CLL) and lymphomas. Outside the lymphohaematopoietic system, CD34 is expressed only on vascular endothelial cells and rarely on solid tumour cells. CD34 has recently been shown to bind to L-Selectin in High Endothelial Venules (HEV) (Baumhueter 1993). This binding has not been reproduced by other labs on the KG-1a myeloblastic cell line (Oxley 1994). Thus, CD34 likely possesses different binding behaviour in different locations. The CD34 cDNA predicts a 105-120Kd protein and an unusually high degree of O-linked glycosylation sites and nine predicted N-linked sites (Simmons 1992). This structure is reminiscent of the heavily O-linked extracellular domains of CD43 and CD45 (Barclay 1993) (Figure 1.3).

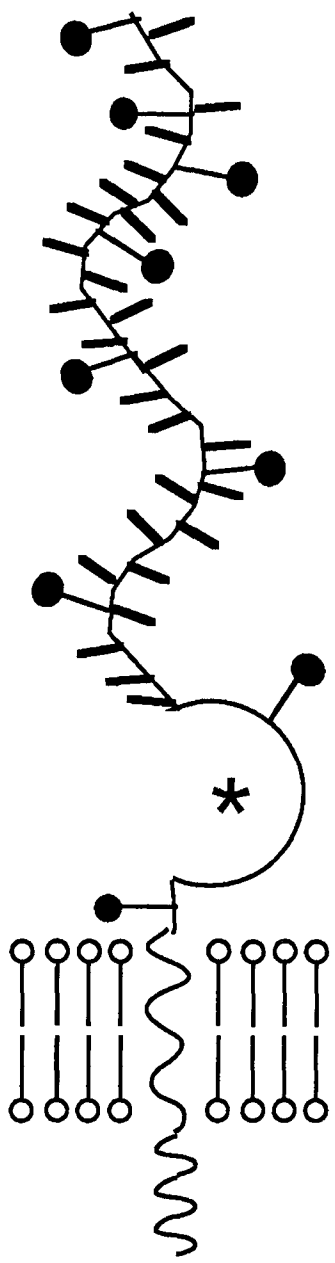


Figure 1.3 The extracellular domain structure of CD34. The structure of CD34 shows only one immunoglobulin-like domain and a long stalk of at least 30 O-glycosylation sites (—) and nine N-linked glycosylation (•—) sites. The structure is reproduced with the permission of N. Barclay from the original version (in Barclay 1993) .

1.4 History of immunoadhesins

As in many areas of biomedical research during the past 15 years, the need for an effective AIDS therapy has produced a broadly useful technology. In early work, the Fc portion of antibody was fused with the *myc* gene to create a "myc-tagged" antibody (Neuberger 1985). The term 'immunoadhesin' was

coined upon construction of the initial CD4-Human IgG1Fc fusion protein for HIV therapeutic studies (Capon 1989). In this study, the constant region from immunoglobulin molecules was chosen because it had previously demonstrated an increased half-life *in vivo* relative to soluble CD4 extracellular domain (ECD) regions alone (and Fab regions of antibody) and thus it was postulated that it could extend the half-life and thus the therapeutic potential of the CD4-ECD ligated to it. Other useful functions of the Fc portion included Fc receptor binding (to facilitate Antibody Dependent Cell-mediated Cytotoxicity or ADCC), complement (C1q) fixation, and placental transfer to potentially protect a fetus from infection. In addition, the known high binding capability of IgG Fc to Protein A sepharose would facilitate purification of large quantities of the desired protein (Capon 1989). That the chimeras would be secreted as homodimers would likely increase their binding avidity against HIV.

The CD4-ECD-Fc DNA construct was made using the entire constant region (CH1-hinge-CH2-CH3) of genomic IgG1 DNA (Ellison 1982) and fusing it to the cDNA coding for the extracellular domains of CD4. Secretion of the soluble recombinant CD4-human Fc construct, unlike mouse ECD-Fc, was not accompanied by κ or λ light chain association (Capon 1989). In addition, the chimeric protein could be purified to homogeneity and the secreted molecule could mimic immunoglobulin structure in its ability to dimerise as well as to bind Fc receptors. Of the three types of Fc receptors known to be expressed on leucocytes, only FcRI (CD64) binds IgG with high affinity, while the others (CD16, CDw32) bind aggregated IgG with low affinity (Burton et al. 1988). A monocyte/macrophage cell line was used to demonstrate that ECD-Fc chimeras bind to the same degree as intact IgG1 to high affinity to Fc receptors (Capon 1989).

The four-domain CD4-Fc had a reduced half-life (~7 hrs.) relative to the two-domain CD4-Fc (~48 hrs.) but still greater than soluble CD4 ECD (sCD4) alone (~15 min.) (Capon 1989) suggesting that the antibody structure itself has evolved to two domains, as this is the structure which has the longest half-life *in vivo*. Interestingly, the CD33 ECD-Fc closely mimics immunoglobulin structure which may suggest why it has been a useful molecule in binding studies and could suggest a role for it in future. The CD4-Fc chimeras were also shown to effectively block HIV infection at ~0.05mM. Notably, the potency of the CD4 analogues was ~fivefold higher than sCD4 and were shown to block infection at a similar level in monocytes and monocytic cell lines (Capon 1989). A later study suggested that in order to maximise therapeutic potential, antibodies (and Fc-chimeras) should contain human V and C regions which have been allotypically matched to the patient to minimise anti-idiotypic responses (Bruggemann 1989).

Subsequent studies on CD4-Fc showed that deletion of the CH1 domain with retention of the hinge region retained dimerisation capability in mouse (Traunecker 1989a) and human (Zettlmeissl 1990) Fc regions fused to a human extracellular domain (ECD). Both groups compared the ability of fusion proteins containing and lacking CH1 to be secreted and found those lacking CH1 to be secreted at about a four-fold lower concentration, likely, as Traunecker proposed, due to a strong interaction of the hydrophobic face of the CH1 domain with the heavy chain binding protein in the absence of immunoglobulin light chains. In addition, Fc chimeras lacking CH1 had been shown to bind C1q (Traunecker 1989a) when those with CH1 could not, although they did bind Fc receptors (Capon 1989). It had also been shown in the analysis of human heavy-chain disease proteins that heavy chains of different subclasses can be secreted without light chains when the CH1

domain is deleted (Seligmann 1979). Zeittlmeisel also constructed CD4-CH2-CH3 proteins which were secreted as monomers in the absence of a hinge region. The same lab went on to demonstrate the use of an immunoadhesin to identify hyaluronate as the receptor for CD44 using fusion proteins which incorporated the hinge region and utilised the CD5 leader sequence in the construct to achieve increased protein expression (Aruffo 1990). CD33 has proved to contain a robust leader sequence which can also be used for genes possessing endogenous sequences which express poorly (Fawcett 1994a).

1.5 CD33 and CD34 as Immunoadhesins

The immunoglobulin-like domains found in CD33 and CD34 suggested that they would be suitable candidates upon which to try the ECD-Fc technology to attempt to probe for their ligands directly using the transient transfection and panning method. This study represents one of the first attempts to employ an ECD-Fc construct like an antibody to identify an unknown ligand and the first attempt with Fc-CD33 and Fc-CD34. Importantly, all previously isolated ligands from the panning technique had used high affinity antibodies as the probe for ligand. To identify the cognate ligands of CD33 and CD34 could give us important information as to their precise role in normal and aberrant cell differentiation. They are often presented together in binding studies, as CD34 is downregulated as CD33 is upregulated on commitment to the myeloid pathway (Andrews 1989).

1.6 CDM8-based transient expression cloning

The CDM8 vector employs several useful features for the expression of cDNAs in transient expression cloning systems. The CDM8 vector (Figure 1.4b) was developed in 1987 by Brian Seed (Seed 1987a) and used in conjunction with the COS cell expression system to achieve high expression of the desired glycoprotein on the cell surface to facilitate the "rescue" of its cDNA by antibody "panning" (Aruffo 1987; Seed 1987a; Seed and Aruffo 1987b). The SV40 genome in the CDM8 vector and in the COS cell, work in conjunction to enable maximal protein expression. The CDM8 based transient expression system has recently been described in detail (Simmons 1993; Simmons and Needham 1991) (Figure 1.4a).

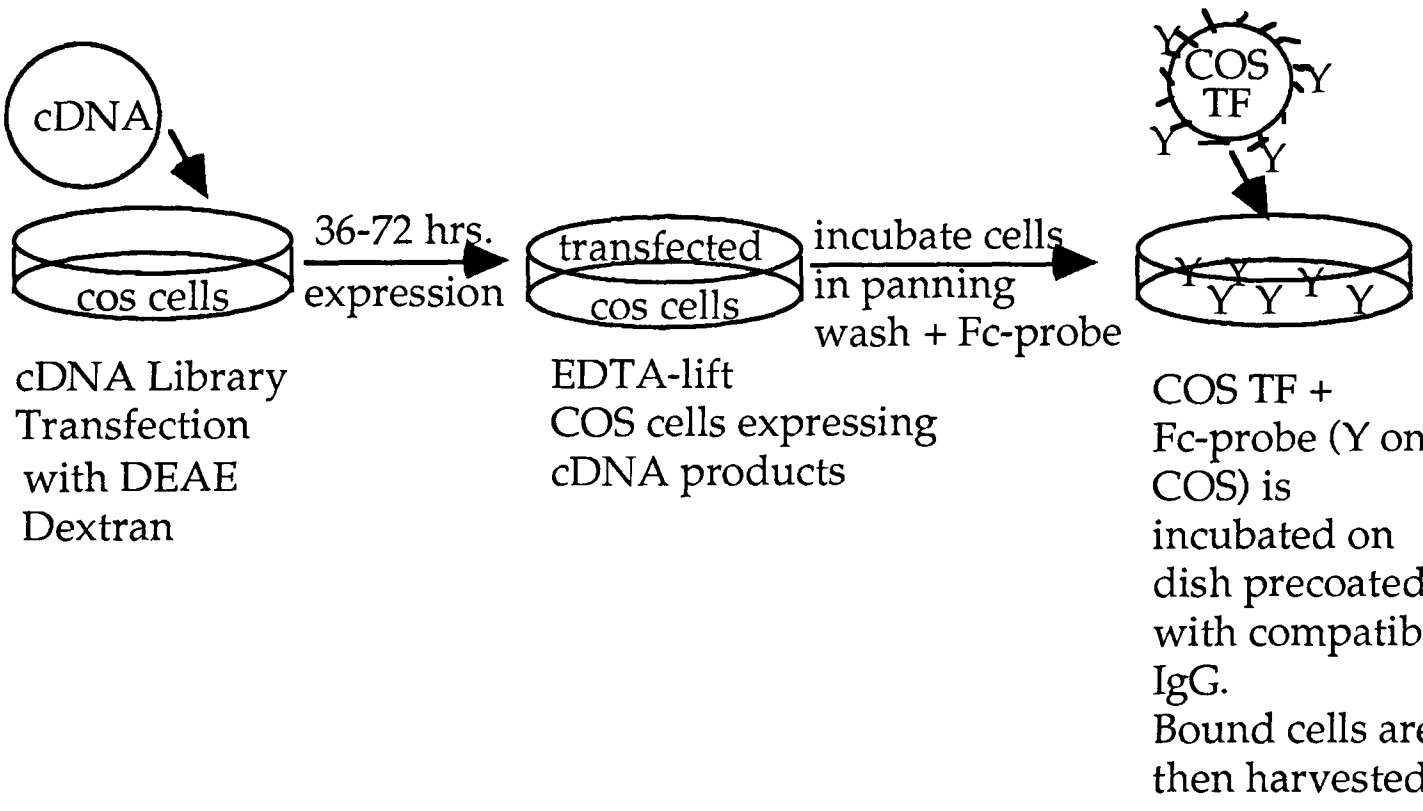


Figure 1.4a. A pictorial description of the panning process. Following cell binding of the cell to the dish, potential ligand-expressing cDNAs are "harvested" and the clones examined for a common pattern of cDNA inserts on mini-prepped cDNA.

1.6.1 A description of frequently used DNA viral vectors

The efficiency and strength of the "transfection/transient expression/panning and rescue" cycle in the system employed in this thesis is dependent on the vector and specific cell type with which it interacts. Mammalian tumor viruses, in particular the papovaviruses, primate SV40 and murine polyoma and herpes viruses, in particular Epstein Barr virus (EBV) have been exploited for cloning. The two advantageous features of these viruses are origins of replication and trans-acting DNA-binding proteins which specifically interact with the origin and the polymerase/primase complex to allow replication of the viral genome to high copy number. As SV40 replicates to a much higher copy number, typically 10^3-10^5 per cell, than EBV, it was chosen in the development of CDM8 (Seed 1987a). A high copy number of the viral vector in the host cell elevates the overall level of transcript production per cell and thus facilitates recovery of the amplified genome from selected cells for reintroduction into *E. coli.*, specifically, the strain MC1061/p3 with CDM8.

The difference in replication potential between EBV-based and SV40 based plasmids can best be illustrated by examining their structural differences. Plasmids containing EBV (Hauer 1989; Kioussis et al. 1987) normally carry both the EBV origin of replication (oriP) and the trans-acting origin amplifier (EBNA1), thus allowing their transfection and amplification in any cell type. SV40-based plasmids by contrast, carry only the origin of replication (SV40 ori, a 350 base pair (bp) fragment of the SV40 genome) and lack the amplifier. Thus, these plasmids must be transfected into cells carrying integrated copies of crippled SV40 genomes which express the SV40 replicator protein large T antigen.

1.6.2 The advantage of SV40-based plasmids

EBV-based plasmids produce only 1-10 copies of plasmid per cell and must often use the high efficiency of phage λ packaging extracts to rescue the vectors. SV40-based plasmids produce a thousand to ten thousand-fold more plasmids per cell thus allowing high expression of specific molecules and direct recovery of episomes from selected cells bypassing any need for additional packaging systems. While SV40 can be toxic to the cell in the long term, cells remain healthy at least 36-72 hours post-transfection which allows for expression, selection by panning and episomal recovery.

All SV40 plasmids share similar features, as outlined above for CDM8 (Figure 1.4b). An SV40 genomic fragment which acts as the origin of replication; a eucaryotic (in this case cytomegalovirus (CMV)) enhancer/promoter; the site of insertion for the cDNA of interest; downstream transcript processing elements: a polyadenylation signal and intron; a prokaryotic origin of replication (pVX) and an appropriate sites for drug selection in *E. coli.* contained in Sup F. CDM8 has the additional advantage of non-palindromic BstXI linkers at the sites flanking the insert, a variety of endogenous restriction sites for ease of cDNA insertion and a polyoma origin of replication which allows the vector to be used in murine polyoma transformed WOP and COP cells (Seed 1987a; Simmons and Needham 1991) which is especially useful for murine antibodies which could potentially cross react with glycoproteins on the surface of the mammalian cells COS-1 and COS-7 (Gluzman 1981) with which CDM8 has usually been used for the cloning of human cell surface molecules. CDM8 also includes an M13 origin of replication which allows production of single-stranded plasmid templates when appropriate strains of *E.Coli.* are superinfected with filamentous f1

phage (eg. M13). Finally, a T7 promoter at the 5' end of the cDNA insertion site allows for *in vitro* production of RNA templates for transcript terminus mapping and production in cells which permit T4 RNA polymerase activity. Alternative versions of CDM8 are described elsewhere (Simmons 1993; Simmons and Needham 1991).

The CDM8 Vector

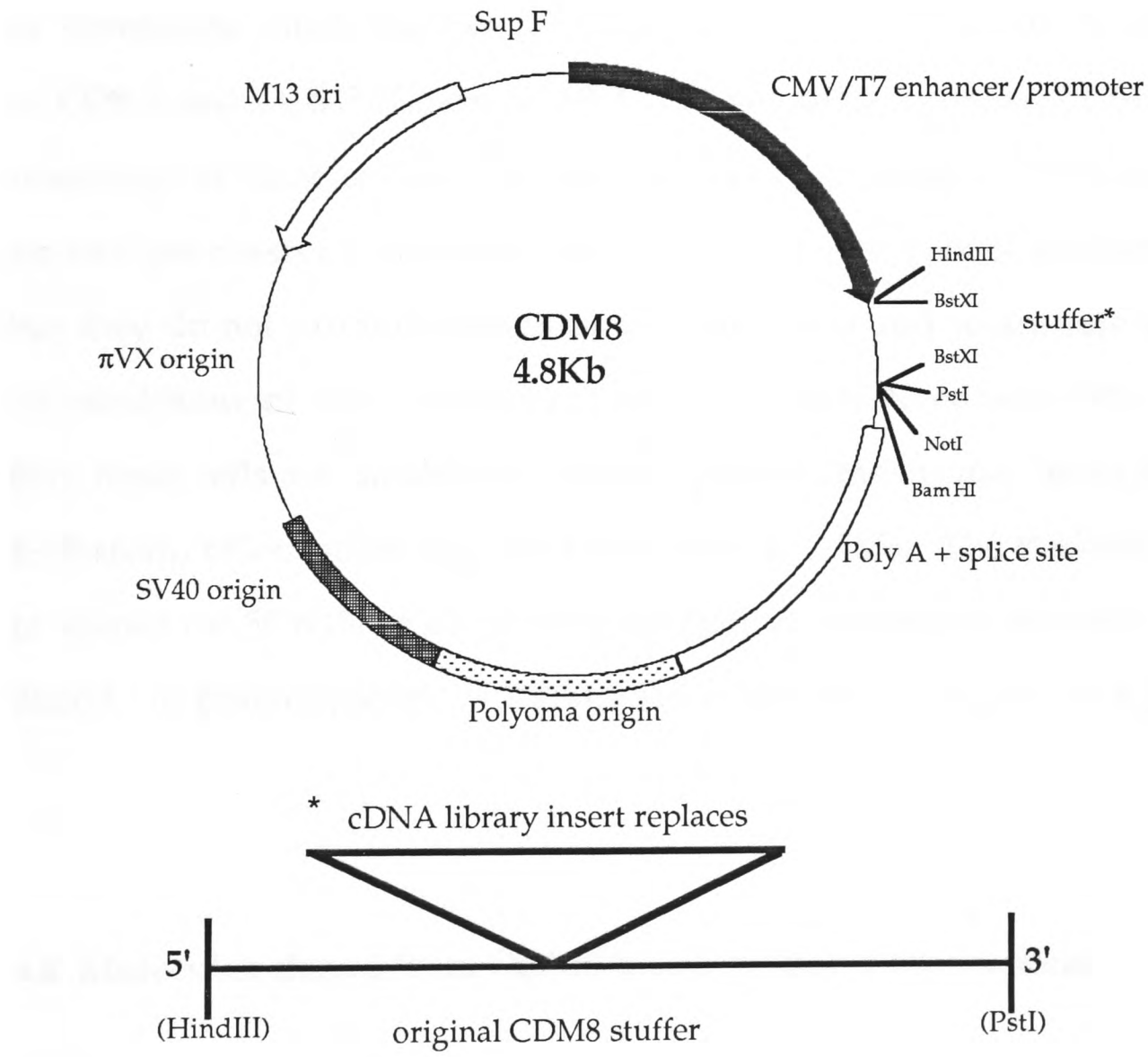


Figure 1.4b. The CDM8 Vector. This vector was used to make all of the cDNA libraries to express the proteins of interest probed with Fc-chimeras. The diagram illustrates the essential features of the plasmid for the COS cell expression system which include: the CMV promoter, restriction sites for cDNA insertion, a Poly A/splice site and SV40 origin of replication.

1.6.3 The advantage of COS cells

As mentioned, all SV40 and polyoma-based vectors replicate to high copy number only in cells containing SV40 or polyoma genomes. These SV40 genomes have been "crippled" by the deletion of several bases at the SV40 origin of replication to preserve the essential elements of the virus while allowing it to integrate with the endogenous cellular genome yet preserve its essential function. The cells were first developed from African green monkey kidney fibroblasts which had been transformed by the SV40 genome and named COS-1 and COS-7 (Gluzman 1981). In our work, we used COS-7 cells. The advantage of these cells is that they express high levels of SV40 large T antigen and permissivity factors which allow large numbers of replicons per cell, but they do not produce infectious viral genomes and so are safe to work with in conditions of low containment (Simmons and Needham 1991) . In addition, these cells are capable of a high degree of transfection using the DEAE-dextran/chloroquine regime (Gorman et al. 1985). The theoretical maxim allows for 50-60% of all cells transfected to express the protein introduced. In practice, levels of 25-30% are reasonable to expect on a regular basis.

1.6.4 Molecules cloned from CDM8-based transient expression

Since the advent of the CDM8-based technology, a large number of cell-surface molecules have been cloned using this method using monoclonal antibodies to identify the correct ligand after several, ideally three, rounds of the transfection-expression-rescue process. These include: the T cell activating/accessory marker CD2 (Seed and Aruffo 1987b); its ligand LFA-3 (CD58) (Seed 1987a); the T cell adhesion protein CD28 (Seed and Aruffo 1987b);

ICAM-1 (Simmons 1988a) and ICAM-2 (Staunton 1989) which recognise LFA-1 (CD11a/CD18); ELAM-1, now referred to as E-selectin, (Bevilacqua 1989); the B cell surface marker CD22α (Stamenkovic and Seed 1990) which mediates platelet and monocyte adhesion; CD44 (Aruffo 1990; Stamenkovic et al. 1989; Stamenkovic et al. 1991) which recognises hyaluronate; ICAM-1 (Simmons 1988a), CD33 (Simmons 1988b) and CD34 (Simmons 1992). VCAM-1, an endothelial adhesion protein for VLA-4 on lymphocytes, was cloned by varying the panning procedure to use cells directly as a probe for ligand (Elices and Goldstein 1990; Osborn 1989). In another variation, the ligand itself (LFA-1) was also used directly as a probe for ligand (Staunton 1989). The use of ECDs ligated to Fc proteins employed in this thesis is a similar attempt to use the surface marker itself as a probe for ligand.

1.7 The phosphorylation of CD33, CD34 and their ligands

CD33 has five potential N-glycosylation sites and a cytoplasmic domain of 82 residues which has been shown to possess phosphorylation potential by virtue of its homology with myelin associated glycoprotein (Bambrick and Braun 1991; Simmons 1988b). Because it is a reliable marker for myeloid cells, CD33 and its ligand have been postulated to play a role in differentiation (Peiper 1988). In previous work, the association of the *src*- related protein tyrosine-kinase (PTK) $p56^{lck}$ with the antigens CD3, CD4 and CD8, led to clues regarding their function in signalling, growth and differentiation (Barber 1989; Rudd 1988). In a similar study, CD34 was shown to be phosphorylated intracellularly on serine residues (Fackler et al. 1990). We reasoned that similar studies could yield information regarding the function of CD33, due to the presence of one tyrosine residue in its carboxy terminus which could be

phosphorylated (in addition to 10 threonine residues and nine serine residues). There is no obvious cysteine motif via which a PTK could associate with the cytoplasmic tail of CD33 in the way that $p56^{lck}$ associates with CD4 (Turner et al. 1990) and CD8 (Barber 1989). However, the protein tyrosine kinases c-*fgr*, and c-*fms* are upregulated as cells differentiate in the monocytic pathway, while differentiation of undifferentiated blasts along the granulocytic pathway to $CD33^+/CD13^+$ cells shows a dramatic increase in *hck* transcripts (Willman 1991). Retroviral infection of erythromyeloid K562 cells with c-*fes*, expressed in early myeloblast progenitor cells before the appearance of other PTKs, results in increased expression of CD13 and CD33 antigen expression (Fang 1993). In contrast to myeloid cell lines with a more differentiated phenotype, such as HL-60 and U937, K562 cells possess significant endogenous levels of phosphotyrosine containing proteins (Giordano et al. 1987; Huhn et al. 1987). This is not unexpected since K562 cells are derived from a Philadelphia chromosome-positive patient with Chronic Myelogenous Leukemia (Lozzio 1975) and carry the 9;22 translocation gene product, $p210^{bcr\text{-}abl}$ (Kloetzer et al. 1985; Naldini et al. 1986). Upon myeloid differentiation, the level of $p210^{bcr\text{-}abl}$ is reduced, a process which would be expected to arrest the proliferative capacity of these cells (Fang 1993). This effect is also shown upon differentiation with hemin (Richardson et al. 1987) indicating that it is likely a general rather than lineage specific effect. Therapeutically, it is possible that CD33 ligands coupled to toxins could be targeted to CD33 expressing cells blocked in the differentiating pathway in an attempt to induce differentiation. The challenge would be to design the toxin to be specific for aberrant rather than normal cells. In normal bone marrow cells, the c-*kit* proto-oncogene product was found to be coexpressed with CD33 on colony-forming cells and a myeloid population likely to be maturing mast cells (Ashman 1991).

1.8 Ligand studies on CD34

CD34 is unusual among known leucocyte antigens in its unusually high level of O-linked glycosylation, 35Kd of the total molecular weight of 110 Kd (Simmons 1992; Sutherland 1992). The only other known leucocyte antigens with a similarly high level of O-linked sugars are CD43, CD45 (Barclay 1993) and P-selectin glycoprotein ligand-1 (PSGL-1) (Moore 1994). CD43 has been shown to be defective in its expression in Wiskott-Aldrich syndrome, likely due to the defective expression of glycosyltransferases (Piller 1989). CD43 has recently been shown to bind ICAM-1 (CD54) (Thomas 1995). CD45RO has been shown to bind to CD22β on B cells (Wilson et al. 1991). Interestingly, both ICAM-1 and CD22 are IgSF members with 5-7 Ig domains and relatively few N-linked glycosylation sites. This suggests that the binding is occuring through a protein-carbohydrate interaction, in that both molecules bind to ligands which carry many O-linkages. CD22, other members of the Sialoadhesin (Sn) family (CD22, CD33, MAG, SMP and Sn) (Crocker et al. 1995) and potentially ICAM-1 have all recently been classified as "I-type lectins" by virtue of the C-X_{3-4}-C motif found in the first or second V domain of each potential candidate from the immunoglobulin superfamily (IgSF) (Powell and Varki 1995). CD34 (Sgp90) in High Endothelial Venules (HEV) has been shown to bind to an L-Selectin Fc-chimera (Baumhueter 1993) as has GlyCAM-1 (Sgp50) (Imai 1993). L-selectin possesses no N-linked sites and binding likely occurs through the calcium-dependent (C-type) lectin domain (Barclay 1993). In contrast, CD34 expressed on the KG-1a cell line, is unable to bind either the L-selectin chimera or native L-selectin (Majdic 1989) which binds to a distinct ligand on these cells (Oxley 1994). L-selectin binds to CD34 on Human Umbilical Cord Endothelial Cells (HUVEC) in a limited way (Norgard et a.l 1993). These studies indicate different binding behaviours for

the native and chimeric L-selectin molecules which may apply to native and chimeric forms of other molecules, particularly those as heavily glycosylated as CD34. While intermolecular binding has often been shown to be the result of a protein-protein interaction, increasing evidence shows, particularly in the heavily glycosylated molecules such as CD43, CD45 and CD34 that protein-carbohydrate interactions are crucial for antibody and intermolecular interactions. Thus, the ability of a particular cell line to correctly glycosylate a molecule by expressing the appropriate glycosyltransferases is of crucial importance in producing a molecule which binds in a physiologically relevant way. This has been shown most directly for CD22β which exhibited decreased binding ability to B cells upon cotransfection of the cDNA in COS cells with α2,6 sialyltransferase (Stamenkovic et al. 1991a). The following experiments will demonstrate the findings and limitations of the techniques employed to assess the binding capability of Fc-CD34 and transfected full-length CD34. CD34 likely possesses different binding behaviour in different cell types which is likely due to differential glycosylation in these cells. In addition, the phosphorylation of CD34 and a novel potential ligand on HUVEC precipitated with Fc-CD34 will be shown.

1.9 A dystrophin homologue is isolated from Fc-CD33 panning

During the panning process with Fc-CD33, a cDNA was isolated which shares homology to dystrophin. The 14 Kb dystrophin gene is necessary for normal muscle function and was first isolated by positional cloning (Koenig et al. 1987). The absence of dystrophin gives rise to Duchenne Muscular Dystrophy (DMD), a progressive, X-linked disease with an incidence of about 1 in 3000. The more benign form of the disease, Becker muscular dystrophy (BMD), is

characterised by the presence of an altered dystrophin protein. In normal muscle, dystrophin comprises about 5% of the membrane cytoskeleton. A large area of current research is focused on how the protein interacts with the plasma membrane, the cytoskeleton and the extracellular matrix (ECM). Because of its interaction with cell-surface glycoproteins in the dystrophin-assciated glycoprotein complex (Ervasti and Campbell 1991), it is thought to function similarly to the integrins with a role in transducing signals from the ECM to the interior of the cell (Brown 1993) (Figure 1.5).

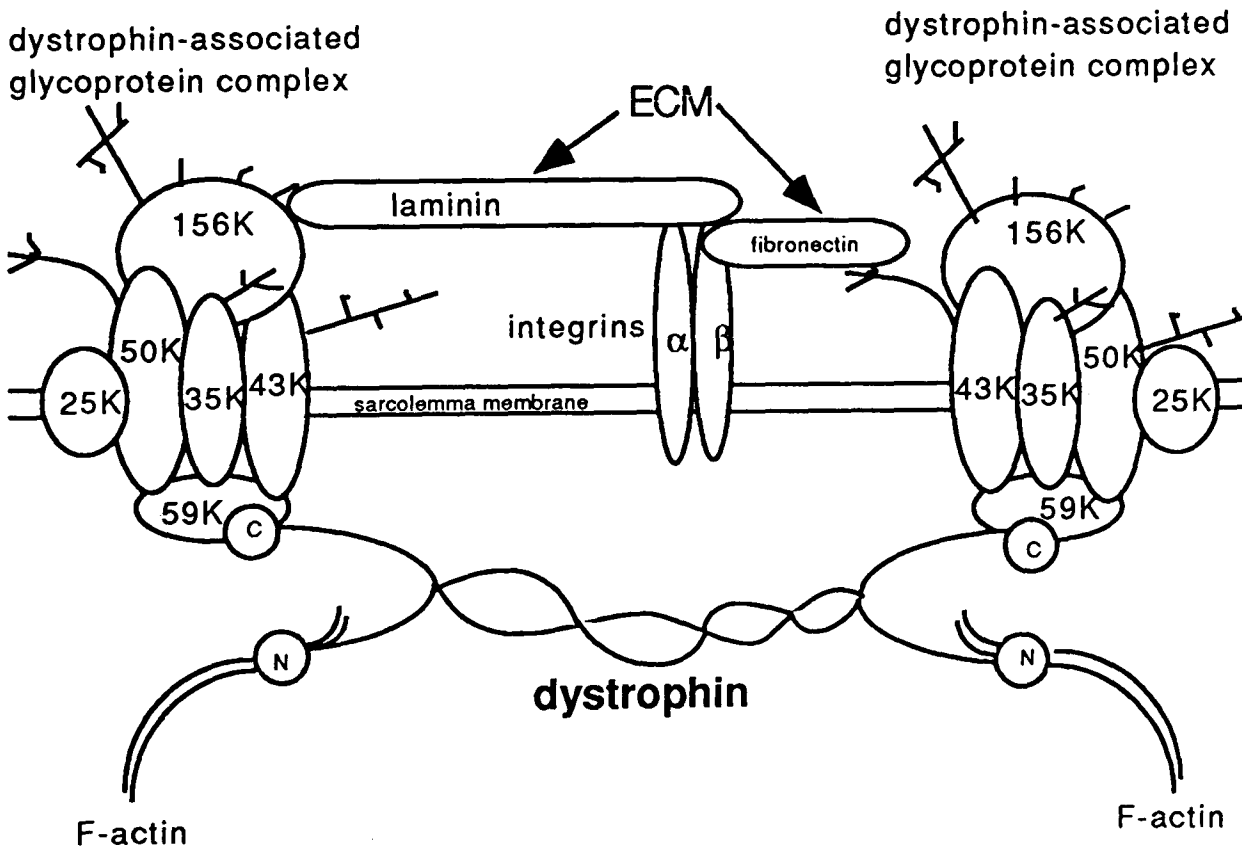


Figure 1.5 Diagram of dystrophin and the dystrophin-related complex The complex shows a model of how dystrophin likely interacts with the cytoskeleton and the cell surface via the dystrophin-associated glycoprotein complex which also associates with laminin and possibly fibronectin in the extra-cellular matrix. Adapted from Ervasti and Campbell, 1991 and Ahn, 1993.

The base sequence of the full length cDNA indicates that the 427Kd dystrophin protein encoded by 79 exons has four contiguous domains including: 1) an N-terminal region of 200-240 residues similar to the

conserved, actin binding domain, recently shown to bind actin; 2) a large domain comprising 24 helical sequences which resemble spectrin repeats, may be elastic and associated in a coiled-coil structure in the cytoskeleton; 3) a cysteine-rich region between residues 3080 and 3360 which shows homology to C-terminal domain of α-actinin in Dictyostelium and may contain two calcium binding sites, and; 4) 420 unique C terminal residues, apart from their relation to Dystrophin Related Protein, which have been suggested to mediate attachment to at least six lectin-binding membrane glycoproteins. C-terminal deletions or alterations are thought to cause the most severe form of DMD, illustrating the importance of the attachment of dystrophin to cell surface glycoproteins. The members of these glycoproteins, known as Dystrophin Associated Glycoproteins (DAG) and Dystrophin Associated Proteins (DAP) form a tight association with dystrophin and provide evidence that dystrophin could interact via these proteins with the ECM, in an integrin-like manner (Ervasti and Campbell 1991). In addition, α5β1 integrin transiently associates with dystrophin in developing myotubes in culture, supporting the suggestion of a transient link between the actin cytoskeleton and the Fibronectin component of the ECM.

Expression of dystrophin in the brain is distinct from that in muscle and important because approximately one-third of DMD patients suffer from mental retardation (Emery 1987). The "brain" form of dystrophin is transcribed from a promoter located at least 90Kb upstream from the muscle promoter. The brain and muscle forms of dystrophin consequently differ only in the first few N-terminal amino acids (Emery 1987). The expression of dystrophin in brain is confined to the post-synaptic regions of cortical neurons and cerebellar Purkinje cells (Lidov 1990). Dystrophin in Purkinje cells differs from brain dystrophin in that it is transcribed from a third promoter located

between the muscle promoter and the second dystrophin exon (Gorecki et al. 1992). "Purkinje cell" dystrophin differs from brain dystrophin by a few amino acids in the N-terminus and accounts for nearly all cerebellar dystrophin.

A series of distal transcripts, termed the apo-dystrophins, also expressed in the brain, have recently been identified (reviewed in (Ahn and Kunkel 1993; Blake 1994). Apo-dystrophin-1 or Dp71 is transcribed from a promoter located between exons 62 and 63, splices on the final exon of dystrophin in a new frame which replaces the three usual amino acids with a 31 amino acid (aa) hydophobic region (Hugnot 1992; Lederfein 1992). Dp71 is predominantly expressed in brain glial cells, liver, and the stomach (Hugnot 1992; Lederfein 1992) and precedes dystrophin expression in embryonic stem cells (Rapaport et al. 1992). Dp116 (apo-dystrophin-2 (apo-2)) is transcribed from a promoter between exons 55 and 56 and encodes the last 946 amino acids of dystrophin. It is expressed exclusively in peripheral nerve and some glial cell lines (Byers 1993). A new protein of about 40Kd, apo-dystrophin-3 (apo-3) shares an identical expression pattern with Dp71 and contains the DAG binding site (Tinsley et al. 1993).

The apo-dystrophins arise earlier in development than dystrophin from the 3' end of the dystrophin gene and often produce alternatively spliced products relative to the dystrophin gene. Apo-dystrophin-3, demonstrating conservation with dystrophin at position 9306-10,432 and alternative splice products of apo-1 in exon 78 are positioned nearest to dystrophin homologue described in this study (Tinsley et al. 1993). The apo-dystrophin cDNA described here, apo-dystrophin-4, demonstrates precise homology to dystrophin from 11,253 to 11,661 and, reversed in an inverted sequence,

precise homology at 13,213 to 13,308. Apo-dystrophin-4 also demonstrates a second, weaker homology with dystrophin in an upstream proximal region from position 5070 to 6090, contained within repeats 14-17, and appears to contain the same hydophobic stretch of 31 aa found in Dp71 in addition to a unique N-terminus.

1.10 Aims of the thesis

This work attempts to illustrate the use and limitations of the panning process to identify ligands for CD33 and CD34 in addition to binding, iodination and immunohistochemical analysis on cell lines. Phosphorylation analyses provide functional information on CD34 and CD33 and their ligands and on their ability to precipitate phosphoproteins. The ability of CD33 to bind multiple ligands directed the characterisation of one potential ligand isolated in the panning process, apo-dystrophin-4, which is described as an interesting molecule in itself. These studies provide useful information regarding the adhesion of CD33 and CD34 which contributes to the basic understanding of their molecular interactions and could be useful in several forms of leukemia therapy.

Chapter 2

Materials and Methods

2.0 Materials

2.1 Monoclonal antibodies, antisera and cells

Antibodies included in this work included the monoclonal antibody supernatant (50 µg/ml) CD33 (MAS 393, clone WM53, Sera-lab, U.K. and WM54, which recoginises the same epitope, Seralab or a gift from S. Watt, IMM) and anti-CD34 ascites (~1 mg/ml) (QBEND 10 and HPCA were also gifts from S. Watt, IMM). Other antibodies to study phosphorylation include: anti-*lyn* (Affiniti, UK), anti-*hck* and anti-*fyn* (Santa-Cruz Biotechnology, Heidelburg), anti-phosphotyrosine (4G10) (Upstate Biotechnology Inc., UK) anti-c-*fgr* (a gift from Mike Waterfield, Ludwig Institute, London), goat anti-mouse, rabbit anti-mouse and goat anti-human IgG Fc immunoglobulin (Sigma). Anti-CD22 crude and purified polyclonal rabbit anti-mouse CD22 antisera were a gift from Paul Crocker (ICRF Labs, IMM) and were raised against murine Fc-CD22 as previously described (Kelm 1994a). The anti-peptide antisera raised against the apo-dystrophin-4 gene include anti-apo-4 P1, P2 and P3 and were all produced by peptide injection into rabbits in collaboration with Dell Watling at the ICRF antibody production facility at Clare Hall, Herts., UK. The cell lines were grown in RPMI media supplemented with 10% (wt/vol) foetal calf serum in an atmosphere containing 5% CO_2. The cell lines KG-1, KG-1a, KG-1b, K562 (Lozzio 1975), HL-60 (Collins 1977), U937 (Sundstrom and Nilsson 1976) and COS-7 (COS) (Gluzman 1981) green monkey kidney cells were obtained from the ICRF Cell Production Unit at Clare Hall, apart from HEL which was the kind gift of Colin Potter (John Radcliffe Hospital, Oxford, U.K.). The anti-*lck*(n) polyclonal antisera was generated against the N-terminal amino acids 39-64 of

the murine *lck* protein ((Marth 1985; Voronova and Sefton 1986); courtesy of Chris Rudd, Dana-Farber Cancer Institute, Boston) and used in immunoprecipitation. The cell lines were grown in RPMI media supplemented with 10% (wt/vol) foetal calf serum. DMEM and RPMI-1640, PBSA, B ($MgCl_2$), and C ($CaCl_2$) and trypsin/EDTA were supplied by ICRF, Clare Hall, Potters Bar, Herts. Other standard solutions appear at the end of this section.

2.1.1 Description of leukemic cell lines used

A description of the myeloid leukemic cell lines used in these studies follows. All Cell Lines were obtained from the ICRF Cell Production Unit at Clare Hall, apart from HEL which was the kind gift of Colin Potter, John Radcliffe Hospital. Descriptions were found in the American Type Culture Collection (ATCC) (Hay 1992). The promyelocytic leukemia cell line HL-60 was dervied from leucocytes from an Acute Promyelocytic Leukemia patient (Collins 1977). U937 is a monocyte-like cell line from a patient with histiocytic lymphoma (Sundstrom and Nilsson 1976). The KG-1 cell line was isolated from a patient with Acute Myelogenous Leukemia and was shown to be completely dependent on colony-stimulating factor CSF for colony formation in soft-gel culture (Koeffler and Golde 1978). KG-1 cells do not show granulocytic maturation in response to DMSO, but mature to macrophages in the presence of phorbol esters. KG-1a is an undifferentiated blast cell variant derived from the Human Acute Myelogenous Leukemia Cell Line (KG-1) which has lost its responsiveness to CSF in culture (Koeffler 1980). KG-1a exhibits T cell-like characteristics including low-level expression of the T cell receptor β mRNA, high-level CD3δ mRNA, intracellular CD3 and

extracellular CD7, although the cells cannot be induced to differentiate further (Furley 1986). Its variability in expressing CD33 likely results from it being an aneuploid cell line in which phenotype can vary over time (Smith 1983). KG-1b has recently been described as another myeloid precursor variant of KG-1 (Watt et al. 1987) and is considered to be more myeloid than KG-1a in that it expresses CD33, while KG-1a is considered to be more of a T cell precursor, as its ability to express CD33 varies (Table 3.1 note and Sue Watt, personal communication).

K562 is an erythroleukemic cell line isolated from a Chronic Myelogenous Leukemia (CML) patient considered to represent an early differentiation stage of the granulocyte lineage and containing a Philadelphia Chromosome 9;22 translocation (Lozzio 1975). The Human Erythroleukemic line HEL, derived from the peripheral blood of a Hodgkin's disease patient who later developed leukemia, expresses HLA antigens, β-2 microglobulin and Fc-receptors (60%) but lack T-lymphoid cell surface characteristics (Martin 1982).

2.1.2 HUVEC endothelial cell preparation and culture

The most commonly used source of endothelial cells *in vitro* are prepared from the human umbilical vein, a large and relatively easily obtainable vessel source. Human Umbilical Vein Endothelial Cells (HUVEC) have the advantage that differentiation can be monitored by analysing various surface markers including CD34 which is lost after one or two passages of the cells. HUVEC require trypsin to break their calcium dependent attachment to tissue culture plastic, which cannot be broken by the 2 mM EDTA/PBSA mix which is sufficient to lift COS cells. Prior to harvesting HUVEC, tissue culture dishes

were coated with an autoclaved PBSA/0.1% gelatin mix which had been heated in the microwave for 2-3 minutes to melt the gelatin. The mix was then cooled with tap water to room temperature. 10 cm dishes were coated with 5 mls each and stored in the incubator until needed, usually no more than a day to prevent the risk of contamination, and aspirated prior to use. Human Umbilical Cords were collected from the maternity suite of the John Radcliffe Hospital and placed immediately after delivery in ice-cold sterile DMEM containing 10 µg/ml each of Penicillin, Streptomycin and Gentamicin. Cell preparations were undertaken 2 - 8 hours after harvest. To release the endothelial cells, the method of (Jaffe 1973) was used. The umbilical vein was rinsed with 70% EtOH at the ends and cannulated at each end with a bevelled polyethylene cannula (Kwill™ filling tubes), secured by ligating the cord with heavy fishing line. The vein was flushed through with 200 mls of PBSA before infusing 20 mls 0.5% Dispase™ (Collaborative Research Inc., Rockville, MD., USA) and sealing the ends. After 15 minutes incubation at 37°C the cord was massaged to mechanically assist in dislodging endothelial cells and the enzyme solution aspirated into a 50 ml polypropylene tube. The vein was washed through with 15 mls of PBSA twice and the washings added to the same centrifuge tube. Cells were pelleted by centrifugation at 800 g for 4 minutes and washed twice in DMEM/10% FCS before suspending in 6 mls of E.C. Medium and seeding the cells into either 3 gelatin-coated wells of a six-well cluster dish (2 mls cell suspension each) or a 10 cm gelatin-coated tissue-culture dish. Typically each well would contain cells at 30% confluency on day 1. Purity could be crudely assessed visually by morphology on phase contrast microscopy.

The culture of endothelial cells on gelatin-coated tissue culture plastic used medium as previously described (Gimbrone 1974). DMEM was supplemented

with 10% FCS, Endothelial Cell Growth Supplement (ECGS, Sigma Chemical Co., St. Louis, Missouri, USA) and Heparin 100 µg/ml. Gentamicin 10 µg/ml and Fungizone 50 µg/ml were added to E.C. medium when it was used to culture populations at high risk of contamination, chiefly primary cultures. HUVEC required a medium change every 2nd day. Cultures were trypsinised and split 1 in 3 when they reached confluence. HUVECs were used in experiments only through the second passage.

2.1.3 YACs and related reagents

Murine λ22 genomic YAC clones were provided by Dr. Jerry Kaplan (Utah School of Medicine, Salt Lake City, UT, USA). Markers were φX174 RF DNA digested with Hinf I (Gibco BRL). All YAC filters and 3' dystrophin genomic phage clones were supplied by Anthony Monaco and Jamel Chelly (Laboratory of Human Genetics, ICRF Labs, IMM, John Radcliffe Hospital, Headington, Oxford, UK). The dystrophin-negative Blondolet muscle cell line was supplied by the ICRF Cell Culture Unit (Clare Hall, Potters Bar, Herts., U.K).

2.1.4 Plasmid vectors

pCDM8 (InVitrogen) and pBluescript (Stratagene) were purchased commercially. Permission to use the Rg vector was granted from Brian Seed, (Department of Molecular Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA). David Simmons and John Fawcett (Cell Adhesion Lab, IMM) constructed the pIG1 vector to produce most of the

soluble recombinant proteins used in this study (Fawcett 1994b; Simmons 1993). Murine CD22 in pCDNA I (InVitrogen) was provided by Paul Crocker (Molecular Hemopoiesis Lab, ICRF, IMM, UK) and was constructed as previously described (Kelm 1994a). Fc-ICAM-1 was a generous gift from A. Berendt and A. Craig, IMM, Oxford and was constructed as previously described (Berendt et al. 1992).

2.2 Methods

2.2.1 Molecular biology methods: introductory remarks

Some basic techniques such as restriction endonuclease digestions are not described here. These methods are well standardised and like other basic molecular biology procedures, were performed using the methods described in the Molecular Cloning Laboratory Manual (Sambrook 1989), Current Protocols in Molecular Biology (Ausubel 1993), or the manual describing antibody use and production (Harlow 1988). The same sources were consulted for other essential information such as the safe handling and disposal of ethidium bromide solutions. Standard reagents and solutions are listed at the end of the section except where it is necessary to include them in the body of a given protocol.

2.2.2 Plasmid preparation

2.2.3 Large scale DNA preparation ("Maxiprep")

An initial mini culture of bacteria transformed with a particular plasmid were grown in a large volume culture of 800 mls to saturation by overnight incubation in a 37°C shaking incubator (250 rpm). The following day, the cells were pelleted by centrifugation in a Beckman J6 centrifuge at 4,200 rpm for 20 minutes and the supernatant discarded. The plasmids were extracted by alkaline lysis via the sequential addition of Solution I (40 mls 10 mM EDTA) - followed by vigorous thumping of the bottle to resuspend the pellet; Solution II (80 mls of 0.2M NaOH - 0.1% SDS solution) - followed by swirling the mixture until it became highly viscous indicating complete lysis; and Solution III (40 mls of 3M KOAc solution) - were added and the bottle gently shaken until floculated lumps were reduced in size to 2-3 mm indicating that neutralisation and genomic DNA and protein precipitation were complete.

The mixture was centrifuged again at 4,200 rpm for 5 minutes and the supernatant poured through a double thickness gauze square into a clean 250 ml centrifuge bottle. The DNA was precipitated from the supernatant by filling the bottle with IsoPropanol, inverting twice and centrifuging again at 4,200 rpm for 10 minutes. The pelleted plasmid DNA was resuspended in 3 mls of high T:E with vigorous agitation or pipeting as necessary. Ideally, the pellet would dissolve immediately.

The final purification of the plasmid DNA was accomplised by Equilibrium Density Gradient Centrifugation ("plasmid banding"). Ethidium bromide

solution, 100 µl of a 10 mg/ml solution, was added to plasmid DNA resuspended in 3 ml high T:E in a universal container. Five grams of optical grade cesium chloride (Sigma) was added to each tube, the tops screwed on and the container shaken vigorously until the cesium was taken up into solution, shearing contaminating genomic DNA to facilitate its subsequent separation from plasmid DNA. The mixture was transferred to a 5 ml polyethylene ultracentrifuge tube via a syringe with a 16 guage needle. The tube would then be filled to the mark with a syringe filled with high T:E, thus avoiding the problem with balancing ultracentrifuge tubes because the entire 5 g of cesium chloride always went into the tube. The balance tubes were sealed on the heat sealing device supplied by the manufacturers before centrifugation in the upright VTi 80 rotor in the Beckman ultracentrifuge T70-M at either 80,000 rpm for 2.5 hours or at 70,000 rpm overnight (> 12 hours). The centrifuge was run down without braking to prevent the banded DNA from remixing.

The top of the tube was vented by the insertion of a needle and punctured with a second needle at the position of the nicked plasmids, the upper band of DNA about 1/3 of the way down the tube with the tip being placed 1mm below the supercoiled plasmid band and then the purified plasmids slowly aspirated with a 1 ml syringe. The aspirate was transferred to a 10 ml polypropylene tube filled with 1M NaCl saturated Butan-1-ol. The tube was vigorously shaken to extract the ethidium bromide from the lower aqueous phase and the upper organic phase was removed with a pasteur pipette connected to a suction line after 5 minutes. The extraction was repeated if the aqueous phase retained any ethidium bromide as this would inhibit DNA purification. The plasmid DNA was then precipitated by the addition of an equal volume, about 1 ml, of 1M Ammonium Acetate and the tube filled

with ethanol. The plasmid DNA precipitate could usually be immediately visualised as white coiled strands and was recovered by centrifugation in a Beckman GPR benchtop centrifuge at 3,000 rpm for 5 minutes at room temperature. The pelleted DNA was washed with 70% ethanol and resuspended in T:E, usually 1 ml, and the absorbance at A_{260} measured and usually yielded 1 mg/ml for pCDM8 based plasmid and 3 - 5 mg from a pUC19 based plasmid, such as pBluescript™ (Stratagene Corporation, La Jolla, CA, USA).

2.2.4 Small scale plasmid DNA preparation ("miniprep")

Small-scale preparation of plasmid DNA was done using the alkaline lysis procedure to examine whether colonies from transformed cells contained clones of interest (Feinberg 1983). Colonies were grown from 6-8 hours in universal tubes until saturating growth was achieved. 1 ml of each culture was harvested into Eppendorf tubes and centrifuged in a benchtop microfuge with a 24 position rotor, the limiting factor in the number of minipreps that could be done. An Eppendorf repeating pipette was used to add reagents to tubes. The cells were pelleted by brief (≈ 20 seconds) centrifugation. The supernatants were discarded and 100 µl of Solution I was added to each tube followed by vortexing for 10 seconds until the pellet was resuspended. 200 µl of Solution II was added followed by 100 µl of Solution III to each tube. The tubes were closed and inverted 3 or 4 times by placing an Eppendorf rack on the tubes followed by centrifugation for 5 minutes. Supernatants were transferred to fresh tubes and an equal volume, 500 µl, of the phenol:chloroform mixture (50:50 v/v) added to all tubes. The tubes were closed and vigorously shaken for 1 minute in an Eppendorf rack "sandwich",

centrifuged for 1 minute and the upper aqueous phase transferred to new tubes. The tubes were filled with Propan-1-ol from a squeeze bottle, sealed and inverted twice followed by centrifugation for 5 minutes to pellet the DNA. Plasmid pellets were washed twice by filling the tubes with 70% ethanol and inverting by placing a ruler on the Eppendorf lids. Pellets were resuspended in 50 µl T:E containing pre-boiled RNAase A at a final concentration of 10 mg/ml. To aid resuspension and destroy any residual endonucleases the tubes were incubated for 5 minutes at 70°C. The DNA was then ready for restriction digest or with well purified pBluescript plasmids, DNA sequencing.

2.2.5 Oligonucleotide synthesis

Oligonucleotides for PCR, probes, and DNA sequencing were synthesised within the laboratory on an Applied BioSystems synthesiser using standard phosphoramidite chemistry with reagents supplied by the manufacturer. Some standard primers such as the universal vector-based primers used for DNA fragments cloned into pCDM8 and pBluescript SK^+ were obtained from the Oligonucleotide Synthesis Service Laboratory at ICRF Central Services, Clare Hall, Potters Bar, South Mimms, Herts., U.K. These were supplied as DNA precipitates which were resuspended at 100 ng/ml for use. Some of the Clare Hall oligonucleotides retained a resin which inhibited specific priming due to different preparation procedures and thus the laboratory synthesised oligos were generally of superior quality.

2.2.6 Oligonucleotide deprotection for PCR

Oligonucleotides were cleaved from the column onto which they were anchored at their 3′ end by use of 1-1.5 mls of NH_4OH solution pre-chilled to -20°C, to prevent explosion, and pulled across the column for 1 hr. with two 5 ml syringes. Syringes were swished back and forth over the column every 20 minutes at room temperature. The oligo was then drawn into one tube and decanted into a screw top tube. Thick parafilm was tied around the top and the oligo was placed in a 55°C water bath for 5 hours minimum or overnight, submerged under lead pigs in case of explosion. Oligos were then put on ice to cool for 15 minutes, carefully unscrewed and decanted into a 10 ml tube followed by the addition of 1 ml of 3 M NaAc. The tube was filled with 100% EtOH and put on dry ice for 10 to 15 minutes. The tube was spun at 3800 for 5 minutes in a Beckman GPR benchtop centrifuge at 3,500 rpm for 10 minutes and washed in two 70% washes. The tube was shaken or vortexed to get rid of salt to yield a small yelow pellet and resuspended in 1 ml T:E to a final concentration of 1 mg/ml. The A_{260} was measured and usually shown to be ≈1. The oligos were then used directly for sequencing or PCR or stored undiluted at -20°C.

2.2.7 Construction of Fc-adhesins.

The primers listed below were used to amplify CD33 and CD34 ECD fragments via PCR for insertion into the pIG1 vector as previously described (Barber 1993; Freeman 1995; Simmons 1993). In the Forward Amplification Primer (FAMP), a Hind III restriction site was included preceded by the appropriate overhang to allow for restriction enzyme digestion immediately following

PCR. The Reverse Amplification Primer (RAMP) includes a Bgl II site followed by a splice donor motif (5' ACTTACCTGT 3') which will splice directly into the splice acceptor site engineered into the intron 3' of the hinge to create an in-frame splice junction for protein processing. The expected sizes were 776 bp and 895 bp for the CD33 ECD and CD34 ECD, respectively. Following PCR of the CD33 and CD34 inserts from the Rg vector, PCR products were resuspended in 11.5 µl T:E and digested in HindIII/BglII and 1.5 µl restriction buffer for 3-4 hours prior to ligation into the PIg1 vector.

2.2.8 Polymerase Chain Reaction (PCR)

The polymerase chain reaction was used to amplify specific fragments of DNA from cDNA libraries or plasmids using a forward and a reverse primer. PCR was always done in a reaction volume of 100 µl. The buffering conditions generally used were those recommended by the manufacturer (Perkin-Elmer) according to the standard protocol for a given reaction which included: 10 µl 10x PCR buffer, 1 µl of each PCR primer (used at a concentration of 1mg/ml), 1 µl of a dNTP solution (containing all 4 nucleotides at a concentration of 25 mM - supplied by Pharmacia - ultrapure dNTP set), 10 µl of template DNA, 0.5 µl *Taq* polymerase (2.5 Units, various manufacturers), and distilled water to a final volume of 100 µl.
The reaction mixtures were overlayed with 100 µl of light mineral oil (Sigma Chemical Co., St. Louis, MO, USA). The DNA templates used included: 1) 10 ng of pure plasmid DNA, either cDNA clones (for constructing soluble forms of cell adhesion molecules or using a captured cDNA as a control), or cDNA libraries; 2) 1st strand cDNA - when reverse transcriptions were performed from 10 µg total RNA, 10 µl of a 100 µl reaction was used immediately

following transcription as described under RT-PCR, and; 3) 200 ng of genomic DNA. The PCR program was generally set into 30 cycles consisting of denaturation at 94°C, primer hybridisation at 37 - 50°C, depending on the manipulation, and extension at 72°C. The length of each incubation step and annealing temperature varied according to the target sequence being amplified and conditions used for individual reactions are described under each type of experiment. Following the run, the reactions were kept at 4°C until 10 µl of each reaction was analysed by electrophoresis on a 1% agarose gel to check for product amplification. Products could be "blunt" cloned into a vector, usually pBluescript, for DNA sequencing and it was necessary to "end-fill" the amplified product to accomplish this. End filling was accomplished by the direct addition under the oil of 1 µl of 25 mM dNTPs and 10 U of Klenow DNA polymerase followed by incubation for 15 minutes at 37°C, the addition of 10 units of polynucleotide kinase and a final incubation for 30 minutes. The remaining Taq buffer was sufficient to ensure the appropriate buffer concentration. The product DNA was then purified from the mineral oil and Taq polymerase by DNA extraction with an equal volume of a phenol/chloroform mix and precipitation by the addition of an equal volume of 1M Ammonium Acetate (NH_4Ac) solution, 10 µl of 10 mg/ml of a linear polyacrylamide (LPA) carrier and 3 volumes or greater of Propan-1-ol. The mix was placed on dry ice for 10-15 minutes. The DNA was pelleted by centrifugation at 14,000 g in a microfuge for 5 minutes and the pellet was washed twice in 70% ethanol, inverting the tube to discard the supernatant. 100 µl of T:E was added to resuspend the pellet for further manipulations. The Mg^{++} optimum and suitable annealing temperature was established with each set of primers but standard conditions were usually suitable for all primers used. Primer combinations were checked when possible for potential dimerisation using the "Amplify" program. Cycling parameters depend on

the primer combination and the product size, if known. 30 cycles was known to give a clear ethidiuim band on an agarose gel. To get this working well on a regular basis without contamination with cDNA, separate stock solutions were made, separate pipetman were used, and in some cases, experiments were done in a separate laboratory to prevent plasmid cDNA contamination. It was not usually necessary to use a laminar flow hood, with the extra precautions taken, however.

2.2.9 Purification of DNA fragments using Geneclean™

For the rapid ligation of PCR products into vector, the Geneclean system (BioRad) was often used to increase the efficiency of DNA purification. To use this system, PCR products were first run out on a 1% low melting point gel. Band was excised from gel and the gel was digested by the addition of 3 volumes of NaI stock solution. Agarose was dissolved by incubation at 5 minutes at 45° to 55°C. Glassmilk suspension was added and incubated for 5 minutes. Glassmilk/DNA complex was pelleted by microfuging the pellet for 5 seconds and removing the supernatant. Pellet was washed 3 times with New Wash and DNA was eluted by the addition T:E into the appropriate volume, usually 100 µl for further manipulations.

2.2.10 End filling of DNA for blunt end ligation

Following PCR, uneven or "sticky" ends would be left on the PCR products which needed to be filled in prior to blunt end ligation into pBluescript via an enzymatic reaction. The following reaction was set up for each fragment.

5 µl fragment (1µg or half of the preparation)

2 µl 10x Klenow Kinase buffer

1 µl Klenow (1U)

1 µl T4 polynucleotide kinase (10U)

1 µl dNTP's (N=G, T, A, C)

10 µl H_2O

20 µl Total

Reactions were incubated at 37°C for 30 minutes. An insert could be transferred from CDM8 to Bluescript directly by virtue of their compatible EcoRV sites. The insert could then be cut out and end-filled with Klenow and dNTP's using the above method. A 50 µl end filling reaction was allowed to proceed for 15 minutes at 37°C. The DNA was then precipitated to purify it and resuspended in 10 µl (~2 µg) using 5 µl (1 µg) for ligation with 100 ng of vector using the ligation mix described below (4 µl Low Salt Buffer, 4 µl Ligation Additions, 5 µl Ligase, 1 µl vector, 5 µl insert, brought to 40 µl with dH_20. If the DNA was purified using the Geneclean II™ kit, each fragment to be ligated was brought to a final volume of 11.5 µl in dH_2O. In preparation for blunt ended ligation, pBluescript SK^+ was digested with either canine or shrimp intestinal phosphatase at 37°C according to standard procedure (Sambrook 1989), precipitated and the purified product diluted to 100 ng/µl ie. 10 µl (1 mg/ml) DNA in 20 µl H_2O. 100 ng vector was used per ligation. Blunt cloning was always done into pBluescript at the EcoRV site which allowed blue/white selection of recombinants by *lac*Z induction with IPTG in the presence of X-*Gal* as an indicator substance. Reactions were incubated at 12°-15°C overnight, transfected into competent TG-1 cells and spread on "high-amp" plates as described below with IPTG/X-*Gal* and following

overnight incubation at 37°C, blue colonies were picked and grown up for small scale DNA preparation.

For "sticky" end ligation to clone the uneven DNA fragments directly, the digested DNA or PCR product was purified by electrophoresis through low melting point agarose on a slide gel as described in "DNA electrophoresis". The insert bands were excised just around the visible portion during brief visualisation with long wave ultraviolet light and the fragments were transferred to Eppendorf tubes. The volume of bands excised from slide gels was estimated to be 12.5 µl and 12.5 µl of T:E added to prevent the gel from congealing during subsequent manipulations. The gel fragments were melted by heating to 70°C for 15 minutes and quickly transferred to 37°C at which temperature all subsequent manipulations and incubations were done. To the melted gel slice of approximately 25 µl, 4 µl 10x Ligation Additions, 4 µl Low Salt Buffer (10x), 1 µl 100 ng DNA vector or 11 µl of a melted diluted gel slice containing 100 ng vector and 5 µl ligase were added for a final volume of 40 µl. During this time a vector control was prepared by assembling a separate tube with all of the above components but making the volume up to 40 µl by adding water or T:E in place of the insert volume. The reactions were mixed well and allowed to proceed for 30 minutes.

During this time, a vial of transformation competent bacterial cells was taken from the freezer and allowed to thaw on ice for 15-30 minutes. TG1 cells were used with pBluescript ligations and MC1061/p3 cells were used with CDM8 ligations. The preparation of these cells is described below. Agar plates were either dried or poured, including the appropriate antibiotics to express the drug resistance genes in either CDM8 or pBS. To prepare the plates, a bottle of LB agar was melted for 5-10 minutes on medium power and cooled to 55°C.

To prepare "amp/tet" plates for CDM8 plasmids, 75 µl of tet and 100 µl of amp were thawed from freezer stocks for a final concentration of 7.5 mg/ml tetracycline and 12.5 mg/ml ampicillin. For pBluescript, "high amp" plates were prepared with a final concentration of 100 mg/ml ampicillin in LB agar. 10 mls of well mixed LB was poured into 10 cm dishes and 15 mls was poured into 15 cm dishes (for Hirt transformations covered in the panning section) and plates were allowed to set for 10-15 minutes. The thawed competent cells were flicked to mix. 5-10 µl (1/10 volume) of ligation mix was mixed with gentle flicking in a 50-100 µl aliquot of thawed competent cells. The mix was allowed to sit on ice for 10-15 minutes followed by a "heat shock" for five minutes at 37°C. The cells were then kept at room temperature until spreading on the agar plate with a flame-sterilised spreader.

A system was developed by David Simmons to allow MC1061/p3 to recover from heat shock and to begin to express antibiotic resistance genes prior to contacting the antibiotics in the plate. 5-10 minutes before plating, 5 mls of drug-free agar was poured on the amp/tet plates and allowed to set during the incubation on ice and the heat shock period. The cells were then carefully spread on the agar to prevent breaking the surface. It was unnecessary to use top agar with the TG1 cells which contained only conventional ampicillin resistance genes. With these cells, 50 µl of a mix of 20 mg/ml X-*Gal* and 20 mg/ml of IPTG in Dimethyl Formamide was spread during heat shock when blue-white colour selection was required. The following day, if colonies were evident, 24-48 were picked and grown for small scale DNA preparation.

2.2.11 TA cloning system

For fast, efficient cloning of PCR products, the TA cloning system (InVitrogen) was used shortly before the end of this study.
The analytical gel run of a few microlitres of the PCR sample yields a calculated concentration of 100 ng of a 1.0 Kb PCR product per 1 µl of PCR reaction mixture. For a 1:1 molar ratio of vector to PCR insert, ligate 50 ng of vector with 500 ng of insert. In order to dilute the PCR product, remove 12.5 µl of PCR sample, add 87.5 µl T:E, and use 1 µl for ligation with the vector in a total ligation volume of 11 µl as follows: Ligation Reaction: 4 µl sterile water, 1 µl 10x ligation buffer, 2 µl pCR^{TM} vector (25 ng/mg), 3 µl diluted PCR product, 1 µl T4 DNA Ligase. Reactions were left on the benchtop overnight or at 37°C and processed as other ligations above.

2.2.12 Transformation and preparation of competent cells

The transformations of plasmid DNA into competent bacterial cells were done according to the method originally described (Hanahan 1983). This technique takes advantage of the finding that suitable strains of bacteria in a mid-log phase of growth are able to take up exogenous DNA after treatment with divalent cations and a brief period of heat shock at 37°C, although the molecular mechanism of DNA uptake is unknown. The bacterial strains used in this study were MC1061/p3 which carries a single copy of episome p3 which expresses amber mutated ampicillin and tetracycline resistance genes and TG1, an F′ strain which carries the lac operon containing a *lacI*DM15 gene to allow for α-complementation by β-galactosidase and thus allows for blue-

white selection of recombinant pBluescript plasmids in the presence of X-*Gal* as an indicator.

To make these bacterial cells transformation competent, a single colony was picked from an amp (TG1) or amp/tet (MC1061/p3) plate and grown in an enriched broth containing a relatively high concentration of divalent cations, in this case TYM. Sequential two-fold dilutions of the culture were made over several hours and the A_{600} monitored until the volume reached 800mls and the absorbance reached 0.5 indicating a large population of rapidly dividing bacteria. The cultures were then tranferred to very clean, preferably new, bottles and rapidly chilled by swirling slightly submerged in an ice bath. The bottles were spun just enough to pellet the cells in a Beckman J6 centrifuge at 3000 rpm at 4°C for 15 minutes. The pellet was gently resuspended by the addition of 100 mls of ice-cold Transformation Buffer I (TFB I) by swirling in an ice bath for 30-45 minutes. The cells were again pelleted by centrifugation at 3000 rpm for 10 minutes at 4°C, decanted and resuspended using the same technique as above with 50 mls of Transformation Buffer II (TFB II). The cells were then aliquoted into 200-500 µl aliquots in Eppendorf tubes and flash frozen in a bath of liquid nitrogen. The tubes were immediately transferred to -80°C and kept for up to six months.

2.2.13 DNA probe preparation

2.2.13.1 Random primed synthesis of DNA probes

This was always done using commercially available kits (Boehringer Corporation, Mannheim, Germany) which follow the method developed by Feinberg and Vogelstein (1983). As a template for labelling plasmid inserts were excised by restriction digest and purified by electrophoresis through a thin low melting point agarose gel. To probe yeast artificial chromosome (YAC) clones, a 5' 546 bp cDNA probe was made by digesting the apo-4 cDNA in CDM8 with Hind III and Hinf I. A 3' 280 bp apo-4 probe was made by digesting the apo-4 gene in CDM8 with HindIII, Pst I and Hpa I which cut at 710 bp. For RNAse protection, a 3' probe was made by ligating the cut fragment into pBluescript. The DNA representing the insert to be used as a probe was visualised under long wave, low-intensity ultraviolet light and excised. The excised band was then reloaded on a fresh gel and repurified. An equal volume (12.5 µl) of water was added to the gel slice containing the probe DNA and the gel was melted and the DNA denatured at 100°C for 5 minutes. The probe was labelled in a reaction according to the manufacturers instructions at 37°C to prevent the gel from congealing. The reaction components included:

Insert DNA (25-100 ng) ~12.5 µl gel mixture

3 µl of dNTPs (bases G, A and T only)

2 µl of 10x reaction buffer supplied

5 µl of α^{32}P-dCTP (specific activity 3000 Ci/mmol)

water to bring volume to 19 µl

1 µl of supplied DNA polymerase (Klenow fragment)

20 µl final volume

The reaction tube was closed after mixing the contents with a pipetman. Reactions were incubated for 1 hour. 100 µl of "high" S:T:E buffer was added to the reaction and the unincorporated nucleotides were separated from the labelled probe DNA by gel filtration through Sephadex G-50 columns pre-equilibrated with S:T:E, by spinning at 1000 rpm for 5 minutes. The activity not retained through the column was the probe DNA and specific activity was then measured by measuring Cerenkov counting, expecting a specific activity of at least 5×10^5 cpm per µl of DNA. Before use all probes were denatured by heating to 100°C for 5 minutes.

2.2.13.2 End labelling of oligonucleotides

For southern blotting, oligonucleotides were endlabelled by a 5′ phosphorylation reaction catalysed by T4 polynucleotide kinase. Labelling reactions consisted of :

15µl	sterile distilled H_2O
1µl	oligo 15-21mer 20 pm/ml (~1mg/ml)
2µl	10x kinase buffer
1µl	T4 polynucleotide kinase (Pharmacia, Uppsala, Sweden)
1µl	γ^{32}P - ATP (3000 Ci/mmol, Amersham, UK)

After 30 minutes to an hour incubation at 37°C, labelled oligonucleotide was purified from excess radioactivity by the use of Sephadex G-25 spin columns or ethanol precipitation as described above. Each probe mix was enough for two filters.

2.2.14 DNA analysis

2.2.14.1 DNA agarose gel electrophoresis

All DNA gel electrophoresis was done through agarose gels cast in T:A:E buffer containing 10 mg/ml of ethidium bromide solution. For many purposes e.g. to check for the presence of amplified products indicating successful PCR reactions, electrophoresis through pre-cast mini-gels on 3 x 2 inch glass slides was used. These gels were stored for up to 2 weeks in an airtight container stacked on paper soaked in 70% ethanol. Since they are very thin gels these have the advantages that small quantities of DNA can be analysed and, for applications such as probe preparation and purification of PCR products for sub-cloning, DNA fragments can be cut from gels in a minimum quantity of agarose. The estimated volume of excised gel fragments was 12.5 µl. If they are stored too long, however, the gel can separate from the slide and sample can be lost.

For more than five samples (e.g. miniprep analysis) or prior to transfer to nylon membranes gels of 1 to 1 1/2cm thickness were cast immediately before use by melting 100 ml of T:A:E/ agarose solution for 3-5 minutes in a microwave oven and cooling under tap water to 50°C before pouring. The standard running buffer was also T:A:E containing 10 mg/ml ethidium bromide. Before loading gels, 1/10th volume of Ficoll-containing 10x DNA loading buffer was mixed with each sample. The voltage applied to gels was: 100V for mini gels and 150V for large gels, running gel until the orange G dye front reached the anodal or lower end of the gel. 1% agarose Mini-gels were run quickly in about 15 minutes and 1% low melting point agarose in 30 minutes at 60-70V to prevent melting and loss of resolution. For all gels

DNA size markers were included, usually λBstXI (4Kb) DNA restricted with BstE II or ϕX174 (1Kb) DNA restricted with Hae III for the sizing of small fragments. DNA fragments were viewed under ultraviolet light with protective goggles and photographed by Polaroid photography. Fragments viewed for the purposes of excising bands for cloning were examined under long wave ultraviolet light to minimise DNA damage.

2.2.14.2 Southern blot hybridisation

DNA samples from RT-PCR were first resolved according to size by electrophoresis as described above. 10 µl of a 100 µl PCR reaction was loaded on to the gel to be transferred for Southern blotting (Southern 1975).
To prepare the gel for transfer, after photographing the gel, the DNA was denatured in a 1 litre solution of 1.5M NaOH / 0.5M NaCl for 1 hour. A piece of Hybond N+ nylon transfer membrane was cut to the exact size of the gel. and several pieces of Whatman 3 paper slightly smaller than the gel as described (Ausubel 1993). The gel was transferred to the membrane and paper, the bubbles were removed, and the apparatus assembled to effect capillary transfer to the membrane using denaturing solution to wick the solution up onto the membrane. Transfer was continued overnight, although several hours was usually adequate. Alkaline transfer was carried out in the same solution as denaturation. Lanes of the membrane were marked with pencil, noting the DNA side. Membrane was neutralised with 2x SSC° for 1 minute and dried under light between two sheets of Whatman paper. Blot can be baked at 80°C for 2 hours although this is not strictly necessary. DNA was immobilised to the blot with crosslinking under U.V. light at 1000 J for 1 minute (Stratalinker, Stratagene Corporation, La Jolla, CA,

USA). The filter was briefly rinsed again in 0.15M NaCl and could be stored wrapped at room temperature or damp in the -20°C freezer in plastic wrap. Because of the small size of PCR products, transfer could be done in 1 hour, in high salt solutions, although it was usually done overnight using conventional conditions to insure complete transfer and to prevent possible damage to the DNA from rapid transfer or if rapid transfer was not convenient.

Membrane prehybridization used the same solution for YAC and RT-PCR filters. Oligonucleotide hybridisations were done with the Church and Gilbert buffer system (Church and Gilbert 1984). Prior to the addition of probe, the blot was prehybridised at 42°C for one hour to block potential nonspecific background signals with the prehybridisation solution. For 500 ml final volume the following mixture was prepared in a 200 ml flask according to the procedure previously described (Monaco et al. 1992) (courtesy of the Human Genetics lab, IMM):

250 ml	50% formamide
100 ml	4x SSC
25 ml	50mM NaPi
1 ml	1 mM EDTA
50 g	10% Dextran sulphate
50 ml	10% SDS
5 ml	10X Denhardts Solution
1.25 ml	Yeast tRNA

All components were mixed except Denhardt's and Dextran Sulfate warming to 40-50°C. Dextran was added slowly and heated while stirring at position 3 - 4 on the hotplate. When dextran was dissolved, the remaining components were added and stirring continued at 45°C until the solution became clear.

The bags were then cut open at one corner and the denatured probe was added to minimise bubble formation to a final concentration of 10^6 cpm/ml. The bags were resealed and hybridisation was carried out overnight, at the optimum Tm based on the purine and pyrimidine content for a DNA probe as calculated blow. *Tm to determine hybridisation temperature.

G/C = 4°C

A/T = 2°C

GCCCAA 20° = T_M

Wash T° = T_M -5°C

For oligonucleotide probes, hybridisation was always carried out at 42°C. Hybridisation was carried out for 3 hrs. or overnight at 42°C for oligo probes. For Washes 2x SSC + 0.1% SDS was used unless more or less stringent conditions were required depending on the activity registered on the blot after the first wash. The buffer was changed immediately after first wash and the activity on the blot monitored. The second wash for 10 minutes at room temperature with shaking was followed by a wash for 1.5 minutes at the appropriate Tm* (the same as the hybridisation temperature) for high activity or a wash 10 minutes at room temperature with shaking. Surplus buffer was blotted on 3 MM paper. Damp filters were mounted on 3MM Whatman paper and wrapped in plastic wrap for autoradiography with fluorescent markers and intensifying screens at -80° C overnight.

2.2.14.3 Denaturing sequencing gel electrophoresis

To determine the DNA sequence of cDNA library clones obtained from panning and PCR, the Sanger dideoxy sequencing mehod was used (Sanger 1977). The standard BioRad glass plate apparatus was used with seals held by virtue of accompanying rubber brackets. Polyacrylamide sequencing gels consisted of 1x T:B:E/5% acrylamide containing 5M urea. For sequencing, 200 mls of gel mix was made up by mixing 20 mls of 10x T:B:E, 25 mls of 40% acrylamide solution (39 linear:1 bis), 100 g urea and making up to 200 mls with dH_2O. The solution was stirred on a stir plate and heated to allow the urea to go into solution. To form a plug at the bottom of the sequencing plates, 1 ml of 10% Ammonium persulphate solution (APS) and 100 µl of TEMED were added to 50 mls of the solution. The mix was then quickly poured into a casting tray in which the assembled sequencing apparatus was stood and the screws were tightened at the bottom to allow the apparatus to stand. After 10-15 minutes, following polymerisation, the rest of the gel mix was polymerised with 1 ml of APS and 100 µl of TEMED and poured slowly between the glass plates as the apparatus was held or rested on a support at its top end. Before the gel set a comb was inserted with the flat side toward the gel between the plates at the top to create a slot into which sharks teeth combs would be inserted. Following polymerisation after about 1 hour or overnight, the top combs were removed and the apparatus was placed in a running tank to which 400 mls of 1x T:B:E were added. The top tank, behind the back sequencing plate, was also filled with 1x T:B:E to cover the gel, the electrodes connected and the gel warmed by running it at 100 W for about 15 minutes until the temperature indicator on the front plate read 50°C. The top well was then flushed out with a large syringe full of 1x T:B:E to remove accumulated urea from the gel and the sharks teeth comb carefully inserted so

that the depth of the teeth in the gel was about 1-2 mm. Denatured DNA sequencing reactions could then be loaded (2 ml if narrow width combs were used, otherwise up to 4 ml) and the gel run at 100 W constant power. As soon as the samples had run into the gel 200 mls of 3 M Sodium Acetate solution were added to the bottom tank - this creates an electrochemical gradient from the bottom to the top of the gel and retards the progress of DNA samples in the gel, the maximum retardation ocurring towards the bottom of the gel (Sheen and Seed 1988). This allows the gel to be run longer to get improved resolution of the larger sized DNA fragments while preventing the smaller, faster fragments from moving off the gel into the bottom tank. This has the effect of increasing the amount of readable sequence and therefore the alternative practice of loading short and long runs was not necessary. A gel run would be terminated as soon as the bromophenol blue dye front had reached the bottom of the gel. Routinely a single reaction would give up to 300 bp of readable sequence.

After the gel was run the plates were separated and the gel lifted off by carefully adhering a sheet of Whatman No. 3 blotting paper to it covering it with Saran™ wrap (Dow Corporation, USA) followed by transfer to a vacuum gel drier, after trimming if necessary. The gel was dried over two sheets of Whatman paper to prevent urea leeching into the drier under vacuum for ≈ 30 minutes to an hour at 80°C until dry to the touch and the Saran wrap could be fairly easily removed. The gel was cut to the size of a cassette and exposed to X-ray film without the use of intensifying screens at room temperature for 8-12 hours or longer if necessary.

2.3 RNA analysis

2.3.1 RNA isolation

To determine whether cDNAs isolated from the panning process were expressed in other tissues, ribonucleic acid (RNA) was extracted from cell lines and selected tissues with guanadinium isothiocyante (GuSCN) and separated from other cell components using the cesium chloride step gradient system according to the method originally established (Chirgwin 1979) and summarised for various tissues (Ausubel 1993). The procedure takes advantage of the property that RNA is the densest cell component and can thus be relatively easily purified. The initial extraction is performed under completely denaturing conditions. With all further manipulations, however, the standard precautions to avoid RNA degradation by RNAses on the hands and on glassware were insured by using gloves, disposable plasticware and using solutions, apart from EtOH, treated with diethyl pyrocarbonate (DEPC). To further prevent contamination, all RNA for the northern blot of the apo-4 gene and some RT-PCR reactions was prepared in a separate lab in a laminar flow hood (courtesy of Andrew McMichael, Molecular Immunology Lab, IMM, Oxford).

Solid tissue samples were immediately frozen in liquid nitrogen and ground with a mortar and pestle prior to RNA extraction to prevent possible RNA degradation during extraction. Solid tissue samples included 1) a skeletal muscle specimin was collected from theatre during a surgical manipulation (courtesy of Miss Hands, surgeon, John Radcliffe Hospital, Oxford) and; 2) a sample of placenta was immediately isolated following newborn delivery (courtesy of the Maternity Unit midwives, John Radcliffe Hospital, Oxford).

For suspension cultures 10^8 cells in log phase were pelleted by centrifuging at 1000 g for 5 minutes followed one wash in media and complete removal of the supernatant. For the adherent cell line Blondolet, 1-2 flasks were used depending on confluence, about 10^8 cells. 5 mls of GuSCN/NaOAc was added to each preparation and pipetted or swirled prior to shearing the DNA. The crude extractions could be stored at -20°C before RNA extraction, although they were usually used immediately. 5.7 M CsCl solutions were prepared carefully measured into tubes of the appropriate size. Normally 4 mls were used, but 12 mls was used for solid tissue samples. Prepared tubes were set aside and covered while the DNA was sheared.

The genomic DNA from the crude preparations was mechanically sheared using a Polytron homogeniser set on the maximum setting for 30 seconds. It was important to wear ear protection and preferably do the manipulation in a separate room to prevent noise pollution. For solid tissue, the Polytron was applied for one minute. Any remaining large tissue was removed by a quick spin at 1000 g prior to layering over CsCl. The sheared solution was carefully layered with a plastic pipet over the CsCl solution to 3mm from the top, using extra GnSCN solution to make up shortfalls in volume. Normally a swing-out SW55 rotor was used, but the SW41 was used for large RNA requirements. The volumes used, amount of CsCl, speed the time required for the run are listed below.

	CsCl solution	**GuSCN** solution	**Speed** (rpm x10^{-3})	**Time** (hrs)
SW55	1.5 mls	3.5 mls	55	8
SW41	4 mls	8 mls	32	12
SW28	12 mls	26 mls	25	20

All centrifugations were done at 22°C and the brake was left off to prevent gradient mixing. Following centrifugation, the upper phase was aspirated to just below the GuSCN/CsCl interface and scored just below it to prevent contamination from the upper phase. The tube was cut at the scored line, and the lower half flicked out and immediately inverted onto a paper towel to wick the remaining liquid away. The pellet was resuspended in 400 µl of DEPC treated water, transferred to a clean Eppendorf and separately phenol and then chloroform extracted as described in small scale DNA preparation. The RNA was precipitated by the addition of 1/10th volume (40 µl 3M NaOAc and 2 volumes 100% EtOH and placed on dry ice for 10 minutes.

The RNA was pelleted by spinning at 14,000 g in a microfuge for 5 minutes at 25°C. The pellet was washed once in 70% EtOH and respun. All traces of EtOH were removed and the pellet was resuspended in 400 µl of DEPC treated water. 5 µl of the preparation was analysed on a 1% agarose gel prepared with DEPC treated water to confirm the quality of the preparation. Absorbance was analysed at 260 nm (A_{260}) and the quantity of RNA based on an absorbance of 1=40 mg/ml.

2.3.2 PolyA$^+$ mRNA isolation using Dynabeads^R

The Dynabead system (Dynal) was used to separate polyadenylated RNA transcripts from total RNA by exploiting the ability of beads precoated with a polyT region to hybridise to the polyA tail in the transcripts. The manufacturers suggested protocol uses 75 µg total RNA as starting material. Of this, approximately 1 to 5% will be mRNA. The approximate capacity of the beads is 2 µg per mg, thus 1 mg beads will separate sufficient mRNA for

any further application. In terms of starting quantity and yield, this procedure is very flexible and can be adapted to suit the experimental design. The kit includes the following solutions:

2x Binding Buffer
20 mM Tris-HCl
2 mM EDTA 1.0 M LiCl
1x Washing Buffer
10 mM Tris-HCl (pH 7.5)
1 mM EDTA, 0.15 M LiCl
1x Elution Buffer
Magnetic particle concentrator - Dynal MPC^{R}-E-1

Adjust the volume of 75 µg RNA to 100 µl with distilled DEPC treated water, or with elution buffer. If the total RNA is more dilute than 75 µg/100 µl then add an equal volume of 2x binding buffer. Heat to 65°C for 2 minutes to disrupt any secondary structure, providing optimal conditions for hybridization. Remove 1 mg Dynabeads^{R} Oligo $(dT)_{25}$ from the resuspended stock tube to an Eppendorf tube in the magnetic rack provided. The beads immediately cling to one side. Remove the supernatant. Transfer the tube to another rack, add 100 µl 2x binding buffer. If the total RNA has already been diluted in binding buffer, this step should be omitted. Add the total RNA to the bead suspension, mix gently and leave to hybridize for 3-5 minutes. Separate using the Dynabead magnet and remove the supernatant. Wash twice using 200 µl washing buffer each time. Be sure to remove all of the supernatant after the last wash - this is particularly important when working in small volumes. Add the desired amount of elution buffer. Heat to 65°C for 2 minutes to disrupt secondary structure and separate immediately using

the Dynabeads. Transfer the supernatant to an Eppendorf tube. The eluted mRNA can be used immediately or stored frozen at -70°C. RNase inhibitors can be added to the preparation.

2.3.3 Poly A+ mRNA purification with cellulose columns

An alternative method of RNA extraction used Oligo(dT)-cellulose columns (Pharmacia) (4 or 8*) Oligo(dT)-cellulose suspended in storage buffer containing 0.15% Kathon^{R} CG to which the polyadenylated RNA would adhere. Solutions provided included:

High-salt buffer 10 mM Tris-HCl (pH 7.4). 1mM EDTA, 0.5 M NaCl.
Sample buffer 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 3.0 M NaCl.
Low-salt buffer 10 mM Tris-HCl (pH 7.4), 1mM EDTA, 0.1 M NaCl.
Elution buffer 10 mM Tris-HCl (pH 7.4), 1 mM EDTA.
Glycogen solution Glycogen in solution at 10 mg/ml.

Poly(A)$^{+}$ RNA was purified from total RNA on a pre-equilibrated oligo(dT)-cellulose spun column according to the manufacturers instructions as follows. The elution buffer was prewarmed to 65°C. The sample was applied to the top of the cellulose bed in the column and allowed to soak in under gravity. Centrifuge the column at 350 x g for 2 minutes. Apply 0.25 ml of High-salt buffer to the column, then centrifuge at 350 x g for 2 minutes. Repeat this wash step with a further 0.25 ml of the same buffer, and centrifuge again. Using the same centrifugation routine, wash the column three times with 0.25 ml aliquots of Low-salt buffer. Lift the column out of the centrifuge tube. Discard the buffer collected in the tube. Place a sterile 1.5 ml

microcentrifuge tube inside the 15 ml tube to collect subsequent effluents from the column. Replace the column with its lower end inside the microcentrifuge tube. To elute bound poly(A)$^+$RNA, wash the column four times with successive 0.25 ml aliquots of Elution buffer, prewarmed to 65°C Centrifuge at 350 x g for 2 min after each application, and collect the entire eluate in the same sterile tube.

2.3.4 RNA electrophoresis

For the purposes of checking the quality of RNA samples by the presence of undegraded ribosomal bands and the presence of a high molecular weight smear, running small samples could be easily done on routine non-denaturing agarose minigels. These gels were run sufficiently quickly to allow resolution of RNA specimens rapidly enough for degradation not to be caused by any RNAses in the gel or running buffer.

When RNA samples were resolved prior to transfer to nylon membranes then the denaturing formaldehyde agarose gels were cast, taking the usual precautions to avoid RNAase contamination of solutions and equipment. Apparatus could be easily treated by soaking in methanol before use and gloves were routinely worn when working with RNA. The buffer used for these gels was based on the ampholytic compound MOPS and is referred to as 10x MOPS buffer. It also contained 5 mM Iodoacetamide added just prior to use from a separate 1M stock solution. Typically, a 400 ml agarose gel would be made and just as the mixture was cooling after melting the agarose 1/60th volume formaldehyde was added and the gel cast.

The gel was transferred to the running tank which was filled with buffer such that the wells remained dry. The RNA samples diluted to the volume of the well in MOPS buffer (1x final) containing 1/6th volume of formaldehyde and heated at 70°C for 15 minutes to ensure denaturation before loading in the dry wells. If total RNA was used samples contained 10 - 25 mg of RNA but if polyA+ selected RNA was separated only 1-2 mg was used. The unused wells were filled with 1x MOPS running buffer containing a small amount of Orange G to indicate the progress of the separation when the gel was run. The gel was then run at 10V/ cm for about 4 hours until the dye front had moved close to the anodal end of the gel. Without photographing, the gel was soaked in 1 litre of 50 mM NaOH (to nick the RNA molecules and improve the efficiency of transfer) for 45 minutes which was then neutralised by the addition of acetic acid to a final concentration of 75 mM. Transfer was achieved using the same solution with a standard arrangement of towels beneath the gel and another stack on top of the nylon membrane which was placed on the surface of the gel. After overnight transfer, the transferred RNA was cross-linked to the membrane by placing the membrane, RNA side up, on a piece of heavy filter paper soaked in 10x SSC in a commercial crosslinking device.

After crosslinking the efficiency of the transfer was checked by methylene blue staining. The filter was immersed in a solution of 0.5M sodium Acetate (pH 5.5) containing 0.005 M methylene blue for 10 minutes. The filter was then rinsed in water until the stain had washed out of the filter except where it was bound to RNA. The stained filters could be photocopied to keep a permanent record of the position of the ribosomal RNA bands for sizing purposes.

2.3.5 Northern hybridisation

Hybridisation of probes to RNA transferred on to nylon membranes was done using random-primed DNA probes according to the same technique employed for southern blotting of DNA .

2.3.6 Preparation of radioactive RNA probes

Antisense RNA probes (riboprobes) were prepared according to a modification of the method first described by (Melton 1984). The template sequence was subcloned into pBluescript which has initiation sites for both T3 and T7 viral RNA polymerases. Before using in labelling reactions plasmids were cut with restriction enzymes such that the plasmid was cut immediately 5′ to the insert sequence coding strand. The riboprobes were then synthesised using the appropriate RNA polymerase so that the antisense strand was synthesised - synthesis being terminated when the RNA polymerase reached cut end of the plasmid. Care was taken to ensure that digestion of template DNA has reached completion prior to double phenol and chloroform extraction of template DNA which was then resuspened in DEPC-treated water and thereafter handled with appropriate precautions to avoid RNAase contamination. Labelling reactions were assembled on ice. To a clean reaction tube were added:

4 µl 5x transcription buffer

1 µl 20 mM DTT

2 µl 3NTP mix (Boehringer Mannheim Corp.)

10 µl α^{32}P]CTP (10 mCi/ml, 400 to 800 Ci/mmol)

The above materials were added before template DNA, followed by,

1 µl placental ribonuclease inhibitor (20 to 40U) (Boehringer Mannheim)

1 µl 0.5 mg/ml template DNA (25 mg/ml final)

1 µl bacteriophage RNA polymerase (5 to 10 U) (Boehringer Mannheim Corp.)

After incubation of the reaction mixture at 37°C for 20 minutes 1 µl of RNAase-free DNAase I was added and incubation at 37°C continued for a further 15 minutes. Riboprobes were purified on spin columns exactly as was done for labelled DNA preparations. The Cerenkov activity of the riboprobes were measured expecting a specific activity of 10^5 cpm/µl of the final preparation. The quality of riboprobe syntheses could be assessed by analysing a 2 µl aliquot on denaturing sequencing type, acrylamide gel and taking autoradiographs, however sufficient counts usually meant that synthesis was complete and the probe was always run alongside the samples to check integrity of the RNA. Riboprobe preparations were stored for up to 1 week at -20°C before use.

Usually sample sizes of 10 - 20 mg of total RNA were analysed in RPA assays. The RNA samples were first precipitated by the addition of 3 M RNAase-free Sodium Acetate to a final concentration of 0.3 M and at least 2 volumes of ethanol and the samples cooled on dry ice. After the RNA was pelleted in the microfuge the pellets were taken up in 35 µl of a mixture containing 1x hybridisation buffer 80% formamide and 2 x 10^5 cpm of each riboprobe - normally 2 probes would be used, the probe for the gene of interest and a GAPDH probe which, since it is a constitutively, ubiquitously expressed gene, serves as a loading control in the analysis. Control samples for the analysis

were 20 µg of yeast tRNA and undigested probe maintained at the incubation temperature to serve as a degradation control. The probe/sample mixture was denatured and then the sealed reaction tubes transferred to the hybridisation temperature of 45°C overnight.

The following day, the unhybridised (unprotected) riboprobe was degraded by the addition of of RNAase A (10 mg/ml) and T1 RNAase (5 mg/ml) in 300 µl of high salt RNAase digestion buffer and incubation at 37°C for 30 minutes. The RNAase was then inactivated by a proteinase K digestion which was done by the addition of 10 µl of 10% SDS solution and 2.5 µl of a 10 mg/ml Proteinase K stock solution (stored frozen until use) followed by further incubation at 37°C for 30 minutes. This reaction was terminated and protein removed from the samples by a single extraction against a 50:50 mixture of phenol and chloroform. The RNA molecules were precipitated from the aqueous phase by the addition of 5 ml of a 10 mg/ml yeast tRNA solution (to serve as a carrier) and 2 volumes of ethanol. After cooling the tubes on dry ice for 15 minutes the RNA in each tube was pelleted in the microfuge for 10 minutes and the samples washed twice in 70% ethanol (the washing steps required vigorous vortexing to achieve desalting which, if not done, affected the quality of the subsequent electrophoresis). The pellets were taken up in 3 µl of RNAase free water and then 8 µl of sequencing stop solution added just before denaturing at 95°C for 3 minutes and resolving fragments on a standard sequencing gel with flat well combs used rather than sharks tooth combs. After the gel was dried down, however, the saran wrap cover was not removed and autoradiographs were made with the aid of intensifying screens at -80°C.

2.3.7 RNAase protection analysis

The protocol given in Ausubel et al (1989) was followed exactly. This technique is used to identify rare transcripts which cannot be visualised on a Northern Blot. Radioactive RNA probes slightly larger than the segment of RNA of interest, due to the vector in which they are inserted, are used to probe a population of RNA. If the RNA of interest is present, it will hybridise to the probe. The hybridised product is then subjected to RNAsine which digests unhybridised RNA, leaving only the signal of interest, which will run at a given size smaller than the probe on a polyacrylamide gel. Riboprobes were synthesised as described above.

2.3.8 Reverse transcriptase-polymerase chain reaction

The reverse transcriptase-polymerase chain reaction (RT-PCR) method used was based on that previously described to detect low levels of dystrophin transcripts in various tissues (Chelly et al. 1988). To synthesise the first strand of cDNA, to 2 μg of total RNA, stored on dry ice, was coprecipitated, as with RNA extraction, with 2 μl primer to prime the RNA and start extension (.2 μg/ml random hexanucleotides (Pharmacia) or 5 μg/μl of oligo-dT (Pharmacia, oligo-dT_{12-17}, 5′ phosphorylated at 5 mg/ml for anchor PCR), 2 μl dNTPs (25mM) and 37 μl DEPC-treated dH_20. The mix was heated for 5-10 minutes at 65°C, spun for 2 minutes at 14,000g and cooled on ice. To this reaction was added 2 μl reverse transcriptase, and 1 μl RNasine™ (destroyed at 65°C). The RNA was then hydrolysed with NaOH and the reverse transcriptase heat inactivated by incubation at 70°C for 10 minutes. The

reaction was incubated for 30-45 minutes at 42°C in a water bath, spun down at 14,000g for 2 minutes and either used immediately or stored at -20°C. 10 µl of the newly synthesised cDNA was then added directly to each reaction using the standard conditions below.

500 ng of each primer. (.5 µl)	Temperature	Time
4 µl Buffer	Annealing - 40-45°C.	30 sec.
0.5 µl amplitaq enzyme (2.5 U.)	Extension - 72°C.	1.0 min.
dH_20 to 50 µl	Denaturing - 94°C.	30 sec.
PCR extension at 72°C.	Run for 30 cycles	

Following PCR, 10µl of each sample was analysed on a 1% ethidium bromide agarose gel and the products visualised by ultraviolet light. If products were visible, the gel was usually subjected to southern blotting with end-labelled oligonucleotides and the products visualised by autoradiography.

2.3.9 Primer extension

The protocol is paraphrased from the published version (Ausubel 1993). The method allows the identification of putative transcription start sites by incubating a complementary oligonucleotide about 80 bp downstream of the 5' end of a given gene with reverse transcriptase in an attempt to produce transcripts which produce a band of RNA at a putative initiation site which is normally found 100-150 bp upstream of the initiating methionine. By running this reaction along known genomic DNA sequence upstream of the 5' end of the gene, a putative position of transcription initiation can be identified. This procedure was used to identify a putative 5′ transcription initiation site upstream of the apo-4 gene by annealing a 40 base antisense

oligonucleotide approximately 80 bp downstream of the 5′ end of the apo-4 cDNA.

The oligonucleotide probe was first end labelled with polynucleotide kinase as described above followed by treatment with RNasine for 10 minutes at 37°C and precipitation with DEPC treated solutions. To anneal the test RNA to the probe, 10 mg of total placental RNA, freshly isolated, was incubated with 1 µl of the 1 mg/ml target oligo, 40 µl of NaAc and two volumes of EtOH, vortexed and placed on dry ice for 15 minutes. The mix was pelleted and washed twice in 70% EtOH with the remaining supernatant removed with a pipetman tip. The RNA was then denatured for 10 minutes at 65°C. The following RT mix was prepared fresh on ice (per reaction):

3.5 µl 4 mM dNTPs

2.5 µl 10x RT buffer

1.25 µl RNAsin

18 µl H_2O

25 µl RT mix was added to each sample and resuspend thoroughly. Resuspension of the RNA and oligo in the RT mix can be monitored by following the radioactivity. 2 µl (40 U) AMV reverse transcriptase was added to each reaction and incubated for 90 minutes at 42°C. The reaction was stopped by adding 1 µl of 0.5 M EDTA to each reaction, and 1 µl of 1 mg/ml pancreatic ribonuclease A with incubation for 30 minutes at 37°C. 100 µl of 2.5 M ammonium acetate was added followed by extraction with 200 µl phenol/chloroform, isoamyl alcohol. The aqueous phase was transferred to a fresh tube and 300 µl ethanol added to precipitate the RNA. The pellet was washed with 500 µl of 70% ethanol/30% water (vol/vol) spun for 5 minutes

and the excess liquid removed with a p200 pipet tip. The pellet was resuspended in 3 µl DEPC treated water followed by 4 µl formamide loading buffer and well mixed. The tubes were boiled for 3 minutes and placed on ice. 3 to 4 µl were analysed alongside cDNA and phage clone DNA reactions prepared according to the sequencing method described above on a 4-5% sequencing gel. The gel was dried as in the sequencing procedure and subjected to autoradiography overnight at -80°C.

2.3.10 *In vitro* transcription and translation of cDNA

The method enclosed in the *in vitro* transcription and translation kit was followed (Promega, UK) and integrated with the published method (Ausubel 1993). This method allows the determination of whether a cDNA can produce a protein by first isolating it from all but the T3 or T7 promoter in its carrier vector, producing a transcript by incubation with reverse transcriptase and then incubating the transcribed product with rabbit reticulocyte lysates and/or wheat germ extracts which contain all of the elements necessary for translation. The materials used included:

pBluescript DNA containing a T7 promoter and the cDNA of interest
Appropriate restriction endonucleases to linearise the template
T:E buffer
5x ribonucleoside triphosphate mix
10x T7 RNA polymerase buffer
Pancreatic ribonuclease inhibitor (e.g., RNAsin from Promega Biotec)
T7 RNA polymerase
Buffered phenol
Isobutanol
10 M ammonium acetate
100% ethanol

In vitro translation kit (Promega)
^{35}S-labeled methionine (1400Ci/mmol)
0.1 M NaOH
10% trichloroacetic acid (TCA)
Amplify™ (DuPont)

The DNA template was prepared by subcloning the apo-4 cDNA into pBluescript which contains the T7 RNA polymerase upstream of the 5′ end. The insert was linearised with PstI or, to generate a second template which truncated the 3′ 270 bp and the inversion, Hpa I. The templates were run out on an agarose minigel and purified using the Geneclean system as described above. The DNA was purified by phenol extraction and ethanol precipitation and resuspended in 50 µl of DEPC treated T:E buffer. The following 25-µl reaction mixture was set up on ice for each template:

8 µl H_2O
5 µl DNA (total 1 µg)
5 µl 5x ribonucleoside triphosphate mix
5 µl 10x T7 RNA polymerase buffer
1 µl RNAsin (30 to 60 U)
1 µl T7 RNA polymerase (5 to 20 U)

The reaction tubes were incubated for 60 minutes at 40°C, spiked with an additional µl of T7 polymerase and incubated for another hour. A 1:10 dilution of DNAse was then added, incubated for 15 minutes at 37°C and the mixture was brought to 100 µl with dH_20. 25 µl of buffered phenol was vortexed and extracted immediately. The aqueous phase was transferred to a new microcentrifuge tube, extracted twice with isobutanol, and 6 µl of 10 M ammonium acetate followed by 70 µl 100% ethanol was added. The RNA was ethanol precipitated on dry ice and washed once with 70% ethanol. The RNA was resuspended in 24 µl DEPC treated dH_2O. 6 µl of 10 M NH_4Ac and 70 µl

ethanol were added, the RNA was reprecipitated and washed once with ethanol. The RNA was resuspended again in 10 µl RNAse-free T:E buffer and samples were run out on a gel to show relative production of product and sizes.

10 µl RNA from each template was then added to the recommended mix in the *in vitro* translation kit (Promega) of Wheat Germ Extracts and Rabbit Reticulocyte Lysates which were stored at -80°C. All kit components were quick-thawed and placed on ice prior to use. Bovine Mosaic Virus was included as a positive control for each set of two templates by adding the following mix:

Wheat Germ Extracts (WGE)
25µl WGE
4µl 1mM Amino Acid Mix (minus methionine)
2µl RNA substrate in H_20
130 mM KAc
1µl RNasin™ Ribonucuclase Inhibitor (40U/µl)
4µl [^{35}S] methionine (1,200Ci/mmol) at 10mCi/ml
Nuclease-free (DEPC treated) water to a final volume of 50 µl

Rabbit Reticulocyte Lysates (RRL)
35µl RRL
7µl Nuclease-free water
1µl RNasin™ RNase Inhibitor (at 40U/µl)
1µl 1 mM Amino Acid Mixture (minus methionine)
4µl [^{35}S] methionine (1,200Ci/mmol) at 10mCi/ml
2µl RNA substrate in H_20
50µl final volume

Reactions were allowed to proceed at room temperature for 60 minutes. 5 µl of each sample was removed and added to 10 µl of 2x SDS protein loading

buffer and tapped to mix. The tubes were heated at 100°C for 2 minutes and spun down. 5 µl of each sample was loaded immediately onto a 10% SDS-PAGE gel and run out alongside molecular weight standards. The remainder of all reactions were stored at -20°C. Following protein gel fixing in stain and destain for 30 minutes each, as described below, [^{35}S] proteins were visualized by fluorography with Amplify followed by autoradiography at -80°C.

2.4 Protein Expression and Adhesion Work

2.4.1 Production of Recombinant Fc-adhesins

Plasmids containing clones coding for recombinant soluble cell adhesion molecules were transfected into COS cells by the DEAE-Dextran method (Seed and Aruffo 1987b). The morning after transfection the medium in the cultures was changed for DMEM/1% FCS. 72 hours post-transfection the medium was collected and replaced with fresh DMEM/1% FCS for another 48 hours of incubation before being discarded. The medium containing Fc-CD33 was centrifuged at 4,200 rpm to pellet any shed cellular debris. Supernatants were then collected and decanted into clean dry bottles. The bottles were placed on a rotating shaker overnight with 500 ml of Protein A- Sepharose beads added. The contents of the bottles were transferrred to a disposable polythene column, recovering the Protein A Sepharose beads with PBSA. The beads were washed with two bed volumes of PBSA supplemented with a further 150 mM Sodium chloride. Bound Fc-adhesins were eluted in fractions by the addition of 1 bed volume of pH 3.5 100 mM acetate buffer into 100 µl 1 M Tris to neutralise the eluate. Ten Fractions were normally

sufficient to elute bound protein. The eluted recombinant Fc proteins were examined by SDS-PAGE analysis on 10% gels and total protein concentration was estimated by measurement of A_{260} with an extinction coefficient of 1.4. If necessary, Fc proteins were concentrated by ultrafiltration, using the Amicon centrifugation system and dialysed against PBSA, if not used within two days.

2.4.2 Large scale manufacture of Fc-adhesins

Plasmids containing clones coding for recombinant soluble cell adhesion molecules were transfected into COS cells by the DEAE-Dextran method. For a large scale transfection 20 x 15 cm dishes of 80% confluent COS cells were transfected using the protocol described in the section on expression cloning. On the day after the transfection the medium in the cultures was changed for DMEM/1% FCS and the cultures allowed to grow for another 3 days. The medium was then collected and the dishes recharged with the same media for a further 48 hours of incubation before being discarded.

Cellular debris was pelleted from the Fc-adhesin-containing medium by centrifugation at 4,200 rpm in the J6 centrifuge. The supernatants were then decanted into clean dry bottles. To capture the Fc proteins, the bottles were placed on a mixer with a magnetic flea and 500 µl of Protein A- Sepharose beads were added to mix overnight at 4°C. The following day, the contents of the bottles were transferrred to a disposable polythene column to allow the Protein A Sepharose beads to be recovered. The beads were washed with two bed volumes of PBSA or until the beads were white, supplemented with a further 150 mM Sodium chloride. Bound Fc protein was then eluted in fractions by the addition of 1 bed volume (500 µl) of pH 3.5 100 mM acetate

buffer 10 times and elution of the unbound protein. The eluted recombinant Fc proteins were checked for quality by SDS-PAGE analysis and an estimation of total protein concentration was taken by a measurement of absorbance at A_{260}. For storage and use in experiments the proteins were then concentrated, using the Amicon centrifugation system. They were then dialysed against PBSA overnight with stirring at 4°C. To dialyse small volumes of protein without losses Fc proteins were incubated in inverted Eppendorf tubes with the lids cut out and a piece of dialysis membrane placed under them. The tube was secured with parafilm and held in a styrofoam rack on PBSA.

2.4.3 Immunoselection and Episomal Rescue

The principles of the method of cDNA isolation with Fc-chimeric probes by transient expression and immunoselection are discussed in the introduction. The practical details of this procedure are based on previously described methods for antibodies (Seed 1987a) which have been modified and employed for Fc-probes (Simmons 1993) as described in the section below.

2.4.3.1 DEAE-Dextran Transfection of COS cells

For the first round of panning or for other procedures requiring transfected cells the following method was used to achieve a high transfection efficiency using the DEAE dextran method (Sussman and Milman 1984). cDNA libraries constructed in pCDM8 were transfected directly into COS cells which had been prepared the afternoon before transfection to achieve 80%

confluence the following day by splitting at 1 x 10^7 cells/15 cm dish. Two dishes 15 cm in diameter were used for each transfection. A transfection mix was prepared in a 50 ml (Falcon) tube consisting of 40 µg of CsCl purified DNA diluted in 1 ml of T:E and swirled to mix. To make the DNA adhere to the COS cells, in a laminar flow hood, 800 µl of DEAE-dextran (10mg/ml) was added to the transfection mix, swirled, and allowed to stand for five minutes at room temperature to allow the DNA to adhere to the Dextran. Any precipitate visible after the incubation meant that the DNA had precipitated in solution and the procedure would have to be repeated with a higher quantity of T:E. 1 ml was in excess, however, and no precipitate was ever seen. 20 mls of DMEM containing 100mM chloroquine diphosphate (Sigma) was then added and mixed by inverting the capped tube. Chloroquine acts as a lysosomal acidifying agent to increase the stability of the transfected DNA as it is taken up into the primary lysosomal compartment, although it also increases the toxicity of the transfection to the cells, limiting incubation time after the transfection mix is added.

The dishes for transfection were then removed from the incubator and the medium was aspirated and replaced with 10 mls of transfection mix. 10% dialysed FCS or Nuserum was added to the original transfections but it was found that minimal amounts of protein remained in the culture dish making it unnecessary to add this to later transfections. The dishes were replaced in the incubator and observed for vacuole formation during the next 2-4 hours. Transfection normally took place after 3 hours, although 2 hours was sufficient for single-clone transfection. DNA uptake by the cells was considered sufficient when vacuoles were evident in the majority of them. If the cells had become spindly or had started to detach, transfection had proceeded further than the optimal limit. At optimal vacuolisation,

transfection was stopped by aspirating the transfection mix and adding 10 mls of a PBSA/10% DMSO mixture and leaving it for exactly one minute. The mix was aspirated and 20 mls of 10% FCS/DMSO was added to allow the cells to recover overnight. The following day, the cells were lifted with 2 mls of a 1/5 mixture of trypsin/PBSA and incubated at 37°C for five to ten minutes. This also allowed residual DMSO to be removed which would prevent cell lifting with divalent cations prior to the panning process. The cells were removed by tapping the dish and separated from each other by gently pipetting with a 5 ml pipet. The cells were replated in 20 mls of the same medium to dilute the trypsin and inactivate it with the endogenous protein. Dishes were prepared for immunoselection the night before or earlier if the plates were frozen as described below. The following day, 36-72 hours post-tranfection, the cells were ready for immunoselection.

2.4.3.2 Preparation of Immunoselection antibody plates

Polystyrene dishes were prepared to allow the Fc portion of a probe adhering to a transfected cell, to be captured by adhering to a secondary antibody on the dish in an attempt to isolate ligand populations. Untreated 10 cm bacteriological plates (Falcon, Oxnard, CA,USA) were coated with affinity isolated goat anti-human IgFc antibodies (Sigma, UK) by adding 5 mls of a 10 mg/ml solution of antibody in 50 mM Tris-HCl (pH 9.5) and swirling or shaking the dish on the benchtop until the surface was uniformly wetted. The dishes were left to stand at 25°C for two hours. The antibody solution was then aspirated and stored at 4°C. The solution could be used a further two times to coat dishes. The dishes were carefully washed twice with PBSA and the remaining protein binding sites were blocked by the addition of PBSA

containing 2 mg/ml of BSA incubated overnight at 4°C. The following day, the solution was aspirated. The dishes were washed once with PBSA and used immediately or aspirated without washing and stored in their original bags at -20°C.

2.4.3.3 Panning procedure

The method described is based on the method originally developed for panning with antibodies (Wysocki and Sato 1978). COS cells 36-72 hours post-transfection were washed once in PBSA and lifted using PBSA/2 mM EDTA incubated at 37°C for 15 minutes. Cells were transferred to 50 ml Falcon tubes and washed once in ice-cold panning wash consisting of PBSA/5% FCS/4 mM EDTA/0.02% sodium Azide and pelleted at 1000 g for five minutes. The pellet was resuspended in 400 uls of panning wash on ice and 10-30 ug of Fc-CD33 or Fc-CD34 was added and mixed with a 1 ml pipet tip. Cells transfected with libraries containing Fc-receptors were pre-incubated with 20 ug of IgG from an irrelevant species (usually goat or mouse) for 20 minutes and washed once in panning wash followed the addition of Fc probe as above. The cells were incubated with the Fc probe on ice for 30 minutes. As a modification of the procedure, cells were incubated on a rotary mixer in the cold room and media was substituted for PBSA in the panning wash. The tubes were then filled with 50 mls of panning wash and the cells pelleted at 1000 g. The cells were resuspended in 5 mls of panning wash, pipetting to achieve maximum cell separation and gently and evenly plated onto panning dishes which had been thawed and/or rinsed once with panning wash and refilled with five mls. The panning plates were then left undisturbed at room temperature for

two hours when cells bound to the Fc probe would bind to the dish via the secondary antibody.

Following the two-hour incubation, unbound cells were removed from the dish with three gentle washes with panning followed by gently aspirating the wash with a pipette. On the third wash, the dish was inspected under a microscope to check for bound cells and the total number of bound cells was counted by carefully moving the dish to see if the cells moved. Bound cells stayed stationary. The maximum number bound cells which was observed was between 5-10 after the second or third rounds of panning. Few cells were observed after the first round. The dish was then carefully removed and either washed again until most of the unbound cells were removed or tipped up to remove the last traces of panning wash. Excess wash was removed with a p1000 tip.

2.4.3.4 Episomal Rescue of Potential Ligands

To rescue episomes from the bound COS cells, 400 µl of Hirt solution (1% SDS/10 mM EDTA) (Hirt 1967) was immediately added to the drained panning plates and swirled to cover the whole dish. A viscous solution became immediately obvious and cells were propped up on their sides at a 45° angle to allow the solution to drain into the corner of the dish for 15-30 minutes. The Hirt solution was then carefully transferred to Eppendorf tubes with a p1000 pipetman to avoid shearing genomic DNA. The genomic DNA was precipitated by the addition of 100 µl of saturated NaCl solution and left overnight in slushy ice. The following day, the genomic DNA and SDS were pelleted at 14,000 g for five minutes at room temperature and the episome or

plasmid containing supernatant extracted with 500 µl phenol and then chloroform extraction. 10 µl of LPA carrier solution was added and shaken for 30 seconds. The tube was filled with 95% EtOH, placed on dry ice for 15 minutes and spun at 14,000 g for four minutes. The residual supernatant was removed with a p200 tip and the pellet was allowed to air dry for 1-2 minutes. The DNA was then resuspended in 50 µl T:E. 15 µl of this mixture was added to 100 µl of MC1061/p3 competent cells pre-thawed on ice for 15 minutes. The cells were gently flicked to mix and left on ice to allow the DNA to adhere to the cells for 15 minutes. The cells were heat-shocked to allow DNA entry for five minutes at 37°C while 10 mls of top-agar was poured onto 15 cm amp/tet LB agar plates and allowed to set. The plates had been prepared by pouring 15 mls of molten LB agar with amp/tet added into a as described in the ligation section above, allowing them to set and drying them in a drying oven for 30 minutes. The transformed into *E.Coli.* was plated out onto the agar plates with a flame-disinfected glass spreader cooled on the agar plate and incubated overnight at 37°C. The next day colonies were counted and at least 1-2 x 10^3 colonies were necessary to continue to the spheroplast fusion step. If growth was not at this level, the Hirt mixture could be retransformed to assess whether the transformation step was at fault.

2.4.3.5 Transfection by Spheroplast fusion

In preparation for rounds two and three of panning, the colonies were amplified to their maximal level in preparation for their conversion to spheroplasts which could be fused with COS cells by polyethylene glycol (PEG) treatment based on the previously described method (Sandri 1983). After counting the colonies obtained from the transformed Hirt solutions, the

plates were allowed to achieve maximum growth throughout the day. To achieve maximal amplification of episomes/bacterium spectinomycin was added to the amp/tet agar prior to pouring the agar dishes. Spectinomycin is a bacteriostatic agent which acts to amplify plasmid number because it cannot prevent the replication of *col*E1 based plasmids. The colonies were then transferred to amp/tet/spectinomycin (100 µl/100 ml of a 100 mg/ml stock) "ATS" LB agar plates by pressing a .45 µm nylon (preferably) or nitrocellulose filter cut to fit inside the 15 cm plate over the colonies until complete adherence was achieved. The filter was lifted and placed, colony side up, on the ATS plate at 37°C until the following morning. This was done in place of the usual procedure with colony amplification in flasks as phage lysis was eliminating colony formation during incubation at 37°C. One 15 cm dish of COS cells at 1 x 10^7 density was then split into six 6 cm dishes in preparation for protoplast fusion the next day to allow for 100% confluence necessary to withstand the spheroplast fusions.

The colonies were removed the following morning from the ATS plates by applying 20 ml of LB supplemented with 12.5 µg/ml amp and 7.5 µg/ml tet to the filter of each plate. Plates were placed on a shaker for 30 minutes, removed and gently mixed with a spreader to remove all bacteria. The mix was then pipetted into 250 ml bottles and spun for five minutes at 5000 g in a JA-14 rotor in the J221 centrifuge. The supernatant was then decanted and the excess liquid aspirated. The pellets were resuspended with a pipet in 10 mls of a mixture of ice-cold 20% sucrose/50 mMTris-HCl, pH 8 and left for five minutes on ice. During this time a stock of chicken egg lysozyme (Sigma) was made up fresh in 250 mM Tris, pH 7.5 with a final concentration of 10 mg/ml. Two mls of this stock was then added to the sucrose/Tris slurry, swirled to mix, and incubated for five minutes on ice to make the bacterial walls

permeable. Percent conversion to spheroplasts was then checked by microscopy. Two mls of 250mM EDTA, pH 8.0 was added and incubated a further 5 minutes on ice followed by the addition of two mls of 50 mM Tris-HCl, ph 8.0 incubated for five minutes at 37°C. In a laminar flow hood, 20 ml of cold DMEM/10% sucrose/10 mM $MgCl_2$ was then added dropwise with swirling at about 2 drops/second and the slurry replaced on ice. It was most convenient to prepare a maximum of two libraries at a time and stagger the timing of the spheroplast preparation for more than this.

In preparation for fusion with polyethylene glycol (PEG) 1450, the COS cells were then removed from the incubator and the medium aspirated from each set of six dishes. A maximum of 12 dishes were used for each fusion. 5.5 mls of spheroplast suspension was added to each dish, the lids were replaced and stacks of three dishes each were made from each fusion to allow two libraries to be spun at once. The dishes were carefully placed on the inside edge of a Beckman centrifuge rotor and spun at 2000 rpm for 10 minutes, run down with the brake off to avoid disturbing the spheroplast layer. The dishes were carefully removed and inspected to insure that a smooth "skin" had attached to the COS cells. The dishes were then placed in two rows of six dishes each in the laminar flow hood with each lid was placed behind its dish and the supernatants were carefully aspirated. With a 25 ml pipet, 1.5-2.0 mls 50% PEG 1450/50% DME (no serum) (w/w) was added to each dish by gentle dropping into the center of the dish or swept around the dish to ensure that the PEG was evenly distributed and the timer was started to allow a maximum of two minutes contact with the PEG solution. After adding PEG to the last dish, the dishes were propped up on their lids and the remaining PEG aspirated. After two minutes, with a 25 ml pipet, 1.5 mls of DMEM alone was added to the center of each dish and allowed to spread radially to wash off

residual PEG. The dishes were propped up on their lids and the media gently aspirated. The wash was repeated, the media aspirated and the dishes replated with 10% FCS/DMEM containing 15 μg/ml gentamycin sulfate to inhibit contamination while the cells were recovering. The cells were left for 4-6 hours at 37°C. After this time the "skins" had visibly loosened and were removed by aspirating the media washing once in 3 mls of DMEM, aspirating again and replating in 3 mls of 10% FCS/DMEM/gentamycin. The medium was changed again the following day and on the second day the cells were lifted with PBSA/2mM EDTA as before and subjected to a round of panning.

2.4.3.6 Analysis of cDNA Clones From Episomal Rescue

With Fc-probes, it became necessary to analyse the cDNA population rescued after the second round of panning, although with antibody panning it had been usual to wait until the third round. It was found that the integrity of the plasmid would sometimes not be retained in the third round and that sometimes the panning efficiency was decreased relative to the second round. To analyse the homogeneity of the cDNA clones selected with Fc-probes, 24-48 random colonies were picked from the transformed Hirt solutions and the rescued plasmids analysed by DNA minipreps. Ideally, increasing numbers of adherent cells would be followed by a common pattern of cDNA inserts emerging from Hind III/Pst I digestion of miniprep DNA from the final round transformation analysed as described above. One-half of the mini-prep DNA was run out and for inserts which appeared in duplicate lanes, the residual other half of the miniprep DNA was transfected by the DEAE-Dextran method into COS cells seeded out into 6 well dishes using scaled down volumes as given in the DEAE-Dextran method above. 36-72 hours

post-transfection these transfectants were then either stained in situ by fluorescent microscopy or subjected to FACS analysis as described below to confirm binding to the Fc probe. No trypsinisation was required for in situ staining but it was necessary for FACS analysis to achieve an even distribution of the cells.

2.4.4 Adhesion Assays to ECM proteins.

An adhesion assay was used to assess whether Fc-CD33 and Fc-CD34 could bind to extracellular matrix proteins (ECM) in the form of full-length or partial peptides including: Collagens I-VII, Type I collagen; Fibronectin - MW 44 Kd; Fibronectin 45 - α-Chymotrypsin 45 Kd Proteolytic Fragment, Gelatin-Binding Region; Fibronectin 120 - α-Chymotrypsin 120 Kd Proteolytic Fragment, Cell-Attachment Region; Fibronectin 634 - Trypsin 30 Kd Proteolytic Fragment, Gelatin-Binding Region; Fibronectin 637 - α-Chymotrypsin 40 Kd Proteolytic Fragment, Heparin-Binding Region; Fibrinogen - fom plasma, MW 341 Kd; Vitronectin - promotes cell adhesion; MW 65 and 75 Kd; Laminin - from placenta, MW 130 and 160 Kd (Calbiochem, U.K.) bound to plastic. 96 well tissue-culture plates (Falcon) were coated with 50-100 µl/well of 10 µg/ml BSA, Human IgG, or anti-human IgFc diluted in 50 mM Tris buffer, pH 9.0 and left to adhere overnight at 4°C. The following day, wells were aspirated, washed 3 times in PBSA.B ($CaCl_2$).C ($MgCl_2$) and coated with 20 µl/well of 10 µg/ml ^{35}S labelled Fc-adhesin. Plates were incubated overnight at 4°C. Wells were washed 3 times in PBSA.B.C., aspirated and lysed in 50 µl/well 1%SDS/.2 M NaOH. Supernatant was collected into vials and adhesion assessed by the addition of 2 ml of scintillant followed by analysis on a Beckton-Dickinson scintillation counter.

2.4.5 Adhesion assay to HUVEC

Primary HUVEC were plated out using a scaled down version of the conditions described under HUVEC cell culture into four 24 well plates and allowed to adhere overnight. The following day, 100 µl of a mixture of IL-1β and TNFα diluted in DMEM/10% FCS was added to each relevant well to a final concentration of 40U/well and 2U/well respectively on each of the two plates. The plates were then incubated with 500 cpm/well of ^{35}S-labelled Fc-chimeras and left to incubate at either 4°C or 37°C for one hour. Input counts were 500 cpm/well, n=3. Wells were washed twice with PBSA.B.C., lysed in 50 µl/well 1%SDS/.2MNaOH, and counted by scintillation. The mean values were then assessed and standard deviations calculated.

2.4.6 Immunoprecipitation and *In Vitro* Kinase Assay

Cell lines in log phase of growth were washed once in RPMI medium with ice-cold phosphate buffered saline. Cells were solubilised in lysis buffer at pH 8.0 containing Digitonin (Sigma, UK) and/or Nonidet P-40 (NP-40, BDH, UK) [1% (vol/vol)], 20 mM Tris-base, 0.5 mM ethylenediamine-tetraacetic acid (EDTA), and 1 mM phenylmethylsulfonylfluoride (PMSF) for 30 minutes at 4°C. Subsequent steps were performed at 4°C. The lysate was centrifuged at 12,000 x g for 10 minutes and precleared for 30-60 minutes with 50-100 µl of protein A sepharose. The protein A was removed by brief centrifugation at 12,000 g and the lysates (20-30 x 10^6 cells/ml) incubated for 1 hr. with 14 µg/ml Fc-adhesin or 5-7 µg/ml antibody. 50 µl of 30% (wt/vol) protein A-Sepharose resuspended in lysis buffer was added to each sample and incubated for 30 minutes and immunoprecipitates then washed twice with cold NP-40 lysis

buffer and subjected to the kinase reaction as previously described (Rudd et al., 1988, Barber, et al., 1989) using 1-10 μCi of [g-^{32}P] ATP (1Ci = 37 GBq; Amersham Intl., U.K.). For reprecipitation of phosphorylated proteins, following the kinase assay, reactions were stopped with a final concentration of 1.5% SDS and boiled for 10 minutes as previously described (Rudd et al., 1988, Barber, et al, 1989) except supernatants were resuspended in 300 μl of lysis buffer with 5-7 μg/ml of target antibody for 30 minutes followed by protein A sepharose as in immunoprecipitation. Samples were washed twice in lysis buffer and subjected to SDS-PAGE and autoradiography as above.

2.4.7 Sequential Depletion of Lysates

Following initial lysis of HUVEC in a 50/50 mix of 1% Digitonin/1% NP-40 the lysate was spun for 10 minutes at 12,000 g and the pellet was lysed again in 1% NP-40 alone in a smaller volume and then subjected to the same sequential depletion and kinase assay as the other samples and compared to the detergent mix for its ability to precipitate phosphoproteins. Sequential depletion was accomplished by adding antibody or Fc protein, incubating for an hour at 4°C, followed by the addition of Protein A sepharose at 4°C. The lysate was then removed and an additional aliquot of Protein A was added followed by a 30 minutes incubation at 4°C. This was repeated up to four times. At the end of the immunoprecipitations, all lysis buffer was removed from the Protein A and all pellets were subjected to the kinase assay.

2.4.8 Immunoprecipitation - stringent wash method

The high salt-low salt washing method was contributed by Paul Crocker, ICRF, IMM. Cells were labelled to as high a specific activity as possible. Solutions are listed at the end of this section. All steps were carried out at 4°C or on ice. Adherent cells were rinsed twice with PBS add 1-2 ml of lysis buffer was added. Cells in suspension were pelleted and resuspended in 1-2 ml lysis buffer for 30 minutes on ice. Nuclei were removed by centrifugation for 10 minutes in a microfuge. The supernatant was ultracentrifuged at 50,000 g at 4°C for 30 minutes. To pre-clear the lysate the tubes were microfuged for two minutes and the supernatant was transferred to a fresh tube. The addition of protein A was repeated. The lysate was divided into 200-300 µl portions with both a positive and a negative control, if possible. Either 5 µgs of the affinity purified mAb of interest, 100 µl of tissue culture supernatant or 5 µl of antisera were added to each tube. The tubes were inverted and incubated at 4°C for one hour. 50 µls of protein A/G-agarose suspension was added and mixed on a rotary mixer at 4°C for 1/2 hour. The pellet was washed once with 1 ml buffer A, twice with 1 ml buffer B and once with l ml buffer C. The lysate was removed with a Hamilton syringe and the pellet was resuspended in 20 µls 1x Laemmli sample buffer containing 1 mM EDTA and 5 mM iodoacetamide. The samples were boiled for 2 minutes, microfuged for 2 minutes, and the supernatant was removed. 10 µls of the sample was loaded on a 10% Laemmli SDS-PAGE gel and run at 150 mV for 1 hour. Samples were analysed reduced and non-reduced if possible. If cells were metabolically labelled, immediately after SDS-PAGE, Amplify was used to intensify the signals.

2.4.9 Immunoprecipitation with Peptide Blocking

To examine whether specific antigens could be blocked with peptide, the aliquot of either purified or crude antisera (≈ 1mg/ml) was incubated for an hour at 4°C in a 1:2 dilution of 10 mg/ml peptide dissolved in 0.1 M $NaHCO_3$, pH 8.3/0.5 M NaCl. Blocked antibodies were then added to lysates at an equal concentration to the test antibody and immunoprecipitated as described above.

2.4.10 Phosphoamino acid analysis

Phosphoamino acid analysis was performed by excising bands from fixed polyacrylamide SDS-PAGE gels in a solution of 50 mM NH_4HCO_3, 0.1% SDS and 5% β-mercaptoethanol for 1 minutes at 100°C, homogenising and eluting with rotation at 37°C for 2 days. Eluted proteins were then precipitated with trichloroacetic acid as described (Boyle 1991). Gel pieces were centrifuged at 12,000 x g for 15-20 minutes. Counts were taken from 100 µl of each supernatant with one 100 µl sample used as a control after adding 100 µl of elution buffer. Samples were then lyophilised, the residue dissolved in 250 µl of deionised H_2O and lyophilised again. The precipitate was washed in acetone and hydrolysed in 6N HCl at 110°C for 2 hrs. The samples were again lyophilised as above. The individual phosphoamino acides were then separated by electrophoresis (1000 V for 30 minutes) at pH 3.5 in a pyridine:acetic acid:water solution (1:10:189 v/v). Ninhydrin was used to detect non-radioactive standards while radiolabelled phosphoamino acids were observed by radiography.

2.4.11 Cell surface labelling

2.4.11.1 Cell surface iodination

For immunoprecipitation analysis of cell surface molecules intact living cells were usually first surface iodinated based on the original method (Emerson et al. 1979). Successful iodination depended on the presence of tyrosines in the ECD of the molecule of interest. Iodination was done enzymatically with lactoperoxidase and glucose oxidase, supplied commercially as Enzymobeads (BioRad Corp., USA), determined to be the superior method to iodinate cells as there is no observed destruction of the labelled protein binding site (Kienhuis et al. 1991). β-deoxyglucose was present in labelling reactions as an oxidising substrate. Enzymobead reagent was rehydrated with 0.5 ml distilled water at least l hour before use. 1% Beta-D-Glucose was prepared in aqueous solution. (2% Alpha-D-Glucose can be used; however, it must be allowed to mutarotate overnight to the Beta configuration.) For the iodination reaction, in a tube were mixed:

0.2 M phosphate buffer pH 7.2	50 µl
Protein sample	10-25 µl (50-500 ng or 10^6 cells)
Enzymobead Reagent	50 µl
1.0 mCi Na ^{125}I	2 µl
1% β-D-Glucose	25 µl

Iodination was allowed to proceed at room temperature for 15-25 minutes with occasional mixing by gently tapping the tube. The reaction was terminated by the addition of 10 mls of PBSA at 4°C to eliminate excess iodine. Cells were washed twice in cold PBSA and centrifuged for 5 minutes at 1500 K. The PBSA was decanted and the tube inverted and blotted. Cells were lysed at 10^7 cells/ml in 1% NP-40 lysis buffer (20 mM Tris-base, 150 mM

NaCl, 0.5 mM ethylenediamine-tetraacetic acid (EDTA), 1 mM phenylmethylsulfonylfluoride (PMSF)) for 30 minutes. Following lysis, lysates were microfuged at 20,000 K at 4°C for 15 minutes to pellet the nuclei and the enzymobeads. The supernatant was pre-cleared overnight at 4°C with 100 µl of a 50% solution of agarose beads and 40 µg/ml of goat anti-mouse IgG antibodies which reduced the number of Fc receptors as well as non-specifically labelled proteins. Beads were spun down and lysate removed the next day. 20 µg/ml of mouse immunoglobulin as a negative control or 60 µg/ml of Fc-CD33 was added to 500 µl of lysate containing 5 x 10^6 cells and incubated for 1 hour at 4°C. An excess of Fc-adhesin was sometimes added due to the established low-avidity binding of Fc-chimeras relative to antibodies (Simmons 1993). Protein A sepharose was then added to precipitatate the antibodies for 30 minutes. The beads were washed two to three times in 1 ml of 1% NP-40 lysis buffer by vortexing for 20 seconds, brief microcentrifugation at 4°C, and removal of the lysis buffer. The final traces of lysis buffer were removed with a Hamilton syringe. The beads were resuspended in 20 µl of 1x sample buffer and boiled for 5 minutes at 100°C to release proteins bound to the beads. Samples were run under reducing conditions by adding 10% β-mercaptoethanol and loaded onto 10% denaturing poyacrylamide gels. SDS-PAGE was carried out for 1 hour at 120 mV. Gels were stained and destained in 10% methanol fixative solution, dried for 1 hour and subjected to autoradiography at -80°C.

2.4.11.2 Whole cell biotinylation

An alternative method to label cell surface molecules was recently developed which allows labelling of the protein with biotin and detection using a biotin-

streptavidin antibody detected with the enhanced chemiluminescence system (Cole et al. 1987). The NHS-LC-biotin was stored frozen at -20°C. To label suspension cells, 10^7 cells washed in PBSA were resuspended in 1 ml of a freshly prepared biotinylation solution (20 mM $NaHCO_3$, 150 mM NaCl, 0.1 mg sulfosuccinimidyl-6-biotinamidohexanoate (NHS-LC-biotin; Pierce, Warner, UK) (pH 8.0), vortexed, and allowed to mix for 30 minutes at RT and and then for 90 minutes at 4°C. To label transfected COS cells in situ, 4 mls biotin solution/10 cm dish was added either on ice for 30 minutes to 1 hour with shaking or at 25°C for 20 minutes followed by incubation on ice with shaking for 30 minutes to an hour. Residual biotin was removed by washing twice in 10 mls PBSA and either spinning suspension cells at 1200 g for 5 minutes or aspirating the supernatant from adherent cells. Biotinylated cells were lysed to a final concentration of 25×10^6 cells/ml in 1% NP-40 lysis buffer and used directly in immunoprecipitation using normal or stringent washing conditions. The lysates were precleared with 100 µl of beads for 1 hour at 4°C , prior to incubation with the polyclonal antibody, and left to incubate 1 hour at 4°C. For polyclonal antibodies and Fc-adhesins, the beads were washed three or four times with 1% NP-40 lysis buffer or high and low salt washes by vortexing and centrifugation at 15,000 g for 10 seconds, resuspended in 30 µl of sample buffer, and boiled for 10 minutes. Minigel and immunoblotting were performed according to the western blotting procedure. Blots were washed, blocked, probed with streptavidin horseradish peroxidase (1/2000 dilution), and ECL detected using the ECL kit (Amersham), with film exposures from 5 seconds to 15 minutes.

2.4.12 Metabolic labelling of Fc-chimeras or cell lines

Recombinant Fc-adhesins were labelled metabolically by labelling the COS cell transfectants. 24 hours post-transfection, the medium in 10 cm diameter dishes of transfected COS cells was replaced with 3 mls of methionine and cysteine-free DMEM/1% FCS and a mixture of ^{35}S Methionine and Cysteine "Translabel" (New England Nuclear products, U.K.) added to a final concentration of 50 mCi/ml. Following 6 - 18 hours (overnight) incubation the supernatants were separated from cellular debris and the protein was purified on a protein A sepharose column and run out on SDS-PAGE, as above.

2.4.13 SDS-PAGE analysis of proteins.

10%, 12% or 5-20% (gradient) polyacrylamide gels were prepared as previously described (Laemmli 1970). Sodium dodecyl sulfate-polycrylamide gel electrophoreses (SDS-PAGE) was performed rapidly at a constant voltage of 120mV on the Hoefer minigel apparatus or a large BioRad (UK), Hoeffer (Germany) or ATTO (Japan) apparatus. The Biorad minigel could be run at 200mV for 1 hour. For the stacking gel 4% gels were used. The gels were made up and run in the SDS-containing buffer shown in the section on buffers and solutions below. If the gel was being run to check the quality of a batch of recombinant protein the gel was next transferred into Coomassie Blue stain for 2 hours after which the gel was destained in fixative. For immunoprecipitations of ^{32}P or ^{125}I labelled cells the gels were simply fixed for 10 minutes, dried down at 80°C under vacuum for 30 minutes and autoradiographs made at room temperature. To facilitate the detection of

^{35}S-labelled proteins after fixing the gel could be treated with a fluorescence intensifying agent (Amplify, Amersham, UK), for 10 minutes prior to drying down, to increase signal intensity. Protein molecular weight standards included: myosin-200K; *E.coli.* β-galactosidase-116K; Rabbit muscle phosphorylase b-97.5K; Bovine serum albumin (BSA)-66K; Hen egg white ovalbumin-43K; Bovine carbonic anhydrase-31K; Soybean trypsin inhibitor-21K; and Hen egg white lysozyme-14K (Bio-Rad). To check the quality and estimate the quantity of a batch of recombinant protein the gel was placed in Coomassie Blue/methanol stain with shaking for 1 hour and destained in fixative destain for 1-2 hours. Autoradiography was performed at -80°C for ^{32}P labelling and room temperature for ^{35}S-labelling.

2.4.14 Western blotting

Western blotting allowed biotinylated proteins to be detected following electrophoretic transfer to a membrane such as PVDF. The procedure for the Western Blot test is paraphrased from the standard method (Ausubel 1993). 1 liter of 1x transfer buffer was prepared which included: 6 g Tris, 28.8 g Glycine, 40% MeOH, diluted to 1 liter with stirring in H_2O and cooled to 4°C. An Immobilon or PVDF filter (Millipore) cut slightly larger than the gel was pre-set with MeOH and soaked in transfer buffer for 15 minutes. Following electrophoresis using coloured protein markers (BioRad), the gel was soaked in transfer buffer for two minutes. The filter pads were also soaked in transfer buffer. Two pieces of Whatman 3 mm paper were briefly dipped in transfer buffer and the blot apparatus was assembled into a sandwich in the order listed below using a BioRad apparatus for minigels or large gels:

Plastic black gridded blot casing
Scotchbrite pad
3 mM filter paper
Polyacrylamide Gel
Immobilon membrane (0.45 um)
3 mM filter paper
Scotchbrite pad
White grid

All the air bubbles were rolled out from the gel/membrane sandwich using a pipet. The blotting sandwich was placed in the tank with the membrane toward the anodal side and the gel toward the cathode side, consistent with the direction of transfer of protein from gel to membrane. The tank was filled with transfer buffer and transfer was initiated at 4°C overnight at 15-18 mV. Transfer could be done in one hour at 60-70 mV with an ice block inserted in the tank. Following transfer, the sandwich was disassembled and on corner cut to show orientation. To visualise transfer, the blot was incubated in PBSA/.01% Ponceau S (Sigma) for one minute and rinsed in water until the bands appeared. The filter was then blocked in blocking buffer for 60 minutes at room temperature on a shaker in 20 mls PBSA, 0.05% Tween 20 and 1% Marvel. Depending on the labelling procedure, antisera was added at a 1:100 in Blocking Buffer, ie. 20 µl antisera in 2 ml blocking buffer. The blot was placed on a shaker at room temperature for 1 hour or overnight at 4°C and washed 2x for 15 minutes or according to the enhanced chemiluminescence (ECL) (Amersham) instructions. ECL reagents and equipment were assembled and taken into the darkroom. Solution A was added to solution B, mixed and, with forceps, protein side of blot was immediately placed over 1-5 mls of solution and left for 1 minute. Excess solution was wicked from the blot by touching a corner to a paper towel. The blot was then wrapped in plastic wrap, the creases were smoothed out so that no excess moisture was

left on the outside of the wrap. The blot was then exposed for a brief period to film using a stopwatch and developed immediately.

2.4.15 N-glycanase digestion

When the N-linked glycosylation status of recombinant Fc proteins was examined deglycosylation was done with N-Glycosidase F. To increase the efficiency of this digestion the protein sample was first denatured by heating to 100°C in PBSA (pH 7.5)/ 10 mM EDTA/ 0.2% SDS. The SDS was "removed" before addition of enzyme by the addition of further PBSA/ EDTA containing NP-40 to a final concentration of 1% - NP-40 forms micelles with SDS. Ten units of N-glycosidase were used per digestion and the reaction allowed to proceed at 37°C for 4 hours before analysis by SDS-PAGE.

2.4.16 Immunofluorescent cell staining

Immunoflourescence analysis was often performed in situ on COS cell transfectants in 6 well dishes to determine whether cDNA isolation by panning had been successful or whether polyclonal antisera stained candidate transfectants. It was also done as a check on the ability of Fc-adhesins to bind ligand on both EDTA-lifted candidate transfectants from panning and cell lines. The medium was aspirated and the cells washed briefly with staining wash, PBSA.B.C./5% FCS/0.02% Sodium Azide. Primary antibody was added, diluted 1:4 in staining wash in the case of monoclonal antibody supernatants, or 1:1000 for mabs as ascites or purified antibodies at 1 mg/ml. Cells in suspension were suspended at a final concentration of 1 x 10^6 cells/ml and

analysed on a Becton Dickinson FACScan. The medium was aspirated and the cells washed in E4 medium/5% FCS. Cells were incubated briefly with 200 µg/ml goat Ig to block Fc receptors (Alexander and Sanders 1977). Primary antibody was added, diluted 1:2 in staining wash in the case of monoclonal antibody (Mab) supernatants, or 2-500 µg/ml Fc-chimera/sample and incubated on ice for 30 minutes with occasional vortexing. Cells were washed twice and supernatant was aspirated. FITC-conjugated goat anti-mouse IgG (for example), diluted 1/500 in staining wash was added to the Mab samples, and goat anti-human Ig-Fc-FITC was added to the Fc-probe samples. After a further 30 minute incubation on ice with occasional vortexing, cells were washed three times in staining wash followed by resupension in 1 ml of PBSA. Cells were analysed in the flow-cytometer immediately, without fixing in 2% formaldehyde, as the fixative has been shown to reduce immunofluorescence, particularly for weak ligand-receptor interactions. A specific reaction of antibody to an efficiently expressing COS cell transfectant resulted in intense flourescence of approximately 30 - 50 % of COS cells, demonstrating the efficiency of the DEAE-Dextran transfection method.

Fluorescence is reported eiter as percent (%) positive fluorescence which reflects the total ligands bound in a given population (ligands/sample) or median fluorescence intensity which reflects the mean number of ligands per cell, with the background values subtracted for each set of data. Percent positive fluorescence can give a more accurate representation of the data as it does not bias the results in favor of larger cells which may have more ligands. Both values are reported where appropriate.

2.4.17 ELISA to test anti-peptide antibody reactivity

To test the reactivity of polyclonal antisera against the peptide against which it was raised, the antisera was incubated on 96 well plates previously coated with the apo-4 peptides according to the following protocol. Peptides were diluted to ~ 1 µg/ml in 0.1M sodium hydrogen carbonate pH 9.6 (typically 1/1000) and 100 µl was added to each well of an IMMULON 2 96 well microtitre plate (Dynatech Ltd., Germany) and left in a wet box overnight at 4°C. The wells were flicked out and washed in 1x Tris buffered saline (TBS), pH 8.2). Wells were blocked by filling with TBS + 2% non-fat dried milk (Marvel). 1% fetal calf serum could also be used for blocking. The plate was left to stand for one hour at room temperature and washed twice with flicking in TBS buffer.

To perform the ELISA, polyclonal antibodies were diluted in TBS/Marvel/0.1% Tween 20 (TMT Buffer)/10%FCS in tenfold dilutions from 1:10 to 1:1000. 100 µl of diluted antibody was added to each well. The plate was then incubated for 30-60 minutes at room temperature. The plate was washed with flicking 3x. The second layer detection antibody, anti-rabbit Ig-HRP (DAKO, Germany) was then diluted at 1/150 in TMT buffer. 100 µl of the secondary detection antibody was then added to each well and left for 30 minutes at room temperature. The plate was flicked out and washed in TBS 3 times. 5 mls of substrate solution, ABTS (2,2′ -azinobis (3-ethylbenzthiazoline sulphonic acid - SIGMA A-188) was prepared at 0.5 mg/ml in 0.1M NaAcetate pH 5.6. 1 ml 30% hydrogen peroxide solution was added to each ml ABTS solution immediately prior to use. 100 µl substrate solution was added to each well and left for a up to five minutes at room temperature or until the

colour developed. The optical density at 406 nm was recorded in a fluorescent plate reader.

2.4.18 Affinity purification of polyclonal antibodies

To test the antibody reactivity of purified polyclonal antibodies, three affinity columns were coated with the peptides against which the antisera were raised and the purified antisera were then eluted. Columns were prepared with the standard procedure using Cyanogen Bromide (CnBr) activated protein A (Pharmacia LKB Biotechnology). CNBr-activated Sepharose 4B is supplied freeze-dried in the presence of additives which must be washed away at low pH (pH 2-3) before coupling the desired ligand. Low pH (pH 2-3) preserves the activity of the reactive groups, which otherwise hydrolyze at high pH. The peptides to be coupled were dissolved in coupling buffer, 0.1 M $NaHCO_3$, pH 8.3 containing 0.5 M NaCl. 5 ml of coupling solution was used per gram of freeze dried CnBr beads. About 5-10 mg protein per ml swollen gel is recommended. Depending on the size of peptide, the formula estimates peptide as 100x smaller than antibody. Each peptide was coupled to beads at 50 mg /ml. Antibody weight molecular density is 10 mg/ml (i.e. for 2 ml of beads - use 20 mg antibody). With peptide use molecular weight (MW). MW of antibody is 160,000.

To prepare the column, 1.5 g of CnBr beads were weighed out and placed in a 50 ml falcon tube. The beads were resuspended in 5 ml of 1 mM HCl, pH 2-3. (1:5000 dilution of stock 11.7M HCl). 1 gram of beads gives about 3.5 ml swollen gel. The gel swells immediately and the additives were immediately washed away by transferring the suspension to a sintered glass filter and washing for 15 minutes with 1mM HCl (~400 mls or 200ml/g of freeze-dried

powder). The peptide ligand (up to 20 amino acids) was dissolved in 50 mls of coupling buffer ($NaHCO_3$, (0.1 M), pH 8.3/NaCl (0.5 M), 5 ml per gram powder) and mixed with the gel in a 50 ml polypropylene (Falcon) tube with rotating end-over-end for 2 hours at room temperature or overnight at 4°C. Remove 100 µl aliquot (pre coupling) was removed to measure the OD_{280}. A magnetic stirring bar should not be used as it could dislodge the activated coupling groups. The following day 100 µl of the post-coupling eluate was removed to check the OD_{280}. The OD of Post vs Pre should be > 95%, otherwise the experiment should be repeated. If the OD was sufficiently different, the beads were transferred to a 50 ml Falcon tube, the supernatant was carefully aspirated and excess ligand was washed away by filling the tube with at least five gel volumes of coupling buffer. Any remaining active groups were blocked by incubation with Tris-HCl buffer (0.1 M, pH 8) or ethanolamine (1 M, pH 9) for 2 hours at room temperature or 16 hours at 4°C. The product with three cycles of alternating pH in a 50 ml Falcon tube, pelleted each time in a Beckman benchtop centrifuge at 1000 g with no brake for two minutes. Each cycle should consist of a wash with Sodium Acetate bufer 1 M, pH 4/NaCl, 0.5 M followed by a wash with Tris buffer, 0.1 M, pH 8/NaCl, 0.5 M. The beads were then loaded onto columns and/or stored sealed at 4°C until ready for use. The beads were never allowed to dry out after ligand coupling.

To couple the antisera to the column, the peptide-coupled beads were resuspended in ~ 20 mls of 0.1M Tris-HCl, pH 8.0/5 M NaCl. Each set of beads with bound synthetic peptide was transferred to a column (BioRad), loading a maximum of two mls and keeping the beads hydrated. The column was washed with 10 mls of 0.1 M glycine, pH 2.5 and neutralized with 10 mls or more of PBSA. The column pH was checked insure it was 8.0. Antiserum

was diluted 1:5 in PBSA, thus 20 mls aliquots of serum were diluted to 100 mls in PBSA. The diluted serum was filtered through a 0.22 µm bottle-top filter to remove debris. The antisera was placed in a bottle above the column with a tube inside leading to the lid of the column below, which had been primed by gently allowing a few drops of serum to run down the side of the tube. The flow rate was adjusted by moving the height of the bottle to about 1 drop/second and the eluate was kept. The serum was clear as expected and the elution took 2-3 hours. The column was washed with 40-50 mls of PBSA by the same flow procedure and the eluate retained. Bound antibodies were eluted by adding 10 mls of 0.1 m glycine pH 2.5 down the side of the column above the beads. The eluate was collected in 1.0 ml fractions containing 100 µl of 0.1 M Tris-HCl, pH 8.0. The column was neutralised with 0.1 M Tris-HCl, pH 8.0 followed by 10 mls of PBSA and the pH was monitored to return to 8.0. The column was capped and stored in PBSA containing 5 mM sodium azide at 4°C. The OD of each fraction of the eluate was checked at 1 = 280 nm to be > 0.7. If it was less than this, something had gone wrong in the procedure or the antibodies had not coupled to the column. The fractions of greatest optical density were pooled, usually about 3 fractions starting at around the third fraction, and placed in dialysing tube hydrated in PBSA. The dialysis tubing was secured and placed in PBSA overnight in the cold room. The following day, the antibodies were concentrated, loading two mls at a time, using Centricon 100 columns (Amicon). The columns were spun at 4500 rpm for 20 minutes or until the desired volume of ~300 µl was achieved. The OD_{280} was checked by the antibodies 1/10 to 50 µl. The OD had to be > 0.14 or preferably 0.2 - 0.5 to use the antibodies in further experiments. An OD of 1.4 = 1 mg/ml protein. The antibodies were stored at 4°C if used within a few days or -20°C or -80°C if the antibodies would not be used for at least two weeks.

2.4.19 Immunohistochemistry

Frozen histological sections embedded in paraffin were stained using the biotin-avidin horseradish peroxidase kit (DAKO, UK), with the staining performed by Margaret Jones and Helen Turley, Department of Cellular Science, John Radcliffe Hospital, Oxford, UK). The method allows detection of binding via the biotin-avidin comples (Hnatowich 1987). Cold tissue sections were incubated 30 minutes with 10 µg/ml of biotinylated Fc-chimera or an antisera and washed once in a Tris/Borate/Saline bath. 100 µl of a streptavidin-biotinylated horseradish peroxidase solution was added to sections which were then incubated 30 minutes at room temperature. Binding was detected with the addition of 100 µl DAB solution/section to detect positive binding shown by brown staining.

2.5 Standard reagents and solutions

2.5.1 Media and standard buffers

Trypsin-EDTA was supplied as a 5x stock (0.5%/10 mM) and diluted in PBSA for use. SDS-PAGE solutions included Running Buffer, 15 g Tris base (5x stock) per litre of solution, 72 g Glycine and 5 g SDS. The resolving gel preparation for a 10 ml 10% gel contained: 3 ml of 30% Acrylamide/bis (29:1) 2.5 ml 1.5 M Tris pH 8.8, 100 µl 10% SDS, 4.4 ml H_2O, 100 µl of 10% ammonium persulphate (APS), 10 µl of TEMED. A 4% stacking gel required mixing 5 mls of 1.3 mls of acrylamide stock included 4% Gel, 2.5 ml 0.5M Tris pH 6.8, 100 µl 10% SDS, 2.1 ml H_2O, 50 µl of 10% APS, 5 µl TEMED.

2.5.2 RNAse protection reagents

Diethylpyrocarbonate
5x transcription buffer
200 mM dithiothreitol (DTT)
3 NTP mix (ATP, UTP, and GTP at 4 mM each)
[α^{32}P] CTP (10 mCi/ml, 400 to 800 Ci/mmol)
Placental ribonuclease inhibitor (e.g., RNAsine from Promega Biotec)
0.5 mg/ml template DNA (support protocol)
Bacteriophage RNA polymerase
2.5 mg/ml template RNAase-free DNase I
10 mg/ml tRNA
25:24:1 phenol/chloroform/isoamyl alcohol
2.5 M ammonium acetate
100% ethanol
75% ethanol/2.5 % 0.1 M sodium acetate, pH 5.2
Hybridization buffer
Ribonuclease digestion buffer
40 mg/ml ribonuclease A
2 mg/ml ribonuclease T1
20% (w/v) sodium dodecyl sulfate (SDS)
20 mg/ml proteinase K (store at -20°C)
RNA loading buffer

Additional reagents and equipment for phenol extraction and denaturing polyacrylamide gel electrophoresis are required. Water and sodium acetate should be freshly treated with DEPC to inhibit RNAase activity and used in all reagents and solutions. DEPC is a suspected carcinogen and should be handled carefully.

5x Transcription buffer

200 mM Tris-HCl, pH 8.0

40 mM $MgCl_2$

10 mM Spermidine

250 mM NaCl

This buffer can be used for T3 and T7 RNA polymerase driven transcription reactions, but not SP6 driven reactions.

5x Hybridisation buffer

200 mM PIPES pH 6.4

2 M NaCl

5 mM EDTA

Working solution 1 part 5x hybridisation buffer to 4 parts deionised formamide

Ribonuclease digestion buffer

10 mM Tris pH 7.5

300 mM NaCl

5 mM EDTA

+ $1/50^{th}$ volume RNAase mix containing 2 mg/ml RNAase A and 0.1 mg/ml RNAase T1, added just before use

2.5.3 Primer extension reagents

Diethylpyrocarbonate (DEPC)

[γ-^{32}P] ATP (10m Ci/ml, 6000 Ci/mmol)

100 ng/ml oligonucleotide

10x polynucleotide kinase buffer

T4 polynucleotide kinase (Boehringer Mannheim, Germany)
4 M ammonium acetate
Ethanol
0.3 M sodium acetate
1x aqueous hybridisation solution
S1 hybridisation solution
75% ethanol/25% 0.1 M sodium acetate, pH 5.2
4 mM 4dNTP mix (Pharmacia)
10x reverse transcriptase (RT) buffer
Placental ribonuclease inhibitor (e.g., RNAsin from Promega Biotec)
AMV reverse transcriptase
0.5 M EDTA
1 mg/ml pancreatic ribonuclease A
2.5 M ammonium acetate
25:24:1 phenol/chloroform/isoamyl alcohol
70% ethanol/30% DEPC-treated H_2O
T:E buffer
Formamide loading buffer

2.5.4 Immunoprecipitation - stringent wash method

Lysis buffer:	1-3% non-ionic detergent (eg 1% NP40)
	20 mM Tris pH 7.5 or 8.0
	150 mM NaCl
	5 mM EDTA
	2.5 mM iodoacetamide
	200 µg/ml soybean trypsin inhibitor
	1 µg/ml pepstatin
	0.5 µg/ml leupeptin
	0.2 µg/ml aprotinin
	1.5 mM PMSF
Buffer A:	10 mM Tris pH 8.0, 500 mM NaCl, 0.5% NP40, 0.5% SDS
Buffer B:	10 mM Tris pH 8.0, 150 mM NaCl, 0.5% NP40, 0.5% DOC 0.05%SDS
Buffer C:	10 mM Tris pH 8.0, 0.05% SDS

Depending on primary antibody species and class, prepare a 20% solution (vol/vol) of protein A agarose or protein G-agarose in PBS.

2.5.5 Standard immunoprecipitation buffer

1% Detergent (NP-40,TX-100, etc.)

20 mM Tris (pH 8-8.3) (Tris-Base)

150 mM NaCl

5 mM EDTA

Phosphatase Inhibitors

0.4 mM Sodium Vanadate

0.5 mM EDTA

10 mM NaF

10 mM NaPPO

2.5.6 Cell culture

PBSA was supplied from the ICRF Central Services Facility according to the following formulation:

137 mM NaCl

2.7 m Cl

8.1 mM Na_2HP

1.5 mM KH_2PO_4 (titrated to pH 7)

Trypsin-EDTA was supplied as a 5x stock (0.5% / 10 mM) and diluted in PBSA for use.

2.5.7 Bacterial culture solutions

TYM broth (per Litre of solution)

20g Bacto-Tryptone

5g Yeast extract

100 mls 1M NaCl

10 mls 1M $MgSO_4$

Autoclave the mixture before use. It can be kept indefinitely at room temperature

Transformation buffer I

per 500 mls of solution

4.95g $MnCl_2$

3.73g KCl

0.73g $CaCl_2$

75 mls glycerol

0.45 μm filtered (not autoclaved) and stored at 4°C

Transformation buffer II

per 500 mls of solution

1.05g Sodium MOPS

0.37g KCl

75 mls glycerol

pH 7.0, 0.45m filtered and stored at 4°C in 50 ml aliquots

Antibiotic stocks

Ampicillin and Spectinomycin made up as 100 mg/ml in water and stored at -20°C. Tetracycline made up as a stock solution of 30 mg/ml in 50% ethanol and stored at -20 °C.

2.5.8 Molecular biology solutions and standard buffers

Tris-equilibrated phenol

Distilled, water free phenol was tris-equilibrated before use by adding several cycles of 200 mls of 1M Tris pH 8.0 to 1 litre of Phenol, shaking vigorously, allowing to settle and aspirating the aqueous phase. The cycles were repeated until the aspirated aqueous phase was pH 8.0. A small amount of aqueous phase must be left behind for storage in a sealed container at 4°C. Before storing 0.4 g of 2-hydroxy quinoline was added to 1 litre of equilibrated Phenol as an anti-oxidant.

Standard Buffers

Stock solutions of filtered 1M Tris-HCl (pH 7.0, 7.5, 8.0, 9.0), 0.5M NaEDTA (pH 8.0), 1M NaCl, 1M KCl, 1M $MgSO_4$, 5M NaOH, 3M NaOAc, 1M Dithiothreitol (DTT), 100 mM ATP, 1M Spermidine and 1M NH_4OAc were maintained. BSA at 10 mg/ml was purchased as a molecular biology grade reagent from Boehringer and stored at -20°C.

T:E		
Standard	High	S:T:E
10 mM Tris (pH 8.0)	20 mM Tris (pH 8.0)	10 mM Tris (pH 8.0)
1 mM EDTA	10 mM EDTA	1 mM EDTA/100 mM NaCl

10x T:B:E

Per litre of solution

108 g Tris base

55 g Sodium Borate

9.3 g Na EDTA (do not pH solution)

Working solution is 0.089M Tris-borate, 0.025 M EDTA (pH 8.3).

50x T:A:E

Per litre of solution

242 g Tris base

57.1 ml glacial acetic acid

100 ml 0.5M EDTA (pH 8.0)

Working solution is Tris-acetate 0.04 M, 0.001 M EDTA and Ethidium Bromide stock solution (10 mg/ml) added to a final concentration of 20 mg/ml.

5x MOPS

per litre of solution

41.8 g Sodium MOPS

16 mls 3M Sodium Acetate

10 mls EDTA

This solution is used for running RNA denaturing formaldehyde gels. MOPS cannot however be DEPC treated (risk of explosion) and therefore must be made up with DEPC-treated water from a clean RNAase free stock. Working solution is 0.04m MOPS pH 7.0, 0.01 M acetate, 0.001M EDTA. 1/60th volume of 37% formaldehyde is added to the gel just before pouring.

DNA sample buffer

per 50 mls

20% Ficoll

0.1g of Orange G dye

in 1x T:A:E /10 mg/ml ethidium bromide

RNA sample buffer

5 µl of RNA sample

2 µl of 5x MOPS buffer

10 µl deionised formamide

3.5 µl of 37% formaldehyde (stock)

and heat to 70°C for 5 minutes

RNA extraction solutions

4M Guanidinium Thiocyanate buffer. For one litre of solution, bring to 7 mls with 3M Sodium Acetate (0.45 µm filtered and stored at 4°C).

Cesium Chloride solution

(5.7M) per 100 mls

126 g of optical grade CsCl solution

2 mls 0.5M EDTA

(DEPC treated and stored in 20 ml aliquots at room temperature)

Church and Gilbert buffer

For one litre of solution, add the following to a small amount of dH_20 with a stirring bar and bring to one litre with dH_20:

27 g Sodium diHydrogen Phosphate solution

83 g diSodium Hydrogen Phosphate

10 mls 0.5M EDTA (pH 8.0)

Working solution is 2 parts of the above phosphate buffer to 1 part of 10% SDS, mixed immediately before use (0.77M sodium Phosphate, 5mM EDTA, pH 8.0)

Plasmid Extraction Solutions

Solution I	per litre of solution	20 mls of 0.5m EDTA (pH 8.0)
Solution II	per litre of Solution	40 mls of 5M NaOH
		50 mls of 20% SDS
	Working solution is 0.2M NaOH, 1% SDS	
Solution III	per litre of solution	200 g Potassium Acetate
		80 mls Glacial Acetic Acid
	Working solution is 5M Potassium Acetate (pH 4.7)	

cDNA synthesis and ligation buffers

RT1 buffer (5x)	0.25 M Tris pH 8.2 (pH 8.8 at 42°C)
	0.25 M KCl
	30 mM $MgCl_2$

RT2 buffer 0.1 M Tris pH 7.5

25 mM $MgCl_2$

0.5 M KCl

0.25 mg/ml BSA

50 mM DTT

10x Low Salt Buffer

60 mM Tris pH 7.5

60 mM $MgCl_2$

50 mM NaCl

70 mM β-mercaptoethanol

2.5 mg/ml BSA

Ligation additions (10x)

1 mM ATP

20 mM DTT

10 mM Spermidine

1 mg/ml BSA

10x Kinase Buffer for labelling probes

0.5 m Tris pH 7.5

10 mM ATP

20 mM DTT

10 mM Spermidine

100 mM $MgCl_2$

1 mg/ml BSA

Linear Polyacrylamide (LPA) Carrier solution

A 5% solution of polyacrylamide was made in water and 0.5 mls of 10% Ammonium Persulphate and 50 µl TEMED were added and left for 1

hour at room temperature. The solution was dilute 1:25 (to 2 mg/ml) for a working solution and aliquoted and store at -20°C.

10x PCR buffer

100 mM Tris pH 8.3 (at 20°C)

500 mM KCl

2 mM $MgCl_2$

0.01% gelatin

DNA sequencing solutions

5x Annealing Buffer

200 mM Tris pH 7.5

100 mM $MgCl_2$

250 mM NaCl

Labelling Mix

2 parts of 1.5 mM each of dGTP, dCTP and dTTP

2 parts of T:E

1 part of 0.1M DTT

0.5 µl of ^{35}S-dATP (Amersham, UK)

1 µl of 0.5% NP-40 per 50 µl labelling cocktail

1µl of T7 DNA polymerase (Sequenase, US Biochemicals) added per 50 ul of labelling mix

Termination nucleotide mixtures

One for each dideoxy-nucleotide containing 80 mM of each dNTP (dATP, dGTP, dCTP and dTTP) and 50 mM NaCl and one containing 8 mm dideoxy-ATP, one containing 80 mM dideoxy-GTP and so on.

Stop solution (Loading Buffer)

95% Formamide

20 mM EDTA

0.05% each of bromophenol blue and Xylene Cyanol

Transfection Solutions

DEAE-Dextran solution

10 mg/ml of DEAE-Dextran (Sigma) was filtered through a 0.45 μm filter and stored at 4°C.

Chloroquine Solution (1000x)

A 100 mM stock was made in distilled water and 1 ml aliquots were stored at -20°C. Chloroquine deteriorates after thawing.

PBSA-10% DMSO

To PBSA add 10% (v/v) DMSO and store at room temperature for up to one week.

PEG solution

PEG-1000 was melted in the microwave for 30 seconds on high power to prevent superheating and aliquoted into 50 ml tubes

which were then frozen. To make PEG solution a 50 ml aliquot is melted by microwaving and added to an equal volume of DMEM. The pH can be titrated back to neutrality with a few drops of 5M NaOH and judged by the colour of the phenol red indicator or with a piece of pH paper. The solution is then filtered through a 0.45 μm filter and stored at 4°C.

Tris and EDTA solutions

For transfection these were made either fresh or by diluting stocks, and filtering through a 0.45 μm filter followed by storage at 4°C.

Protein Electrophoresis Buffer

Running Buffer	15 g Tris base
(5x stock) per litre of solution	72 g Glycine
	5 g SDS

Chapter 3

Fc-CD33 and Fc-CD34 Cloning and Binding Studies

3.0 Specific aims

The aim of this chapter is to describe the rationale for constructing Fc-CD33 and Fc-CD34, the method of their construction, and some of the subsequent binding studies in which they were employed. The emphasis of this chapter is on Flow cytometry studies, adhesion assays and immunohistochemistry comparing several Fc-chimeras. Surface labelling with Fc-CD33 is also included. Panning data on Fc-CD33 and Fc-CD34 are emphasised in subsequent chapters.

3.1 Vectors used in Fc construction

Several vectors have been used in the construction of fusion proteins including: pBR322 (Zettlmeissl 1990) which required a splice donor to be engineered in the insert prior to ligation into the vector; pHT4-YK12 in which human ECD/mouse IgG1 was inserted apparently without a splice donor (Traunecker 1989a); pSV2 dhfr (dihydrofolate reductase) and pRMH for CD4-hinge-CH2-CH3 proteins with a synthetic splice donor engineered into the protein (Gregersen 1990); the CDM8 based expression vectors which incorporated a synthetic splice donor sequence to allow the reading frame to be preserved between CD4 and the various Ig exons to which it was fused (Zettlmeissl 1990); and the CD5-IgG1 (Rg) vector used to make Fc-CD44 which required an endogenous splice donor sequence in the ECD DNA to achieve correct splicing (Aruffo 1990). A subsequent version of Rg (Rg_2) included an endogenous splice donor sequence at the 3′ end of CH1 which would splice into the 5' hinge splice acceptor region (Hollenbaugh et al. 1992). An endogenous splice donor sequence allowed secreted Fc proteins to dimerise

consistently, as the previous version of Rg could produce a mixed population of monomeric and dimeric proteins.

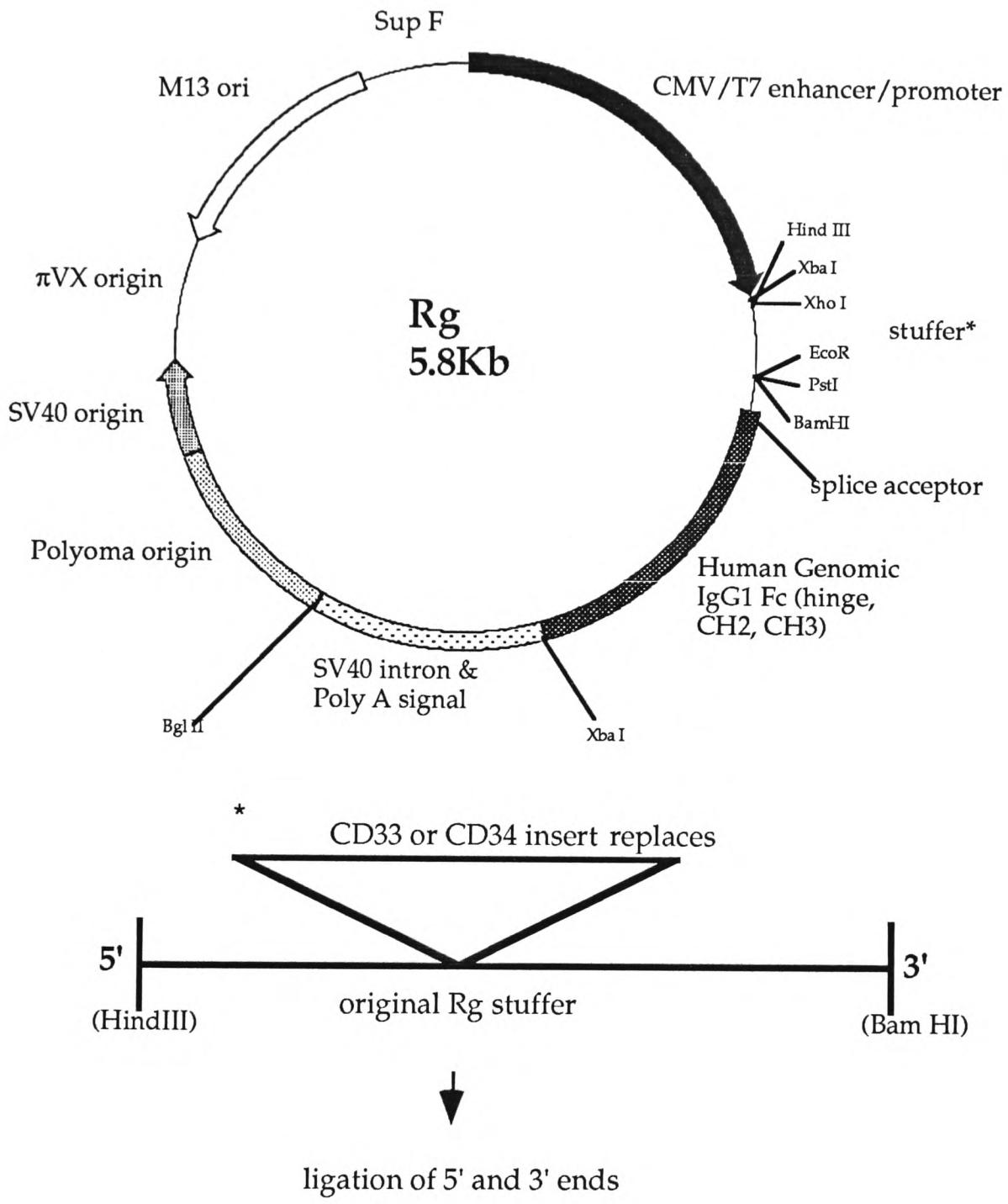


Figure 3.1 The CDM8-Fc (Rg) Vector. This vector was used in the initial construction of Fc-CD33 and Fc-CD34. The Rg vector retains all of the elements of CDM8 with the addition of a genomic fragment of IgG1 Fc at the 3' end of the insert to allow for the expression of Fc proteins in COS cells. The commonly used restriction enzymes for inserting ECD fragments are in parentheses.

The pIG1 vector was designed as an alternate to the CDM8-Fc (Rg) vector and is similar to it apart from the Not I restriction site immediately 3' of the IgG region in place of the Xba I site in Rg, restriction sites from the pBluescript vector and a slightly longer intron 5' to the hinge region to allow for more efficient splicing between CH1 and the hinge region. A splice donor motif is created following insert ligation just upstream of the Bam HI site when a 10 bp splice donor motif is included in the PCR primer. The endogenous splice donor motif allows for efficient splicing and more consistent dimer formation of soluble proteins. A potential disadvantage of a pure dimeric population is that homomultimers or aggregates are more likely to form over time which could lead to artefactual results, particularly in binding assays. The binding studies which follow all used Fc-protein which was less than two weeks old.

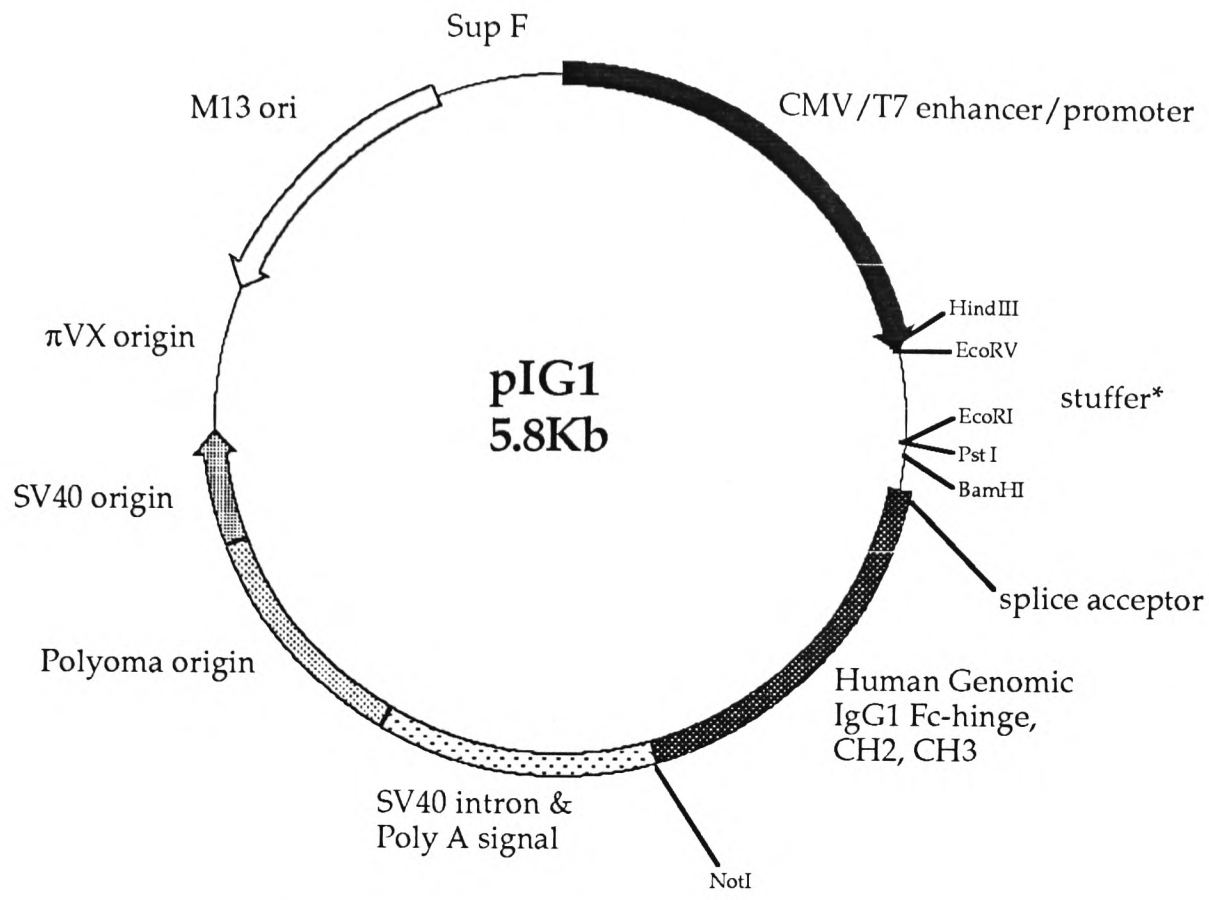


Figure 3.2 The pIG1 vector. The schematic diagram shows the essential features of the vector used to construct Fc-CD33 and Fc-CD34 ECD chimeras using the pIG1 vector. Like Rg, the vector retains all the elements of CDM8 but includes pBluescript restriction sites and requires a splice donor to be engineered into the 3' end of the PCR product for insertion into the vector. Note the replacement of the Xba I site with the Not I site at the end of the IgG1 fragment, relative to Rg.

Prior to deciding on the Forward Amplification Primer (FAMP) design, the potential strength of each ECD leader sequence to express protein was evaluated

based on its sequence. If the ECD contained a strong leader sequence, as did CD33 and CD34, the FAMP was designed to include the appropriate restriction site adjacent to the 5' nucleotide of the leader sequence.

3.2 CD33 and CD34 as immunoadhesins

This study represents one of the first attempts to use an ECD-Fc construct like an antibody probe to identify an unknown ligand, and the first attempt with Fc-CD33 and Fc-CD34. Previously isolated ligands employing the panning technique had used high affinity antibodies as the ligand probe. The construction of Fc-CD33 and Fc-CD34 in Rg was followed by the construction of both Fc-CD33 and Fc-CD34 in pIG1.

Initially, both CD33 and CD34 ECD cDNA fragments were produced by PCR and ligated into the Rg vector in collaboration with David Simmons. I performed the transfections and protein expression for all subsequent binding experiments with both Fc-CD33Rg and Fc-CD34Rg constructs, and subsequent pIG1 constructs. Thus, PCR primers for Fc-CD33Rg are not shown. Following construction of the Fc-CD33Rg, the CD33 ECD DNA was sequenced to check the clone for possible errors introduced in PCR with the following single-stranded oligonucleotide primers (Table 3.1).

Table 3.1 CD33 oligonucleotide primers

No.	Position	Oligonucleotide (5′ - 3′)
1.	@1	CTG CTT CCT CAG ACA TG
2.	@180	TTC ATG GTT ACT GGT TC
3.	@300	CCA GTA GGA ACA ACT GC
4.	@490	CCT TAC CTG CTC TGT GTC
5.	@640	ACC TGA CCT GTC AGG TGA AG
6.	CDM8-FSP	GAC GAA TTG ACC GTT ATC G
7.	CDM8-RSP	CAT AAT ATG TCC ACC AAC TTA TC

3.2.1 Fc-CD33 and Fc-CD34 constructs in the Rg vector

The CD33 ECD includes a V and a C2 domain while the CD34 ECD includes a long polypeptide with many potential O-linked sites followed by an Ig-like domain (reading from the N-terminus). The primers listed below were used to amplify CD34 ECD fragments via PCR from the full length DNA in CDM8 for insertion into the Rg vector (Table 3.2). In the Forward Amplification Primer (FAMP), a HindIII restriction site was included preceded by the appropriate overhang to allow for restriction enzyme digestion immediately following PCR. The Reverse Amplification Primer (RAMP) included a Bgl II restriction site (to allow for digestion and ligation into the linearised Rg vector).

Table 3.2 Oligonucleotide Primers Used for Fc-CD34Rg PCR (5' - 3')

CD34 FAMP+HindIII	ATCT AAGCTT GTC TCT CTA GGA GCC GCG CGG
CD34 RAMP+Bgl II	ATCT AGATCT TCT TTT GGG AAT AGC TCT GGT GGC

The CD34 ECD 895 bp PCR fragment (Figure 3.3a) was cut with HindIII and Bgl II and the cut product precipitated and run out on a low melt gel. The purified fragment was cut out from the gel and ligated into an Rg plasmid linearised with Hind III and Bam HI, as Bgl II and Bam HI produce compatible ends when digested. Following ligation, the ligation mix was heat shocked into MC1061/p3 cells and spread on amp/tet agar plates to examine colony formation. Selected colonies were prepared according to the small-scale preparation of DNA protocol and examined for inserts following digestion with Hind III and Pst I. The expected size of the insert from the 5' end of the signal sequence to the Pst I site at the 5' end of the hinge DNA was 1255 bp. The CD34 ECD minipreps following ligation in Rg reflect the appropriate size for the CD34 ECD (Figure 3.3b).

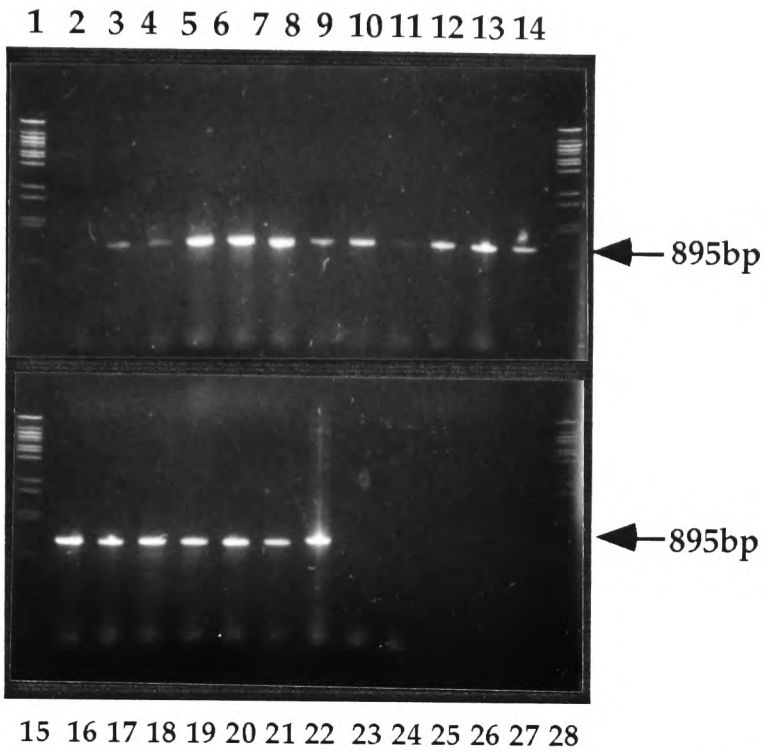


Figure 3.3a CD34 ECD PCR products on 1% agarose gels. Lanes 1,14, 15 and 28 show size markers; Lane 2 shows CD34 5' forward and 3' reverse primers alone and lanes 3-22 show the amplified ECD of CD34, indicated by the arrows, following 30 cycles of amplification. Lanes 23-24 are primers only and Lanes 25-27 are empty.

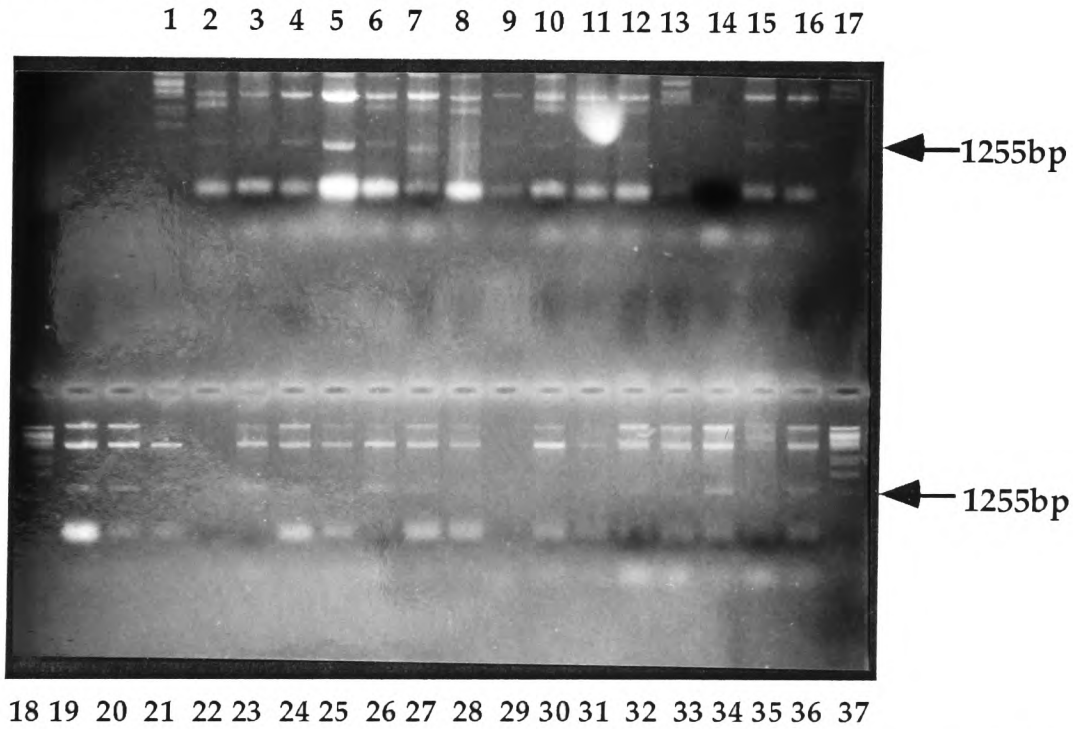


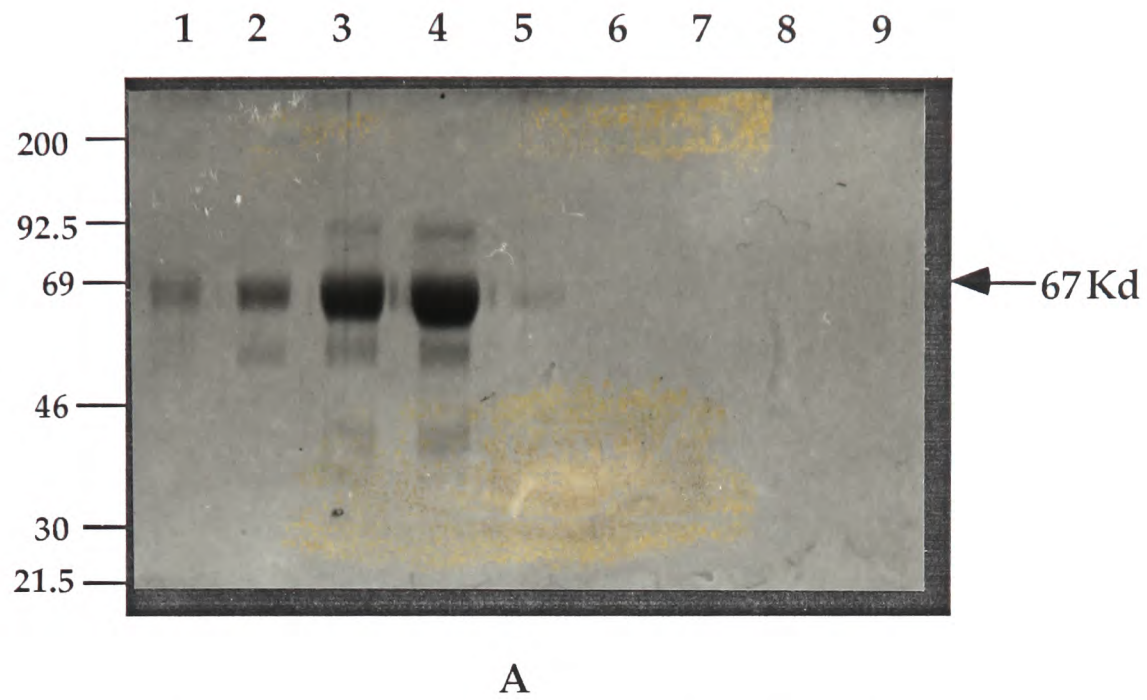
Figure 3.3b. Mini-prep DNA of CD34 ECD inserted into Rg. Lanes 1,17, 18 and 37 are DNA markers. Remaining numbered lanes represent plasmid DNA cut with Hind III and Pst I to release the insert, labelled with arrows, from the vector.

The CD34 ECD sequence was checked for accuracy following PCR using primers at various points along the extracellular domain (Table 3.3).

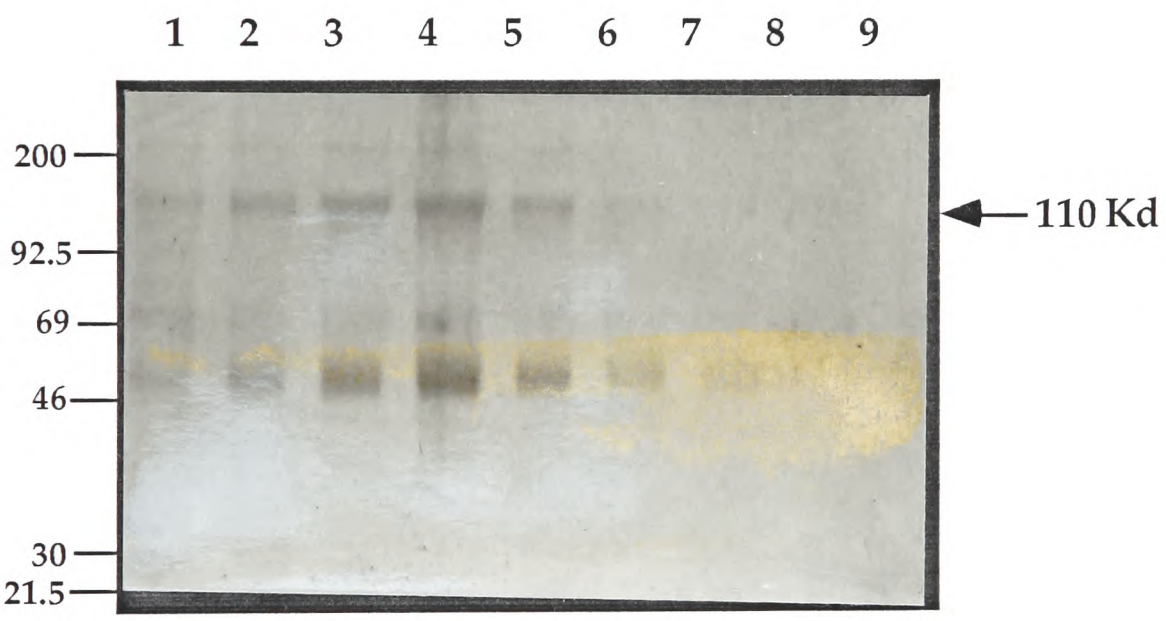
Table 3.3 CD34 primers

No.	Position	Oligonucleotide (5′ - 3′)
1.	@1	GTC TCT CTA GGA GCC
2.	@52	GAG GGC CCA GGA TGC
3.	@183	CTA CAA ATG TAT CCT AC
4.	@421	AAC CTT GAA GCC TAG
5.	@766 (Jn primer)	GGC CAA CAG AAC AGA AAT TTC

After checking the Fc constructs for accuracy, Fc-CD33 and Fc-CD34 were transfected into COS cells. Soluble Fc proteins were harvested from the media, purified from protein A Sepharose columns and concentrated using Centricon filters (Aruffo 1990; Simmons 1992). The A_{280} was checked prior to using the purified Fc proteins in an assay. The reduced vs. nonreduced fractions were then compared with molecular weight standards using SDS-PAGE to ensure that the sizes of the soluble Fc proteins were accurate.



A



B

Figure 3.4. SDS-PAGE Analysis of fractions of Fc-CD33 and Fc-CD34 in Rg. Arrows indicate Fc-CD33 (A) and Fc-CD34 (B). Sequential fractions of protein from the media of 10 15cm dishes of transfected COS cells. Fractions were eluted from a 1.0 ml protein A sepharose column. Lanes 1-9 represent protein recovered in 500 µl aliquots of 1M glycine (pH 3) eluted into 100 µl of 1M Tris. 5 µl of eluted protein was loaded into each lane of the 10% SDS-PAGE gel. Fractions were reduced with 10% β-ME prior to loading. The yield/10 15 cm dishes was approximately 0.5 mg protein.

The light chain expression shown at 15-25 Kd (Figure 3.4A) and heavy chain at about 50 Kd (Figure 3.4A and B) is due to the elution of FCS Ig at a slightly slower rate than the Fc protein fractions. Subsequent to the binding studies in this chapter, greater Fc purity was obtained by first depleting the FCS of immunoglobulin by preincubating it with protein A sepharose before adding it to the COS cells the day following transfection (suggested by Paul Crocker). As predicted, the ECD of CD33 is migrating at about 67 Kd (Figure 3.4A) and the CD34 ECD at about 110 Kd (Figure 3.4B). The presence of immunoglobulin did not interfere with CD33 and CD34 binding in subsequent assays and may even have provided an additional blocking function in cells expressing Fc receptors.

To assess the dimerisation capability of Fc-CD33, the chimeric protein was run under nonreducing (NR) and reducing (R) conditions (with β-mercaptoethanol). The Fc-protein was demonstrated to form dimers as shown by halving its ECD weight from 92 Kd (NR) to 46 Kd (R) upon cleaving the disulfide bonds in the dimeric form (Figure 3.5). The ^{35}S-labelled R and NR fractions show three bands, the bottom two of which likely represent incompletely glycosylated immature forms of CD33 still in the cytosol (Figure 3.5 NR, Lane 2 and R, Lane 2). The band that runs at 67 Kd in the NR gel could also represent a population of monomeric Fc protein.

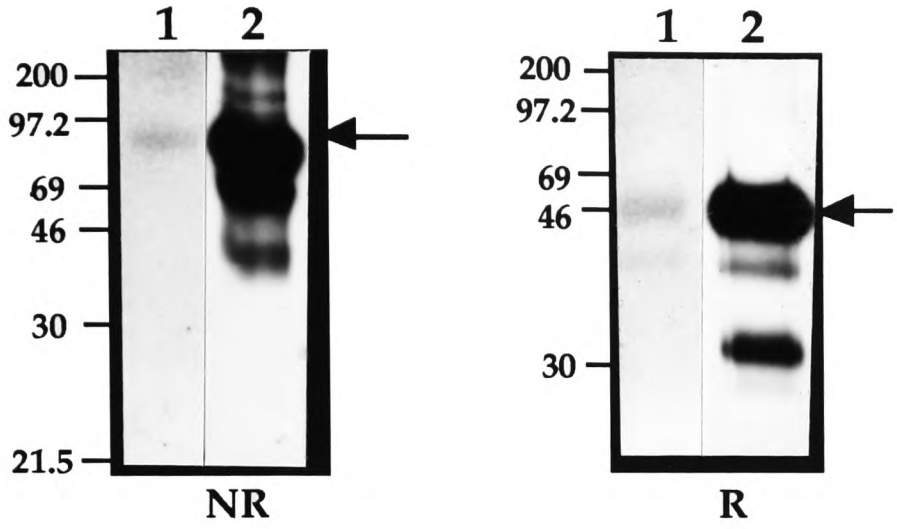


Figure 3.5 ^{35}S-Labelled Fc-CD33Rg under nonreduced (NR) and reduced (R) conditions. Both gels are 10% SDS-PAGE. (NR) and (R): Lanes 1 - Fc-CD33 unlabelled; Lane 2 - Fc-CD33 pulsed with ^{35}S-Methionine following transfection runs at 92 Kd in the NR gel and 46 Kd in the R gel, demonstrating that dimerisation occurs. All fractions were purified on protein A sepharose columns. Smaller ^{35}S-labelled bands in Lane 2 likely represent partially glycosylated forms of Fc-CD33. In both gels the arrow points to the fully glycosylated form of Fc-CD33.

3.2.2 N-Glycanase digestion of CD33

N-Glycanase was also used to determine whether or not Fc-CD33 was undergoing glycosylation in the COS cell (Figure 3.6). Although this enzyme digests only N-linked sugars, it provides a useful tool to monitor the extent of glycosylation on the extracellular domains of Fc-chimeric molecules. An important concern was whether COS cells were able to glycosylate overexpressed proteins from transfected cDNAs in the same conformation as native protein as COS cells were thought to lack certain sialyltransferases. The shift to a lower molecular weight shown by the N-glycanase digestion of Fc-CD33 demonstrated that it was likely that all of the N-linked sites were being utilised in the glycosylation of the

protein (Figure 3.6). Each glycan is roughly equivalent to 3 Kd and CD33 has 5 N-linked sites/molecule. The deglycosylated form of CD33 showed a shift of about 15 Kd on SDS-PAGE reflecting the probability of accurate glycosylation. Large quantities of purified protein would allow a more precise answer to this question.

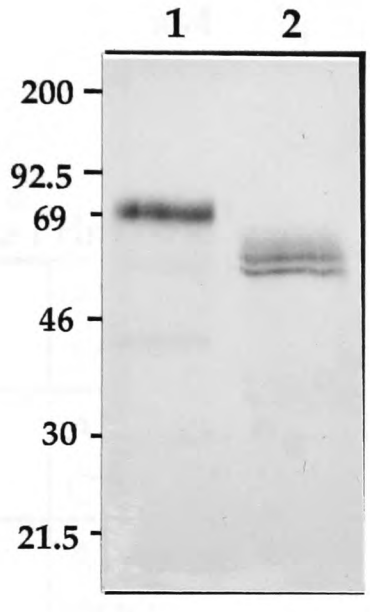


Figure 3.6 N-Glycanase Digestion of Fc-CD33 demonstrates that all of the correct glycosylation sites are being used. The 10% SDS-PAGE gel shown was run under reducing conditions. Lane 1- native Fc-CD33Rg without N-glycanase; Lane 2-Fc-CD33 deglycosylated with N-glycanase. Deglycosylation of Fc-CD33 demonstrates that the protein is being glycosylated with N-linked sugars before export from the COS cell. This analysis shows that CD33 is being glycosylated to approximately physiological levels, although the sugars could potentially be in a different conformation.

3.2.3 CD33 and CD34 extracellular domain PCR products

The primers listed below were used to amplify CD33 and CD34 ECD fragments via PCR for insertion into the pIG1 vector (Table 3.4). In the Forward Amplification Primer (FAMP), a Hind III restriction site was included preceded

by the appropriate overhang to allow for restriction enzyme digestion immediately following PCR. The Reverse Amplification Primer (RAMP) includes a Bgl II site followed by a splice donor motif (5' ACTTACCTGT 3') which will splice directly into the splice acceptor site engineered into the intron 3' of the hinge to create an in-frame splice junction for protein processing. The expected size for the CD33 ECD was 776 bp and 895 bp for the CD34 ECD (Figure 3.7a).

Table 3.4 Oligonucleotide Primers Used for PCR (pIG1) (5' - 3')

CD33 FAMP+HindIII	ATCT AAGCTT ACA TGC CGC TGC TGC TAC TGC
CD33 splice RAMP+BglII	ATCT AGATCT ACTTACCTGT GAA CCA GTC CTG CTC TGG
CD34 FAMP+HindIII	ATCT AAGCTT GTC TCT CTA GGA GCC GCG CGG
CD34 splice RAMP+BglII	ATCT AGATCT ACTTACCTGT TCT TTT GGG AAT AGC TCT GGT GGC
Universal Reverse Junction Sequencing Primer	CTG CCC TGG ACT GGG GCT GC

Following PCR, the product was gel purified, excised, cleaned, digested with the appropriate enzymes, precipitated and ligated into the pIG1 vector.

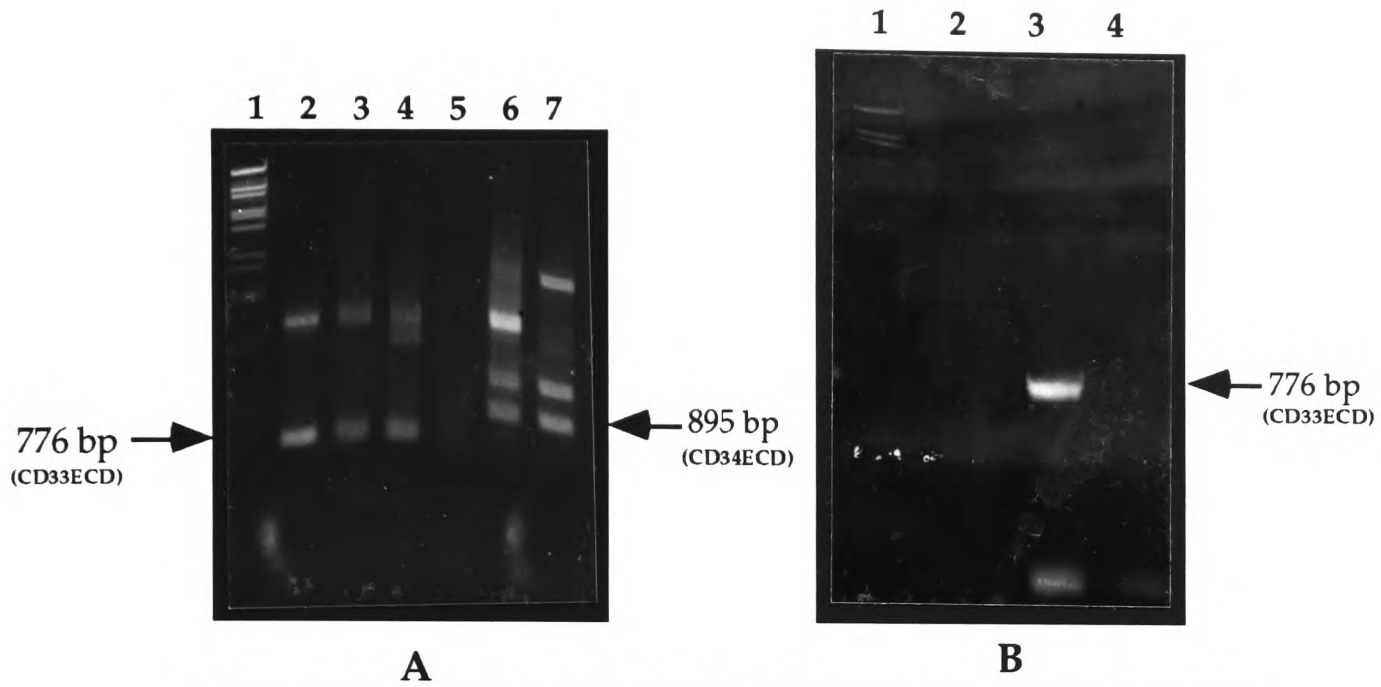


Figure 3.7A: CD33 and CD34 ECD PCR products. Products appear on a 1% agarose gel to check for size and purity. Lane 1 on A and B, DNA size markers. On A and B the forward primer used was CDM8 forward sequencing primer (FSP) and the appropriate splice Reverse Amplification Primer (RAMP) which terminated at the respective 3' region of the ECD for each insert. Templates include on **A**: Lane 2, CD33 plasmid (pCD33) in CDM8 (10ng); Lane 3, Fc-CD33 in Rg; Lane 4, pCD33 (100ng); Lane 5, space; Lane 6, Fc-CD34 in Rg; Lane 7, pCD34 (100ng). **B:** Lane 3, pCD33 (10ng); Lane 4, FSP and splice RAMP alone. The ECD products following 30 cycles of amplification are marked with arrows.

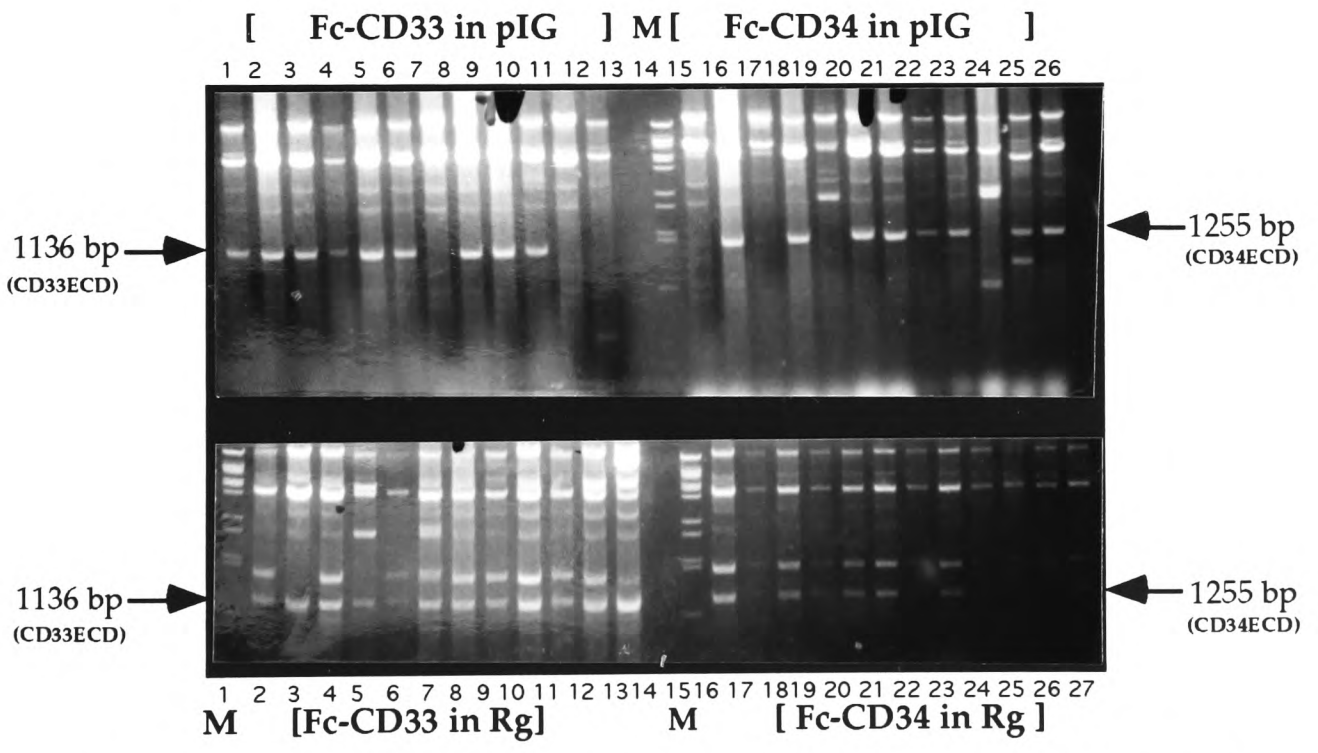


Figure 3.7B: Comparison of Mini-prep DNA of CD33 and CD34 ECDs inserted into pIG and Rg. Lanes containing λ BstXI DNA markers are labelled with an "M". Brackets denote respective minipreps. Rg and pIG plasmid DNA were cut with HindIII and Pst I to release the insert, labelled with arrows, from the vector. Rg minipreps show an extra fragment at about 900 bp which could be a replicon which arose during recombination. The insert sizes are 360 bp larger than the PCR products to account for the extra IgG1 fragment before the Pst I site.

Miniprep DNA containing inserts for CD33 and CD34 was then isolated (Figure 3.7 B). An extra 360 bp was added to the size of the initial CD33 and CD34 PCR product inserts cut out with Hind III and Pst I as this is the length of genomic DNA from the Bam HI site to the Pst I site at the 3' CH1 intron/5' hinge border in pIG1. An unexpected fragment of about 900 bp occured in both Fc-CD33 and Fc-CD34Rg and one in Fc-CD34 pIG1 miniprep DNA (Figure 3.7B, lower half and Lane 25, upper half) which is likely a replicon. A replicon is part of the original vector which can split off during recombination and delete the insert or outstrip the growth of stably integrated inserts resulting in a mixed population in a single E.Coli. colony. This effect can result from instability of the genome in MC1061/p3 cells during recombination. A replicon could inhibit protein production from transfected DNA which could account for the lack of protein produced from the above minipreps. At the time of performing the experiment, the extra fragment was not known to be a potential replicon.

Several representative clones were selected from the miniprep DNA and amplified for large scale DNA preparation. This "maxiprepped" DNA was then transfected into COS cells using the standard DEAE dextran method (Seed 1987a; Simmons 1993) and then pulsed with methionine to check for expression levels of Fc proteins. The media was recovered, debris spun away, and the Fc-chimeric proteins separated from the media using a 1 ml protein A column. The fractions were analysed at A_{280} followed by SDS-PAGE to determine quantities of protein in each sample for protein expressed from Rg constructs (Figure 3.1) and pIG1 constructs (Figure 3.8). The presence of a replicon would likely have inhibited protein expression following transfection. Another explanation could be that the intron 5' of the hinge had to be lengthened relative to the early version of pIG1,

as correct splicing was not always obtained. The pIG1 vector was shown to be marginally superior to Rg in expressing protein, likely due to more consistent dimer formation in soluble proteins. This and early publication restrictions on Rg made pIG1 the vector of choice for future experiments. A comparison of Fc-CD33 and Fc-CD34 (pIG1) with that of Fc-ICAM-1 (Rg), a construct which shows reduced expression in both vectors, is shown below (Figure 3.8). Fc-ICAM-1 is shown for a size comparison as different proteins cannot be compared quantitatively.

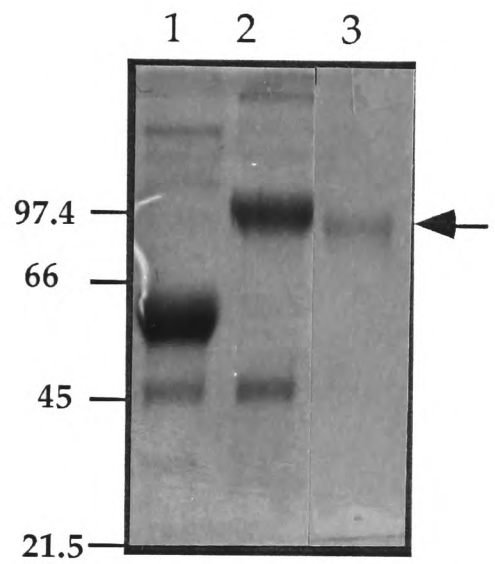


Figure 3.8 Fc-CD33 and Fc-CD34 in pIG1 with Fc-ICAM-1 in Rg. Lane 1- Fc-CD33 (pIG1); Lane 2 -Fc-CD34 (pIG1); Lane 3 - Fc-ICAM-1 (Rg). Fc-CD33 and Fc-CD34 produced a higher level of protein than Fc-ICAM-1 (Rg) given the same transfection conditions. The 10% SDS-PAGE gel was run under reducing conditions by adding 5%β-ME. Arrow points to the ECD of Fc-ICAM-1.

Because successful production of soluble Fc proteins in the pIG1 vector did not occur until near the end of the first year of the course of this D.Phil. study, many studies using the Rg vector were repeated with the pIG1 vector. The functional differences between proteins produced from both vectors were negligible, despite a partial monomeric population produced from Rg. Dr. Simmons was first able

to successfully produce protein at a high level with Fc-CD33 and Fc-CD34 in pIG1. Thus, in the interest of time, his construct was used to pursue binding studies with the Fc-CD33 and Fc-CD34 proteins. I performed all transfections and protein purification in both vectors.

3.3 CD33 ligand binding studies using Fc-adhesins

3.3.1 Characteristics of myeloid cell lines

We reasoned that a ligand for CD33 was most likely expressed on the myeloid cell lines which expressed CD33 itself as it had been suggested that CD33 could play a role in differentiation (Griffin 1986; Peiper 1988). We thus chose cell lines which represent various stages in the differentiation pathway. The stages of differentiation represented by these cell lines are diagrammed below (adapted from Koeffler 1983). A table follows, listing the cell lines in order of their position on the differentiation pathway, the level of CD33 and CD34 expressed on each cell line, whether or not it expresses Fc receptors, and whether it can be further differentiated (Table 3.5).

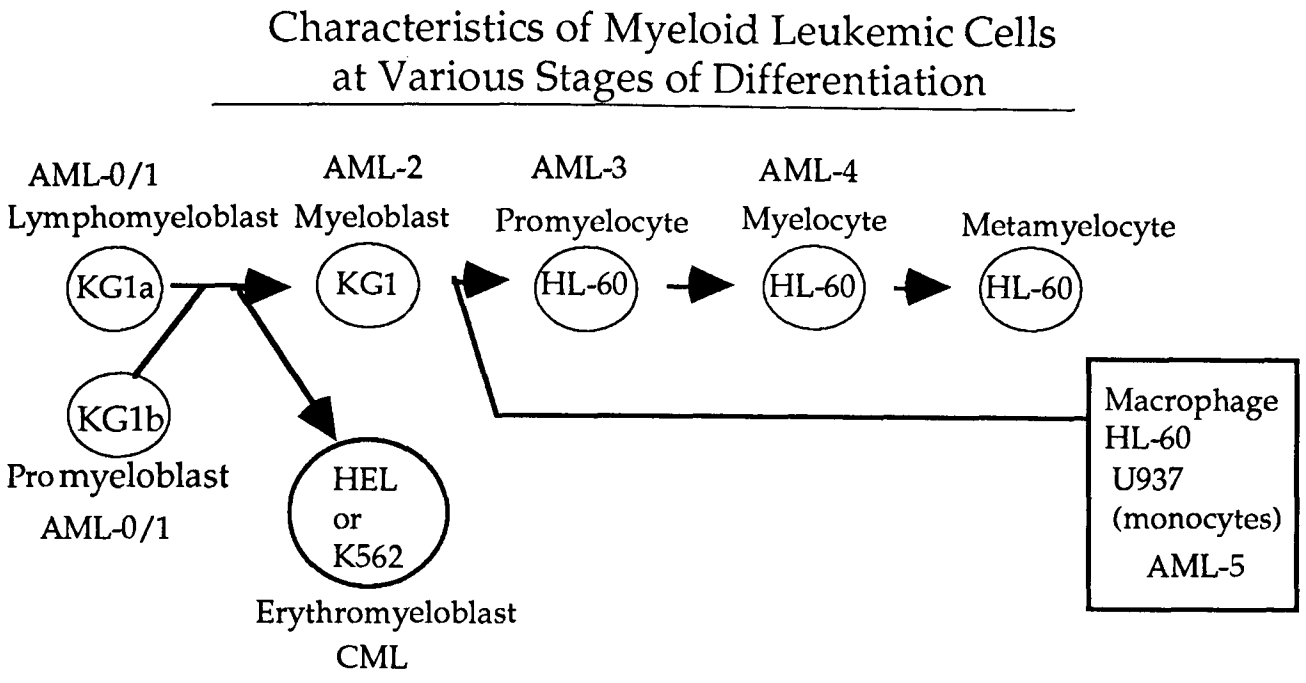


Figure 3.9. The stages of differentiation of myeloid cells.

CD33 and CD34 Expression in Various Leukemic Cell Lines

Cell Line		CD33	CD34	FcRI p75	FcRII p40	FcRIII p50-80	Differentiable?
KG-1b	Myeloblasts (cryptic TCR) (GM-CSF responsive)	+	+	-	-	-	Yes
KG-1a	T cell-like/Pre-Myeloblasts (GM-CSF unresponsive)	+/-*	+++	-	-	-	No
KG-1	Myeloblastic (GM-CSF responsive)	+	+	-	-	-	Macrophages
K562	Early Myeloid and Erythroid Blasts (erythromyeloblasts)	+	-	-	+++	+/-	Fetal Erythroblast
HEL	Early Myeloid and Erythroid Blasts (erythromyeloid)	+	-	-	+++	+/-	Fetal Erythroblast
HL-60	Promyelocytic Metamyelocytic Myelocytic	+	-	++	+	+/-	Granulocytes; macrophages
U937	Monocyte-like	+	-	++	+	+/-	Macrophages

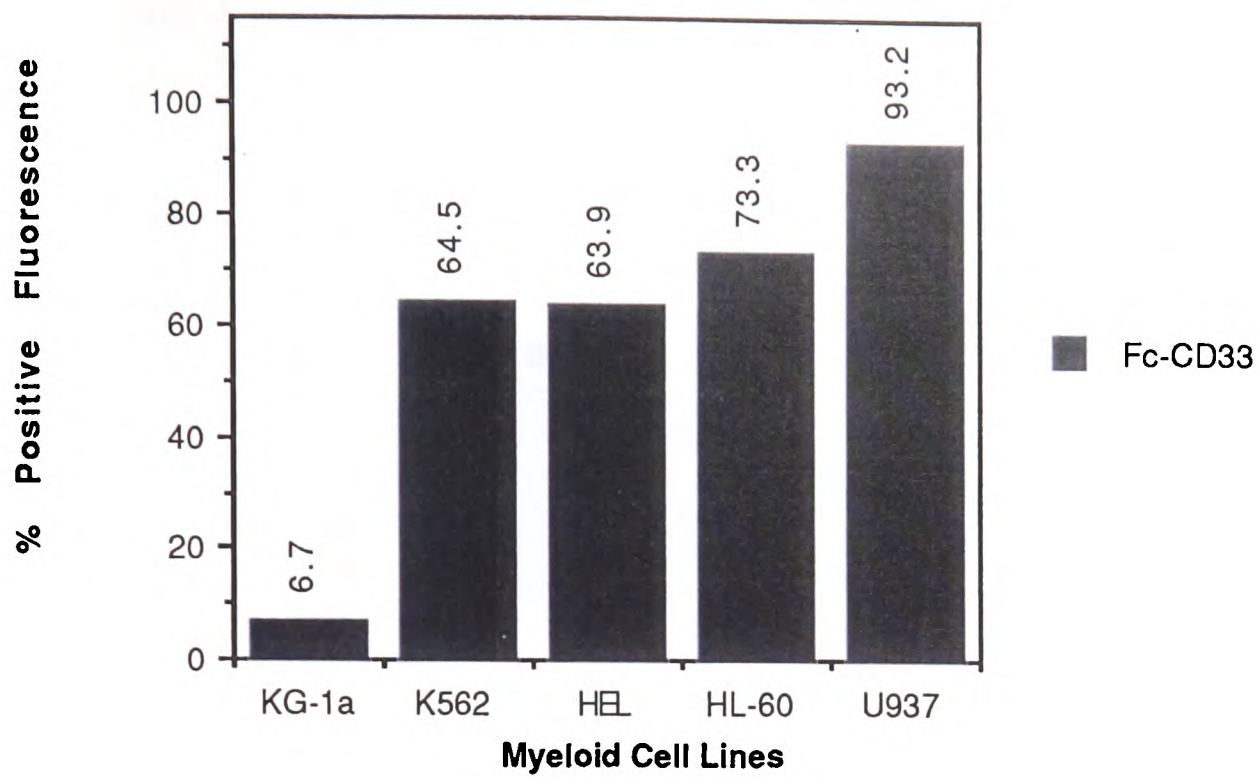
Table 3.5.The characteristics of myeloid cell lines listed by position in the differentiation pathway.
*Reports vary on whether KG-1a expresses CD33 (MY9). Early reports said it was negative (reviewed in (Griffin 1985) and exhibited T cell-like characteristics (Furley 1986). More recent FACS analysis with different monoclonal antibodies showed about 35% of cells reacted positively (Koller 1989). In immunoprecipitation, MY9 strongly precipitated CD33 (Peiper 1989). Clearly, differences exist between labs in the ability of CD33 to be detected in this cell line.

By providing homogenous populations of cells during different stages of myeloid maturation, these leukemic cell lines provide a model system to study the induction of myeloid leukemic cell differentiation. The specific origins and characteristics of the cell lines appear in Chapter 2.

3.3.2 Fc-CD33 binding increases with differentiation

Initially, the binding of Fc-CD33 was compared with that of an anti-CD33 monoclonal antibody (Mab) (MAS 393, clone WM53, Sera-lab), known to bind to these myeloid cell lines, using Flow Activated Cell Sorter (FACS) analysis (Figure 3.9A-B). This technique was critical to determine levels of binding at various stages along the differentiation pathway. Fc-receptor binding was controlled for in the routine manner by using an initial blocking wash with 20 μg/ml of goat immunoglobulin. The Flow Cytometry curves (Figure 3.9B) are presented to reveal the correspondence of Fc-CD33 binding to CD33 expression along the myeloid differentiation pathway represented by the cell lines. As Figure 3.9A-B demonstrate, the highest levels of binding were shown for HL-60, U937 and HEL, with the highest values for median fluorescent intensity shown for HL-60 and HEL (Figure 3.9C). The Y axis represents cell number while the X axis represents Fluorescence Intensity (Log_{10}).

A.



B.

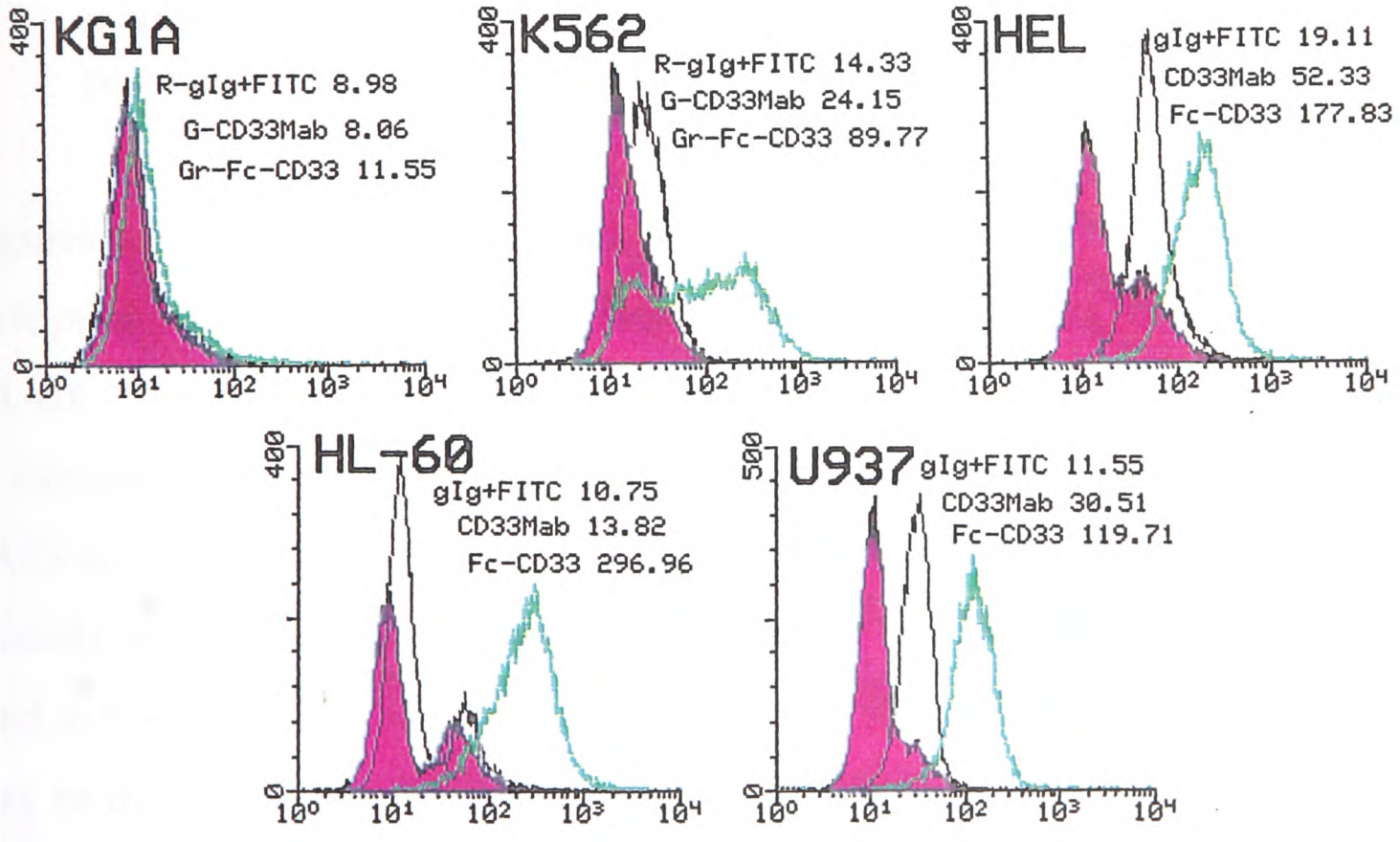


Figure 3.9A and B. Fc-CD33 ligand binding increases in more differentiated myeloid cell lines. Flow Cytometry analysis shows that Fc-CD33 binding is maximal on the monocytic and myelocytic cell lines U937 and HL-60 as well as the erythroleukemic lines HEL and K562. 200 µg/ml Fc-CD33 was used in each sample.

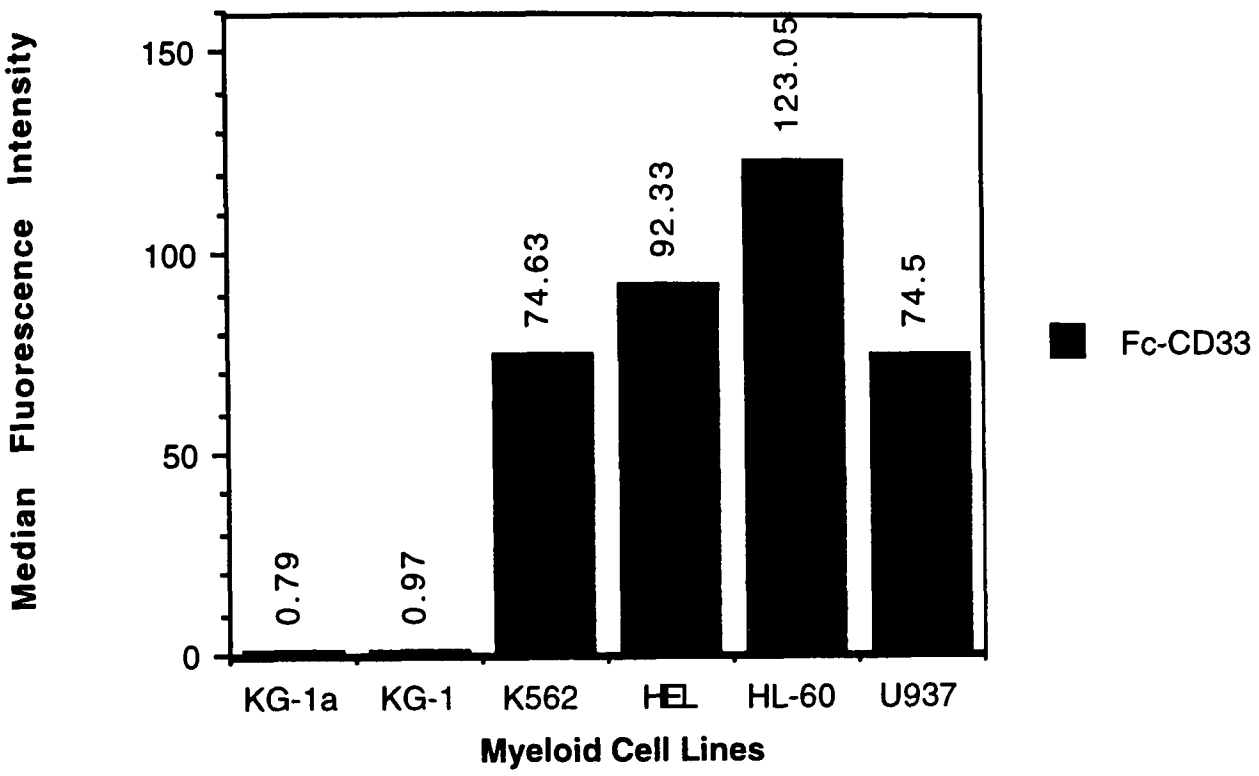


Figure 3.9C. Median Fluorescence intensity demonstrates a binding peak in HL-60 which drops off with monocytic differentiation. A comparison of the mean values of two experiments shows that Fc-CD33 binding increases to a peak in HL-60 and drops off in the U937 promonocytic population which may be partially differentiated.

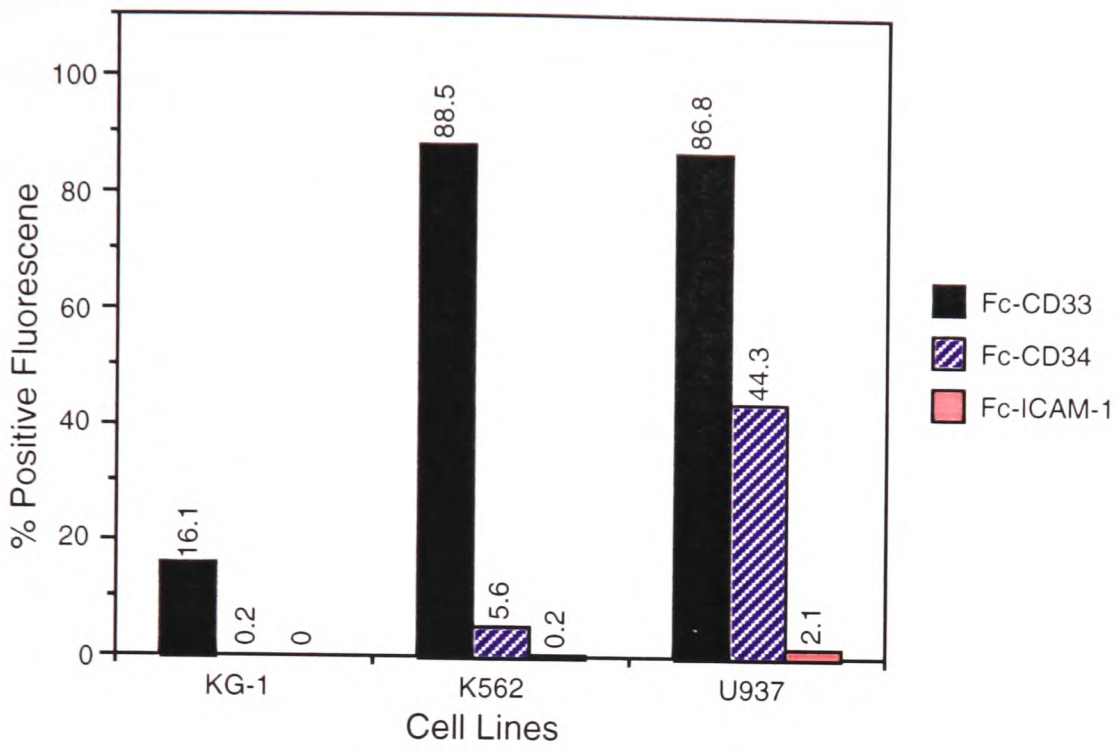
Figures 3.9 A and B show that there is a clear increase in the binding of Fc-CD33 to myeloid cell lines as terminal differentiation of CD33 is approached (U937 and HL-60). This binding pattern closely parallels the progressive appearance of CD33 at various stages along the differentiation pathway as previously shown with FACS analyses using anti-CD33 Mabs on these myeloid cell lines (Koller 1989). Notably, there is a difference in the ability of the erythroleukemic cell lines to bind to Fc-CD33; HEL binds ~20% more strongly than K562 Figure 3.9C). This may be due to the inablility of HEL to produce the full number of erythroid precursor proteins (Rutherford 1981). HEL expresses a higher level of CD33 than K562 as shown by antibody staining (Figure 3.9B). In this cell line it is possible that the ligand for CD33 is expressed proportionally to CD33. In HL-60, however,

a much lower level of CD33 expression is shown than for U937 yet both bind Fc-CD33 at a high level (Figure 3.9B). At the final stages of thesis preparation, it was discovered that all HL-60 supplied by the ICRF unit at Clare Hall since 1988 are likely to be mouse cells or a mouse/human mixture, so the Fc-CD33 binding shown is likely to be against a mouse molecule which may have a human homologue in *bona fide* HL-60. This accounts for the poor staining of CD33 Mab on HL-60. It is also possible that CD33 can recognise several ligands in tandem and independently of CD33 expression which may account for the differential in staining capacities between CD33 Mabs and Fc-CD33.

3.3.3 Fc-CD33 binds specifically among Fc-adhesins

Saturating levels of Fc-CD33, Fc-CD34 and Fc-ICAM-1 fusion proteins were then used to confirm the binding behaviour of Fc-CD33 along the differentiation pathway. The use of three Fc proteins allowed for two additional controls to verify that the binding of any particular Fc protein was not due to the presence of Fc receptors. This was also controlled for with the blocking reagent goat immunoglobulin (gIg). Cell lines KG-1, K562 and U937 further demonstrate the ability of Fc-CD33 ligand binding to increase along the differentiation pathway (Figure 3.10A-B).

A



B

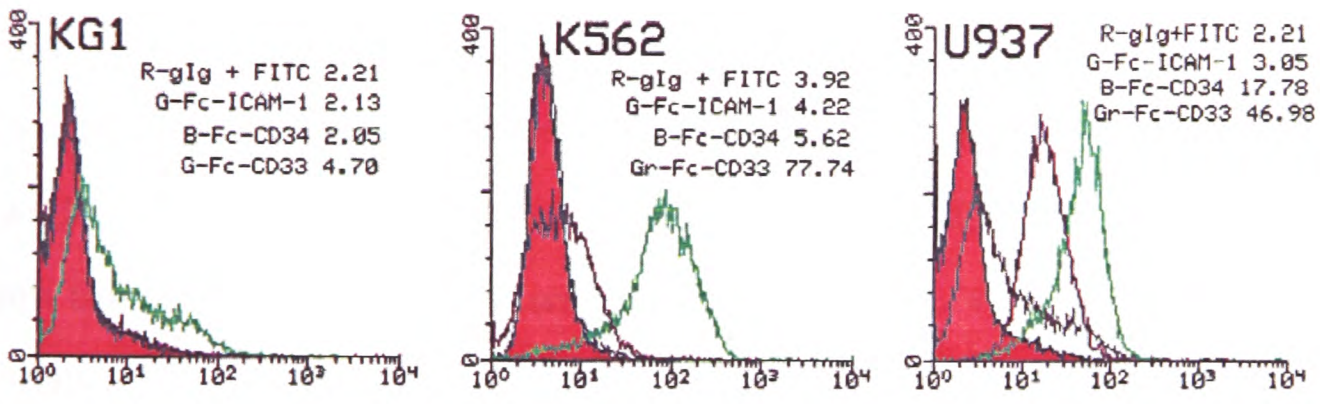


Figure 3.10A and B. Fc-CD33 binds with increased avidity compared with other Fc-adhesins. Fc-CD33, Fc-CD34 and Fc-ICAM-1 were used to probe cell lines at the **A.** early (KG-1), **B.** middle (K562) and **C.** later (U937) stages of myeloid differentiation. The lower level of binding shown for Fc-ICAM-1 and Fc-CD34 compared with Fc-CD33 demonstrates the specificity of Fc-CD33 to bind more differentiated cell lines and shows that binding is not due to Fc-receptors. Fc-CD33 shows the highest level of binding to all three cell lines among the three probes tested. 200ug/ml Fc protein for test and 200ug/ml goat immunoglobulin (gIg) was used for control samples. Values shown represent percent positive fluorescence or the average number or ligands per sample.

Additional binding studies were undertaken to test the binding behaviour of Fc-CD33 to specific overexpressed proteins using the panning method on transfected cDNA libraries previously made from myeloid cell lines, including U937, HL-60, HEL and K562. The results are further described in Chapter 6. In tandem with the FACS analyses, cDNA libraries from these cell lines were subjected to the panning process to attempt to isolate a single ligand for CD33 by probing with the Fc-CD33 protein. Panning results with the K562 and U937 libraries produced some positive adherence and further binding studies using the FACS analysis were carried out on these cell lines to attempt to quantitate the binding behaviour of Fc-CD33.

3.3.4 Fc-CD33 binds ligand in a proportional manner

In an attempt to quantitate relative binding affinities and avidities of Fc-CD33 for K562 and U937, a titration study was performed with Fc-CD33. A comparison between the cell lines demonstrates that U937 binds more avidly at an earlier stage than K562 (20 μg/ml) but that the cell lines bind at an almost identical level at 100-200 μg/ml) (Figure 3.11 A and B). This suggests the presence of an additional high-affinity ligand on U937 which allows for increased binding at an earlier stage. These analyses show that even 200 μg/ml of Fc-CD33 is insufficient to produce saturating binding in both U937 and K562 (Figures 3.12A and B).

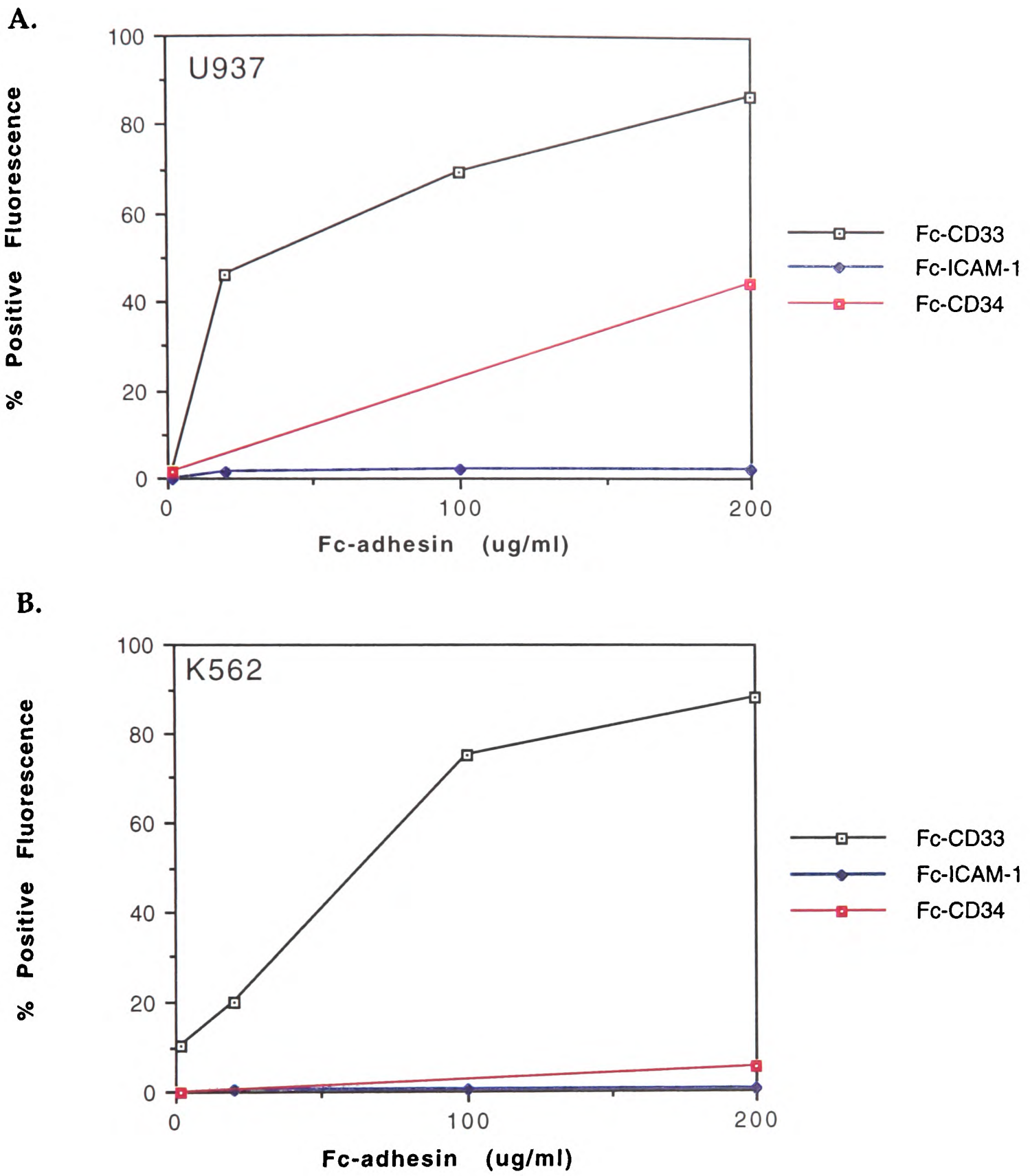


Figure 3.11A-B. Fc-CD33 demonstrates increased avidity in ligand binding to U937 (A) and K562 (B) with the addition of high concentrations of Fc probe. Increasing binding using increasing quantities of Fc-CD33 in binding against U937 (**A**) and K562 (**B**). A high-affinity ligand is suggested in U937 by the early binding levels. Cells were blocked with 200 µg/ml goat Ig prior to incubation with Fc-CD33 to prevent Fc-receptor binding. 20, 100 and 200 µg/ml Fc-adhesin or 200 µg/ml goat Ig was used in each sample. Values plotted represent percent positive fluorescence or the total number of ligands per population.

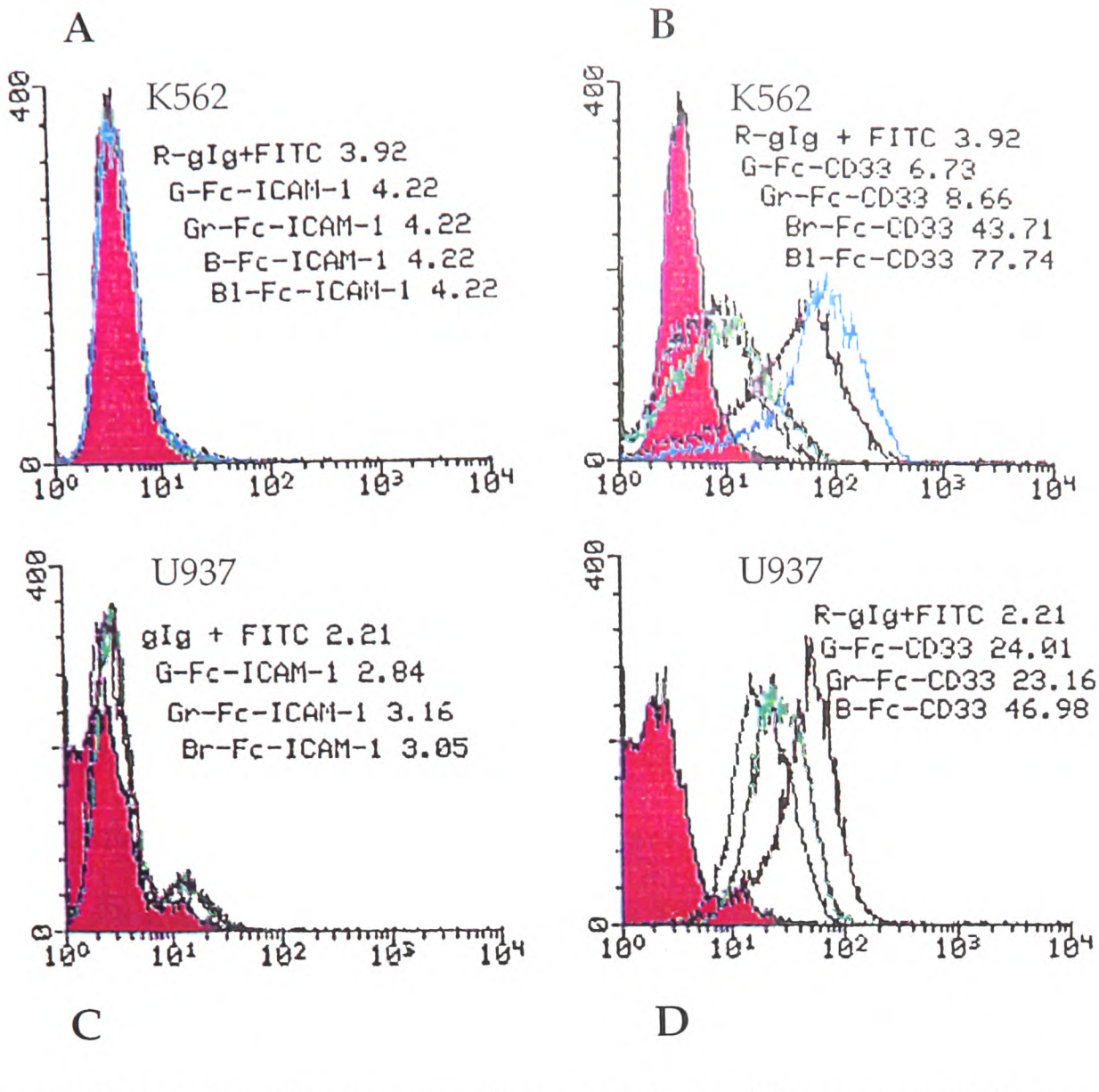


Figure 3.12 Median Fluorescence Intensity shows higher overall fluorescence for K562 than for U937 in ungated cells. This study demonstrates the direct correlation between an increased level of Fc-CD33 and Fc-CD33 binding to K562. A titration of 20, 100 and 200 μg/ml Fc protein and 200 μg/ml goat Ig was used per sample. Curves show median fluorescence for each concentration listed from top to bottom in increasing order of concentration. Graphs and FACS curves demonstrate that Fc-CD33 binding is concentration dependent.

Figure 3.12 shows that K562 has a higher overall level Fc-CD33 binding than U937 in titration analysis in ungated cells. Taken together, these binding studies demonstrate a high degree of binding of Fc-CD33 to U937, K562 and, to a much lesser extent KG-1a (Barber 1993). The possibility that Fc receptors were binding to the Fc-CD33 in the original FACS analysis has been controlled for with two control Fc-immunoadhesins in addition to the goat-anti mouse

immunoglobulin used as a blocking reagent. In the presence of these controls, Fc-CD33 retained distinct binding to several cell lines in a differentiation specific manner.

3.3.5 Surface iodination of cell lines

In an attempt to ascertain the size of a potential ligand for Fc-CD33, several cell lines were surface iodinated and immunoprecipitated with Fc-CD33. Figure 3.13 demonstrates the ability of Fc-CD33 to precipitate increasing numbers of ligand subunits as myeloid differentiation progresses. A weak band appears in KG-1b at 58 Kd (Lane 2); three faint bands appear in K562 at 54Kd, 58Kd and 61Kd and a very faint band at 100 Kd (Lane 4); U937 displays three bands at 54Kd, 69Kd and 97-110Kd and HL-60 displays a a heavily labelled broad band from 54-61 Kd, one at 23 Kd and a broad band at 97.5-110 Kd (Lane 8). The heavily labelled bands in HL-60 are all distinct from the control lane except for a slight overlap at 61 Kd. As expected, the most apparent ligand appeared to be precipitated by the cell lines most strongly expressing CD33.

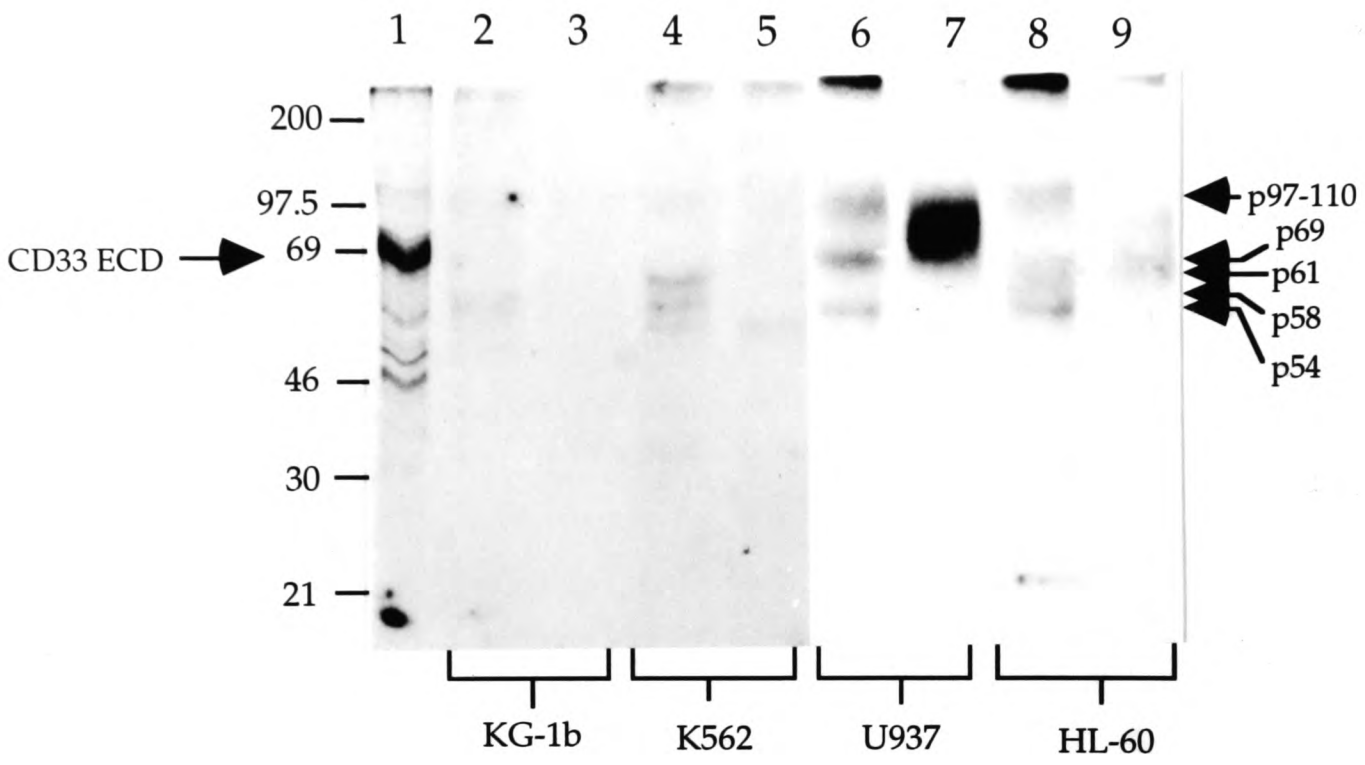


Figure 3.13 Fc-CD33 precipitates various ligand subunits as myeloid differentiation progresses. Lane 1 - ^{35}S labelled Fc-CD33 protein run as size marker; arrow points to Fc-CD33-ECD; Lanes 2, 4, 6 and 8 - Fc-CD33; Lanes 3, 5, 7 and 9 -mouse Ig. KG-1b shows a single ligand subunit at 51Kd (Lane 1) and K562 shows three subunits at 54, 58 and 61Kd (Lane 4). Additional larger subunits at 97-110Kd are shown in U937 (Lane 6) and HL-60 (Lane 8). Exposure time is three days at -80°C. This gel reinforces the finding that Fc-CD33 binding increases upon increasing differentiation.

Clearly, the binding behaviour of Fc-CD33 varies depending upon the cell line used with it, and thus position in the differentiation pathway, to which it binds. The iodination data above is suggestive that at least three proteins are precipitable with Fc-CD33 from K562 as well as HL-60. The bands appear to be heavily glycosylated and their average sizes range from 54-69 Kd and 97-110 Kd.

3.3.6 Fc-CD33 and Fc-CD34 show no specific ECM binding

Both CD33 and CD34 are expressed in the Bone Marrow and could potentially bind to components of the extracellular matrix during cell differentiation. Previous studies had demonstrated the ability of murine erythroleukemia (MEL) cells to bind selectively to the ECM protein fibronectin via cell surface fibronectin receptor (Patel and Lodish 1984). The fibronectin receptor was then shown to be downregulated upon differentiation of the cells into reticulocytes which would allow the release of the cells from the marrow (Patel and Lodish 1987). An initial binding study was thus undertaken to assess the ability of Fc-CD33 and Fc-CD34 to bind to a panel of extracellular matrix proteins (ECM) in the form of full-length or partial peptides (Calbiochem, Table 3.6) bound to plastic. In this assay, ^{35}S labelled Fc-CD33 and Fc-CD34Rg was added to plates which had previously been coated with extracellular matrix proteins. Figures 3.15A-B and 3.16A-B showed no significant binding for Fc-CD33 or Fc-CD34 to ECM in these assays. It may be significant that Fc-CD34 shows marginal binding to Fibronectin in 3.16A. The previous FACS data suggests that Fc-CD33 binding is more significant intercellularly, however.

Table 3.6 Extracellular Matrix Proteins for Fc-CD33 and Fc-CD34 Binding Studies

Protein	Source/Function
Collagen I	Human Membrane/major fibrous stuctural elements of cartilage, tendon, skin, bone, lung and blood vessels that serve to maintain cells in discrete tissues. Collagens also interact with other ECM components. Type I collagen is the most abundant and exists as a heterodimer.
Collagen II	HumanMembrane/see Collagen I
Collagen III	Human Membrane/see Collagen I
Collagen IV	Human Placenta/see Collagen I
Collagen V	Human Membrane/see Collagen I
Collagen VI	Human Placenta/see Collagen I
Collagen VII	Human Membrane/see Collagen I
Fibronectin	Human Plasma/agent for promoting cell attachment to commonly used substrates; MW 44 Kd.
Fibronectin 45	Human Plasma/α-Chymotrypsin 45 Kd Proteolytic Fragment, Gelatin-Binding Region.
Fibronectin 120	Human Plasma/α-Chymotrypsin 120 Kd Proteolytic Fragment, Cell-Attachment Region.
Fibronectin 634	Human Plasma/Trypsin 30 Kd Proteolytic Fragment, Gelatin-Binding Region.
Fibronectin 637	Human Plasma/α–Chymotrypsin 40 Kd Proteolytic Fragment, Heparin-Binding Region.
Fibrinogen	Human Plasma/Essential for Blood Clotting,; MW 341 Kd
Vitronectin	Human Plasma/promotes cell adhesion; MW 65 and 75 Kd.
Laminin	Human Placenta/promotion of cell, platelet, Type IV Collagen, heparin and glycosaminoglycan adhesion and promotion of neurite regeneration and binding; MW 130 and 160 Kd.
anti-Human Ig Fc	Goat antisera/an antibody specific for the Fc portion of the chimeras used as a positive control.
Human Ig	Human Serum/ a negative control antibody

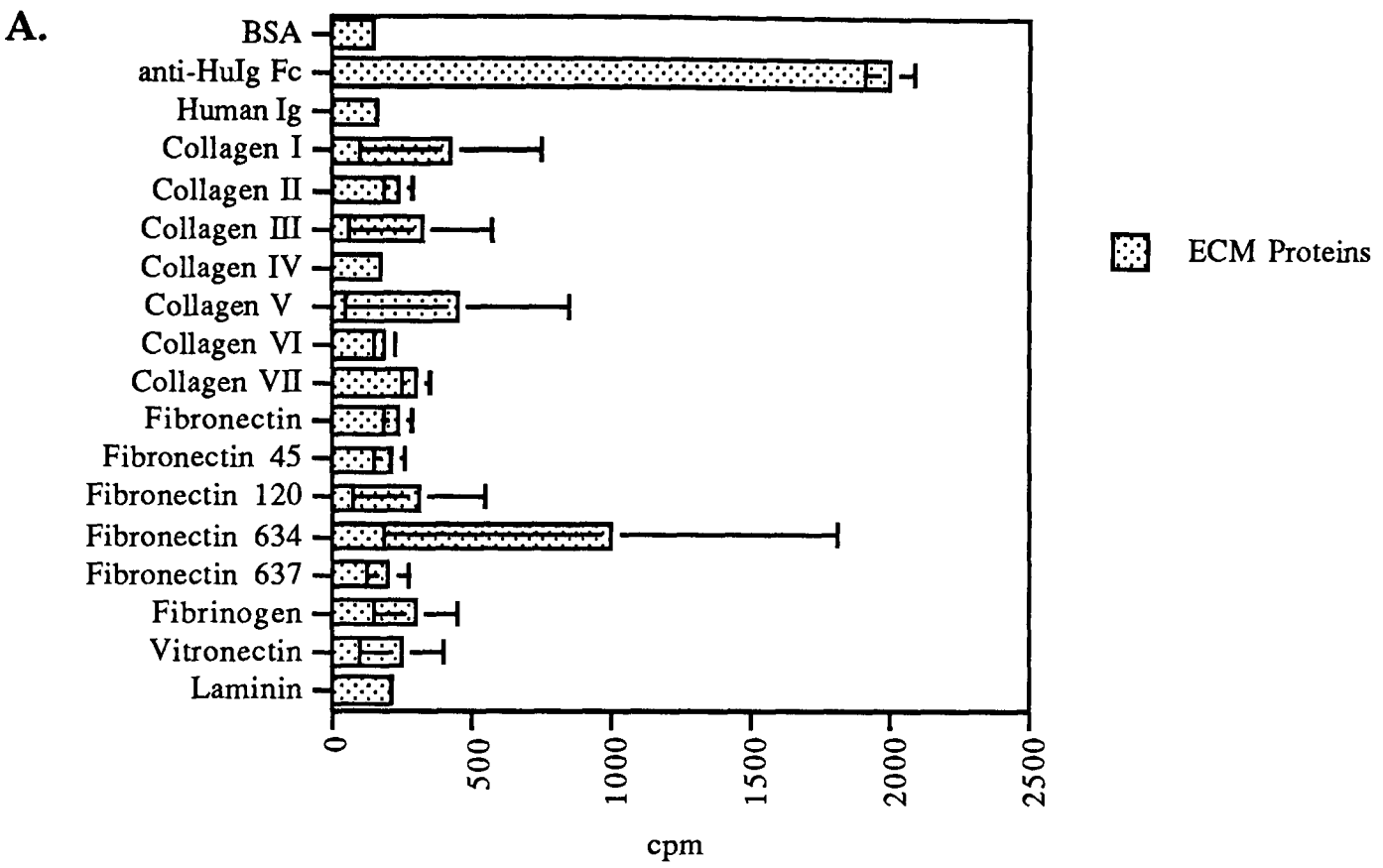


Figure 3.14 A and B. ^{35}S Labelled Fc-CD33 shows no specific binding to a panel of Extracellular Matrix Proteins. A. Marginal binding above negative control proteins BSA and Human Ig is shown for Collagen VII, Fibronectin, and Laminin. **B.** A second experiment with selected ECM proteins. The positive control for the Fc portion of the chimera is anti-Human Ig Fc. Mean values shown represent n=3. Error is represented by the Standard Deviation. Input counts were 1.9×10^5/well.

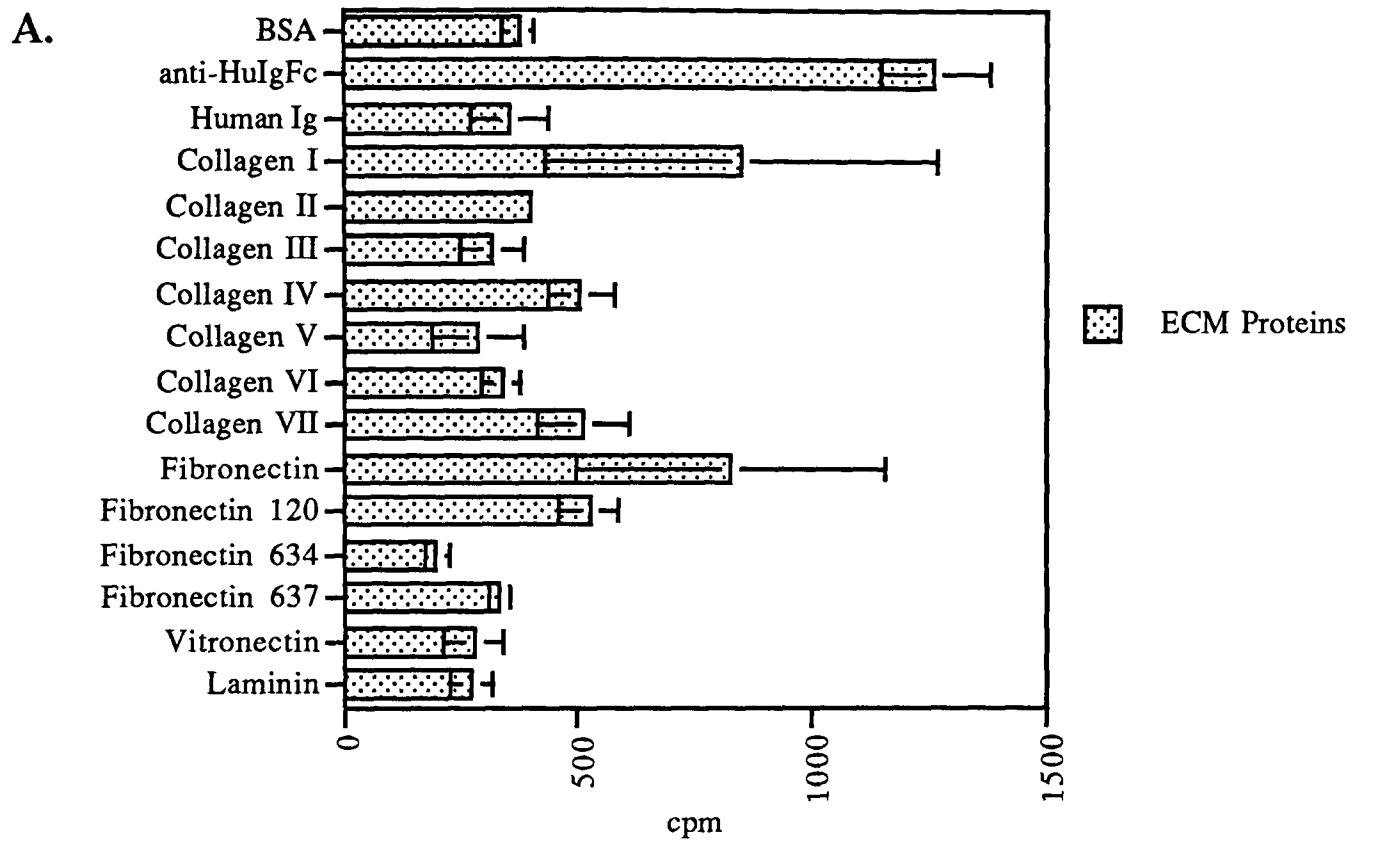
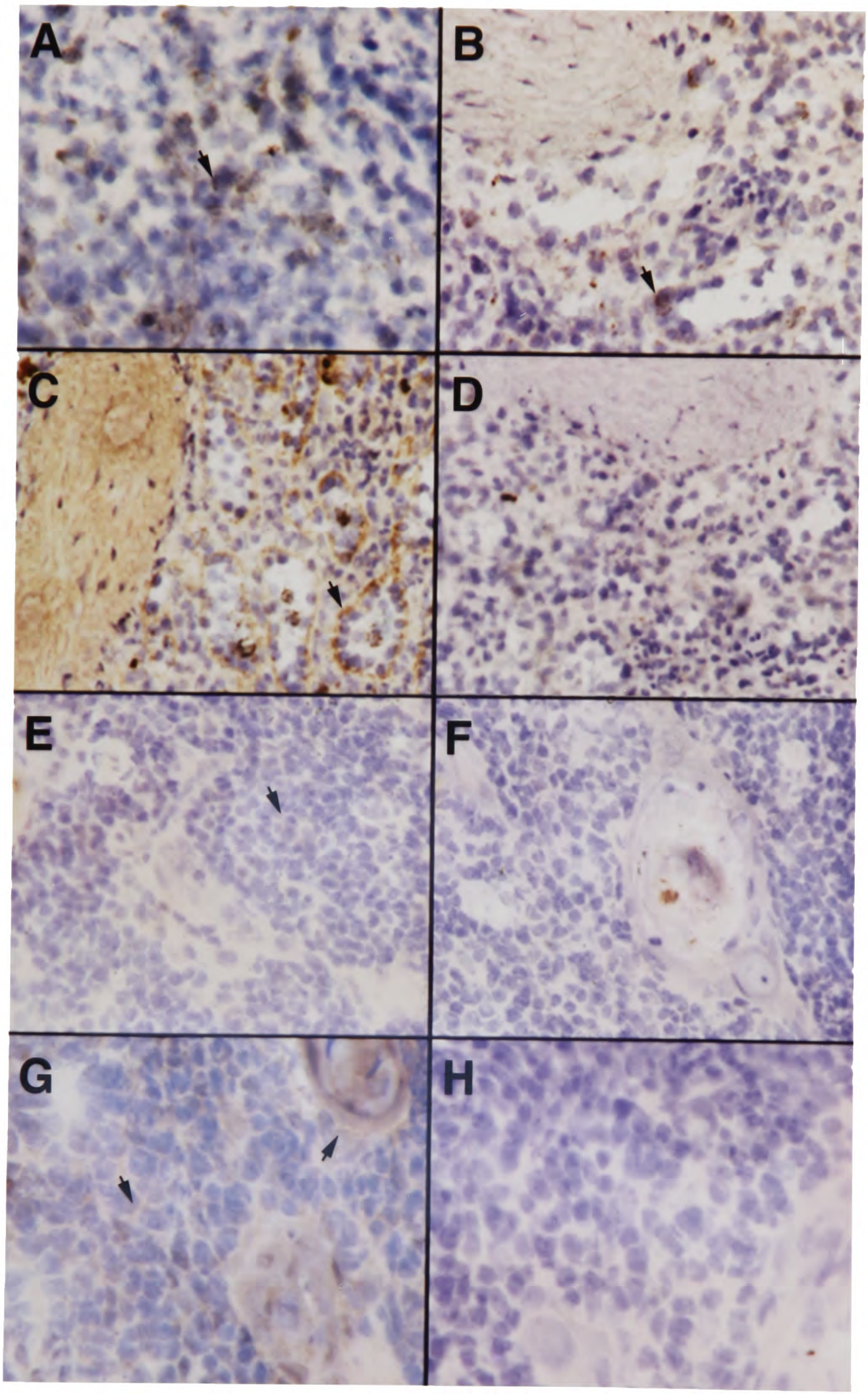


Figure 3.15 A and B. CD34 shows no specific binding to Extracellular matrix proteins. A. Fibronectin shows some binding above background. **B.** No significant binding above background is shown for proteins which showed marginal binding in A. Mean values represent n=3, error represents standard deviation. Input counts of ^{35}S-labelled Fc-CD34 (Rg) were 1.7×10^3 counts per well.

3.3.7 Immunohistochemical analysis of CD33 and CD34

Fc-CD33 and Fc-CD34 were subjected to immunohistochemical analysis using the avidin-biotin detection system on frozen tissue sections from spleen (Figure 3.16A-D) and thymus (Figure 3.16E-H) (courtesy of Margaret Jones, Department of Pathology, Department of Cellular Science, John Radcliffe Hospital, Oxford, UK). Fc-ICAM-1 and Fc-CD43 are presented as positive and negative controls. In splenic red pulp, which consists mainly of monocytes, macrophage and endothelium, Fc-CD33 showed probable staining of monocytes, but essentially negative staining for polymorphonuclear leukocytes (PMNs) (Figure 3.16A) and very weak staining on thymus (Figure 3.16E). Fc-CD34 showed some staining in spleenic endothelium (Figure 3.16B) and negative staining in thymus (Figure 3.16F). Fc-ICAM-1 showed positive staining on splenic sinusoidal endothelium and connective tissue (Figure 3.16C) which may be consistent with an ability to bind some "I" domain-containing subunits, such as collagen VI, a component of both endothelium and the extracellular matrix in a divalent-cation independent manner as has recently been shown for collagen/collagen VI binding, although Mn^{++} was required for αL and αM transfectant binding to ICAM-1 fusion proteins (Kamata et al. 1995). In thymic cortex containing immature lymphocytes, Fc-ICAM-1 showed some positive staining of thymic epithelium, particularly on Hassal's Corpuscles and connective tissues in thymus (Figure 3.16G) and no staining of lymphocytes in either tissue, consistent with the FACS data. A divalent-cation independent mediated monocyte-endothelial binding was also recently suggested by the use of anti-CD44 antibodies (Toyama et al. 1993). CD43 was essentially negative in both tissues (Figure 3.16D and 3.16H) apart from some weak macrophage staining in spleen.


A
B
C
D
E
F
G
H

3.4 Discussion

Following the cloning and expression of Fc-CD33 and Fc-CD34 in the pIG1 vector, it was shown that Fc-CD33 shows increased binding to more differentiated leukemic cell lines following the trend presented in the model below (Figure 3.17). HL-60 is in parentheses due to the report that the HL-60 supplied is likely of mouse or possibly mixed mouse/human origin. It may contain a homologous ligand, however, as strong binding was produced to Fc-CD33 and, in later adhesion assays with *bona fide* HL-60, the trends shown here were shown to be accurate (personal communication, Paul Crocker). Iodination data suggests that CD33 can bind to ligands ranging in size from 60-69Kd and 97-110Kd with the possibility of a higher molecular weight ligand. No specific Fc-CD33 binding was shown to ECM proteins suggesting that CD33 ligands are not expressed in the extracellular matrix. These data were reinforced by immunohistochemical analysis which showed some binding of Fc-CD33 to monocytes and macrophages in spleen, consistent with the expression of ligands on promonocytic cells (Figure 3.16A). Fc-ICAM-1 bound to the ECM in spleen in a divalent cation-independent manner which may suggest that it can bind a non-αL or αM "I" domain as collagen can bind α2 on collagen VI.

Subsequent to the FACS analyses presented here, a solid-phase adhesion assay was performed with Fc-CD33 bound to plastic and incubated with several myeloid cell lines. In this assay, Fc-CD33 was shown to bind strongly to U937 and HL-60, but not to K562 (HEL was not tested). It was also demonstrated that desialylated CD33 transfectants bind to HL-60 cells and erythrocytes (RBCs) in a sialic acid-dependent manner while native CD33 transfectants fail to bind either

cell (Freeman 1995). Both HEL and K562 express large amounts of sialic acid yet some antibodies fail to recognise common antigens on both cell types, such as glycophorin A (Rearden 1985), attributed to the inability of the cell lines to express the adult (β) form of hemoglobin (Martin 1982; Rutherford 1981) which likely gives rise to distinct sialic acid (sia) determinants. CD44 and CD43 are examples of surface markers expressed only on cells which express foetal and not adult hemoglobin. It is likely that CD33 also requires a sialylated ligand with a specific conformation of sialic acids. Given the variability of the cell-surface phenotype of some cell lines over time (Smith 1983) and given the highly differentiable capacity of HEL and K562 (down the erythroid pathway with hemin, down the myeloid pathway with DMSO), it is possible that HEL and K562 possess a CD33 ligand, but that time has altered its sia conformation. CD22β and Sn have shown different binding capacities to different sialic types (Kelm 1994b). Thus, Fc-CD33 may recognise a distinct glycoprotein to that found on RBCs with a similar conformation of sialic acids. Alternatively, it may recognise a known protein which has not yet been characterised.

The discrepancy between K562 binding in the FACS and solid phase assays could exist because Fc-CD33 can access cryptic ligand(s) in the FACS analysis which are inaccessible in the solid phase assay. Such a cryptic ligand may only be accesible to Fc-CD33 binding in a FACS analysis on cells such as K562. The adhesion assay may show binding to a distinct or exposed ligand with a preference for cells which express it at a higher level. That CD33 transfectants are unable to bind HL-60 *in vivo* could indicate that transfected CD33 is bound by neighboring sia residues in cis-interactions (Crocker et al. 1995; Freeman 1995). These cis-interactions imply that the expression of endogenous sialyltransferases could

play a role in regulating CD33 binding. In B cells, α2,6-sialyltransferase is upregulated upon B cell activation (Erikstein et al. 1992) which increases cis-interactions with CD22 and thus decreases intercellular binding (Braesch-Anderson 1994; Hanasaki 1995). A similar phenomenon could exist in monocytes in that upon activation, α2,3-sialyltransferase (which could normally provide low-affinity cis-ligands for CD33) would be downregulated, thus freeing cell surface CD33 to bind intercellularly to high affinity ligands. Given that COS cells overexpress α2,3-sialylated glycoproteins, relatively small sia binding proteins such as CD33 would be much more susceptible to inactivation of intercellular binding due to cis-interactions and thus may not represent the *in vivo* circumstance. However, CD33 shares a similar structure with CD2 in that it contains only two Ig domains. CD2 has been demonstrated to interact intercellularly with LFA-3 (Seed and Aruffo 1987b). Sialoadhesin, with 17 Ig domains, is able to interact intercellularly above the glycocalyx and is thus not inhibited by neighboring sias, as proposed in the "rainforest" model (Crocker et al. 1995). Activation of CD33 bearing cells could also produce an alteration of cell surface charge which could free sia-bound CD33 to interact intercellularly. The binding data with both the FACS analysis and the solid phase assay suggest that circumstances exist in which CD33 is not inhibited in intercellular binding by sialic acid.

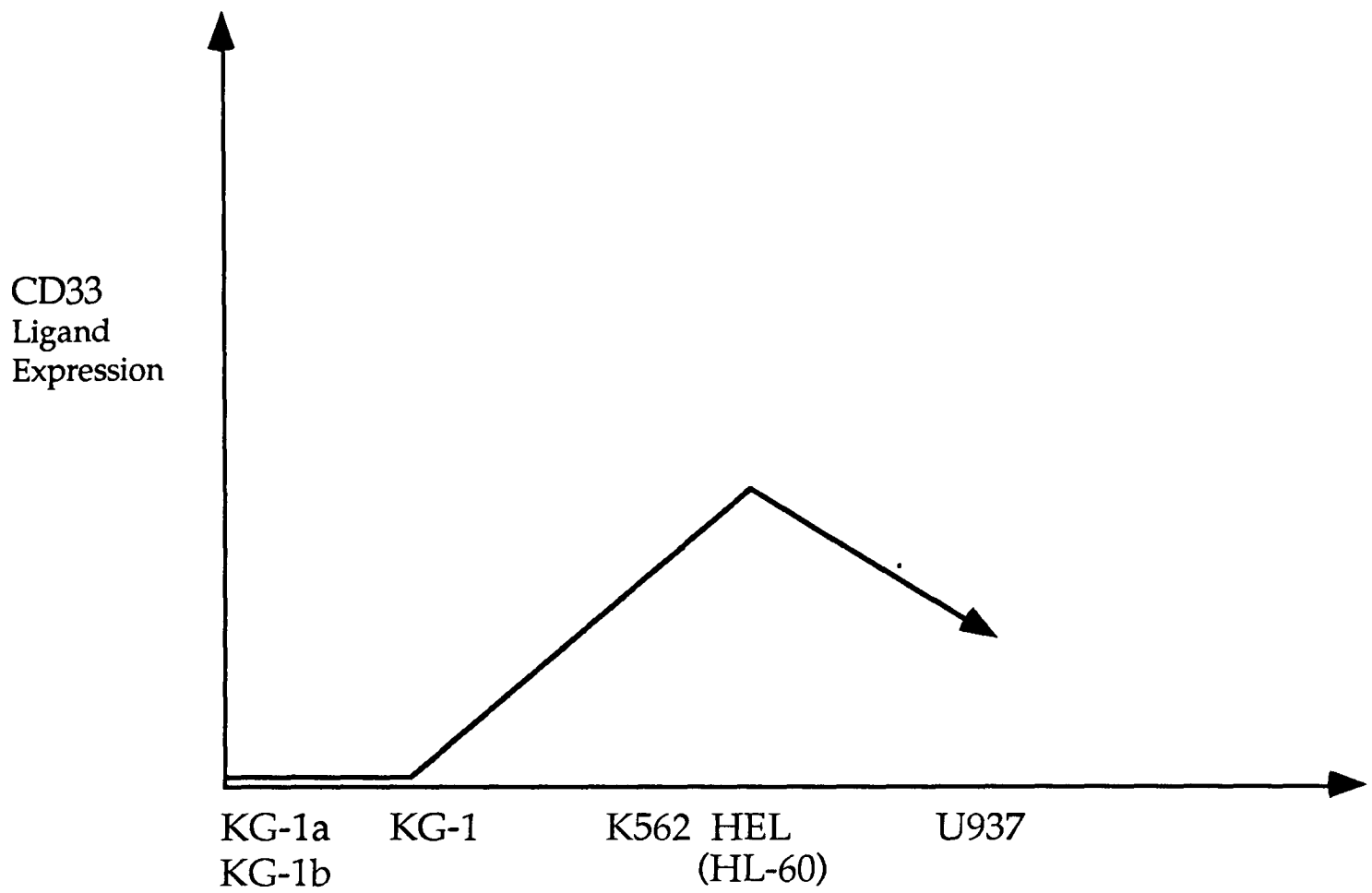


Figure 3.17 Proposed model of CD33 ligand binding

Chapter 4

Phosphorylation Studies on CD33 and its Ligands

4.0 Specific aims

The aim of this chapter is to establish the potential of CD33 and a ligand to be phosphorylated. This was explored using the experimental technique of the *in vitro* kinase assay. This technique can establish whether phosphoproteins which potentially associate with the cytoplasmic tail can be precipitated with Mabs against the antigen of interest. Both CD33 Mabs and Fc-CD33 precipitated phosphoproteins under these conditions. CD33 was shown to associate with the *src*-related tyrosine kinases *fgr, fyn, hck,* and *lyn* and may itself be phosphorylated under certain conditions.

4.1 Phosphorylation in cell signalling

The ligation of ligand to receptor often results in triggering of intracellular molecules which bring about the transcription of new genes. The detection of intracellular signals via phosphorylation has become an important tool for studying the transduction of signals from the outside to the inside of, and within, the cell. The role of phosphorylation in cell signalling has been shown to be important in associations of intracellular tyrosine kinases with the cytoplasmic tails of CD3, CD4 and CD8 which associate with the T cell receptor (TCR) (reviewed in (Rudd et al. 1991) and CD22 and CD19 which associate with the B cell Receptor and thus important in the growth and differentiation of both cell types. Members of the Sn family have recently been shown to include CD22, and may thus undergo a similar mode of activation. A similar mode of association has not yet been shown for

myeloid cells but the five forms of the low affinity Fc receptor CDw32 are exclusively present on monocytes and may provide provide a similar mechanism of activation for the myeloid arm of the immune system. CD33 is strongly expressed on myeloid cells and shares close homology with MAG and other members of the recently characterised Sn family. It was thus important to test the possibility that CD33 could be phosphorylated to identify clues regarding its potential function *in vivo*.

4.1.1 CD33 possesses homologous YXXL/V motifs

Phosphorylation depends on an association between kinase and substrate catalysed by the presence of ATP. Phosphoproteins identified in the phosphotransferase assay can indicate any of four modes of phosphorylation: 1) that the surface marker precipitated is autophosphorylated on serine, threonine and/or tyrosine residues and does not associate with a kinase; 2) that an associated intracellular kinase, usually tyrosine or serine/threonine, is phosphorylating the precipitated molecule but is not phosphorylated itself or; 3) the kinase associating with the precipitated surface molecule is capable of autophosphorylation as well as phosphorylation of the surface marker; 4) that a tyrosine kinase associates with the surface marker and autophosphorylates but does not phosphorylate the marker, or; 5) that an artefactual phosphoprotein has been precipitated with the surface marker of interest. Combinations of these associations could take place within the same molecule at different sites. The association of CD3/CD4 and CD8 with the tyrosine kinase p56*lck* is an example of the third type of phosphorylation--the cytoplasmic tails of the surface markers do not contain kinase domains or sometimes even tyrosines but they

are capable of associating with a tyrosine kinase, and, in the case of CD3 and CD8, being phosphorylated by it (Barber 1989).

Motifs predictive of the type of phosphorylation one can expect with particular molecules have become useful indicators for predicting modes of phosphorylation. The immunoreceptor tyrosine-based activation motif (ITAM) is often found in CD3/T cell receptor and B cell receptor proteins which associate with tyrosine kinases upon receptor ligation. A YXXL sequence, often $YXXLX_nYXXL$ is critical for the homology (Reth 1989). A variation of this motif has been shown for CD5 ($DNEYXXPX_8YXXL$), the first four residues of which are homologous to the autophosphorylation site in p56*lck* and p59*fyn* (Rudd et al. 1993). Though MAG and CD33 share partial homology in their cytoplasmic tails, only CD33 contains the YXXL motif, suggesting that they are phosphorylated via partially or completely different pathways.

4.1.2 CD33 shares MAG homology at a Major Phosphorylation Site

Initial phosphorylation studies on MAG in oligodendrocytes suggested that it was phosphorylated on tyrosine residues (Bambrick and Braun 1991). Subsequently, the initial events of myelination have been shown to involve *fyn* tyrosine signalling via MAG (Umemori et al. 1994). Initial alignments of MAG and CD33 showed that homologous tyrosines were contained in the cytoplasmic region which suggested that CD33 may be phosphorylated in a similar manner. Following the identification of MAG as a member of the sialoadhesin (Sn) family, a second alignment was performed on the cytoplasmic tails of members showing homologous serines or

tyrosines in the cytoplasmic region (Figure 4.1a). Given the homology of tyrosine 341 in CD33 to tyrosine 620 (Figure 4.1a, position 85), the major phosphorylation site in MAG, it is possible that CD33 becomes phosphorylated in the same manner. In the highlighted long form of MAG (L-MAG), tyrosine 620 interacts specifically with the the SH2 domains of phospholipase C (PLC-γ) thought to be distinct from the site which interacts with *fyn* (Jaramillo et al. 1994). The L-MAG sequence (SEELAEY) (Figure 4.1a, position 93) is similar to one of the epidermal growth factor autophosphorylation sites (TAENAEY) known to interact with SH2-containing signal transduction molecules (Downward et al. 1984; Salzer et al. 1987). This region shares a marked homology with SMP, the human and "A" form of murine CD33 and a limited homology with CD22β, more apparent in the "downstream" alignment (relative to CD22β) (Figure 4.1b). That human and murine CD33 align with this region suggests that they could interact with SH2 domains in a similar way. Adjacent to the motif homologous to the EGFR, however, is the motif YAEI in MAG (and in CD33h) (Figure 4.1b, position 87) similar to pYEEI, known to be the optimal binding motif by SH2 domains in p56*lck* and p59*fyn* (Songyang et al. 1993). The motif for CD33 in this region is YSEV (S and E are polar and I and V are hydrophobic) indicating that this region may also be phosphorylated by an SH2 domain other than that in PLC-γ. It is thus possible that MAG and CD33 could alternate between PLC-γ and *fyn* phosphorylation in this region, particularly since MAG lacks the usual YXXL motif.

In the cytoplasmic tail alignments shown below, homologous regions of potential tyrosine and serine amino acids within relevant members of the Sn family are highlighted in an "upstream" (Figure 4.1a) and "downstream" (Figure 4.1b) alignment relative to the position of the other members of the other members of the

Sn family to CD22β. CD33 was also analysed for the presence of an ITAM. Most compelling is that the highlighted tyrosine at position 59 (Figure 4.1a) and position 119 (Figure 4.1b) correspond to tyrosine 324 in CD33 and is contained in an $EXXYXXLX_{14}YXXV$ motif which is similar to the $(D/E)XXYXXLX_{(6\text{-}11)}YXXL$ activation motif. Of the six tyrosines in CD22β, four are in YXXL regions which align to at least two regions of the other Sn sequences, thus necesitating an upstream and downstream alignment. CD33 aligns most readily to the YTTL (Figure 4.1a) and YSEL (Figure 4.1b) sequences in CD22β both of which were recently found to increase the activity of the SHP (PTP1C) phosphatase *in vitro* following CD22 ligation by mIg (Doody et al. 1995). SHP is known to negatively regulate signalling (Pani 1995). At least three of the YXXL sequences in CD22 are therefore likely to act as immunoreceptor tyrosine-based inhibitory motifs (ITIMs) rather than ITAMs. The ITIM consensus sequence has been shown to be (T/S)XX(p)YXXL (Thomas 1995). In contrast, the potentially phosphorylated sequence in CD33 is EXXYXXL which more closely resembles an ITAM. It could, however, act as as an ITAM or an ITIM under different circumstances.

```
          1                                                  50
CD22β     .KLQRRWKRT QSQQGLQENS SGQSFFVRNK KVRRAPLSEG PHSLGCYN.P
  SMP     ...RRKKGAG SPEVTPVQPM AGPGGDPDLD LRPQQVRWL. ..RGAMERWA
L-MAG     ...RRKKNVT ESPSFSAGDN PPVLFSSDFR ISGAPEKYES ERRLGSERRL
hCD33     KTHRRKAART AVGRNDTHPT TGSASPKHQK KSKLHGPTET SSCSGAAP.T
mCD33A    ..CRRKTTKL SV.....HMG CENPIKAHQQ DSKVHSNPEN PRPLQKDS.P

          51                                                 100
CD22β     M.MEDGISYT TLRFPEMNIP R......... .TGDAESSEM QRPPRTCDDT
  SMP     LGVKEGSGAP QEVTPTSHPP MKPTRGPL.. .EDPPEYAEI RVK.......
L-MAG     LGLRGEPPEL DLSYSHSDLG KRPTKDSYTL SEELAEYAEI RVK.......
hCD33     VEMDEELHYA SLNFHGMNPS K......... .DTSTEYSEV RTQ.......
mCD33A    QEQSSVHTKI SLDFMGGKP. .......... ....QEYSEI ..........
 EGFR                                       TAENAEYLRV
          101                                                150
CD22β     VTYSALHKRQ VGDYENVIPD FPEDEGIHYS ELIQFGVGER PQAQENVDYV
  SMP     .......... .......... .......... .......... ..........
L-MAG     .......... .......... .......... .......... ..........
hCD33     .......... .......... .......... .......... ..........
mCD33A    .......... .......... .......... .......... ..........

          151
CD22β     ILKH
  SMP     ....
L-MAG     ....
hCD33     ....
mCD33A    ....
```

Figure 4.1a Amino Acid "upstream" Alignment of the Cytoplasmic Tails of members of the Sn Family. The sialoadhesin family sequences shown include, L-MAG (Salzer et al. 1987), Schwann Cell Myelinating Protein (SMP) (Dulac et al. 1992), CD22β (Wilson et al. 1991), human CD33 (Simmons 1988b) and the murine type A CD33 (Tchilian 1994). This alignment was prepared using the "Pileup" program from the Genetics Computer Group (GCG) v. 7.3 (Genetics Computer Group 1991). GapWeight is 3.00 and GapLengthWeight is 0.100 for the alignment. All YXXL motifs and potential serine phosphorylation sites on CD33 are highlighted. The region of highest homology between sequences is between residues 85-91, the region of L-MAG which shares homology with the autophosphorylation site in the EGF receptor known to interact with SH2 domains. The four YXXL motifs on CD22 are also highlighted.

```
        1                                                            50
 CD22β KLQRRWKRTQ SQQGLQENSS GQSFFVRNKK VRRAPLSEGP HSLGCYNPMM
 CD33h KTHRR..... .......... .........K AARTAVGRND TH........
mCD33A ..CRR..... .......... .........K TTKLSVHMGC EN........
 L-MAG ..RRK..... .......... ........KN VTESPSFSAG DNPPV.....
   SMP ..RRK..... .......... .........K GAGSPEVTPV QPMAG.....

        51                                                          100
 CD22β EDGISYTTLR  FPEMNIPRTG DAESSEMQRP PRTCDDTVTY SALHKRQVGD
 CD33h .......... .PTTGSASPK HQKKSKLHGP TETS...... SCSGAAPTVE
mCD33A .......... .PIKA..... HQQDSKVHSN PENP...... RPLQKDSPQE
 L-MAG .......... LFSSDFRISG APEKYESERR LGSER..... RLLGLRGEPP
   SMP .......... .PGGDPDLDL RPQQVRWLRG AMER...... WALGVKEGSG

        101                                                         143
 CD22β YENVIPDFPE DEGIHYSELI QFGVGER.PQ AQENVDYVIL KH.
 CD33h .........M DEELHYASL. NF.HGMN.PS KDTSTEYSEV RTQ
mCD33A .........Q SSVHTKISL. DF.MGGK.PQ .....EYSEI ...
 L-MAG ........EL DLSYSHSDLG KRPTKDSYTL SEELAEYAEI RVK
   SMP .........A PQEVTPTSHP PM.KPTR.GP LEDPPEYAEI RVK
  EGFR                                 TAENAEYLRV
```

Figure 4.1b "Downstream" Alignment of Cytoplasmic Tails of Relevant Sn Family members. Members with most similarity in structure are clustered together. CD33 is similar to CD22 in that they share a YXXL motif but similar to MAG in that they share homology with the EGF SH2 binding domain. The four YXXL motifs in CD22 allow it to align with CD33 in at least two different places. This alignment is referred to as "downstream" because it aligns downstream to CD22 relative to the previous alignment. Homologous serines which may be sites of potential phosphorylation are also highlighted. Only the downstream alignment of CD22β includes the tyrosine critical for the EGFR homology with the other aligned sequences at this position. The EGFR homology may also point to an additional autophosphorylation/SH2 interaction for CD22 at this site.

4.1.3 Importance of spacing between YXXL motifs

The distance between YXXLs is important for molecules such as CD3ζ which recognise kinases carrying two SH2 domains, such as ZAP-70 (Hatada et al. 1995). Distance is likely less important for molecules recognising kinases which possess only one SH2 domain. The two domains of ZAP-70 recognise the $YXXLX_7YXXL$ Reth motif, however, molecules such as CD22 have YXXL motifs with spacing of $YXXLX_{30}YXXL$, $YXXVX_{11}YXXL$ and $YXXLX_{16}YXXL$. Only the first CD22 YXXL motif listed recognises the murine SHP SH2 domain, while both YXXL's in the last motif recognise SHP SH2 domains (Doody et al. 1995). YASL in CD33 is homologous to the penultimate YXXL in CD22 and may thus be capable of recognising SHP as well.

4.1.4 Other motifs identified

Other important motifs were identified within the CD33 tail as follows (Figure 4.1b). A Casein kinase II (serine/threonine) phosphorylation site (S/TXXD/E) in the TSTE sequence (position 133) which is within the domain which aligns with the EGFR TAENAEY motif. Two Protein kinase C phosphorylation sites (S/TXR/K) at position 2 (THR) and position 68 (SPK). Finally, three myristoylation sites were found, including (GSNDTH) (GAAPTV) and (GMNPSK). These sites could facilitate the attachment of the CD33 tail to the inner face of the plasma membrane and thus help it to interact with the appropriate non-receptor tyrosine kinase which attaches itself to the membrane via the same mechanism.

4.1.5 PXXP motifs and kinase associations

Interestingly, the biproline PXXP motif required for SH3 domain recognition is, among the Fc receptors, exclusively present as a tetraproline motif in the cytoplasmic tails of FcγRIIB1-3 (CDw32), as in the HIV-1 Nef protein, recently shown to be required for *hck* and *lyn* recognition in U937 cells infected with HIV-1 (Saksela et al. 1995). It is possible that the presence of tetraproline motifs in FcγRIIB could contribute to the replication of HIV-1 in cells bearing these receptors, particularly the monocytes, through a similar association with *hck*. If *hck* associates with CD33 it could also prove important in the HIV association. FcγRIIA and C possess only a biproline motif. CD16 do not possess PXXP motifs.

4.2 ITAMs and ITIMS in kinase signalling

The unique signalling role among Fc receptors played by FcγRII has been shown by recent work on ITAMs and ITIMs. CD16 (low affinity Fc receptor) and CD64 (high affinity Fc receptor) possess neither PXXP nor YXXL motifs. FcγRIIA and C posess homologous ($YXXLX_{12}YXXL$) motifs recently shown to act as an ITAM in B-cell signalling. FcγRIIB1-3 have one YXXL motif downstream of the tetraproline motif contained in a 13 amino acid sequence recently shown to modulate B-cell signalling (Muta et al. 1994) by acting as an ITIM (Van den Herik Oudijk et al. 1995). It is possible that single YXXL/I sequences could be more likely to act as ITIMs *in vivo* than the classical bi-YXXL motif. Alternatively, sequences similar to the bi-YXXL motif could be modified slightly by replacing the I or L with a V, which has a similar structure, and still acheive binding to an SH2 domain. Recent structural work to

crystallise the two SH2 domains of ZAP-70 shows that each L is critical to anchor CD3ζ into the cleft of each SH2 domain to achieve binding and full phosphorylation (Hatada et al. 1995). Such binding could also be achieved with a similar amino acid. When one of the YXXL sequences was abrogated in CD3, some phosphorylation was seen but at a reduced level. The presence of one YXXL sequence in CD33 suggests that it could associate with a kinase with one SH2 domain rather than two as required for CD3ζ or with two using YSEV as one motif.

4.3 Phosphorylation studies on CD33 and a potential ligand

A phosphotransferase (kinase) assay was performed on cell lines representative of various stages of myeloid differentiation to assess the ability of the cytoplasmic tail of CD33 to be phosphorylated as well as that of a potential ligand of CD33 precipitable with Fc-CD33. Cells were lysed in a 1% mixture (50/50) of Digitonin and NP-40, immunoprecipitated and subjected to a kinase assay as previously described (Rudd 1988; Barber 1989). As the following figures demonstrate, it appears that CD33 is increasingly phosphorylated as the differentiation pathway progresses (Figure 4.2) while its ligand varies in slightly in its phosphorylation between cell lines (Figure 4.4). This study represents that phosphoproteins can be precipitated with anti-CD33 Mabs and Fc-CD33 and could indicate a potential role for phosphorylation in the growth and differentiation of cells expressing CD33.

4.3.1 CD33 phosphoproteins increase with differentiation

Figure 4.2 demonstrates that the phosphorylation of CD33 increases upon increasing differentiation of myeloid cells, reflected in the heavily phosphorylated sample at 56-59 Kd immunoprecipitated with the anti-CD33 antibody in monocytic U937 cells (Lane 10). By contrast, KG-1a, which represents the earliest stage of differentiation in these cell lines, demonstrates no phosphorylation of CD33 (Lane 4), reinforcing the data showing that these cells do not express CD33. Some phosphorylation of CD33 was demonstrated in the more diffentiated KG-1 cells (Lane 2).

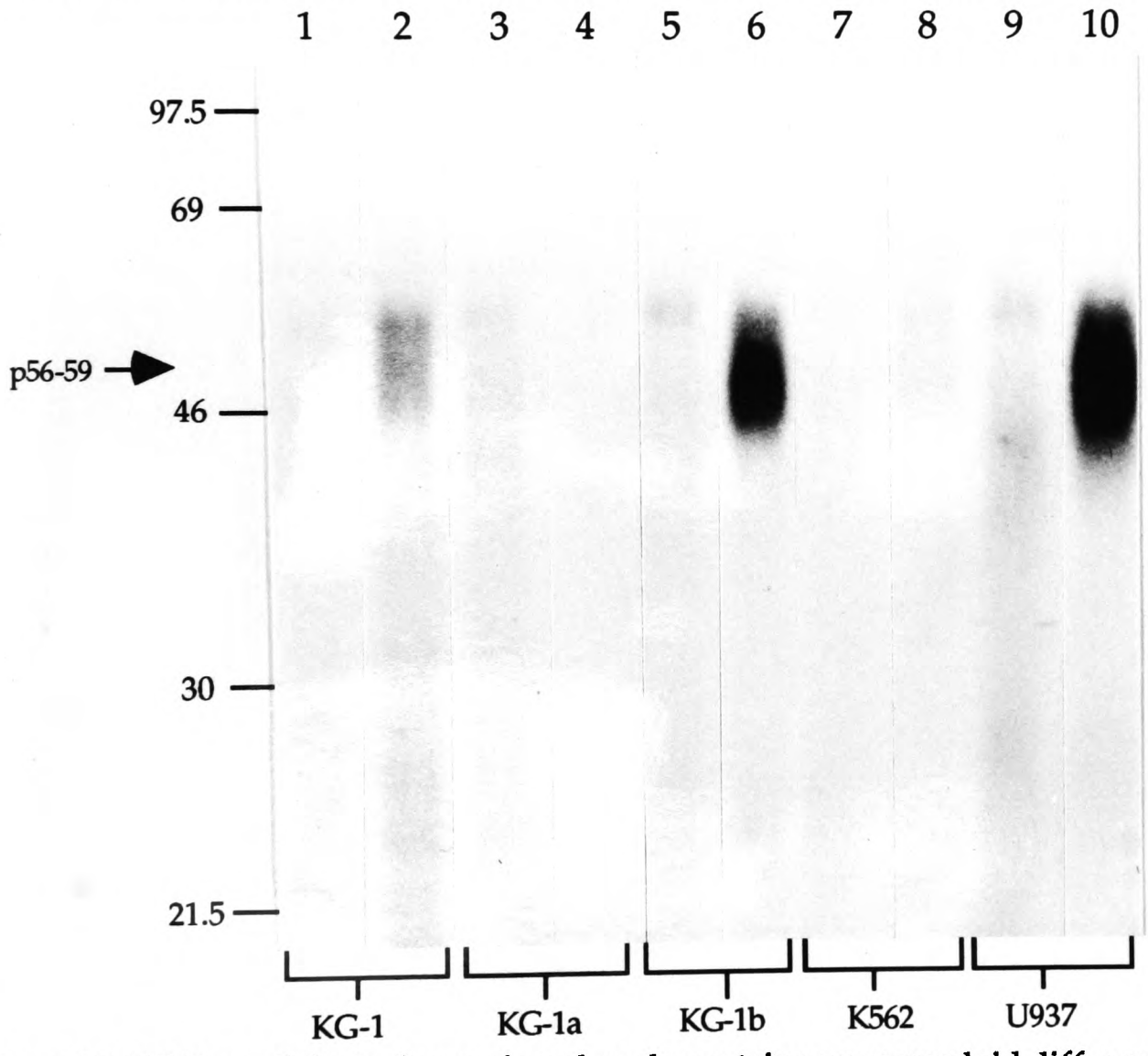


Figure 4.2 CD33 precipitates increasing phosphoproteins upon myeloid differentiation. Increasing levels of CD33-associated phosphorylation are present in anti-CD33 antibody immunoprecipitations subjected to a kinase assay, shown as cells progress from KG-1 (Lane 2) to KG-1b (Lane 6) to pronounced phosphorylation in promonocytic U937 (Lane 10). No phosphorylation is shown in KG-1a (Lane 4) and faint phosphorylation is shown in K562 (Lane 8), No phosphorylation is shown in the goat anti-human Ig control lanes (Lanes 1, 3, 5 7and 9). Lanes 1-2 - KG-1; Lanes 3-4 - KG-1a; Lanes 5-6 - KG-1b; Lanes 7-8 -K562; Lanes 9-10 - U937. Samples are shown on a 12% SDS-PAGE gel .

A marked level of phosphorylation is seen for KG-1b, (Lane 6) the promyelocytic precursor variant of KG-1, which expresses CD33. The level of CD33 phosphorylation shown for KG-1b is consistent with previous observations which demonstrated that, despite cytoplasmic expression of the CD3 δ and ε chains, the cell surface phenotype is more myeloid than lymphoid (Watt et al. 1987). Given the level of CD33-associated phosphoproteins, indicating a relatively high level of CD33 on the cell surface, these data also suggest that that KG-1b may be more differentiated than was previously thought, thus suggesting the position following KG-1 in which it is placed. The KG-1b data also suggest that phosphorylation of CD33 plays a significant role in the early stages of differentiation. Relative levels of CD33 cell-surface expression between KG-1, KG-1a and KG-1b have not been shown, as KG-1b was not tested in the original profile with anti-CD33 Mabs (Koller 1989). Preliminary data on HL-60 demonstrates a similar phosphorylation pattern to U937 with three less phosphorylated bands at 40-59 Kd (data not shown). An unexpectedly faint degree of phosphorylation is seen with K562 (Lane 9) which might be expected considering the marked phosphorylation of the CD33 ligand in K562 if a reciprocal phosphorylation mechanism between marker and ligand were active. The broad band seen in the U937 cells suggests that several phosphoproteins may associate with CD33 (Lane 10). These data suggest that CD33 becomes more heavily phosphorylated along the myeloid pathway. Thus, the phosphorylation of CD33 could play a direct role in the growth and differentiation of myeloid cells.

4.3.2 *Src*-related kinases associate with CD33

CD33 immunoprecipitates from U937 lysed in NP-40 alone were then tested to determine whether potentially associated phosphoproteins could withstand a more stringent detergent, as phosphoproteins which are precipitated under these conditions more likely result from stable associations (personal communication C.E. Rudd). CD33 was shown to precipitate a strong band at p56/59 and, in order of decreasing activity, p32, p66 and p110 under these conditions (Figure 4.3A, Lane 2).

To determine whether CD33 was associating with an SH2 domain containing *src*-related kinase, a reprecipitation analysis was performed on U937 cells (Figure 4.3B). Kinased proteins were eluted from the protein A sepharose under denaturing conditions and reprobed with anti-*lyn* , anti- *fyn* , anti-c-*fgr*, monoclonal anti-CD33 and anti-phosphotyrosine antibodies. Anti-phosphotyrosine antibodies confirm that CD33 is being phosphorylated on tyrosine residues (Lane 4) as well as by *c-fgr* (Lane 6), *fyn* (Lane 5) and, to a lesser extent, *lyn* (Lane 3). CD33 Mabs reprecipitated no signal at 56/59 Kd, confirming that the proteins immunoprecipitated at 56/59 Kd are*src*-related kinases (Lane 7). The higher level of phosphorylation shown for anti-phosphotyrosine relative to the individual kinases suggests that several kinases associating with CD33.

Figure 4.3C shows that the 56/59 Kd protein is precipitated at a greater level than the 32 Kd, 66-67Kd or the 110Kd protein in CD33 kinase assays (Lane 2). In addition, anti-*hck* (Lane 5) is also shown to associate with CD33 and is able to precipitate slightly greater levels of tyrosine-phosphorylated proteins than anti-phosphotyrosine antibodies (Lane 4). Monoclonal CD33 antibodies reprecipitate no

phosphorylated protein controlling for antibody cross-reactions (Lane 3). *Hck* and *c-fgr* appeared to predominate in several repetitions of the reprecipitation followed by *fyn* and then *lyn*. These results suggest that CD33 associates with several *src*-related kinases which potentially compete for available YXXL and YXXV sites on the CD33 cytoplasmic tail.

In several kinase assays, a weakly phosphorylated band appeared at 66-67Kd suggesting direct phosphorylation of CD33. An unexpected finding appeared when non-heat inactivated serum was used to grow the cells. In this assay, the kinase activity at 56-59Kd was markedly decreased, and that of the 66-67Kd band increased (Figure 4.3D, Lane 2). Heat-inactivated serum is thought to stimulate better cell growth, potentially through aggregated IgG which binds FcRs with a higher affinity than unaggregated IgG. Thus, cells grown in FCS which was not heat-activated could more closely approximate resting cells. Although monoclonal CD33 antibodies could not reprecipitate this band, polyclonal CD33 antibodies would be useful to examine whether this band represents autophosphorylated CD33. This effect suggests a reciprocal regulatory mechanism exists between associated kinases and potentially autophosphorylated CD33.

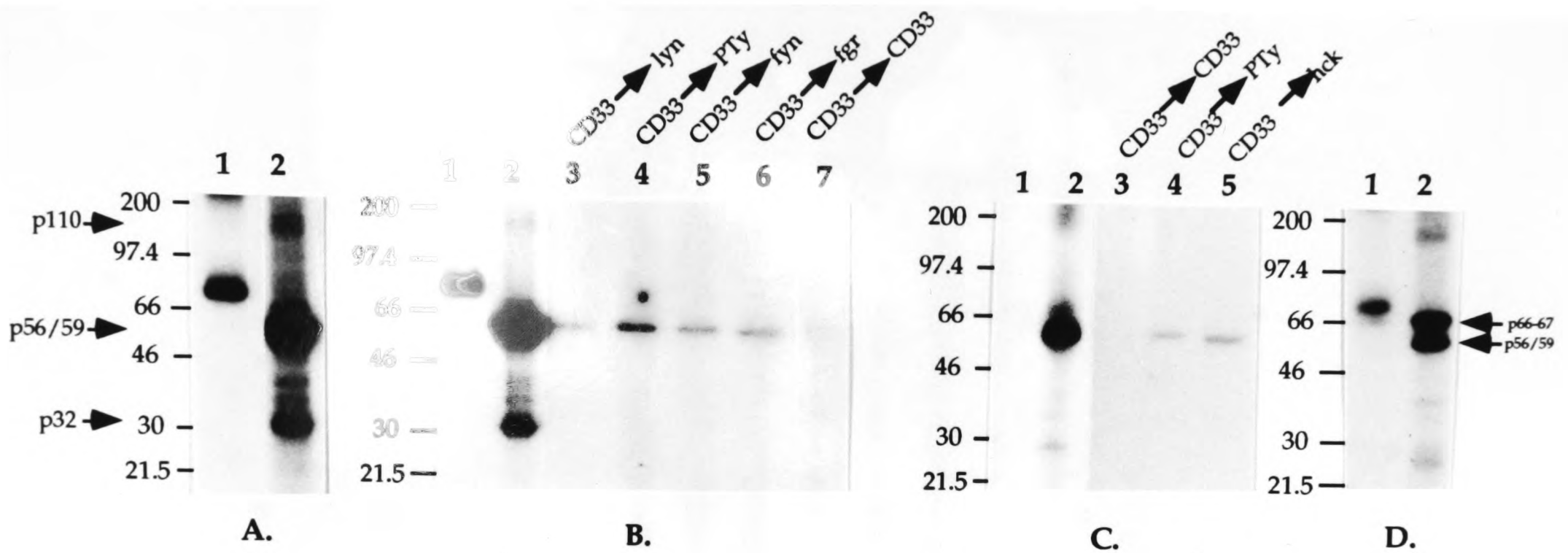


Figure 4.3. CD33 is phosphorylated by several *src*-related non-receptor tyrosine kinases in U937. A. A phosphotransferase assay was performed on immunoprecipitations of goat anti-mouse Ig (Lane 1) and monoclonal CD33 antibodies (Lane 2). CD33 precipitates phosphoproteins at p32Kd, p56-p59 Kd, p66-67 and a faint one at 110Kd (11 C, Lane 2). **B.** CD33 can associate with *fyn, c-fgr,* and *lyn* on tyrosine residues. CD33 kinase assay supernatants were denatured and reprobed with the *src*-related tyrosine kinases anti-*lyn* (Lane 3) anti-phosphotyrosine (Lane 4) anti-*fyn* (Lane 5) and anti-*c-fgr* (Lane 6). Anti-CD33 antibodies show that CD33 does not coprecipitate material at 56/59Kd (Lane 7). The 72 Kd phosphoprotein precipitated with goat anti-mouse Ig is independent of CD33 phosphoproteins. **C.** *Hck* can also associate with CD33. Kinase assays were performed on immunoprecipitates using goat anti-mouse Ig (Lane 1) and anti-CD33 antibodies (Lane 2). Denatured kinase assay supernatants were reprobed with anti-CD33 (Lane 3); anti-phosphotyrosine (Lane 4) and anti-*hck* (lane 5) antibodies demonstrating tyrosine phosphorylation and *hck* activity (Lane 3). **D.** CD33 precipitates increased quantities of p66-67 and decreased quantities of p56/59 (Lane 2) when cells are grown in non-heat inactivated serum.

4.4 Immunoprecipitation of phosphoproteins with Fc-CD33

Following phosphorylation experiments with anti-CD33 antibodies, we asked whether Fc-CD33 might, upon binding to a potential ligand, produce distinct phosphorylation activity. Figure 4.4 shows that Fc-CD33 precipitates the most heavily phosphorylated bands at about 56 Kd, 59 Kd and 100 Kd from KG-1a (Lane 2), K562 (Lane 6), and HL-60 (Lane 10). KG-1 (Lane 4) and U937 (Lane 8) more weakly precipitate the band at 59 Kd and a faint amount of material at 56 Kd. U937 lacks the 56 and 100 Kd band (Lane 8) while KG-1 shows it (Lane 4).

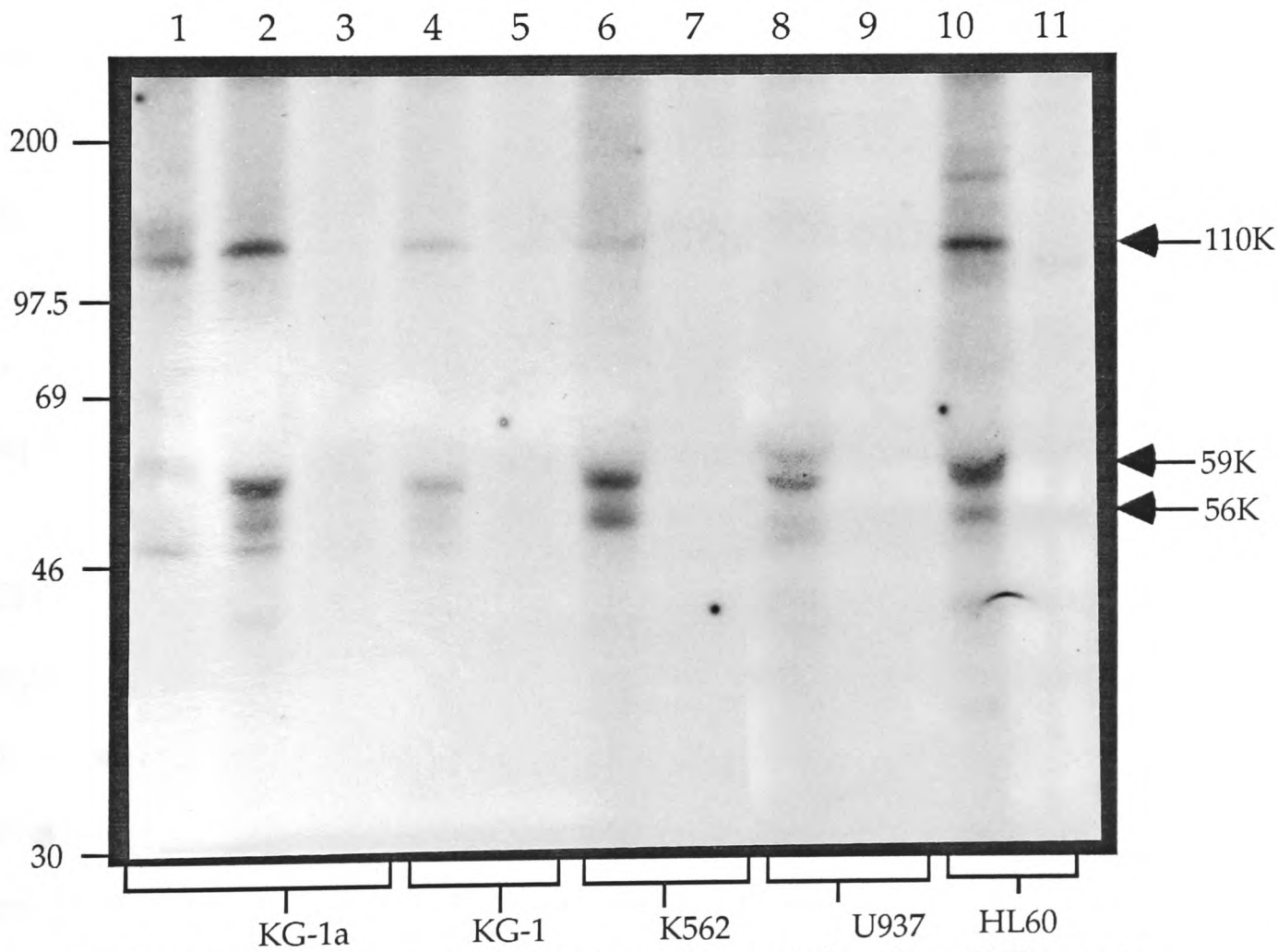


Figure 4.4. CD33 ligands precipitate phosphoproteins in cell lines as differentiation progresses. Fc-CD33 Phosphorylation activity is most pronounced in KG-1a (Lane 3), K562 (Lane 7) and HL-60 (Lane 10) and less significantly in KG-1 (Lane 4) and U937 (Lane 8). Fc-CD34 precipitates distinct phosphoproteins in KG-1a (Lane 1) and no phosphoroteins are present in the mouse Ig negative controls (Lanes 3, 5, 7, 9, and 11). Arrow points to the band at about 56-59 Kd common to all cell lines. The 10% SDS-PAGE gel was exposed for two weeks at -80 ℃.

The bands at 56/59Kd could be the *lyn* tyrosine kinase which has been identified in myeloid cells (Torigoe et al. 1992a). That only the 59 Kd band is present in U937 could mean that *fyn* is the associated kinase, as *lyn* usually occurs as a p53/56 Kd doublet (O'Connor et al. 1992). Alternatively, isoforms of *lck* have been found in K562 and HL-60 but not U937 (Rouer et al. 1994) *Lck* could selectively associate with a CD33 ligand in HL-60 and K562 only. p59/64 *hck* (some sources report it as p56/59 *hck* (Hesketh 1995)) could also potentially associate with a CD33 ligand as it is upregulated upon terminal myeloid differentiation (Willman 1991).

To identify each of the above potentially associated kinases precisely, it will be necessary to perform reprecipitation or western blotting experiments using antibodies raised against the N-terminus of each potentially associated *src*-related kinase (Rudd et al. 1993) as was done in Figure 4.7 with the K562 cell line. The above data suggest that the ligand for CD33 can be phosphorylated or that it associates with a molecule that is phosphorylated and that this pattern of phosphorylation is fairly consistent between cell lines.

Fc-CD34 was immunoprecipitated with KG-1a cells (Lane 1) and showed three phosphorylated bands at 47 Kd, 60 Kd and 97.5-105 Kd. It is interesting to note that the phosphorylation patterns of the least differentiated KG-1a (Lane 2) and most differentiated HL-60 (Lane 10) are very similar, while cell lines in the middle of the differentiation pathway vary somewhat.

Notably, U937 demonstrated the highest percent positive fluorescence in binding to Fc-CD33 in the FACS analysis (Figure 3.10 A and B) and it also shows the highest level of phosphorylation activity with CD33 and significant phosphorylation with

Fc-CD33. It is possible that the phosphorylation of the ligand of CD33 occurs in inverse proportion to the phosphorylation of CD33 and vice-versa. A caveat is the marked contrast between the comparatively low level of CD33 ligand phosphorylation in U937 (monocytic) (Lane 8) which is in inverse proportion to the expression level of CD33 on this cell line (Figure 3.10b). When comparing this with the very high level of CD33 phosphorylation in U937 (Figure 4.1) it could be argued that the phosphorylation of CD33 appears to increase as that of its ligand decreases in the U937 cell line.

4.5 Fc-CD33 ligands associate with *lck* in K562

The panning results with Fc-CD33 gave the greatest number of positives when probing a transfected K562 cDNA library. Thus, this cell line was chosen to attempt to identify the kinase which precipitated with Fc-CD33 ligand using a rabbit anti-*lck* antisera raised against the N-terminal portion of the *lck* molecule and anti-phosphotyrosine antibodies for reprecipitation analysis from the original phosphotransferase pellets precipitated with Fc-CD33 from K562 cells.

Cell line 56 Kd	Pre I A 567bp*** 56 Kd++	I A 263 bp 56 Kd	I B 165bp 56Kd	I C 186bp 56Kd	II A * 180bp 56Kd	II B 69bp 56Kd
Healthy BM/PBL + Remission	-	+ (7%)	+ (9.2%)	+ (1%)	+ (67%)	+ (15.4%)
AML-0+	-		+ (29.1%)**			+ (6.5%)
AML-1	-		+ (21.5%)			
HL60 (My)	-	-	-	-	+	+
K562 (My/Ery)	+	-	+	-	-	-
U937 (My)	-	-	-	-	-	-
Daudi (B)	+	+	+ (91%)	+ (9%)	-	-
Raji (B)	+	+	+ (91%)	+ (9%)	-	-
CEM (T)	-	-	+ (34%)	-	+ (43%)	-
Jurkat (T)	-	-	+ (34%)	-	+ (43%)	-
Molt 4 (T)	-	-	+ (34%)	-	+ (43%)	-

Table 4.1 Type I and Type II variants of *lck* in Healthy and Leukemic Cells
* major form identified in T cells
** increase in this type not observed in AML-2-5; comparable high expression of Type I B also observed in B-ALL and T-ALL.
+ AML-0 and AML-1 are undifferentiated blasts
++ all weights are approximate based on the known weight of *lck* and the assumption that alternatively spliced forms share the same weight as the known form.

Six forms of the *lck* molecule have now been identified, at least two of which are upregulated in AML (Table 4.1) (Rouer et al. 1994). Percentages represent the densitometric quantity of each form of *lck* in each cell. Only two forms of the six known have been sequenced; thus, the possibility of varying sizes of isoforms cannot be ruled out. Four forms of *lck* are generated by a proximal (Type I) promoter and two forms by a distal (Type II) promoter. Normal T cells mainly express*lck* from the distal Type II promoter, malignant cells express *lck* from both promoters and non-lymphoid cells express *lck* from the proximal Type I promoter. The major transcript of *lck* found in T cells and T cell leukemias, is Type IIA. The proximal promoter, 25

Kb upstream, has been demonstrated to generate variants of *lck* at low levels in HL-60 (Type IIA and IIB) while U937 lacks all *lck* expression. K562 generates a very high level of the 567 bp immature precursor Pre IA and a low level of the 165 bp Type IB. The major Type IIA RNA transcript of *lck* is 180 bp. All forms of *lck* are thought to precipitate proteins at 56 Kd on SDS-PAGE .

Fc-CD33 in both vectors was used to attempt to reprecipitate phosphoproteins from K562 (Figure 4.5). From Fc-CD33Rg immunoprecipitates, both the anti-phosphotyrosine (Lane 5) and anti-*lck* (Lane 6) antibodies reprecipitated a major band at about 47-56 Kd and a minor one at 110 Kd. Fc-CD33 in the pIG1 vector reprecipitations showed mainly the 110Kd species with both the anti-phosphotyrosine antibody (Lane 4) and a slightly stronger band with the anti-*lck* antibody (Lane 5) with a very faint band in both lanes at 47-56 Kd. The differences between Fc-CD33 in precipitating ligand associated with *lck* could be due to slight precipitation of the Fc-CD33 probe in solution or slight differences in the vectors which produce a conformational change in the recombinant molecules produced. Alternatively, the reprecipitated *lck* proteins may have degraded slightly in the assay preventing binding to the antibodies.

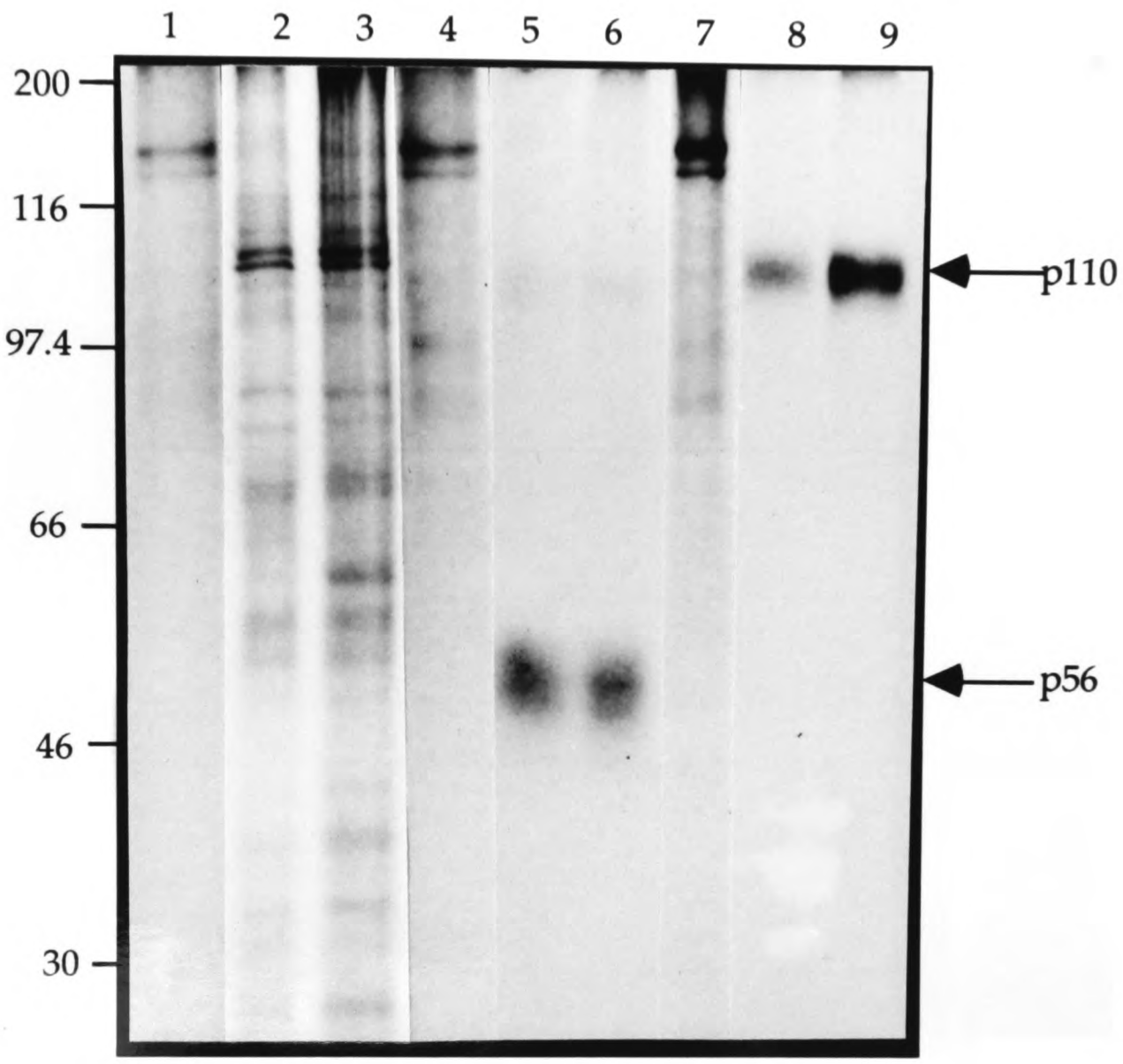


Figure 4.5. Phosphoproteins at 110 Kd and 56 Kd can be reprecipitated with the anti-*lck* antisera from phosphorylated Fc-CD33 immunoprecipitates. Lane 1- rat Ig, negative control; Lane 2 - rat-anti-phosphotyrosine antibody; Lane 3 - anti-phosphotyrosine antibody; Lane 4 - Fc-CD33-Rg; Lane 5 - Fc-CD33Rg reprecipitated with the anti-phosphotyrosine antibody; Lane 6 - Fc-CD33Rg reprecipitated with the anti-*lck* antibody; Lane 7 - Fc-CD33 (pIG1); Lane 8 - Fc-CD33-pIg1 reprecipitated with the anti-phosphotyrosine antibody; Lane 9 - Fc-CD33 (pIg1) reprecipitated with the anti-*lck* antibody. Arrows point to the reprecipitated phosphoproteins. 12% SDS-PAGE gel was exposed for 3 days at -80°C.

The proteins reprecipitations likely represent the Type pre IA and IB forms of *lck* described above in K562. The broad bands precipitated suggest that a slight amount of degradation of the protein could have taken place in the reprecipitation. It is also

possible that the forms of *lck* produced in other cell lines could vary slightly in weight, although different promoters should not alter the weight of the protein product.

The cDNA of the K562 transcripts of *lck* have not been sequenced, so it is possible that the coding region of *lck* is different from the Type IIA form. It has been proposed that the 13 exons of the *lck* gene can undergo exon shuffling (Rouer 1989). Thus, it is possible that this form has an altered weight. The high band is likely the band that is commonly seen when nearly all of the nonreceptor tyrosine kinases are immunoprecipitated. It may be an inactive kinase (C.E. Rudd personal communication) or it could be the p110 catalytic subunit of PI-3 kinase which has been shown to bind to the SH3 domain of p56*lck* (Prasad et al. 1993). Attempts to clone a band in this position have proven unsuccessful (C.E. Rudd, personal communication) but its identity could be shown with additional reprecipitations using the appropriate antibodies. Alternatively, it is possible that the anti-*lck* antibody cross reacted with an unrelated phosphoprotein. This is unlikely considering the specificity of the antibody used.

This is the first known demonstration of phosphorylation activity precipitated with the anti-*lck* antibody in K562 cells. Notably, upon remission, patients revert to expressing only the Type II A transcript of *lck* (Rouer et al. 1994). These data argue that *lck* may play a role in the function of a ligand for CD33 in leukemic K562 cells. A potential association of *lck* with the CD33 ligand should be examined in other myeloid cells known to express *lck* such as HL-60.

4.6 Discussion

These results demonstrate that CD33 associates with increasing levels of phosphoproteins as the differentiation pathway progresses and reveals that CD33 can precipitate proteins at p32, p56/59, p66-67 and p110. CD33 reveals a potential ITAM in its cytoplasmic tail (Figure 4.1a and 4.1b) and can precipitate appreciable quantities of phosphoproteins in the myeloid precursor KG-1b and promonocytic U937 (Figure 4.2). The homology of CD33 with the phosphorylated lectin-like molecules CD22 and MAG suggest that phosphorylation could play an integral part of its function intracellularly. CD33 ligands can also precipitate phosphoproteins *in vitro* which could associate with their cytoplasmic tails. Inverse proportions of these associated proteins are precipitated by Fc-CD33 and CD33 Mab in KG-1a, K562 and U937 which suggests that phosphorylation could serve to regulate the mechanism of activation between CD33 and its ligands.

Reprecipitation analysis confirms that c-*fgr*, *fyn*, *hck* and *lyn* are capable of association with CD33 (Figure 4.3B and C). CD33 can precipitate a band at 66-67Kd which may be an autophosphorylated form of the protein (Figure 4.3D). Recent findings suggest that more immature hematopoietic cells express fewer kinases (*c-yes*, *fyn*, *lyn*) and that *hck*, *src* and *lck* expression is acquired upon further differentiation (O'Connor et al. 1992). In addition, the HIV-1 Nef protein was recently shown to be required for *hck* and *lyn* recognition in U937 cells infected with HIV-1 (Saksela et al. 1995). That *hck* associates with CD33 could also implicate it in a role for HIV association in U937. CD33-associated phosphorylation in promonocytes suggest that it may also play a role in the $CD33^+$ antigen-presenting cells including

monocytes, possibly dendritic cells, and macrophages (reviewed in (Peters et al. 1996).

CD33 has been shown to behave like a lectin (Freeman 1995) which could be consistent with a role for it in activating its ligand in intra or intercellular contact, as other lectins (eg. Con A and PHA) have been shown to play this role in T cell activation. Conversely, it is possible that the ligand for CD33 could activate it upon cell adhesion. Fc-CD33 ligands precipitate phosphoproteins at 56/59 Kd and 110 Kd which strongly suggests an association with the *src*-related kinases (Figure 4.4). The band at 56Kd could be one or more of the *src*-related tyrosine kinases p53/56*lyn*, p55c-*fgr* or p56/59*hck* which have been identified in myeloid cells (Torigoe et al. 1992a; Ziegler et al. 1987) In K562, *lck* is identified as one of the *src*-related kinases to associate with CD33 ligands.

A model is presented with potential modes of interaction for *hck, c-fgr, fyn* and *lyn* with CD33 (Figure 4.6, the CD33 ECD is adapted from (Barclay 1993)). The first molecule of CD33 shows a generalised mechanism of binding between the SH2 domain of either *c-fgr, fyn, hck* or *lyn* and either the YXXL or YXXV regions of CD33 (Figure 4.6A). The second molecule of CD33 attempts a more precise mechanism of the way in which the kinases interact with CD33 based on the known structural model of *lck* in which the SH2 and SH3 domains have been shown to form a dimeric hydrophobic pocket on activation (Eck et al. 1994) and the recent structural work on ZAP-70 in which the two SH2 domains of the kinase bind specifically to the Y and L in the YXXL motifs contained within the ζ-chains of CD3 (Hatada et al. 1995) (Figure 4.6B).

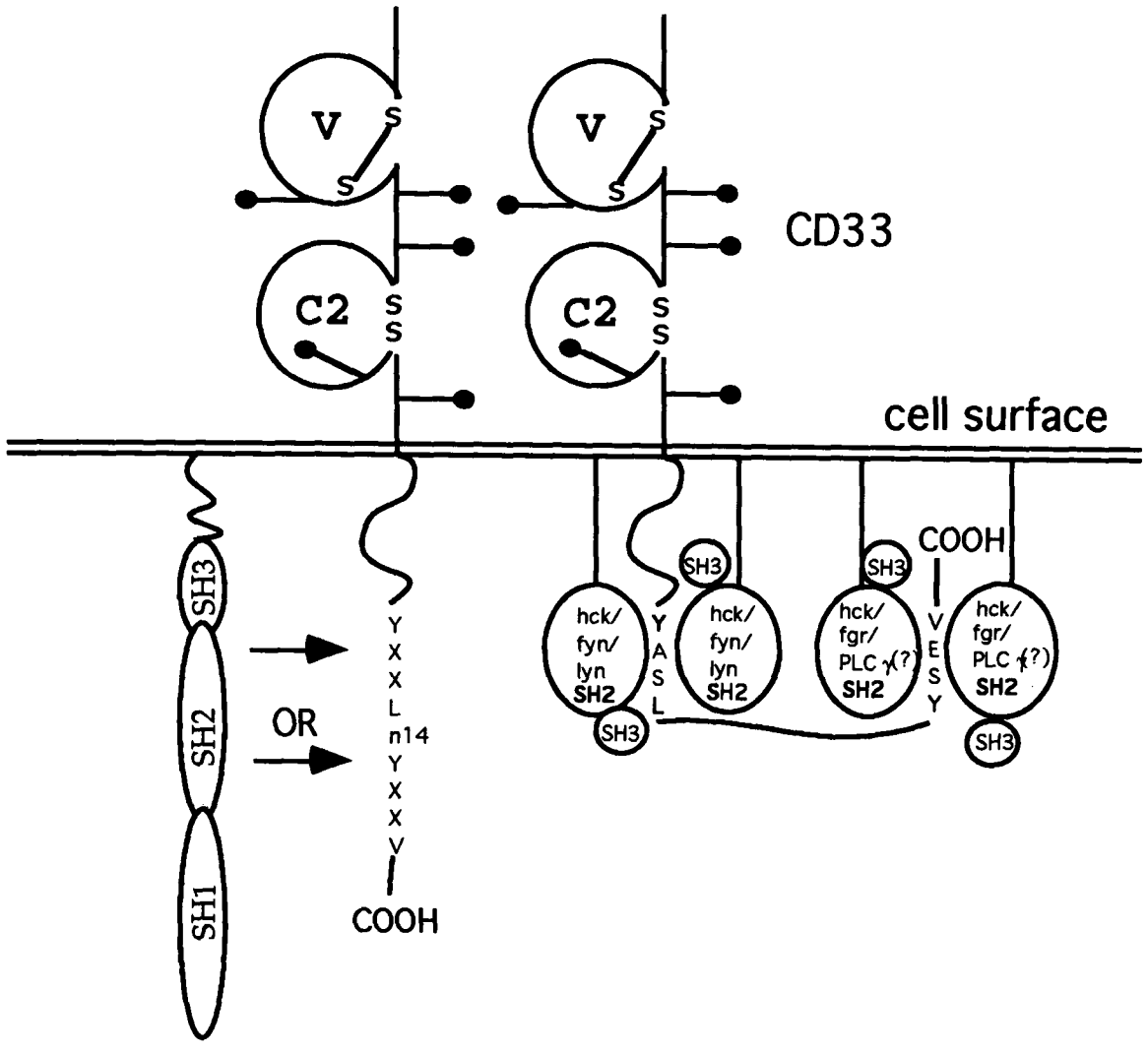


Figure 4.6 A model for the interaction of SH2 domains with the YXXLn14YXXV regions in CD33. The model shows how SH2 domains from a number of *src*-related tyrosine kinases can associate with the cytoplasmic tail of CD33 following phosphorylation. CD33 is adapted from the original model (Barclay 1993).

Recent information suggests that there is a negligible difference in the affinity of SH2 domains for YXXL, YXXV or YXXI motfs (J. Ladbury, personal communication). The kinases identified in this study have only one SH2 domain each which suggests that, if the kinases form homodimers upon activation, one of the SH2 domains interacts with the cytoplasmic tail, while the other serves to provide another hydophobic region to allow maximal activation following a competition of the protein tyrosine kinases for available CD33 sites of interaction.

Chapter 5

CD34 Binding and Ligand Studies

5.0 Specific aims

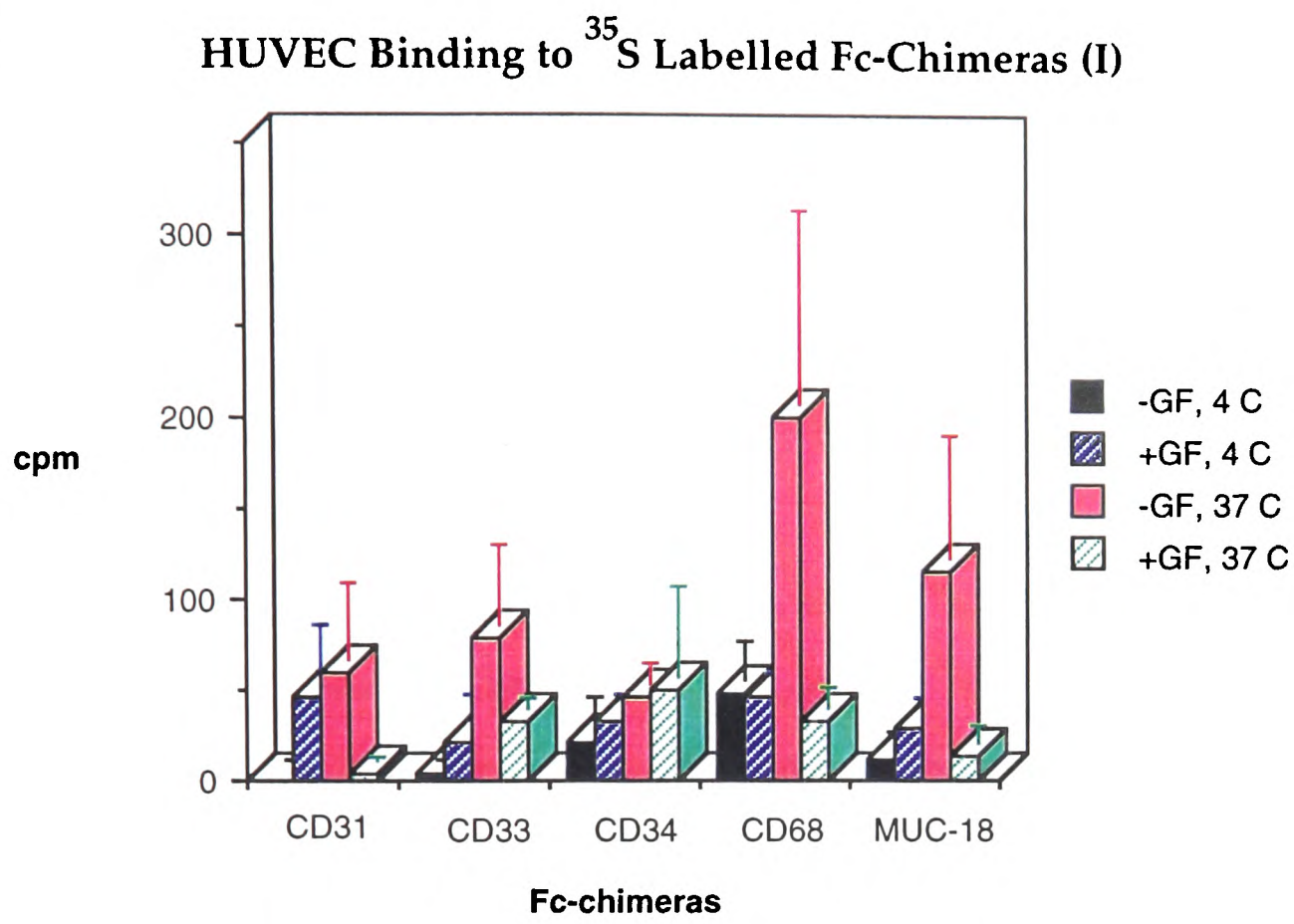
The aim of these experiments using Fc-CD34 was to assess potential binding of the chimeric molecule using the approaches of panning, cell surface and metabolic labelling, *in situ* adhesion, and phosphorylation studies. The results of these methods will be described to assess the adhesion ability of Fc-CD34 and some potential characteristics of its ligands.

5.1 Binding studies on Fc-chimeras to HUVEC

To assess whether different ^{35}S-labelled Fc proteins could bind to HUVEC, a comparative binding study was performed at 4°C and 37°C in the presence and absence of the growth factors (GFs) Tumour Necrosis Factor (TNF) and Interleukin-1β (IL-1β). TNF-α is a pleiotropic cytokine which acts as both a positive and a negative regulator of myeloid cell proliferation and differentiation (Degliantoni et al. 1985). It has also been shown to inhibit stem cell factor (SCF) induced proliferation of bone marrow progenitor cells *in vivo* (Rusten 1994). Endothelial cells express TNF receptors and TNF activation induces the expression of VCAM-1, ICAM-1 and E-selectin in HUVEC (Cockerill 1994). Both IL-1 and TNF are secreted by mononuclear phagocytes upon stimulation with inflammatory agents and IL-1 can also be produced by other cell types (reviewed in (Dinarello 1985)). In HUVEC, treatment with IL-1 or TNF has been shown to stimulate migration of neutrophils across endothelial monolayers (Furie 1989). We thus used these cytokines to examine the ability of several available Fc-proteins to bind to HUVEC in the presence and absence of IL-1β and TNFα. To assess the Fc-CD34 binding to HUVEC in the context of other Fc-chimeras, primary cells

were isolated, grown to confluence and divided into four 24 well plates. The following day, two plates had IL-1β and TNF added. The plates were then incubated with various ^{35}S-labelled Fc-chimeras and left to incubate at either 4°C or 37°C for one hour. Wells were washed twice, lysed and counted by scintillation. The values were then assessed for binding behaviour (Figure 5.1).

A.



B.

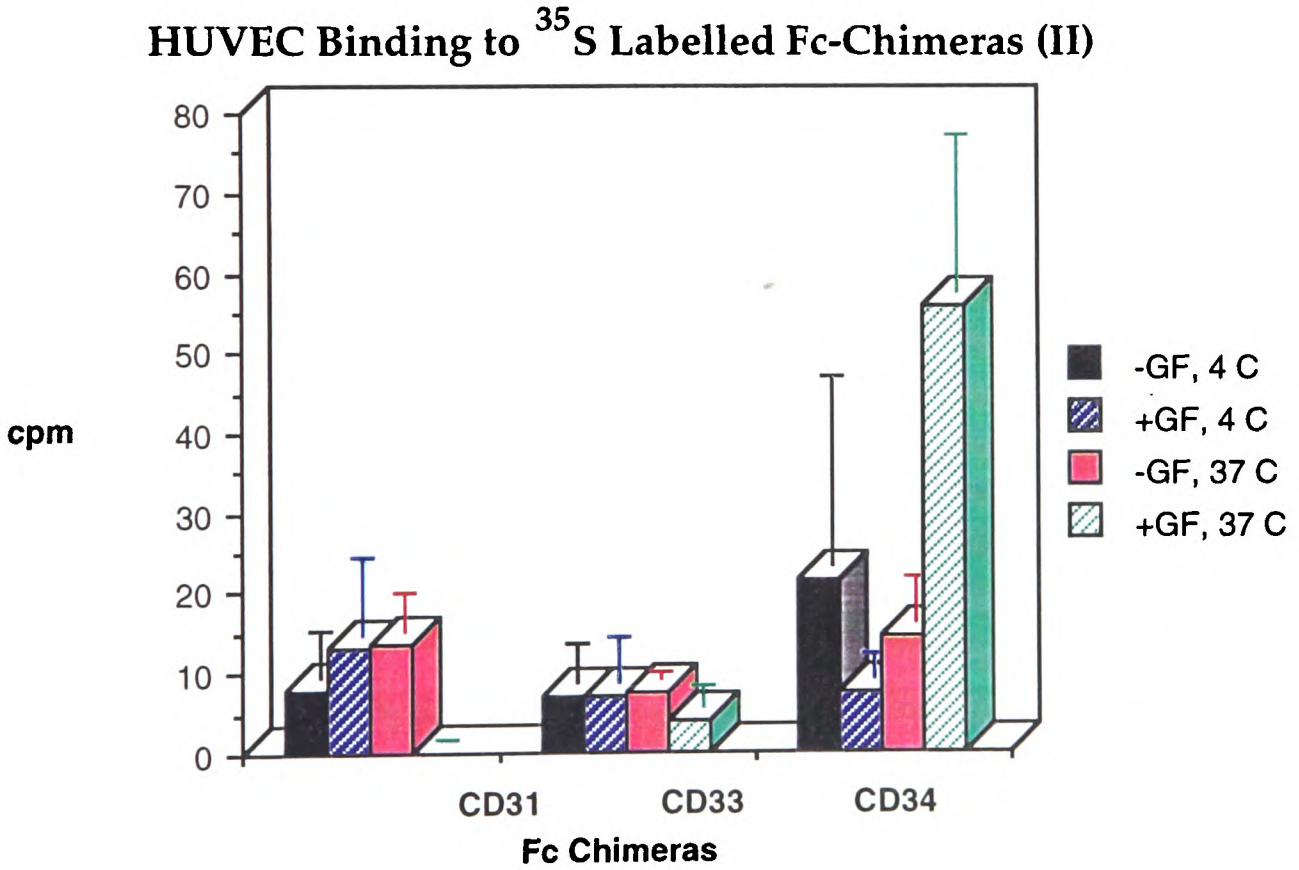


Figure 5.1A and B. CD34 demonstrates binding to HUVEC at 37°C in the presence of growth factors. Primary HUVEC were incubated in the presence and absence of the growth factors IL-1β and TNFα prior to the addition of ^{35}S labelled Fc-proteins. Fc's were incubated for 1 hr. at 4°C and 37°C. Cells were then lysed and counted by scintillation. Input counts were 500 cpm/well. n=3 for each mean value and the error shown represents standard deviation (SD). Mean values have had background counts subtracted.

In both Figure 5.1 A and B only mean values above 50 cpm, prior to subtracting error, are taken to be potentially positive as this represents 10% of the input counts. In both experiments, Fc-CD34 demonstrates mean binding at 26.0 and 46.0 cpm after subtracting SD in the presence of growth factors at 37°C. All other Fc-CD34 binding is read as negative. In contrast, the other Fc-proteins demonstrate the highest mean value of binding in the absence of growth factors at 37°C, the maximum shown by the heavily O-glycosylated mucin Fc-CD68 at 148.0 cpm (Holness and Simmons 1993) followed by Fc-MUC-18 at 83.0 cpm (Lehmann 1989). It is interesting to speculate that the activation produced by TNF and IL-1 could produce a negative charge to repel binding of these anionic molecules. These assays provided sufficient evidence that Fc-CD34 was capable of some binding to HUVEC to pursue further studies with these cells.

5.2 Fc-CD34 binds to a 110 Kd protein on iodinated HUVEC

Since CD34 is expressed on HUVEC it was reasoned that it could recognise a potential ligand on these cells. Thus, to compare the binding characteristics of several Fc proteins, HUVEC were iodinated *in situ* and subjected to sequential immunoprecipitations with Fc-CD31, Fc-CD33, Fc-CD68, Fc-MUC-18 and an anti-CD31 Mab (JC70A) as a positive control, as CD31 is strongly expressed on HUVEC (Figure 5.2). The detergent used for all studies in this chapter is a 50/50 mix of 1% NP-40/1% Digitonin unless otherwise specified. Lysate was limited, so following the immunoprecipitation of the first Mab or Fc probe, a second Fc probe was added.

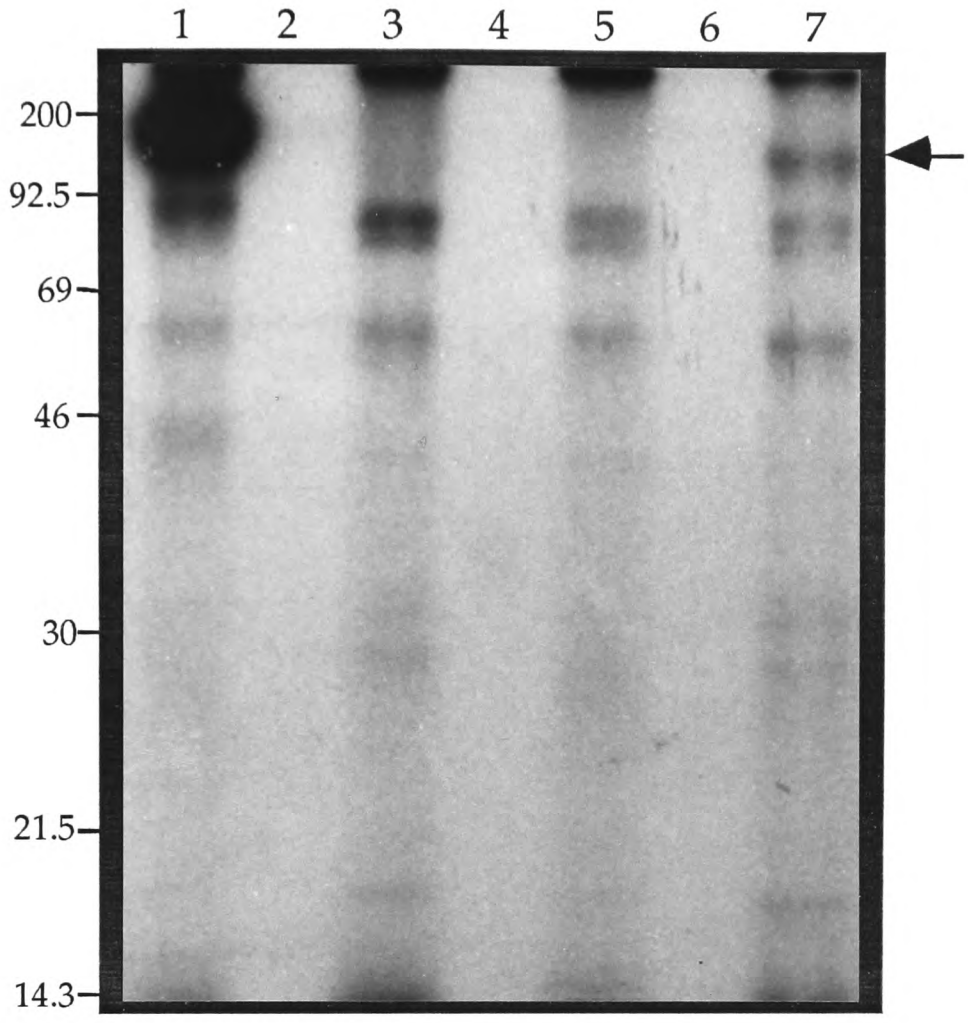


Figure 5.2 Fc-CD34 is shown to uniquely label a 110Kd band in immunoprecipitates of ^{125}I labelled HUVEC. Following immunoprecipitation, Fc-CD34 reveals a unique band (Lane 7) not shown with Fc-CD33 (Lane 5) or Fc-CD31 (Lane 3). Lanes 1,3 and 5 were incubated 1hr. with probe; Lane 7 was probed with a second antibody and incubated overnight. Lane 1 - anti-CD31 Mab, JC70A; Lane 3 - Fc-CD31; Lane 5 - Fc-CD33; Lane 7 - Fc-CD33 reprobed with Fc-CD34; Exposure time was 13 days at -80°C.

JC70A precipitated a heavily labelled unique band at 120Kd (Lane 1) while Fc-CD31 (Lane 3) and Fc-CD33 (Lane 5) precipitated no unique bands. Fc-CD34 was added to the lysate containing Fc-CD33 and was shown to precipitate a unique band at about 110Kd (Lane 7) when compared with the Fc-CD33 probe alone (Lane 5). Because the band immunoprecipitated with the Fc-CD34 molecule roughly corresponded to the weight of CD34, it was thought that CD34 could bind in a homophilic manner. Thus, Fc-CD34 binding properties were further tested using ^{35}S metabolic labelling and immunoprecipitation.

5.3 Fc-CD34 binding to HUVEC is divalent cation-dependent

Calcium dependence has been shown to be important in ligand binding for several molecules to endothelium including L-selectin (Norgard 1993). We therefore tested the ability of CD34 to bind in a calcium dependent or independent fashion by iodinating HUVEC *in situ* followed by a wash in media containing Ca^{++} and Mg^{++} and one in PBSA. The cells were then lysed for immunoprecipitation in the presence or absence of 0.4 M EDTA added to the lysis buffer (1% 50/50 NP-40/Digitonin). Figure 5.3 demonstrates that Fc-CD34 precipitates unique bands at 66Kd and 110 Kd in the absence of EDTA (Lane 2), approximately the same size as the 70Kd and 110 Kd phosphoproteins immunoprecipitated from HUVEC (Figure 5.7 A and B) and the 110 Kd protein shown from iodination (Figure 5.2). These proteins are not precipitated by Fc-CD34 in the presence of EDTA (Lane 4) nor are they present in the control lane with (Lane 4) or without (Lane 1) EDTA added. A weakly labelled band is precipitated at 70Kd in the control without EDTA as is a fainter band just above but distinct from the 110 Kd band (Lane 1).

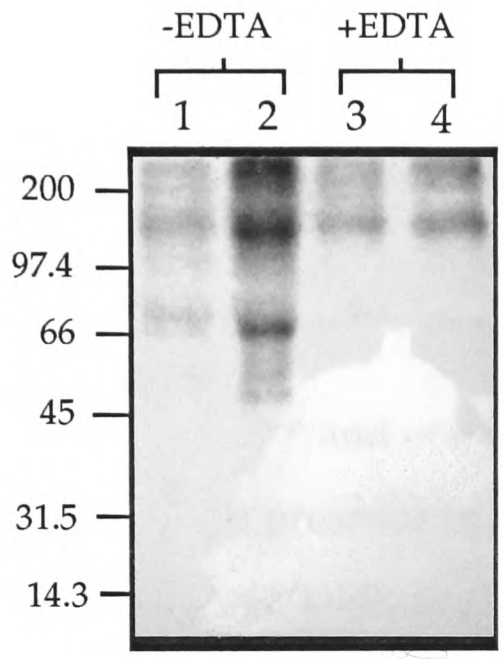


Figure 5.3 Fc-CD34 precipitates potential ligands at 66Kd and 110Kd in the absence of EDTA on ^{125}I labelled HUVEC. Lanes 1 and 3, rat immunoglobulin; Lanes 2 and 4, Fc-CD34Rg. Lanes 1 and 2 were incubated in the absence and Lanes 3 and 4 in the presence of 0.4M EDTA. Fc-CD34 binding to HUVEC thus appears to be divalent-cation dependent. Exposure time is 2 days at -80°C.

EDTA will chelate all divalent cations (Ca^{++}, Mg^{++}, Mn^{++}) and further work would be needed to distinguish which the divalent-cation dependent 66 and 110Kd proteins precipitated by Fc-CD34 bind most tightly. This experiment could be repeated in the presence and absence of TNFα and IL-1β to test for a higher level of binding on activated endothelium.

5.4 Fc-CD34 shows no homophilic binding to myeloid cell lines

To determine whether Fc-CD34 was binding homophilically and to assess a ligand size, CD34 expressing cell lines (KG-1, KG-1a, KG-1b) were

metabolically labelled with ^{35}S and immunoprecipitated with Fc-CD34 and compared with those using the anti-CD34 Mab QBEND10 and control antibodies. QBEND10 is known as a class II anti-CD34 Mab because its carbohydrate epitope is not neuraminidase-sensitive but it can be removed by O-sialoglycoprotease (Sutherland 1992). In this experiment, Fc-CD34 demonstrated distinct binding from anti-CD34 Mabs (Figure 5.4). In all three cell lines QBEND10 precipitated a unique band at about 110 Kd (Lanes 2, 6, and 10) when compared with the negative control (Lanes 1, 5 and 9). Fc-CD34 in the Rg vector precipitated faint bands or those that also appeared in the control lanes (Lanes 3, 7 and 11). In contrast, Fc-CD34 in the pIG1 vector precipitated bands unique from the negative control at 50 and 60Kd in KG-1 (Lane 4). In KG-1a (Lane 8) and KG-1b (Lane 12), these bands were also precipitated but the 50 Kd band is also represented in the control. All immunoprecipitates had equivalent amounts of protein and antibody. Thus, the discrepancy in binding between cell lines could be explained by different ligand proteins expressed on each cell line or perhaps differences in glycosylation between each protein bound. This assay definitively shows that Fc-CD34 does not bind homophilically to these $CD34^+$ cell lines.

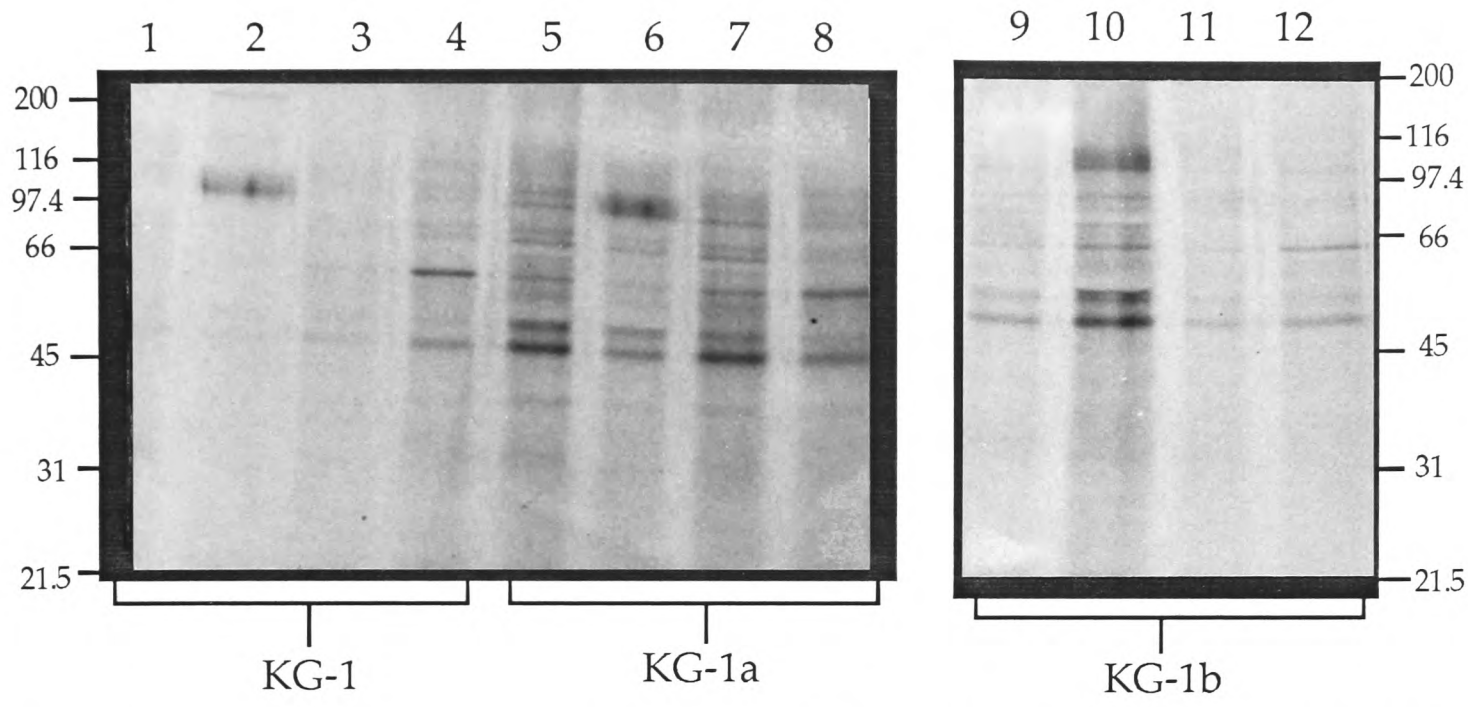


Figure 5.4 Fc-CD34 does not bind CD34 in ^{35}S Labelled immunoprecipitates of cell lines. Lanes 1, 5 and 9 - goat anti-mouse immunoglobulin; Lanes 2, 6 and 10 - anti-CD34 Mab (QBEND10); Lane 3, 7 and 11 Fc-CD34 (Rg); Lanes 4, 8 and 12 Fc-CD34 (pIG1). Figure represents a 4.5 day exposure. 2 x 10^6 cells were labelled with 100 μCi of ^{35}S. Fc-CD34 precipitates bands distinct from anti-CD34 Mab showing that binding is not homophilic in this assay.

5.5 Fc-CD34 binds at a similar level to all candidate cDNA products

As CD34 is a mucin (with an Ig-like domain) it is likely that its binding behaviour is dependent upon correct glycosylation of the transfected molecule to have any relevance to its behaviour *in vivo*. Antibody binding data from anti-CD34 antibodies raised against various epitopes in transfected cells (Watt et al. 1993) would suggest that the glycosylation of CD34 transfectants is similar to native CD34. COS cells lack the α2,6 sialyltransferase, however, (Sgroi 1993; Stamenkovic 1991) which makes it unlikely that the sialylation of Fc-CD34 is physiologically relevant for all cell types. This, and the general low-affinity behaviour of Fc-chimeras may

account for the inability of Fc-CD34 to bind consistently to candidate cDNA libraries in the panning exercise.

Briefly, panning addresses the ability of a probe, either an antibody or Fc-ECD chimera, to bind to discrete proteins on the surface of transfected COS cells. Two to three repetitions of panning should sufficiently enrich the panned cDNAs of interest to produce positive fluorescence and/or staining with the initial Fc-ECD in Flow Cytometry. Panning with the Fc-CD34 constructs was carried out with a number of myeloid cDNA libraries. Bone Marrow and HUVEC cDNA transfectants gave the highest number of cells panned in the first and second rounds (3-10 per dish after the second and third round). As with CD33, however, a completely homogenous clonal population was not achieved, but several candidates were subjected to staining from an original pool that showed at least 25% size homogeneity at the miniprep stage.

Several of the candidates identified from minipreps were shown to have coligated to the stuffer originally present in CDM8 which is normally completely excised prior to the insertion of cDNAs when the cDNA library is made. The stuffer was excised with the appropriate restriction enzyme(s) and the insert was religated using the endogenous linker and sequenced to check that the stuffer had been removed prior to transfecting the candidate again. The absence of stuffer made no difference to the ability of Fc-CD34 to bind to these candidates (Figure 5.5).

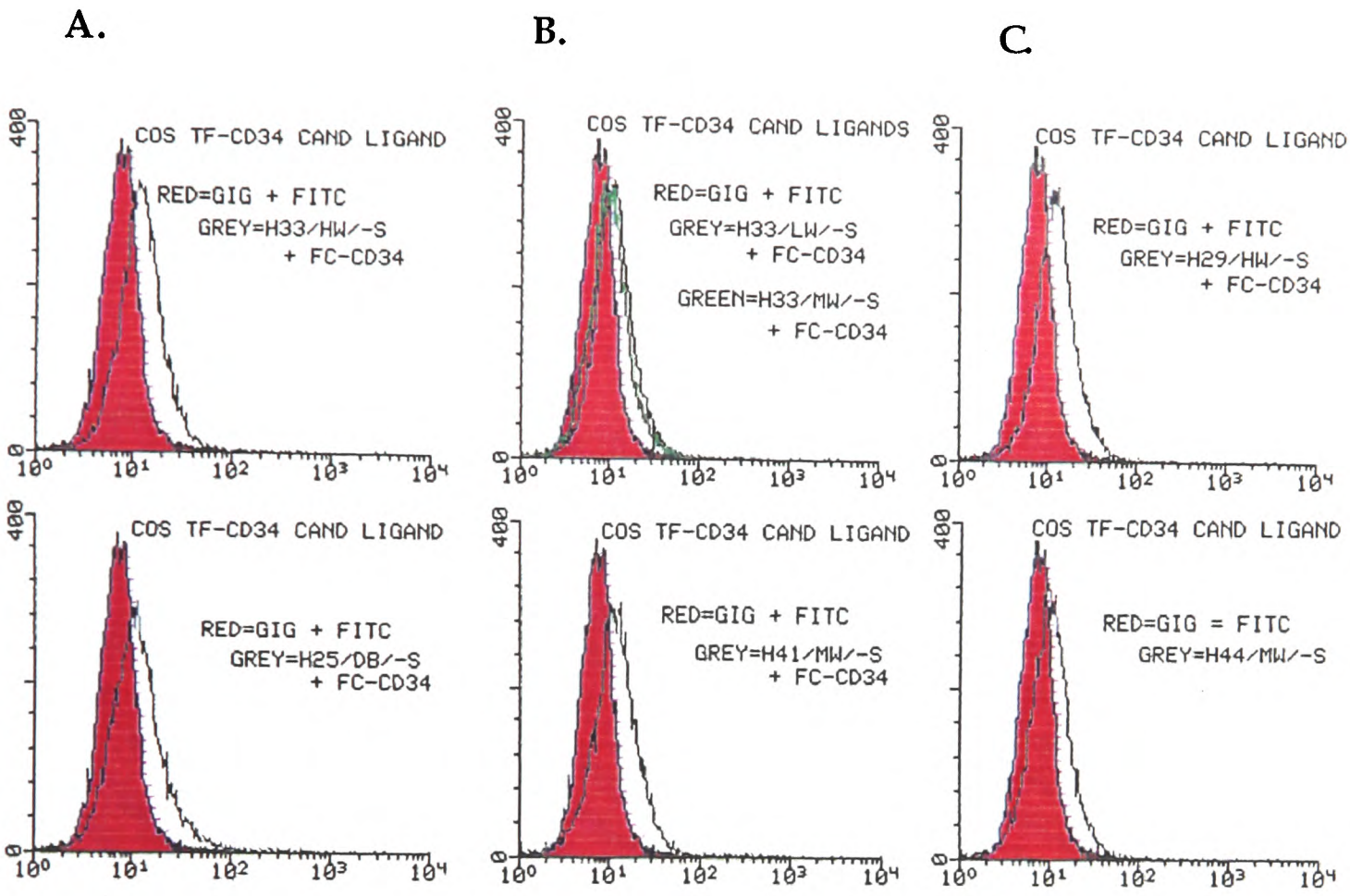


Figure 5.5 Candidates screened using Fc-CD34 against HUVEC cDNA Transfectants show no significant binding. In each quadrant, the negative control is sham transfected COS stained with goat Ig to block and then goat anti-human Ig (FITC). Total staining for the negative control is 2.5%. Candidate staining levels represent the total for that sample minus the negative control staining (red peak). All candidates shown have the stuffer removed. **A.** H33/high mol. wt. - 13.9%; **B.** grey - H33/low mol. wt., green - H33/medium mol. wt. 7.1%; **C.** H29 - 15.9%; **D.** H25 - 15.2%; **E.** H41 - 15.4%; **F.** H44 - 12.2%. A 1:50 dilution of the Fc-proteins was used for each sample staining.

The level of staining obtainable with Fc-CD33 on full-length pCD33 and Fc-CD34 on full-length pCD34 COS cell transfectants was assessed for homophilic binding (Figure 5.6 A and B). FACS staining was about 22% positive for both

Fc-CD33 and Fc-CD34 against their respective transfectants suggesting that each can bind homophilically to transfectants.

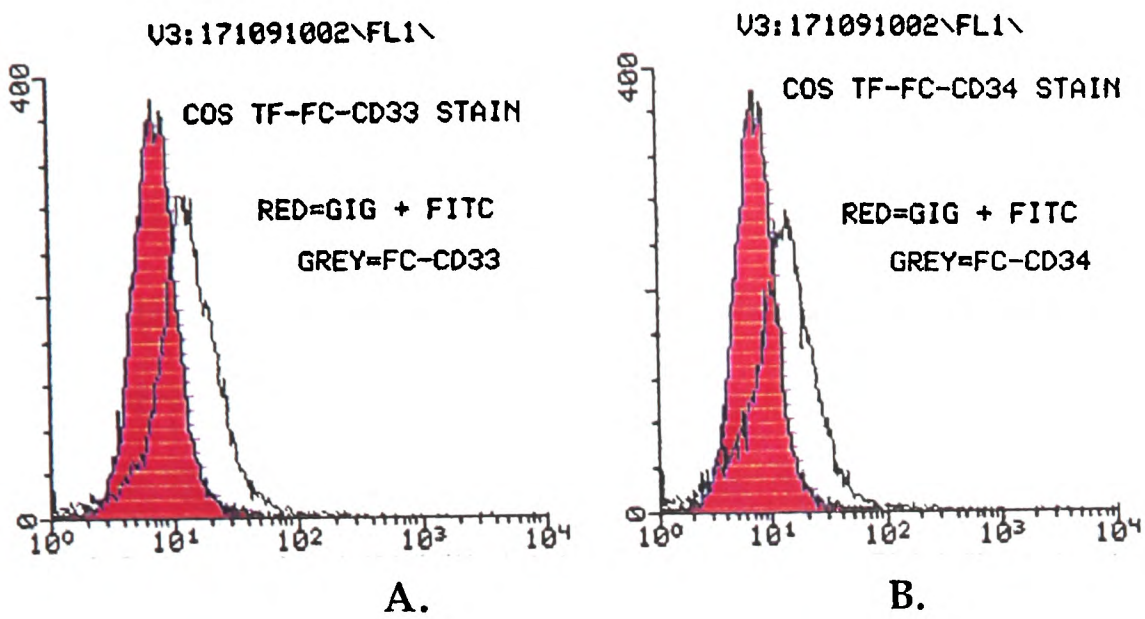


Figure 5.6 Fc-CD33 and Fc-CD34 show no homophilic binding to COS transfectants. A comparison with Fc-CD33 binding to pCD33 is shown for comparison. **A.** Fc-CD33 binding to transfected pCD33 and **B.** Fc-CD34 binding to transfected pCD34 show about 22% positive staining. A 100 μg/ml of the Fc-proteins was used for each sample staining.

Due to the probability that correct CD34 glycosylation is not reproduced by COS cells, and a heterogenous population of potential ligands from panning, it was concluded that an expression vector system was an inappropriate way to isolate a ligand for Fc-CD34. Concurrently, adhesion assays with Fc-CD31 demonstrated that a clustering effect was necessary with these chimeras to get high avidity binding even with a known ligand interaction (Simmons 1993). Thus, direct binding studies were replaced by studies on phosphorylation in which HUVEC could be probed with anti-CD34 Mabs and Fc-CD34 to compare binding interactions and subsequent phosphorylation of the cytoplasmic tail of CD34 and a potential ligand for it.

5.6 Fc-CD34 precipitates a 70Kd HUVEC phosphoprotein

Following the initial binding studies on iodinated HUVEC showing a distinct band at 110 Kd, we tested the possibility that a potential ligand binding to Fc-CD34 in HUVEC could be phosphorylated. CD34 itself had been shown to be phosphorylated by protein kinase C (Fackler et al. 1990) and potential phosphorylation of the ligand could give a clue regarding its identity and function and whether its intracellular phosphorylation behaviour was similar to that of CD34.

The phosphotransferase assay on primary HUVEC was performed with Fc-CD34 and an anti-CD34 Mab to examine the phosphorylation behaviour of the cytoplasmic tail of a CD34 ligand on HUVEC (Figure 5.7 A and B). Any phosphoproteins precipitated could indicate that 1) the ligand itself is phosphorylated; 2) a phosphoprotein, potentially a kinase, associates with the cytoplasmic tail to phosphorylate it or to itself be phosphorylated, or; 3) an irrelevant phosphoprotein is precipitated with the protein of interest. The latter possibility should be minimised by detergent and washing conditions. Studies comparing the CD34 Mab ICH3 with Fc-CD34 (Rg) on primary HUVEC revealed that a discrete band at 69 Kd and, less intensely, one at 116-120Kd, were phosphorylated with Fc-CD34 (Figure 5.7 A, Lane 3) both of which were not present in the control (Lane 1). ICH3 recongnises a sialylated carbohydrate epitope which is neuraminidase and O-sialoglycoprotease sensitive making ICH3 a class I, sialic acid-dependent anti-CD34 Mab (Sutherland 1992). The finding suggested that a ligand for CD34 exists on HUVEC which can coprecipitate phosphoproteins. HUVECs contain no Fc receptors and do not express CD33 making it unlikely that the phosphoproteins are due to an interaction with the Fc portion of the chimera. Precipitation with ICH3

showed a less intense but similar phosphoprotein precipitated at about 69Kd on longer exposure (Lane 2), again providing suggestive evidence that CD34 could be homophilic and heterophilic in its binding.

The phosphotransferase experiment was repeated using identical detergent conditions but using the anti-CD34 Mabs QBEND10 and 4G10 and Fc-CD34 (pIG1) (Figure 5.7 B-C). In addition, a sequential depletion of lysates of immune complexes was performed to again test the level of phosphorylation of the phosphoproteins precipitated by CD34 and Fc-CD34 and to attempt to get a high level of purity of phosphoproteins for phosphoamino acid analysis. Following initial lysis of HUVEC in a 50/50 mix of 1% Digitonin/1% NP-40, the pellet was lysed again in 1% NP-40 alone in a smaller volume and subjected to the same sequential depletion and kinase assay as the other samples. Sequential depletion was accomplished by incubating the lysate with antibody or Fc protein followed by the addition of Protein A sepharose. The lysate was then removed from the protein A pellet and fresh protein A was added for additional incubations (Figure 5.7 C).

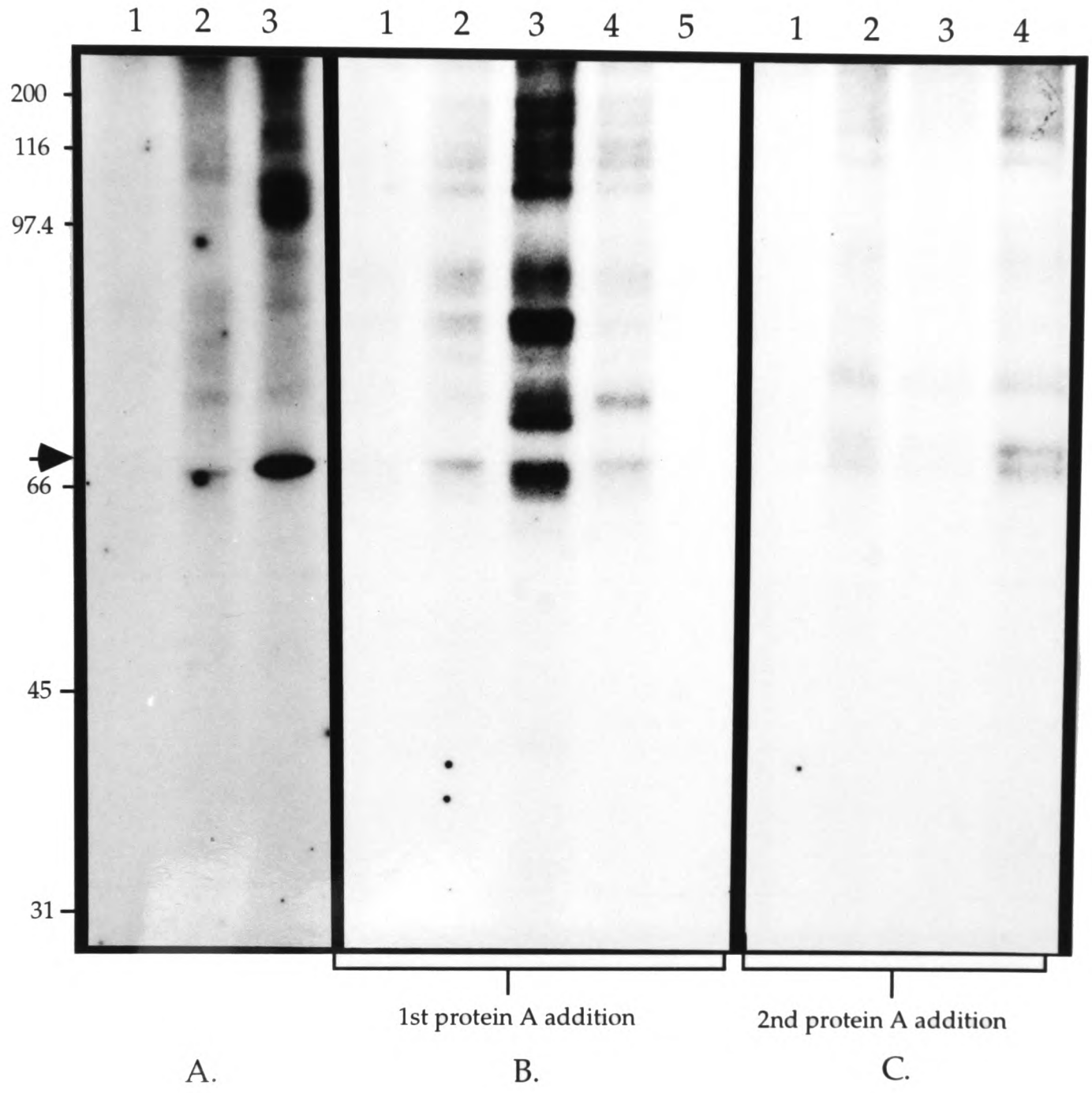


Figure 5.7 Phosphotransferase Assays on HUVEC comparing kinase activity between CD34 and its ligand. Immunoprecipitations were repeated for several samples by the sequential addition of protein A sepharose (PAS) prior to performing the kinase assay. **A.** Lane 1 - mouse immunoglobulin; Lane 2 - anti-CD34 Mab (ICH3); Lane 3 Fc-CD34 (Rg). **B.** Lane 1 - mouse Ig; Lane 2 - anti-CD34 Mab (4G10); Lane 3 - anti-CD34 Mab (QBEND10); Lane 4 - Fc-CD34 (pIG1); Lane 5 - Fc-CD34 (NP-40). **C.** Lane 1 - Fc-CD34 (NP-40, 4th PAS addition); Lane 2 - anti-mouse, reprobed with Fc-CD34 (4th PAS addition); Lane 3 - anti-4G10 reprobed with Fc-CD34; Lane 4 - Fc-CD34 (50% Digitonin/50% NP-40) (4th Prot. A addition). Exposure is 17 hrs. at RT. Fc-CD34 precipitates a phosphoprotein at 70 Kd distinct from CD34 phosphorylated bands. The 110 Kd band is likely phosphorylated CD34.

Fc-CD34 precipitated siginificantly less kinase activity in NP-40 alone than the 50/50 mix, although faint bands are represented by the mixed detergent by the fourth depletion (Figure 5.7 B, Lane 5 and C, Lane 4). A striking difference was shown between the ability of the anti-CD34 Mabs 4G10 (Figure 5.7 B, Lane 2) and QBEND10 (Lane 3) to precipitate phosphoproteins as compared with Fc-CD34 (pIG1) (Lane 4) which precipitates a relatively faint band at 70 Kd and one at 80Kd. The simplest explanation for this is that QBEND10 is a higher affinity antibody than either ICH3 or 4G10, although it could be an "activating" antibody and thus induce a higher level of phosphorylation upon binding to CD34. It is likely that the 70Kd band precipitated with Fc-CD34 (pIG1) in Figure 5.7 B corresponds with the 69Kd band precipitated by Fc-CD34 (Rg) previously (Figure 5.7 A, Lane 3). In relation to the other antibodies tested, QBEND10 is not only a higher affinity antibody but also one which interacts with an epitope potentially important in the phosphorylation of CD34 and/or associated proteins. That this carbohydrate epitope is sialic acid-dependent could suggest a role for sialic acids in an interaction with ligand in binding and potential activation of $CD34^+$ cells. To confirm the identity of the most heavily labelled bands from the QBEND10 and Fc-CD34 precipitates, phosphoamino acid analysis was undertaken in collaboration with Dr. H. Lahooti in the laboratory of Dr. Malcolm Parker (ICRF, Lincoln's Inn Fields, London).

5.7 CD34 and ligand phosphoproteins are labelled on T and Y

Because some of the phosphoproteins precipitated by QBEND10 also appeared to be precipitated by Fc-CD34, phosphoamino acid analysis (Boyle 1991) was carried out on the 70Kd, 80Kd and 90Kd phosphoproteins from both HUVEC

immunoprecipitations to determine the extent of phosphorylation on serine, threonine or tyrosine residues. Previous data on CD34 has shown that CD34 is itself phosphorylated on serine and threonine residues (Fackler et al. 1990). Phosphoamino acid analysis was performed by excising bands from fixed polyacrylamide SDS-PAGE gels, homogenising and eluting the fixed proteins. Eluted proteins were then precipitated and individual phosphoamino acids were separated by electrophoresis alongside serine, threonine and tyrosine standards (Boyle 1991). Radiolabelled phosphoamino acids were observed by autoradiography.

Figure 5.8 shows that the 90Kd band is the most heavily phosphorylated on tyrosine residues followed by the 70Kd and then the 80Kd band. Densitometric analysis would allow for quantitation of relative amounts of phosphorylation for each precipitate. Tyrosine was recently reported to be the major form of CD34 phosphorylation upon incubation of KG-1 and KG-1a cells with the QBEND10 and ICH3 class I anti-CD34 "adhesion-inducing" antibodies (Majdic 1994). In the same study, incubation with QBEND10 was shown to produce homophilic aggregation in KG-1a and, to a lesser extent, KG-1 via the β-2 integrin (ICAM-1/LFA-1) but not the β-1 integrin (VLA-4, VLA-5, and CD29) pathway. The elevated level of phosphorylation represented

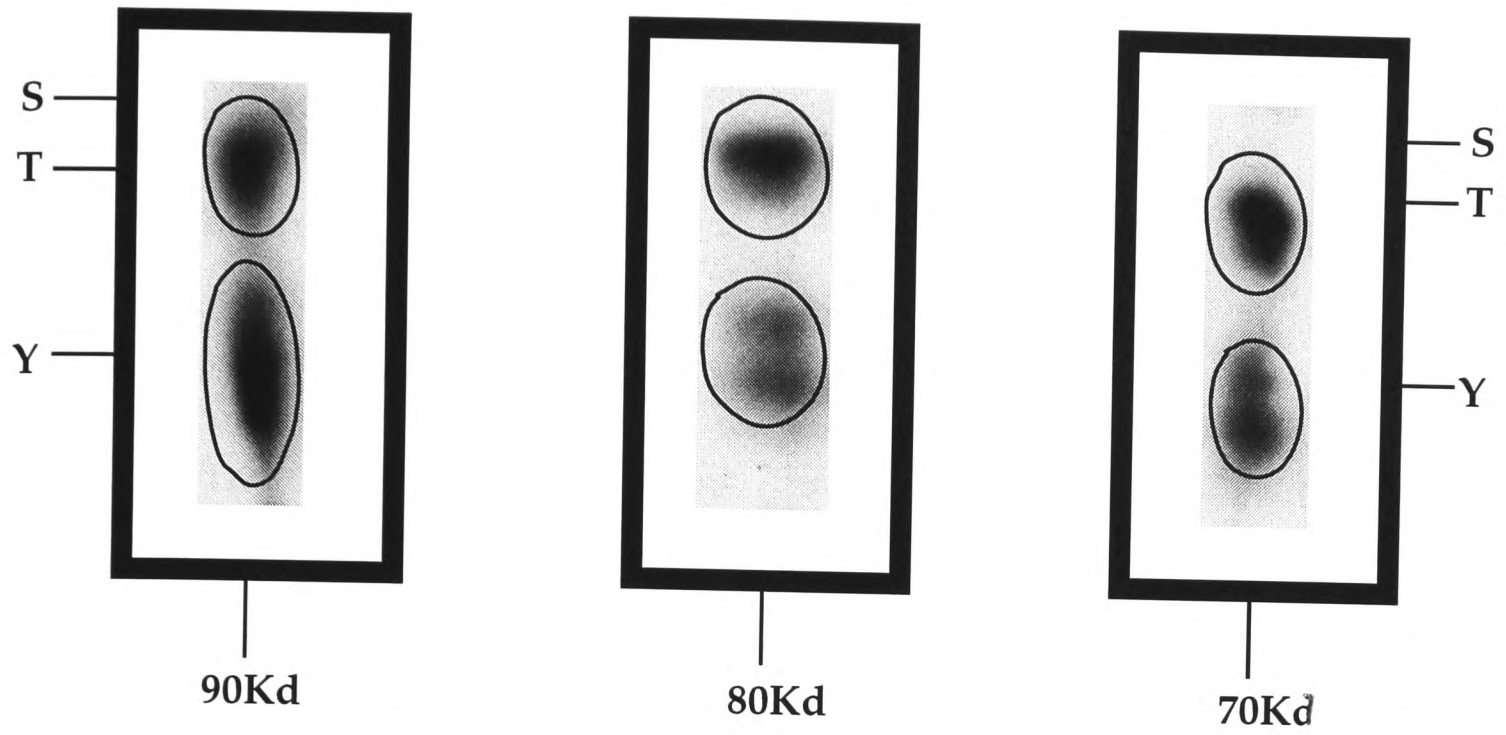


Figure 5.8. Phosphoamino Acid Analysis of Phosphoproteins precipitated with anti-CD34 Mab (QBEND10). Bands from the original QBEND10 immunoprecipitate (Figure 5.7 B, Lane 3) were excised from the gel and subjected to Phosphoamino Acid Analysis. All of the three major phosphoproteins precipitated with the QBEND10 antibody appear to be phosphorylated on threonine and tyrosine residues.

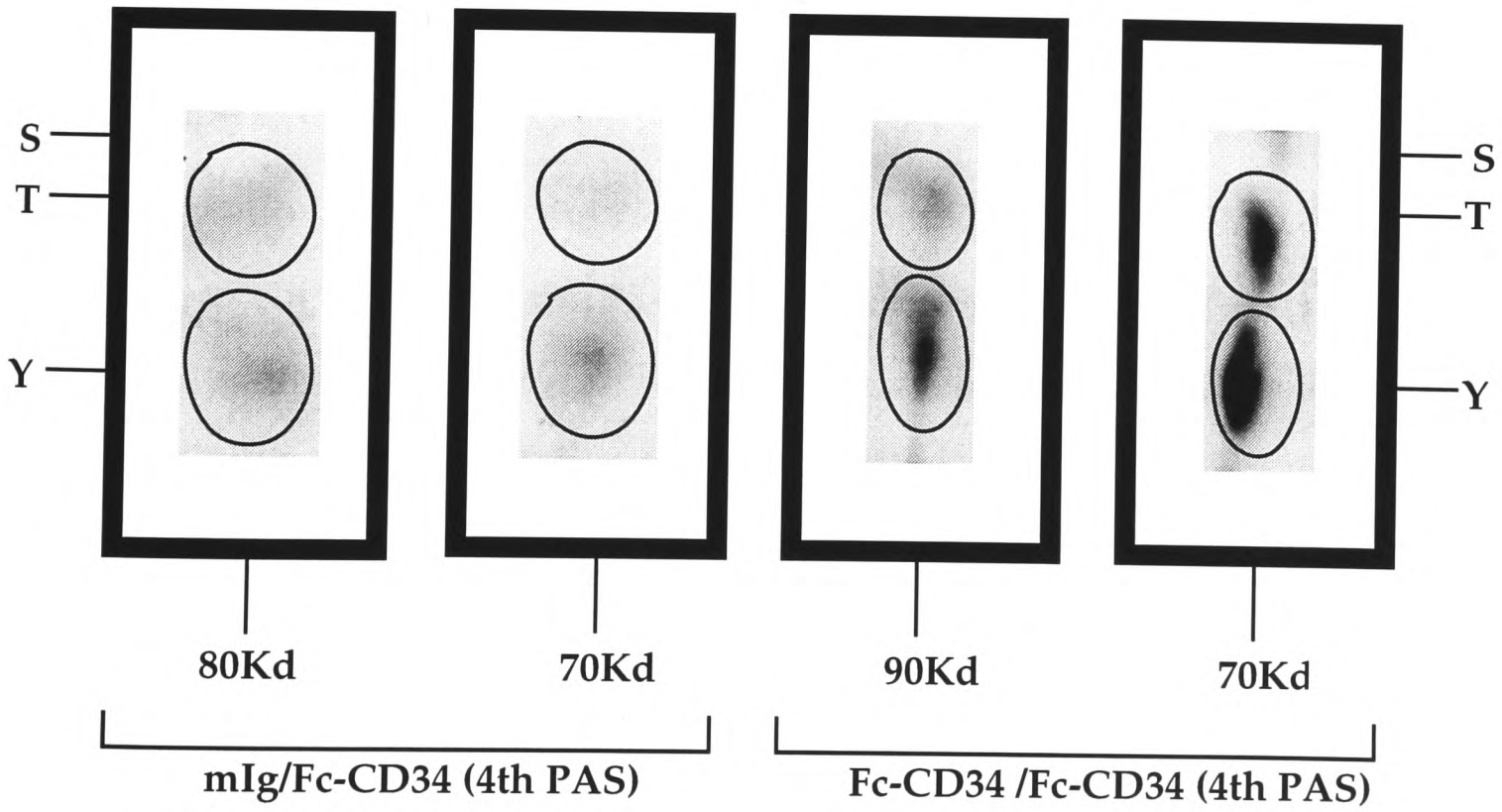


Figure 5.9. Phosphoamino Acid Analysis of Fc-CD34 Precipitated Phosphoproteins. Bands were excised from lanes representing the fourth addition of protein A sepharose and subjected to electrophoresis to separate individual phosphoproteins. The results indicate that the major level of phosphorylation occurs on tyrosine and threonine residues and that this activity persists even after the lysate is depleted by the addition of sequential lots of protein A sepharose. In both cases, the 70 Kd band retains the highest degree of phosphorylation with a marginally higher amount on tyrosine.

using QBEND10 and, to a lesser degree, ICH3 antibodies in HUVEC could indicate that CD34 undergoes a similar level of signal transduction and increases binding affinity upon binding of the QBEND10 epitope to CD34 in these endothelial cells.

The 70 Kd and 80 Kd phosphoproteins from the 4th addition of Protein A Sepharose to mouse Ig/Fc-CD34 (Figure 5.7C, Lane 2) and the 90 Kd and 70 Kd bands from Fc-CD34 alone (Figure 5.7C, Lane 4) were also analysed using this method (Figure 5.9). In this analysis, although there was an expected reduction of kinase activity relative to the first addition of Protein A Sepharose, it is still clear that for the proteins analysed below, the main phosphorylation occurs on threonine and tyrosine residues. In addition, in both precipitations, the 70 Kd band exhibits the highest degree of persistent phosphorylation. These results are suggestive that CD34 and its ligand are phosphorylated with similar proteins. The marked reduction of phosphorylation activity in the 70 Kd band from the mIg/Fc-CD34 phosphoprotein relative to the 70 Kd band from the Fc-CD34 alone, particularly in the level of tyrosine still present after four lysate depletions, is consistent with the high-affinity antibody depleting the kinase activity earlier than the Fc probe.

5.8 Discussion

Ligand binding studies with Fc-CD34 suggest that it can bind homophilically to pCD34 transfectants but not to metabolically labelled cell lines. However a 110Kd band was immunoprecipitated from HUVEC. The phosphorylation data obtained by probing HUVEC lysates with Fc-CD34 and anti-CD34 Mabs

also suggest that CD34 and a ligand may precipitate similar phosphoproteins. Fc-CD34 precipitates a previously unidentified phosphoprotein at 70Kd from HUVEC lysates. Phosphoamino acid analysis suggests that this band is principally phosphorylated on tyrosine and threonine. This is consistent with the three tyrosines and five threonines available for phosphorylation in the cytoplasmic tail of CD34. Phosphoproteins in a similar position appear upon binding CD34 with Mabs which recognise sialic acid-dependent epitopes and with the Fc-CD34 protein which is itself heavily O-glycosylated which suggests that these proteins could potentially be phosphorylated following a protein-carbohydrate interaction. Peptide sequencing or mass spectrometry of the phosphoprotein as well as reprecipitation analysis could provide information as to the identity of the phosphorylated molecule. As L-selectin is known to be a ligand for CD34 in HEV, the phosphorylation studies here provide evidence for the signalling properties of CD34 and a distinct ligand on HUVEC.

Chapter 6

Characterisation of a Potential CD33 Ligand

6.0 Specific aims

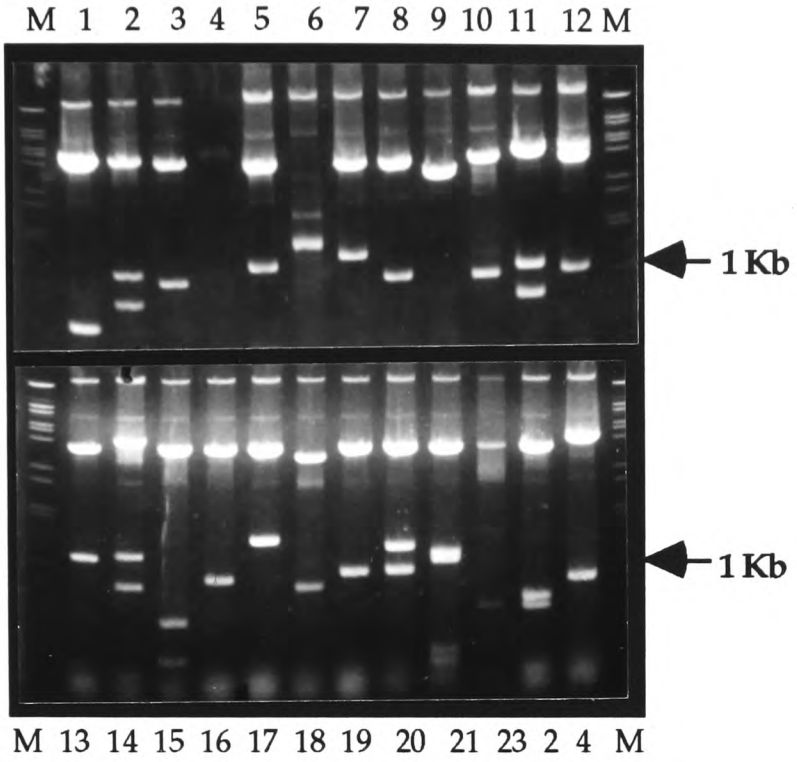
The aim of this chapter was to use the panning method to isolate candidate cDNA products which bound Fc-CD33 and then to test their binding efficiency as potential ligands. The majority of the work focuses on characterising the apo-dystrophin-4 (apo-4) cDNA and its potential role as a ligand for CD33. The limitations of panning to isolate ligands for Fc-chimeras are also addressed.

6.1 Fc-CD33 as a probe with the panning method

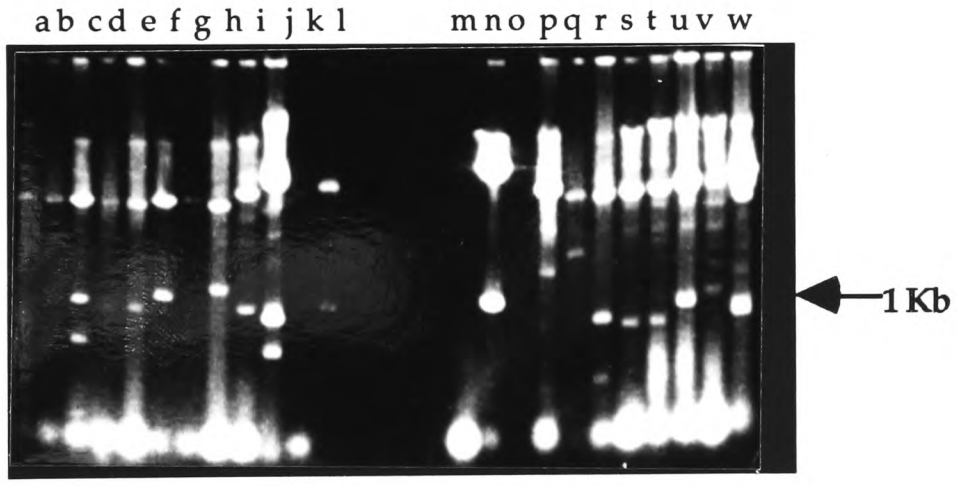
The CDM8 and pIG1Fc expression vectors were used to isolate candidate ligands of interest from cDNA libraries transfected into COS cells, lifted and probed with Fc-CD33, and then incubated on "panning" plates previously coated with goat anti-human Fc (Simmons 1993). Each round of panning was designed to allow for the enrichment of the population of candidates which were expected to bind to the Fc chimera used as a ligand probe.

Panning was attempted with Fc-CD33 using transfected U937, HL-60, K562, KG-1, KG-1a, KG-1b and placenta cDNA libraries. K562 and placenta gave the most promising panning results with Fc-CD33, showing at least five COS cells panning by the second round. In retrospect, this may be due to the high level of potentially CD33-binding sialic acid on K562 cells which was likely present on COS cells transfected with its cDNA. The same libraries were employed with Fc-CD34, in addition to HUVEC, which gave the most promising results and was pursued for a candidate ligand. Panning efforts were pursued several times with little success beyond the second round. This was likely a result of amplification of transfected *E.Coli.* on amp/tet plates to prevent lysis by bacteriophage,

thought to be lysing the cultures normally amplified in flasks. Following several attempts at panning, 20 novel cDNAs from K562 and placenta cDNA libraries were selected as candidates based on their high representation, performing 24-48 minipreps after each round. The 'panel' of candidates is shown below as cut miniprep cDNA (Figure 6.1A-B). Most of the candidates had a size of about 1 Kb, a well-represented size in cDNA libraries (Simmons 1992).



A. Representative Second Round Candidate minipreps from placental cDNA



B. Twenty Candidate cDNAs for Fc-CD33 and one (w) for Fc-CD34. Top four Fc-CD33 staining candidates are: b-P2,3; f-apo-4; l-ICAM-1; n-ICAM-1.

Figure 6.1A-B Representative minipreps from Fc-CD33 Panning. A. Inserts obtained from a placental cDNA library probed with Fc-CD33 on the 2nd round of panning and digested with HindIII/PstI. **B.** Nineteen candidate cDNAs panned with Fc-CD33 and one panned with Fc-CD34 which were subjected to staining for Fc-CD33 binding. Most clones are near 1Kb in length.

6.2 FACS staining of "candidate" cDNAs with Fc-CD33

Following transfection of each of the twenty cDNAs, these candidates were then subjected to a 'blind' screen using the Fluorescence Activated Cell Sorter (FACS) analysis to assess binding with Fc-CD33. Several candidates appeared to be positives with the highest level of staining at 38%. Most candidates were isolated from the 2nd or 3rd rounds of panning using the placental cDNA library. Previous staining showed that several candidates produced levels of staining which differed by only 5-10% making true positives difficult to distinguish. By contrast, the previously shown FACS analysis of myeloid leukemic cell lines using 33Fc was much higher (up to 97%). Low levels of staining with the transfectants was attributed to low transfection efficiency of the cDNA and the low affinity binding of the ECD-Fc constructs (other labs have reported binding of several ECD-Fc proteins to be far lower than antibodies or other ECD-Fc's) (Hollenbaugh et al. 1992). The existence of several ligands for each molecule, as suggested by iodination experiments with Fc-CD33, would require the transfection of all of the putative ligand chains to achieve full ligand binding efficiency.

Fc-CD33 binding was assessed in a FACS analysis to transfectants with the four major candidates isolated using the panning process as well as to CD33 (Figure 6.2). All candidates were isolated from the placenta library, a good source of diverse genes, on the second or third round of panning. The staining is shown for all candidates together and for each candidate individually. Two of these candidates proved to be ICAM-1 and 250bp of sequence for the other did not match known sequence. In one transfection out of at least five attempts, 36% binding was achieved to ICAM-1, and 9.6% to apo-4Rg. The highest percentage of staining that one can expect with antibodies to transfected molecules is usually

about 10-30%; ECD-Fc proteins bind with lower affinity and thus 2-10% staining was considered positive (personal communication, David Simmons).

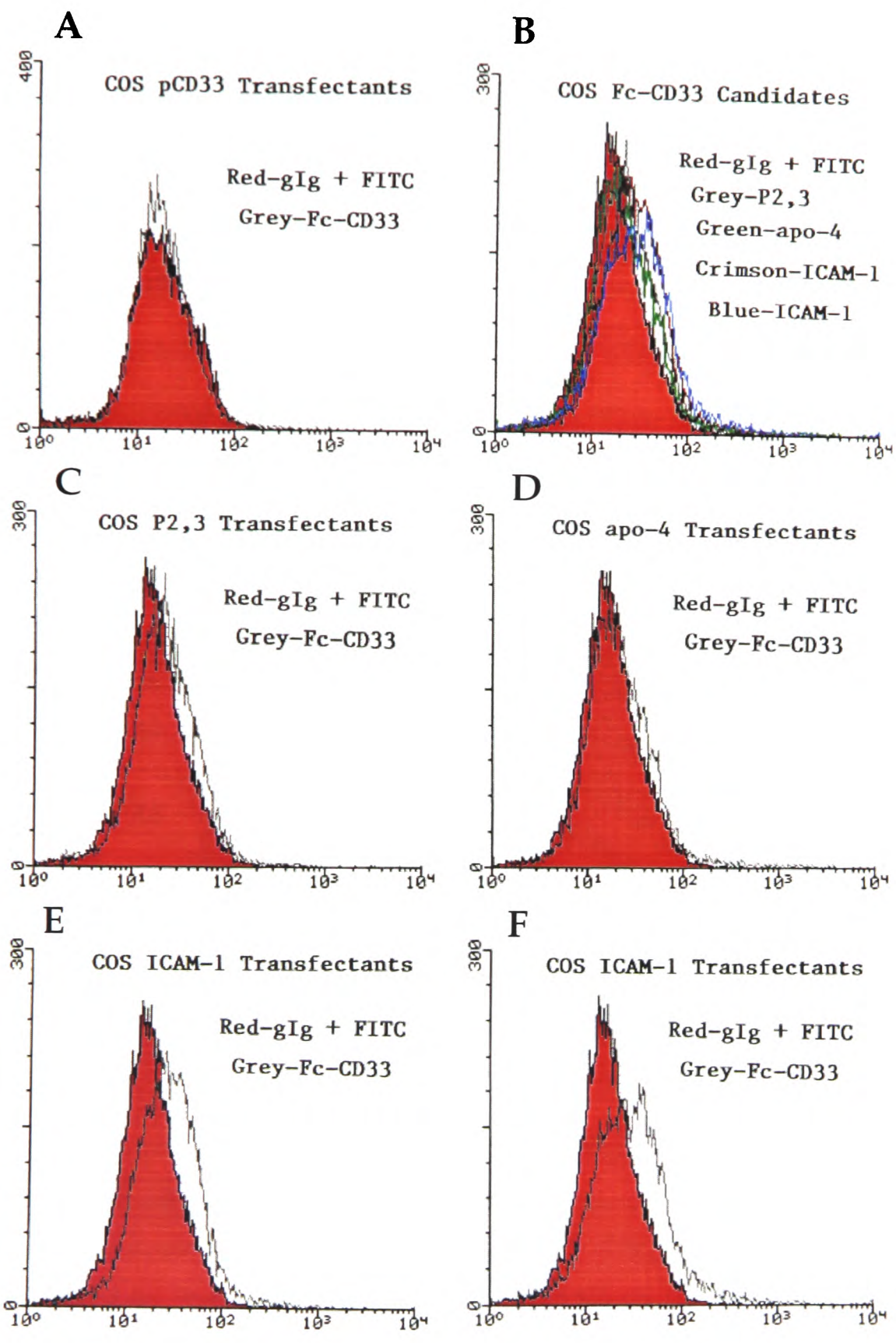


Figure 6.2 FACS Analysis of CD33 candidates shows low avidity binding in COS transfectants (TF). A-F - shaded peak, goat anti-human Ig FITC alone; A. grey peak Fc-CD33 binding to CD33TF; B. Fc-CD33 staining of four candidate ligand transfectants, compared in descending order; individual candidate TF binding to Fc-CD33 is shown in C. to P2, 3TF, 2.2%positive; D. to apo-4, 2.4% positive; E. to ICAM-1, 3.3% positive and F. to ICAM-1, 6.4% positive.

Based on the staining data, several candidates were sequenced. Two of these were revealed to be ICAM-1 cDNA and a cDNA clone homologous to the 3' end of the dystrophin gene which was subsequently characterised as apo-dystrophin-4 (apo-4). ICAM-1 was viewed sceptically because several labs had isolated it using Fc-chimeras and it was thus thought to bind non-specifically (D. Simmons, personal communication), although thus far it has been shown to bind to LFA-1 (Simmons 1988a), hyaluronan (McCourt 1994), and CD43 (Rosenstein 1991). To directly test the ability of Fc-CD33 to precipitate surface labelled proteins, K562 and COS transfected with the apo-4 cDNA and ICAM-1 were iodinated and precipitated with Fc-CD33 (Figure 6.3).

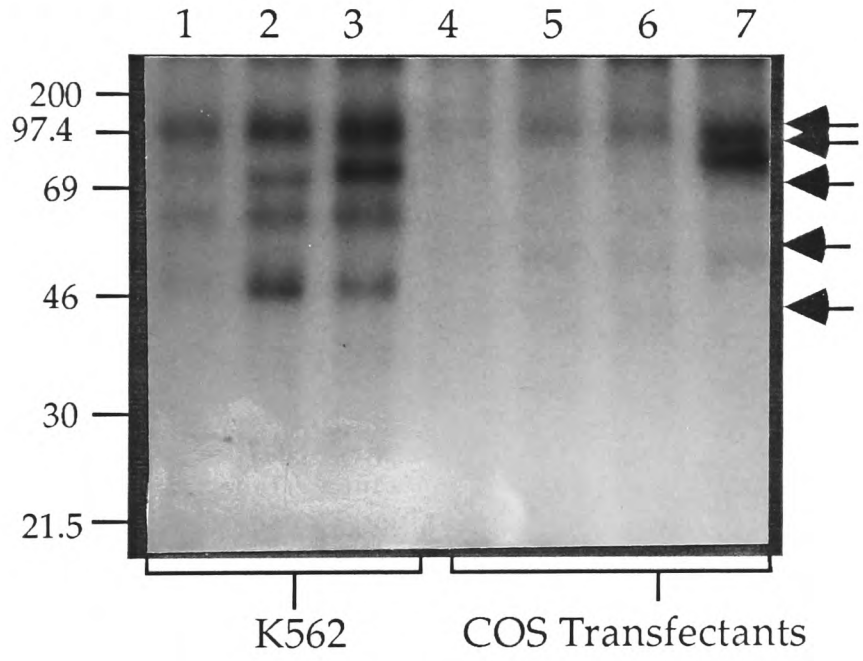


Figure 6.3 Fc-CD33 precipitates similar proteins from K562 and transfectants. K562 (Lanes 1-3), and apo-4 (Lanes 4-6) and ICAM-1 (Lane 7) transfectants were iodinated and immunoprecipitated with mouse Ig (Lanes 1 and 4) and Fc-CD33Rg (Lanes 2, 5 and 7) and Fc-CD33pIg1 (Lanes 3 and 6). Fc-CD33 precipitates bands at 46 Kd, 55 Kd, 69-70 Kd and 100Kd in K562 (Lanes 2-3), a 95-100Kd doublet ICAM-1 transfectants (Lane 7), and a fainter 100Kd band from apo-4 (Lanes 5-6) as well as signals at 50 and 66Kd (Lane 5) more faintly precipitated with Fc-CD33pIg1. Proteins from cell lines are often slightly larger than those from transfectants due to glycosylation differences.

Native ICAM-1 precipitates at about 85-110Kd (Simmons 1988a). Fc-CD33Rg precipitated a heavily labelled doublet at 95-100Kd from the ICAM-1 transfectant

(Lane 7). In this assay, ICAM-1 was being used for comparison to the apo-dystrophin-4 cDNA, however, some of the sample was lost in the iodination, otherwise an ICAM-1 negative control would have been included. Fc-CD33Rg (Lane 5) and Fc-CD33pIg1 (Lane 6) precipitated a band at about 100Kd from the apo-dystrophin-4 cDNA and fainter bands at average weights of 45, 50 and 66Kd not seen in the control lane (Lane 4), suggesting that apo-4 could be a low-affinity ligand for CD33. The 66Kd band is a bit more heavily labelled in Lane 5, suggesting that Fc-CD33 Rg is more efficient at precipitating apo-4. Under stringent washing conditions, Fc-CD33Rg (Lane 2) and Fc-CD33pIg (Lane 3) precipitated a unique band from K562 at 46Kd and others, although they appear to be similar to those in the control lane. The bands from Fc-CD33pIG are identical to those precipitated with Fc-CD33Rg except that the 70Kd band is more heavily labelled which likely accounts for its slightly higher position on the gel (Lane 3).

While Fc-CD33 was not shown to bind ICAM-1 in a further metabolic labelling experiment, the FACS staining and iodination were suggestive that it could be a ligand given the right conditions. CD33 is sialylated and ICAM-1 has been shown to bind sialylated proteins such as CD43 (Rosenstein 1991). CD33 COS transfectants will only bind if they are desialylated and it is probable that neighboring sialic acids on the COS cell will inhibit CD33 binding (Freeman 1995). *In vitro*, however, it is possible that Fc-CD33 is capable of binding the sialic acids on a transfected ICAM-1 molecule. During the panning process, no one candidate ever emerged as a strong positive for any Fc-adhesin tested (CD31, CD33, CD34, CD68, MUC-18) and, ultimately, Fc molecules, which approximate cell-cell interactions rather than antibody-ligand interactions, were deemed to have an affinity too low to use the panning method to search for ligands. Based on the iodination data shown here and in Chapter 3, several candidate ligands

probably exist for CD33, as is the case for CD34 (Baumhueter 1993; Oxley 1994) and hyaluronan (Aruffo 1990; McCourt 1994). Recent literature suggests that molecules which bind with low affinity via a protein-carbohydrate interaction (eg. the selectins) are good candidates for having multiple ligands (Baumhueter 1993). As the panning system provided constraints for proving any ligand definitively, the apo-dystrophin-4 clone was thus further characterised as a potential low-affinity ligand for CD33 and a potentially interesting clone which could play a role in both muscle cells and hematopoietic cells as has recently been shown for Dp71 (apo-dystrophin-1) also originating from the 3'end of the dystrophin gene but utilising a different frame from dystrophin in the final exon (Hugnot 1992; Lederfein 1992).

6.3 An Fc-CD33 candidate ligand is termed apo-dystrophin-4

To compare the apo-dystrophin-4 cDNA to known dystrophin and related cDNAs, the entire cDNA was sequenced in both forward and reverse directions and translated into all three frames (Figure 6.4) and the oligonucleotides used to sequence it and to perform PCR appear in the appendix. The cDNA sequence begins at position 1, with additional upstream genomic sequence at -233 to -1 obtained with clones from the 3' intronic region of exon 78, later found to match the region of apo-4. The structure of the 3' approximately 546bp was compared with the sequencing of genomic phage clone DNA from the 3' region of dystrophin (Kunkel et al. 1985) to confirm the precise point of a 136bp inversion discovered at the 3' end of apo-4 . The second reading frame includes both the putative N-terminus of apo-4, an upstream exon identified with GRAIL and the hydrophobic frame of exon 79 spliced onto Dp71.

```
      ......TTTCCTATTCAATGTATAGTGCACCAAAGGTCAATTCAAGAGTTTATTATTATT
-239   --------+---------+---------+---------+---------+---------+---------+ -180
      ......AAAGGATAAGTTACATATCACGTGGTTTCCAGTTAAGTTCTCAAATAATAATAA
            F   P   I   Q   C   I   V   H   Q   R   S   I   Q   E   F   I   I   I

      ATTTTCAACCCAAGTAAAAGCAGAGAGAAAATAGCCACCTCCACCATAGCCTCAGAAGCA
-179   --------+---------+---------+---------+---------+---------+---------+  -120
      TAAAAGTTGGGTTCATTTTCGTCTCTCTTTTATCGGTGGAGGTGGTATCGGAGTCTTCGT
      I   F   N   P   S   K   S   R   E   K   I   A   T   S   T   I   A   S   E   A
                                      start GRAIL exon    transcripton start site
                                                  |                   |
      AGCCAACAGCCTGAAACAGCTTTGAAATGAAAAGTTGGTGTGGCGGTGATGGTGGCAGTG
-119   --------+---------+---------+---------+---------+---------+---------+   -60
      TCGGTTGTCGGACTTTGTCGAAACTTTACTTTTCAACCACACCGCCACTACCACCGTCAC
      S   Q   Q   P   E   T   A   L   K   *   K   V   G   V   A   V   M   V   A   V

      ATAATGGTGACCGATGGTTGGGTGCTGGTGATGGTAGTGGTAGTTGTGAAGGTGGTGATG
 -59   --------+---------+---------+---------+---------+---------+---------+    0
      TATTACCACTGGCTACCAACCCACGACCACTACCATCACCATCAACACTTCCACCACTAC
      I   M   V   T   D   G   W   V   L   V   M   V   V   V   V   V   K   V   V   M

                                              (----N----)                 ---
      GTGGTTTGATTGATAGTAAAAAAAATGTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT
   1  --------+---------+---------+---------+---------+---------+---------+    60
      CACCAAACTAACTATCATTTTTTTTACAAGCAATTATGTTCATCTCTCATTCATTAGTTA
      V   V   *   L   I   V   K   K   M   F   V   N   T   S   R   E   *   V   I   N

      -N----
      CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT
  61  --------+---------+---------+---------+---------+---------+---------+   120
      GTTAGTGAGTATCGGTTCCACCTTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
      Q   S   L   I   A   K   V   E   K   M   Y   P   I   M   E   Y   S   C   S   D

      AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT
 121  --------+---------+---------+---------+---------+---------+---------+   180
      TCTTTAGAACACGAATAGATACCTTAAGAAAACTATATATAAATGTAACCCTTGGACTTA
      R   N   L   V   L   I   Y   G   I   L   L   I   Y   I   Y   I   G   N   L   N

      GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA
 181  --------+---------+---------+---------+---------+---------+---------+   240
      CATCGAACTGTAAAAAGGTACATTTGTGGTCATCGGACTAGGTTGTAATTCGACTATGAT
      V   A   *   H   F   S   M   *   T   P   V   A   *   S   N   I   K   L   I   L

      ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA
 241  --------+---------+---------+---------+---------+---------+---------+   300
      TGTTTGTTGCACATTACCGAAGTAATTATTCCGAAACGAAGAAGGACCTTTGACCACTTT
      T   N   N   V   *   W   L   H   *   *   G   F   A   S   S   W   K   L   V   K

      (----N----)
      AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG
 301  --------+---------+---------+---------+---------+---------+---------+   360
      TTAGTTTGGAACAACACATGTGGGAGCTACGTCGAAGACACAACAGAAGTGGGTCTTTAC
      N   Q   T   L   L   C   T   P   S   M   Q   L   L   C   C   L   H   P   E   M
```

```
    GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA
361 ---------+---------+---------+---------+---------+---------+ 420
    CCCTTACTAAAGGGTTTACCGTTTCTTTGTCTCACTACGATAGATAGACGTGGAAAACAT
    G  N  D  F  P  N  G  K  E  T  E  *  C  Y  L  S  A  P  F  V
                          begin exon 79
                                   |
    AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG
421 ---------+---------+---------+---------+---------+---------+ 480
    TTCAGACAGAAAGAAAGAGAAACAAAAGGTCCTGTGTTACATCCTTCAGAAAAGGTGTAC
    K  S  V  F  L  S  L  C  F  P  G  H  N  V  G  S  L  F  H  M

    GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA
481 ---------+---------+---------+---------+---------+---------+ 540
    CGTCTACTAAACCCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTTCCT
    A  D  D  L  G  R  A  M  E  S  L  V  S  V  M  T  D  E  E  G

    GCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA
541 ---------+---------+---------+---------+---------+---------+ 600
    CGTCTTATTTACAAAATGTTGAGGACTAAGGGCGTACCAAAAATATTATAAGTATGTTGT
    A  E  *  M  F  Y  N  S  *  F  P  H  G  F  Y  N  I  H  T  T

                                   (----N----)
    AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG
601 ---------+---------+---------+---------+---------+---------+ 660
    TTCTCCTAATCTGTCATTCTCAAATGTTCTTTATTTAGATATAAAAACACTTCCCATCAC
    K  R  I  R  Q  *  E  F  T  R  N  K  S  I  F  L  *  R  V  V

    GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG
661 ---------+---------+---------+---------+---------+---------+ 720
    CATAATATGACATCTAAAGTCATCAAAGATTCAGACAATAACAAAACAATTGTTACCGTC
    V  L  Y  C  R  F  Q  *  F  L  S  L  L  L  F  C  *  Q  W  Q

    GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTG
721 ---------+---------+---------+---------+---------+---------+ 780
    CAAAATGTGCAGATACGTTAACATGTTTTTTCAATATTCTTTTGATGTACATTTTAGAAC
    V  L  H  V  Y  A  I  V  Q  K  S  Y  K  K  T  T  C  K  I  L

    ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA
781 ---------+---------+---------+---------+---------+---------+ 840
    TATCGATTTATTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAAATTTTTAAAT
    I  A  K  *  L  A  I  S  L  Y  G  T  H  F  G  L  F  K  N  L
                inversion start site
                         |
    TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
841 ---------+---------+---------+---------+---------+---------+ 900
    ATTGTCAATATTTCTTTCTTAATATTTCCTTTTTCTTTTATTGCGTTACCTGTTCACCAC
    *  Q  L  *  R  K  N  Y  K  G  K  R  K  *  R  N  G  Q  V  V

    AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA
901 ---------+---------+---------+---------+---------+---------+ 960
    TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT
    K  L  *  T  Q  V  C  T  I  I  R  N  T  P  K  P  K  *  G  R
```

```
     AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961  ---------+---------+---------+------  996
     TTATCGTACTCTTCGGCACAAACTACAATTAATTAA
     N  S  M  R  S  R  V  *  C  *  L  I
```

Figure 6.4. The full-length apo-dystrophin-4 cDNA and upstream genomic sequence translated. Genomic sequence upstream of the apo-dystrophin-4 cDNA is shown from -233 to -1, the start of the apo-4 sequence. The sequence was subjected to a MacVector alignment and was homologous with the 3' end of the dystrophin cDNA, up until the inversion at 860. A search of the 3' 136bp of apo-4 alone showed that it was precisely homologous to 3' dystrophin sequence 1.62Kb downstream in the 3'UTR in the reverse orientation and was thus an inversion of the 3'UTR and genomic sequence. Three potential starting methionines are highlighted at +25, +88, and +100. and a potential CAAT box is underlined at +57. These M's are in phase I of apo-4 alone, but appear in phase II here due to the presence of upstream sequence. The longest open reading frame which should be obtainable from this sequence begins with the M at +88 (33aa). The beginning of exon 79 at 453 is also underlined. For a predicted protein in frame 2, a putative transmembrane domain is underlined and putative N-glycosylation sites which follow the Asn-Xxx-Ser/Thr motif are shown (in parentheses when utilised if nonsense suppression occurs). A cryptic polyadenylation site is underlined at +990. The upstream exon identified by GRAIL appears from -90 to -5 shown by the broken underlining.

The 23 stop codons in the apo-4 sequence prompted a search for additional nucleic acid subsequence sites which could provide a clue as to how apo-4 was translated. The sites identified follow: 1) **Splice sites**-Potential donor and acceptor sites were found near the start of the cDNA suggesting that the gene is unspliced; 2) **Cap sites** - Several Cap sites I and II were identified. A Cap site I at -3bp upstream from a CAAT box at 57 may indicate a transcription start site; 3) **CAAT boxes** - a CAAT box was found at +57, -31 bp upstream from a potential initiating methionine at +88bp and -39 bp upstream from the "strong" initiating methionine at +100bp, based on Kozak sequences (Kozak 1987) -- a CTF-NF1 transcription factor binding site overlaps with this CAAT box to reinforce the

proposal that this could be a potential promoter region. No TATA box is present at -40, but one is found on the genomic DNA at -219bp. If it is a true TATA box, it most likely acts on sequence upstream of apo-4; 4) **Poly A sites -** two AATAAA Poly A sites have been observed in the cDNA at +545 and +631 as well as a cryptic AATTAA site at 990, which could potentially be used; 5) **Poly T regions-** Poly T regions have been found downstream of both Poly A sites, providing evidence that these sites could be used for cleavage following recognition of the Poly A signal; 6) **Inverted Repeats (Palindromes)-**Seven palindromes, which have been shown to occur near inversions, ranging from 8-10bps were found in the cDNA and seven more in the upstream genomic sequence; 7) **Direct Repeats-** signal insertion sites for transposable elements were searched in relation to the inversion. 13 direct repeats have been found in the sequence with a minimum stringency of eight bp (maximum 11bp). Notably, a direct repeat of 8bp exists at 861 and 3 bp downstream of the inversion and appears as

TTATAAAGAAAG^AA**TTATAAAG.** An 8bp direct repeat exists immediately 5' to the inversion and forms a palindrome with genomic sequence. A detailed list of the subsequence sites found appears in Appendix III. Nucleic acid subsequence sites not found in apo-4 include AP2, GCN4, Homeo, MalT box, Octamer, Pu box, and Sp1 sites. Sp1 sites are most often identified with housekeeping genes.

6.3.1 Apo-4 shares a similar inversion with globin Gγ(γβ)^0

The globin Gγ(γβ)^0 gene was one of the first disease causing genes which contains two deletions and an internal inversion in Indian Aγδβ Thalassaemia (Jones et al. 1981). The apo-dystrophin-4 cDNA described here contains 1.62Kb and 657bp deletions and the insertion of a 136 bp inversion at the 3' end (with an 8-11 bp overlap at the 5′ end of the inversion when compared with genomic and known cDNA sequence) (Figure 6.5). In genomic DNA, the 136bp sequence occurs as the reverse and complement of the inversion, 1.623Kb downstream.

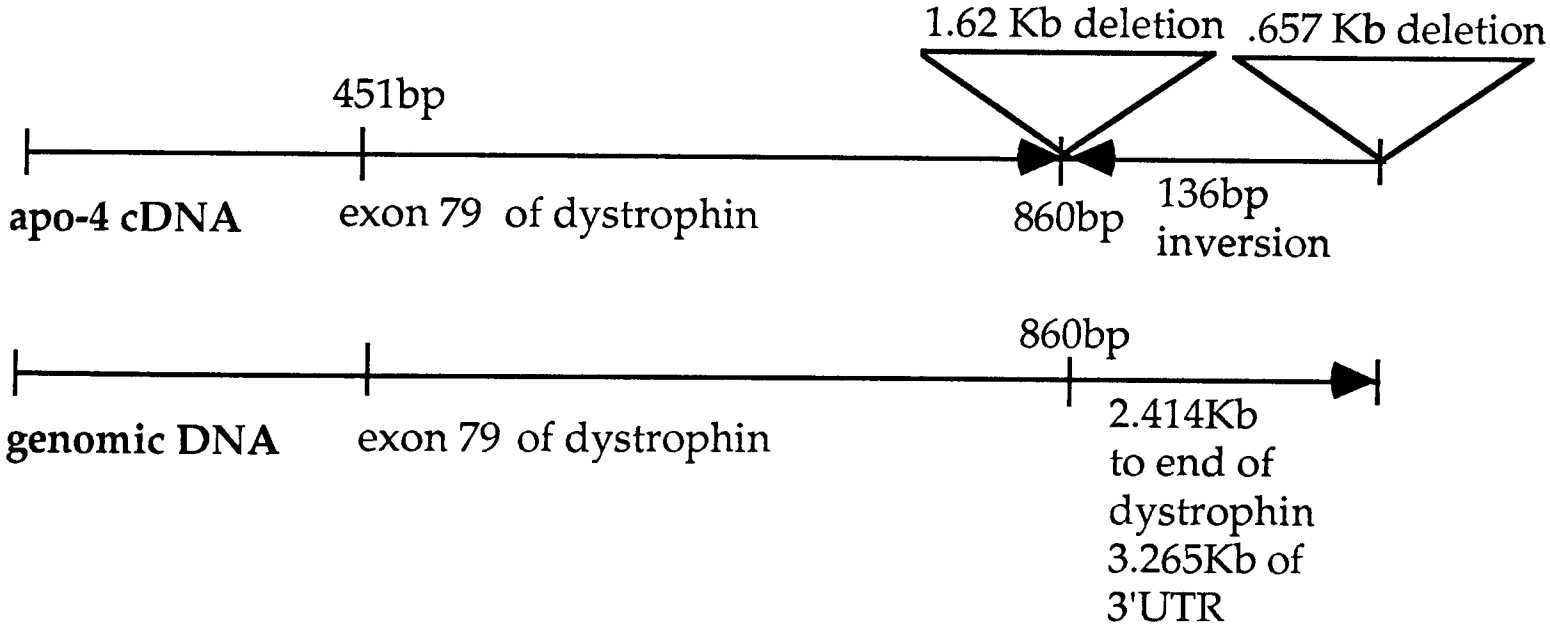
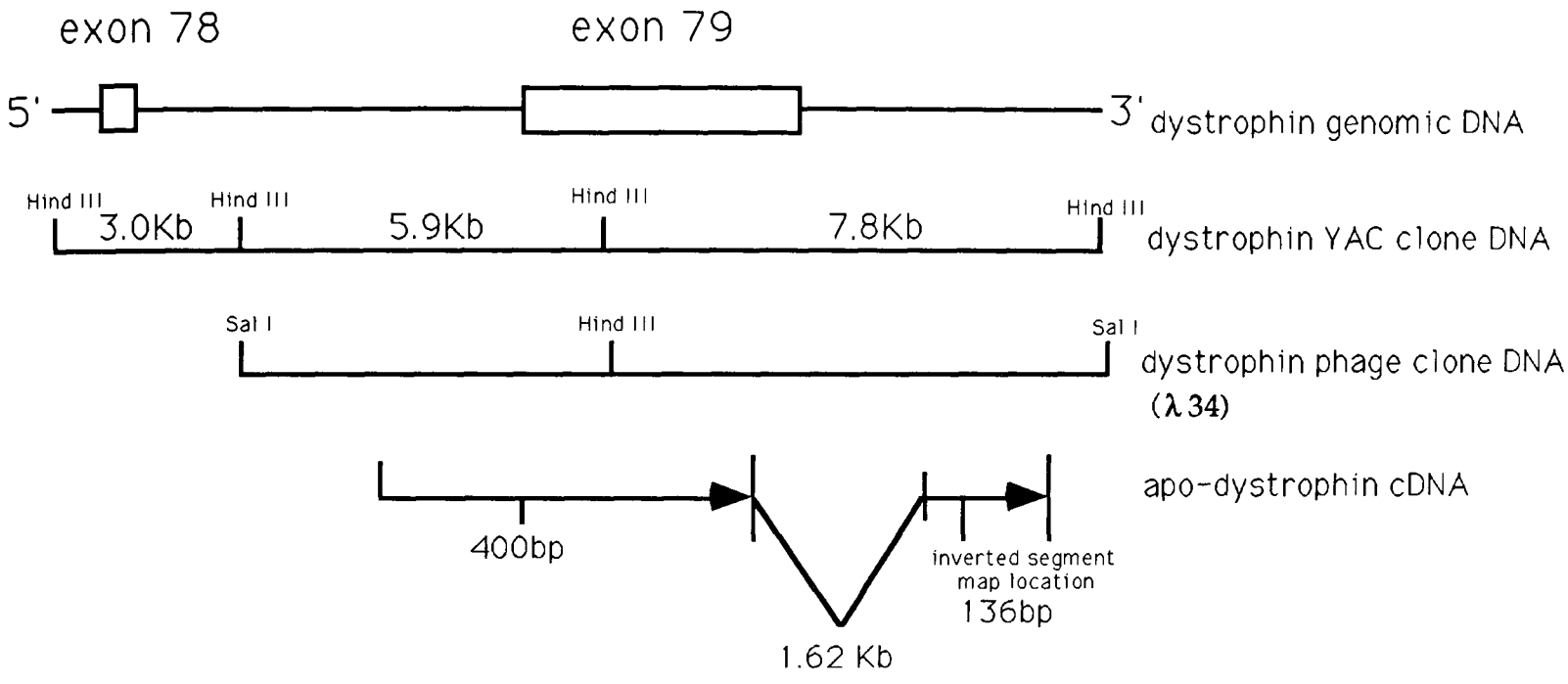


Figure 6.5 Gene Structure of apo-4 cDNA compared with the dystrophin 3' UTR. The above figure shows a basic comparison of the differences between the apo-4 cDNA and genomic sequence, homologous with the dystrophin 3'UTR, diagramming the deletion and inversion at the apo-4 3'end.

6.3.2 Yeast Artificial Chromosome (YAC) hybridization

Due to the unusual structure of apo-4, work was carried out in collaboration with Dr. Jamel Chelly in Dr. Anthony Monaco's lab to map the full length apo-4

(Figure 6.8) and 5' 451 bp upstream of exon 79 and 3' 546 bp fragments to genomic YACs (Figure 6.7) (Monaco et al. 1992). The YACs probed covered the region from exon 78 and its 3' intron, to exon 79 in genomic dystrophin DNA, with the appropriate fragment sizes (Figure 6.6). Restriction sites are indicated.



*cDNA map is not precisely drawn to scale

Figure 6.6. Comparison of YAC and phage genomic dystrophin DNA and apo-dystrophin-4 cDNA clone. The above map indicates the relative sizes and positions of YAC clones in the region found homologous to the apo-dystrophin-4 cDNA. The fragment sizes indicated represent one HindIII YAC fragment each. Full length and partial DNA probes from the apo-dystrophin-4 clone were constructed to confirm that the clone was represented in the region identified in a homology search. A 5' 451 bp probe hybridised to the 5.9 Kb Hind III fragment and the 7.8 Kb Hind III fragment while the 3' 546 bp localised to only the 7.8 Kb Hind III fragment, as expected (data not shown). The intron between exons 78 and 79 is 4.8 Kb. The λ34 phage clones were used in PCR experiments to check whether the inverted region could be reproduced in genomic DNA. Exon 79 of dystrophin occurs 451 bp downstream of the apo-dystrophin-4 cDNA. The 451 bp region was hybrised to human and murine genomic YAC clones to confirm its location on genomic DNA in the region of exon 79 of dystrophin.

Both the known 3' 546 bp and previously uncharacterised 5' 451 bp fragments hybridised, confirming that this was a novel cDNA, homologous with the 5' region of exon 79, the final carboxy terminal exon of the dystrophin gene. In addition, 451 base pairs of intronic DNA exist 5' to exon 79 in apo-4.

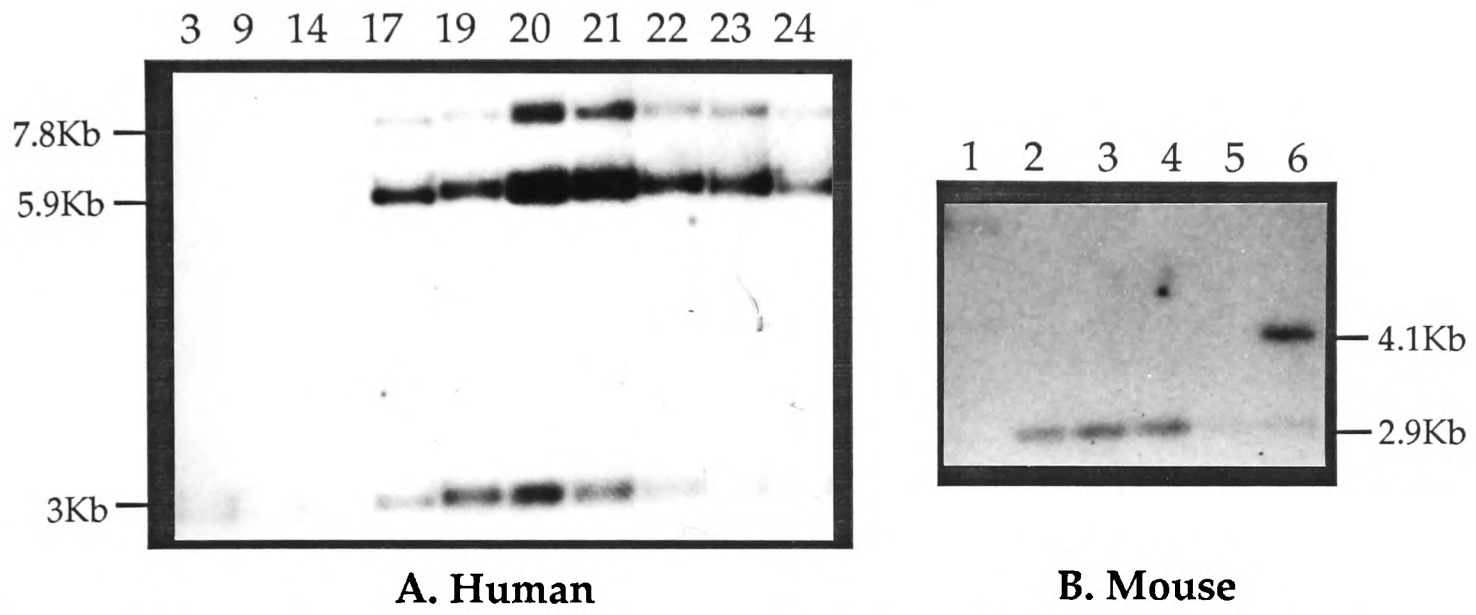


Figure 6.7 Human Yeast Artificial Chromosome (YAC) Clone Hybridization verifies the position of the full-length apo-dystrophin-4 cDNA on genomic DNA. A. human and **B.** mouse YAC clone fragments were probed with the full length 1Kb human apo-dystrophin-4 cDNA. The apo-4 probe was excised from CDM8 using HindIII and Pst I. **A.** Human YACs from the 3′ end of a genomic dystrophinYAC contig (Monaco et al. 1992) include clones upstream of exons 78 and 79 (3, 7, and 14) as controls and clones 17-24 which rinclude a 3.0 Kb (exon 78), 5.9 Kb and 7.8 Kb (both exon 79) Hind III fragments. The apo-dystrophin-4 cDNA overlaps the human 5.9 and 7.8 Kb fragments. **B.** The mouse YACs cover the 3' half of the genomic dystrophin DNA (Y. Ishikawa-Brush and A.P. Monaco unpublished). **A. Lanes 3, 9 and 14 -** No hybridization is shown to YAC fragments upstream of exon 79; **Lanes 17-24 -** hybridisation is shown to both the 5.9 Kb (5' end of apo-dystrophin-4 cDNA) and 7.8 Kb (3' end of apo-dystrophin-4 cDNA) YAC fragments which contain exon 79. Hybridisation to the 3 Kb fragment (Lanes 17-21) may represent cross-hybridisation from the probe with the ori and amp regions of the pYAC4 vector (Larin et al. 1991) or hybridisation to repetitive regions of DNA in intron 78, as apo-4 is not homologous to exon 78, but may be homologous to its intron. **B. Lane 6** - the full-length probe hybridises to the 4.1 Kb murine YAC fragment containing the complete region of the apo-dystrophin-4 cDNA. **Lanes 1 and 5-** no hybridisation is shown to a fragment upstream of exon 79. **Lanes 3-5** - Some intron hybridisation is shown to the 2.9 Kb fragment by the probe which may also represent hybridisation to the inverted region of the apo-dystrophin-4 clone. These results show that both the 3' and the 5' ends of the cDNA hybridise genomic mouse DNA in the region of exon 79. There are no Hind III sites in the genomic mouse DNA which overlap the apo-dystrophin-4 cDNA, so all hybridisation is in the 4.1 Kb fragment. The 3' region hybridises to both mouse and human YACs confirming that at least part of this region is well-conserved and provides evidence that the 5' region of the cDNA hybridises to the expected region in genomic DNA.

Notably, both the 5' 451bp and the 3' 546bp apo-dystrophin-4 cDNA fragments hybridised to mouse YAC clones, indicating that both fragments shared some homology to mouse in the region of exon 79 and the region previously considered intronic upstream of exon 79 (Figure 6.8).

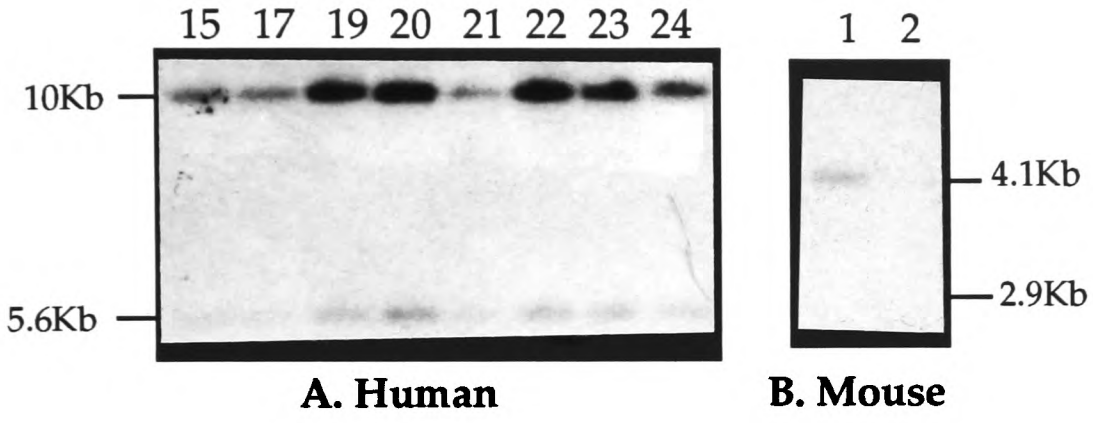


Figure 6.8 The 5' 451 bp of the apo-dystrophin-4 cDNA hybridises to both mouse and human YAC clones demonstrating its conservation between species. **A.** Human and **B.** Mouse YACs were probed with a 451bp fragment from the 5' end of the apo-dystrophin-4 cDNA. The probe was excised from the CDM8 vector using Hind III/Pst I and the 451 bp fragment isolated with Xmn I. The human YACs represent 5.6 (3′) and 10 Kb (5′) fragments cut with Hind III and Sal I which include exon 79 of the complete dystrophin genomic DNA. **A.** Hybridisation is shown to the human "retrofitted" (Eliceiri et al. 1991) 5' 10 Kb genomic fragment containing the 451 5' bp of the apodystrophin cDNA upstream of exon 79. The band at 5.6 Kb could represent hybridisation of the probe to intron upstream of exon 79 which acts as an exon in apo-4. **B. Lane 1**- the mouse 4.1Kb fragment also hybridises indicating that some conservation between species exists in the 5′ 451 bp region of the apo-dystrophin-4 cDNA. **Lane 2 -** no hybridisation is shown to the murine 2.9 Kb fragment which represents the region downstream of exon 79. This analysis verifies that the apo-dystrophin cDNA clone can hybridise in the expected regions of genomic DNA in both mouse and human YACs. The hybridisation of the 5' region of the cDNA lends weight to the possibility that this region could contain an intron which could be expressed as an exon and is conserved between species. The figures represent one filter hybridised with the same probe, washed under normal stringency and exposed overnight at -80°C.

6.3.3 Homologous regions in cDNA, genomic and mouse DNA

To assess the level of species conservation of any proposed new exons, mouse and human genomic DNA were aligned with the apo-4 cDNA using GCG Pileup (Genetics Computer Group 1991). It was found that the 5' 1- 451 bps are in the 5' intronic region of exon 79 DNA relative to the full-length dystrophin mRNA (Koenig et al. 1988) and nucleotides 453-860 overlap the three translated amino acids of exon 79 in dystrophin (33 aa in the shifted frame used in Dp71) and its remaining untranslated region (UTR) (Figure 6.9). Nucleotides 861-997 are in the inverted region and also normally occur in the reverse orientation at 13,163-13,309 bp in the 3' UTR of the dystrophin cDNA, 1.623 Kb downstream of the inversion beginning at 860. In other words, apart from the inverted region of 136bp at the 3' end, the apo-dystrophin-4 cDNA is completely homologous to genomic DNA from the same region. Thus, only the alignment of the human and mouse genomic clones are shown (Figure 6.9) and include 233 bp of genomic sequence upstream of the start of the apo-4 cDNA. It shows 96% homology to the full-length human cDNA beginning 43 bp upstream of exon 79 to 860 (1100 in Figure 6.9) and approximately 57.2% overall homology to the expressed 5' apo-4 region 451 bp upstream of exon 79 and 48.5% homology to the proposed exon, termed here as 78.3. The exon identified by GRAIL in the apo-4 cDNA is from 149-243 and shares a 37.9% homology with the mouse DNA shown. Another 75bp of mouse genomic DNA should be sequenced however, to have a quantitatively identical number of nucleotides in this region and thus determine a more precise alignment. The potential mouse exon identified by GRAIL at 465 shares 64% homology with the human sequence over the next 78 nucleotides to the first available splice site, lending weight to the idea that this region contains additional exons which could be spliced into a functional mRNA.

```
          1                                                            50
Mgen1073  .......... .......... .......... .......... ..........
Hapo1234  ctagtttcct attcaatgta tagtgcacca aaggtcaatt caagagttta
Consensus ---------- ---------- ---------- ---------- ----------

          51                                                          100
Mgen1073  .......... .......... .......... .......... ..........
Hapo1234  ttattattat tttcaaccca agtaaaagca gagagaaaat agccacctcc
Consensus ---------- ---------- ---------- ---------- ----------

          101                              begin GRAIL exon @149 150
Mgen1073  .......... .......... ttcACAGgCT tAAgCAGCca gtAAATGAcA
Hapo1234  accatagcct cagaagcaag ccaACAGcCT gAAaCAGCtt tgAAATGAaA
Consensus ---------- ---------- ---ACAG-CT -AA-CAGC-- --AAATGA-A

          151                                                         200
Mgen1073  AtT....... .......... .........T AtgtGgtAgt cAgGtcactG
Hapo1234  AgTtggtgtg gcggtgatgg tggcagtgaT AatgGtgAcc gAtGgttggG
Consensus A-T------- ---------- ---------T A---G--A-- -A-G-----G

          201                                          apo-4 5'end 250
Mgen1073  TGCTGGTaAT GGTgaTctTA GcaGgcAgAG aaGGTGgTaG TGaTTTGATa
Hapo1234  TGCTGGTgAT GGTagTggTA GttGtgA.AG gtGGTGaTgG TGgTTTGATt
Consensus TGCTGGT-AT GGT--T--TA G--G--A-AG --GGTG-T-G TG-TTTGAT-

          251          M1                                             300
Mgen1073  GtaAaagtgt AgAcTaTaCa acAgaAtAAa TAcAagtatA GTAA......
Hapo1234  GatAgtaaaa AaAaTgTtCg ttAatAcAAg TAgAgagtaA GTAAtcaatc
Consensus G--A------ A-A-T-T-C- --A--A-AA- TA-A-----A GTAA------

          301                              M2          M3      350
Mgen1073  .......atc caaCAAaGTG tgAAAGgTGT gTgCCATtAc acAtctTTCt
Hapo1234  aatcactcat agcCAAgGTG gaAAAGaTGT aTcCCATcAt ggAataTTCc
Consensus ---------- ---CAA-GTG --AAAG-TGT -T-CCAT-A- --A---TTC-

          351                                                         400
Mgen1073  cG........ GtgATaagag cCTTgTCTAT GaAgTTC... TGAgATgTgT
Hapo1234  tGttctgata GaaATcttgt gCTTaTCTAT GgAaTTCttt TGAtATaTaT
Consensus -G-------- G--AT----- -CTT-TCTAT G-A-TTC--- TGA-AT-T-T

          401                                                         450
Mgen1073  TaggAagatG AAtCatcAat TtaCaT.... TTcTcCCcat cAAAtgaCAc
Hapo1234  TtacAttggG AAcCtgaAtg TagCtTgaca TTtTtCCatg tAAAcacCAg
Consensus T---A----G AA-C---A-- T--C-T---- TT-T-CC--- -AAA---CA-

          451              begin mouse GRAIL exon                     500
Mgen1073  cAtgCTGATC CAgtATTAAG CTaATACTAA C.....ACca tgcAatGCTT
Hapo1234  tAgcCTGATC CAacATTAAG CTgATACTAA CaaacaACgt gtaAtgGCTT
Consensus -A--CTGATC CA--ATTAAG CT-ATACTAA C-----AC-- ---A--GCTT

          501                                                         550
Mgen1073  CATTAAcAAG GaTTTGCTTC TTgCTaGAAA tgGGT..AAA AaCggACtgT
Hapo1234  CATTAAtAAG GcTTTGCTTC TTcCTgGAAA ctGGTgaAAA AtCaaACctT
Consensus CATTAA-AAG G-TTTGCTTC TT-CT-GAAA --GGT--AAA A-C--AC--T

          551                                                         600
Mgen1073  GgTcTGTAtA CCtTCaATGC AGCTTaTGTG TTGTCTTttC C..tgAAatG
Hapo1234  GtTgTGTAcA CCcTCgATGC AGCTTcTGTG TTGTCTTcaC CcagaAAtgG
Consensus G-T-TGTA-A CC-TC-ATGC AGCTT-TGTG TTGTCTT--C C----AA--G

          601                                                         650
Mgen1073  GtAATGAcTc CCaAtAgtGg cAAccAgggG tacaATaCT. ......TGCA
Hapo1234  GgAATGAtTt CCcAaAtgGc aAAgaAacaG agtgATgCTa tctatcTGCA
Consensus G-AATGA-T- CC-A-A--G- -AA--A---G ----AT-CT- ------TGCA
```

```
                651                                                  exon79 700
Mgen1073    CacTTTGTAA A....cTCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT
Hapo1234    CctTTTGTAA AgtctgTCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT
Consensus   C--TTTGTAA A-----TCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT

                701                                                  750
Mgen1073    AGGAAGcCTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT
Hapo1234    AGGAAGtCTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT
Consensus   AGGAAG-CTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT

                751                                                  800
Mgen1073    TAGTtTCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC
Hapo1234    TAGTaTCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC
Consensus   TAGT-TCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC

                801                                                  850
Mgen1073    TCCTGATTCC CGCATGGTTT TTATAATATT CgTACAACAA AGAGGATTAG
Hapo1234    TCCTGATTCC CGCATGGTTT TTATAATATT CaTACAACAA AGAGGATTAG
Consensus   TCCTGATTCC CGCATGGTTT TTATAATATT C-TACAACAA AGAGGATTAG

                851                                                  900
Mgen1073    ACAGTAAGAG TTTACAAGAA ATaAAATCTA TATTTTTGTG AAGGGTAGTG
Hapo1234    ACAGTAAGAG TTTACAAGAA AT.AAATCTA TATTTTTGTG AAGGGTAGTG
Consensus   ACAGTAAGAG TTTACAAGAA AT-AAATCTA TATTTTTGTG AAGGGTAGTG

                901                                                  950
Mgen1073    GTAcTATACT GTAGATTTCA GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA
Hapo1234    GTAtTATACT GTAGATTTCA GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA
Consensus   GTA-TATACT GTAGATTTCA GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA

                951                                                 1000
Mgen1073    ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAAA AGTTAaAAGA
Hapo1234    ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAAA AGTTAtAAGA
Consensus   ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAAA AGTTA-AAGA

               1001                                                 1050
Mgen1073    AA...ACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT
Hapo1234    AAactACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT
Consensus   AA---ACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT
                                              begin inversion@1100
               1051                                                 1100
Mgen1073    GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGAt
Hapo1234    GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGAa
Consensus   GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGA-

               1101                                                 1150
Mgen1073    TgtaAActaA Agtgtgcttt AtAAAAaAAg ttgtTtataA AaaccctAa
Hapo1234    TtatAAaggA A.......aa AgAAAAtAAc gcaaTggacA AgtggtgaAg
Consensus   T---AA---A A--------- A-AAAA-AA- ----T----A A-------A-

               1151                                                 1200
Mgen1073    acaaacACaC AcGcacaCAC AcacAcacac AcacaCaCAc AcaCAcAcTG
Hapo1234    ctgtgaACtC AgGtgtgCAC AattAtcagg AacacCcCAa AacCAaAgTG
Consensus   ------AC-C A-G----CAC A---A----- A----C-CA- A--CA-A-TG

               1201                                                1243
Mgen1073    AGGcAGcAca ttgtTttGcA ttacTtTagc gTGTatcaTA t..
Hapo1234    AGGtAGaAat agcaTgaGaA gccgTgTttg aTGTtaatTA att
Consensus   AGG-AG-A-- ----T--G-A ----T-T--- -TGT----TA ---
```

Figure 6.9. The alignment of Human cDNA and Mouse genomic DNA in the region of Apo-dystrophin-4 (Hapo) sequence. Hapo1234 is the 997bp of the human apo-4 cDNA, beginning at +239 plus 233bp of upstream genomic sequence. Mgen1073 represents the genomic mouse DNA sequenced in the apo-4 5' region with some upstream sequence. Apo-4 aligns with genomic DNA apart from the inversion at 1100, where mouse DNA is 35.6% homologous. 96.4% homology is shown at the beginning of the splice acceptor site for exon 79 at 667, to the inversion at 1100. The proposed exon 78.3 ranges from 318-425 and shows 48.5% homology to mouse DNA from the same region. The potential exon identified by GRAIL from 318-425 near the start of the apo-4 cDNA shows 37.9% homology to mouse. The first three starting methionines in apo-4 are underlined between 251 and 350 and designated M1, M2 and M3.

The 45.5% homology shown for exon 78.3 is markedly less than the homology in the region of exon 79, yet it is not inconsistent with the approximately 40% homology in the N-terminus between human and mouse for other cell surface molecules, including a proximal extracellular domain of phagocytic glycoprotein-1 (Pgp-1, 42%) (Zhou et al. 1989) and the N-terminus of the leukocyte common antigen (LCA, 50%) (Thomas et al. 1987).

6.4 GRAIL identifies an exon upstream of apo-dystrophin-4

Because the apo-4 sequence contained stop codons and was from a region which normally acts as an intron, we attempted to identify potential new exons by submitting the apo-4 sequence to a GRAIL search to localise intron/exon borders based on known probabilities. Both the full length mouse sequence (1044bp) and human sequence (1234bp) which included the ~230bp of sequence upstream of the apo-dystrophin-4 cDNA were submitted to a GRAIL database search for the potential identification of new exons. While no new exons were found in apo-4, the mouse sequence identified a potential exon at position 291 (465 in Figure 6.9) with a score of 0.141 and the human sequence identified a potential exon at

position -91 (149 in Figure 6.9) with a high score of 0.936 in the same frame as that predicted for apo-4. These studies suggested that the apo-4 gene was unspliced.

6.4.1 A possible upstream transcription start site is found

Primer Extension was performed to attempt to identify the 5' end of the mRNA transcript transcribed from apo-4 and a potential upstream transcription start site to determine whether the 5' end of apo-4 contained a truncated open reading frame (Fig. 6.10A). For the primer extension reaction, an antisense 36 oligonucleotide primer was designed to anneal 80 bp downstream of the 5' end of the apo-dystrophin-4 cDNA with the structure: (5'CCTTGGCTATGAGTGATTGATTGATTACTTACTCTCTACTTG3'). The same primer was used to sequence apo-4 cDNA in pBluescript and λ34 genomic phage clones to precisely map the distance of a transcription initiation site from the 5' end of the apo-4 cDNA. Transcription start sites vary in distance from the initiating methionine, but most are from 50-150bp upstream (D. Hochhauser, personal communication). The genomic clones had not previously been sequenced in this region, which covered part of the intron between exons 78 and 79 in the full-length dystrophin cDNA. In a comparison with genomic DNA alongside the cDNA, the appearance of a band indicated the end of the RNA transcript and a (proximal) start site for transcription at 70 bp upstream of the 5' end of the cDNA, -70 in the given sequence, near the end of the putative upstream exon identified with the GRAIL search (Figure 6.26A) and within range to act on any of the first three methionines in apo-4. Another band, which may correspond to an additional (distal) start site, occurs roughly 250 bp upstream of the proximal start site. This putative promoter may act upon a potential exon upstream of apo-4 and its upstream exon. A TATA box was not found in what

should be the promoter site, although a CAAT box was identified as well as a GC rich region. When the apo-dystrophin-4 upstream sequence was submitted to the Eucaryotic Promoter Database (Bucher 1990), region -139 to -177, 69-108bp upstream of the putative initiation of transcription, aligned with 70% identity to the Hs arginase liver promoter. Thus far, five promoters, not all with TATA boxes, have been isolated along the length of the dystrophin gene, from tissue including brain (Gorecki et al. 1992), muscle, cerebellum, sciatic nerve, and general non-muscle (reviewed in Ahn and Kunkel 1993).

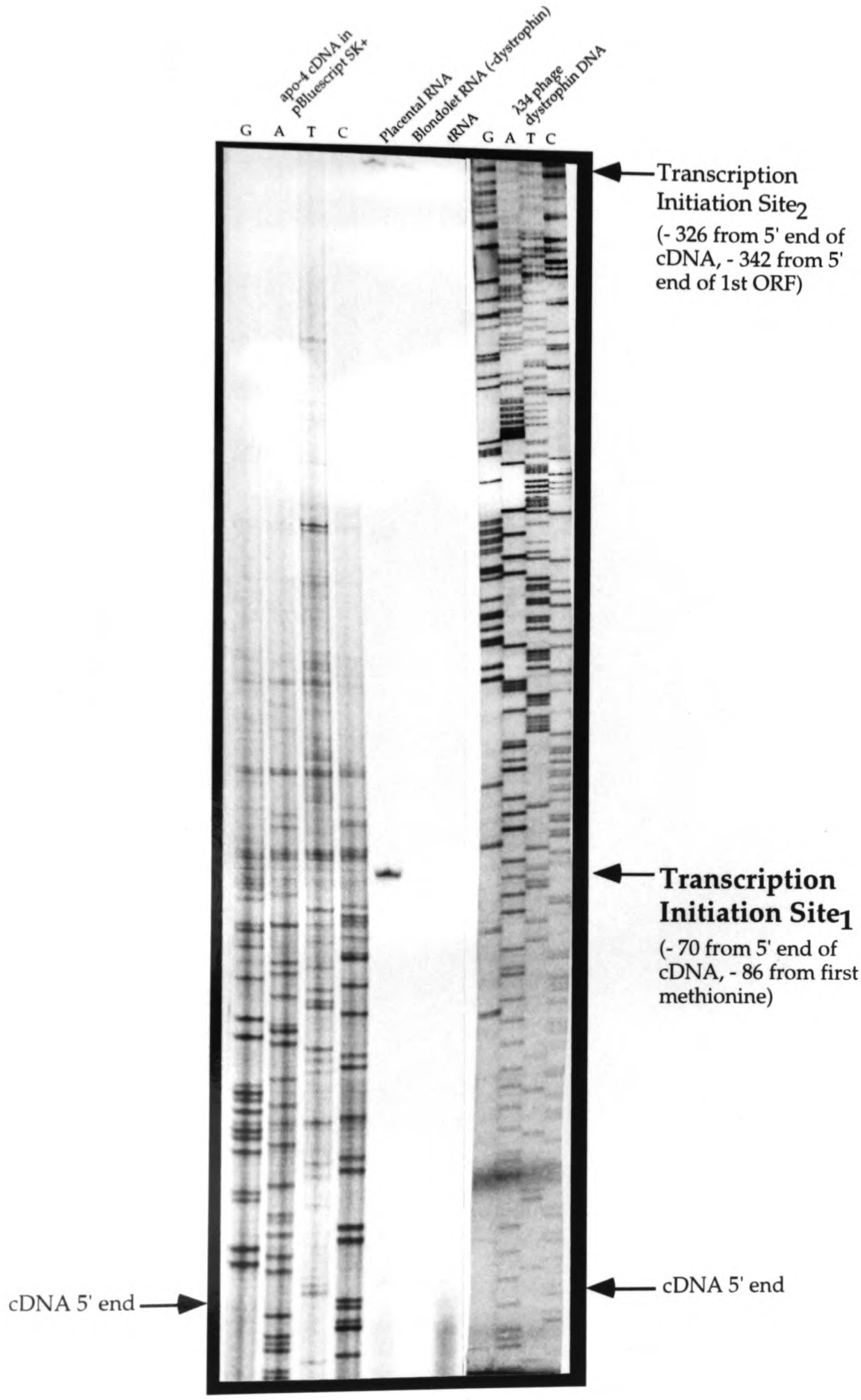


Figure 6.10A. Primer Extension shows a putative transcriptional start site for the apo-dystrophin-4 cDNA. One putative transcription start site is shown in genomic DNA at -70 bp from the 5' end of the apo-dystrophin-4 cDNA, -86 from M1 and absent in the Blondolet muscle cell line which lacks dystrophin. A second potential transcription start site appears at approximately -326 bp, present in minor amounts in the cell line. Neither site appears in the control tRNA lane. Exposure time was 12 hrs. at -80 °C.

6.4.2 The start site contains a potential Pol II binding region

Promoters for RNA polymerase II genes which do not contain obvious TATA boxes fall into two classes (Smale and Baltimore 1989). The first is the GC-rich promoter, found primarily in housekeeping genes which usually contain several transcription start sites spread over a large region and several potential binding sites for the Sp1 transcription factor (Dynan and Tjian 1983). The second class includes the remaining promoters, many of which are not constitutively active but are regulated during differentiation or development and initiate transcription at only one or a few tightly clustered start sites. Drosophila homeotic genes are in this class as are genes that are regulated during mammalian immunodifferentiation, such as the terminal deoxynucleotidyltransferase gene (TdT) (Landau et al. 1984); the TcR β-chain genes (Anderson et al. 1988); the *lck* gene (Garvin et al. 1988), the λ_5 gene (Kudo et al. 1987); and the $V_{pre\text{-}\beta}$ gene (Kudo and Melchers 1987). The transcription start site and surrounding sequences characterised in the apo-4 5′UTR suggest that the promoter element is in the second class. This class of promoter is characterised by at least 40-50% homology to a 17 bp "Inr" sequence which flanks the transcription initiation site found in TdT (Smale and Baltimore, 1989) (Figure 6.10B).

```
          -70 bp from 5' end of apo-4
                        |
  Inr = GCCC TCAT TCTG GAGAC
apo-4 = GCGG TGAT GGTG GCAGT - 48% perfect homology with Inr
                                 71% match on type of base
                                   (purine vs. pyrimidine)
```

Figure 6.10B. An Initiation motif found in apo-4 may act as an Inr promoter in place of a TATA box. The above sequence begins at -70 in the apo-4 5′UTR and shares nearly 50% homology with the "initiation response element" (Inr) shown to act in place of a promoter.

6.5 Northern blotting shows a major apo-4 transcript of 1.62 Kb

To test whether the apo-4 gene could be expressed at high levels in selected tissues and cell-lines, a poly(A) selected blot was prepared with 5-10 μg RNA per lane from placenta and leukemic cell lines using oligo dT cellulose columns. The RNA was purified in a separate laboratory to reduce the risk of contamination (courtesy of Andrew McMichael, IMM). The blot was probed with a full length apo-dystrophin-4 ssDNA and compared with previously isolated RNA made using oligo dT Dynabeads. Successful hybridisation was shown on RNA fom apo-4 COS cell transfectants from Dynabead RNA but not to K562 or the dystrophin-negative cell line Blondolet (Figure 6.11), although the K562 and Blondolet had slightly less RNA and had been previously thawed a few times. No hybridisation was demonstrated against 10 μg of oligo dT column Poly A selected RNA from panel of T, B and myeloid cell lines and only very little against the positive control COS transfectant poly A+ RNA, indicating that the RNA, although it mostly appeared of high quality on coomassie blue staining of the blot, was not hybridising well (data not shown). This shows that the full length transcript is produced in apo-4 transfected COS cells at high levels, as expected.

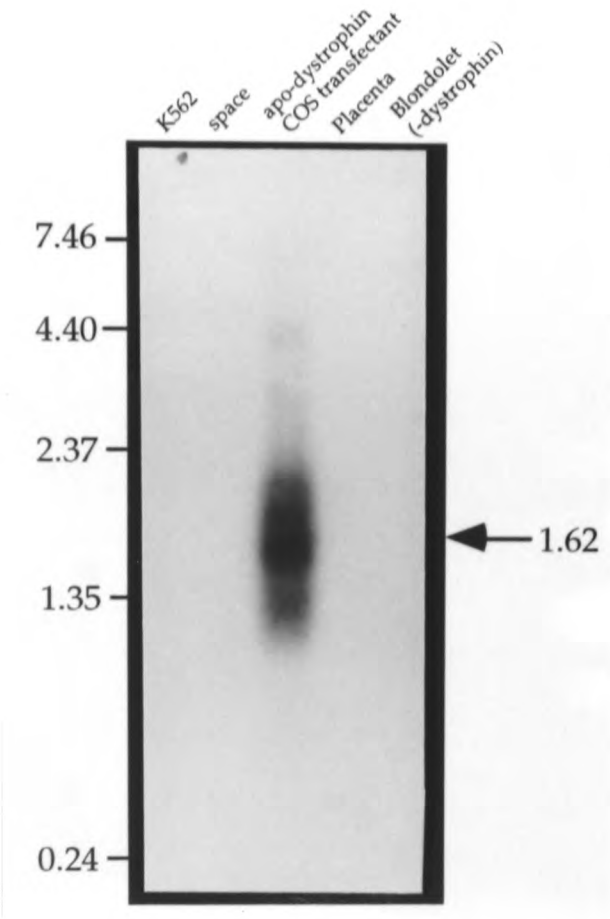


Figure 6.11 The apo-dystrophin-4 cDNA hybridises to RNA from apo-4 transfectants. 10 µg of Poly A selected RNA was probed using a full-length apo-dystrophin-4 DNA probe. CDM8 has an endogenous Poly A site 625 bp from the Pst I site at the 3' end of the insert. Transcripts range in size from 1.2-1.62Kb with a predominant transcript at 1.62 Kb. The apo-4 cDNA contains endogenous Poly A sites at 546 and 632 and the slightly altered sites TATAAA at 849 and 863 bp and AATTAA at 990bp which may account for the smaller signals. The K562 cell line, Placenta and dystrophin-deleted Blondolet cell-line did not hybridise to the probe. The placental RNA was total, as insufficient quantities were available for Poly A selection. The blot was washed twice in 2X SSC, once in 1X SSC and exposed for five hours at -80 °C.

A 5' 451 bp apo-4 probe also did not hybridise to a commercially available blot with 10µg Poly A selected RNA from heart, lung, muscle, kidney, liver and placenta (data not shown). Thus, apo-dystrophin-4 transcripts are likely to be restricted and not expressed at a high level.

6.5.1 RNAse protection reveals no protected transcript

To determine whether the inversion from apo-4 cDNA could be detected in 10 μg of total RNA, a 290bp cDNA probe designed from apo-4 cDNA cloned into pBluescript SK+. The insert was then digested with Hind III and Hpa I to excise the 5' 700bp from the apo-4 insert and the remaining 290 bp were religated. After scaling up the DNA, the new insert in pBS was digested with Xho I to obtain a probe containing a 290 bp insert and 130 bp of vector sequence, for a 420bp probe. Following RNAse protection with this probe, no unique signal was seen in the expected region above the 180 bp GAPDH probe, which protected the expected fragment of 120 bp (Figure 6.12). Unique signals of about 80 bp were seen on the gel, but they were considered too small to be significant.

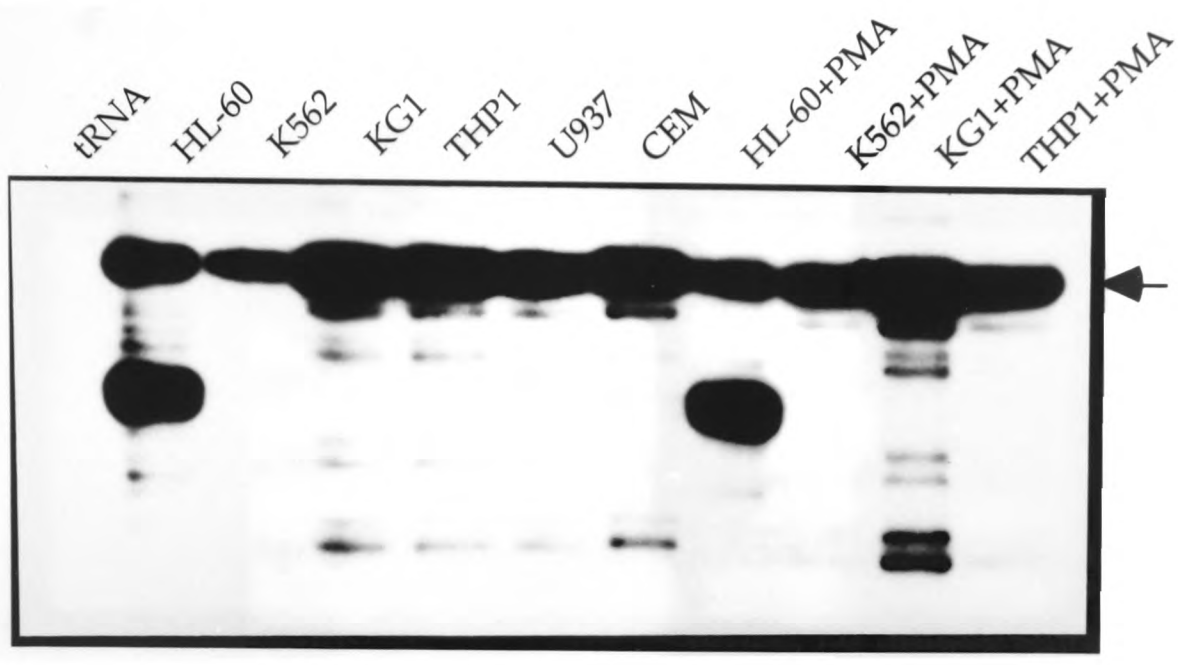


Figure 6.12. RNAse protection on 10 μg of total RNA shows hybridisation to a GAPDH probe but not to a 3' apo-4 probe. The arrow points to a strongly hybridised GAPDH fragment at 120 bp and demonstrates undegraded RNA, but the expected 290 bp signal was not detected for the 3' apo-4 probe (lanes not shown). Probes were labelled with ^{32}P and gel was exposed for 24 hrs. at -80 °C..

To examine whether the 5' 468 bp of apo-4 could be detected in RNAse protection, a probe was prepared by digesting with Pst I and Xmn I to leave the 5' 468 bp, but the protection assay revealed no hybridised product. This confirmed the hypothesis that apo-4 was a very rare transcript which could only be detected via RT-PCR, as was the case for the apo-dystrophin transcript Dp71 (Hugnot 1992) and many of the splice variants of CD44 (Jackson et al. 1992).

6.5.2 RT-PCR analysis produces evidence for apo-4

To attempt to reproduce a full-length apo-4 gene, isolate potential splice products and/or to reproduce the inversion, oligonucleotide primers were designed around the regions of the novel cDNA known to diverge with published dystrophin cDNAs (Figure 6.13). Initial experiments were done in collaboration

with Jamel Chelly in the Human Genetics Lab, IMM, who first demonstrated that there is a low level of transcription of tissue specific genes in every cell type by identifying rare "ectopic" or "illegitimate" dystrophin transcripts in lymphoid tissue which could not be detected by Northern blotting or RNAse protection using RT-PCR (Chelly et al. 1988). Dr. Chelly also developed the quantitative PCR technique used in this study (Chelly et al. 1990). The primers were then incubated with 1st-round cDNA made from freshly isolated RNA (hereafter referred to as "RNA" to distinguish it from genomic DNA or a cDNA library) from various tissues in Reverse Transcriptase polymerase chain reactions (RT-PCR), with the apo-4 cDNA used as a control in only one of the three experiments shown. Splice products which would include the inversion were sought by designing a primer which was just downstream of and overlapped the inversion (R2o) and this was used in combination with F1 and R1 (Figure 6.14). The initial RT-PCR using dystrophin-negative patient samples and placental genomic DNA was performed in Dr. Monaco's laboratory without the apo-4 cDNA with newly synthesised primers and pipetmen which had not been used with apo-4 cDNA to prevent potential apo-4 contamination in negative controls. Genomic DNA was used to test whether the transcript could be detected in normal DNA and provided an additional control of DNA without reverse transcriptase added. The initial experiment demonstrated a reproduced inversion in RNA from adult brain, KG-1 and THP1 cell lines and no product in dystrophin-negative patient cell lines or with primers alone, despite an extended annealing time of 30 minutes (Figure 6.15A). A further experiment was then performed with similar samples with the addition of commercially available placental DNA, more cell lines and the apo-4 cDNA control. Positive products from the second experiment were Southern blotted alone (Figure 6.15C) and with the products from the first experiment (Figure 6.15B). Southern blots were probed with an 18-21 bp oligonucleotide, usually internal to those used to

produce the product. All combinations of primers were checked with the "Amplify" program and none were potential dimer formers. The RT-PCR primers included: F1 (@80) GATTGATAGTAAAAAAAATG;

F2 (@712) CAATGGCAGGTTTTACACGTC;

R2 (@475) GGAAAAGACTTCCACATTGT;

R2.3o (@875) CTTTTTCCTTTATAATTCTTTC and;

R1 (@986) CATCAAACACGGCTTCTCATGC as pictured (Figure 6.13).

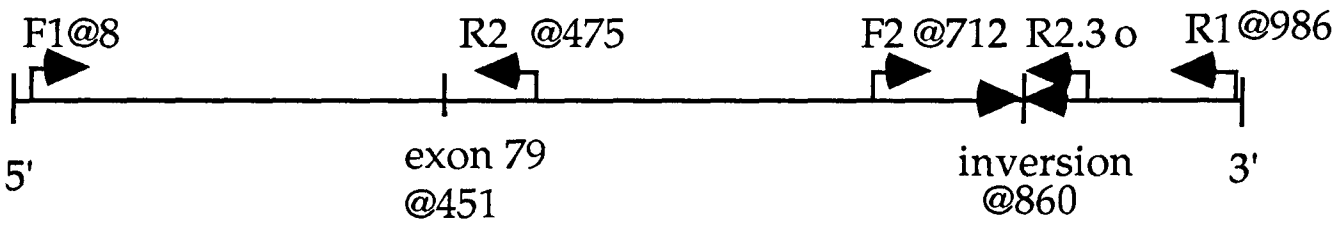


Figure 6.13 Position of RT-PCR primers to analyse potential splice products and to attempt to reproduce the inverted region. Three sets of PCR primers were designed to reproduce the 5' unique 451 bp (F1+R2), the 3' 284 bp including the inversion (F2+R1) and 978 bp of the full-length apo-4 cDNA (F1+R1) or smaller, potentially spliced products.

6.5.3 Evidence for splicing and a reproduced inversion

Four regions of the apo-dystrophin-4 cDNA were explored using Reverse Transcriptase PCR followed by Southern blotting. The first blot looks only at potential splice products and an attempt to reproduce the full-length apo-4 while the next three examine splice products, whether the inversion could be reproduced and whether an unspliced 5' end of apo-4 could be reproduced:

1) F1+R2o were used to reproduce the apo-4 5' end to the inversion and to probe for splice products containing the inversion. For these reactions, newly isolated cDNA made from freshly isolated RNA was used in the absence of the

apo-4 plasmid (Figure 6.14). F1+R2o (R2o is 3′ of the inversion except for the last 5/22 bp) products provided suggestive evidence that apo-4 could be spliced in addition to full-length reproduced products. In a second experiment, however, only the full-length products were reproduced. As expected, the lane with primers alone yielded no products. However, a full length product was not reproduced despite having reproduced a region immediately over the inversion. Products from F1+R2o which blotted with the F2.2R probe were obtained from K562 (0.85. 0.65, 0.4, 0.2 Kb), K562+PMA (0.85. 0.65, 0.4, 0.2 Kb), HL-60 (1.0, 0.85. 0.65, 0.4, 0.2 Kb), HL-60+PMA (1.0, 0.85. 0.65, 0.4, 0.2 Kb), CEM (0.85. 0.65, 0.4 Kb), placental RNA (0.85. 0.65, 0.4 Kb), U937 (0.85. 0.65, 0.4 Kb) (Figure 6.14). In subsequent assays, only the .85 Kb product was reproduced indicating that it is more highly represented in the RNA than the potential splice products. Shorter F1+R2 fragments were detected in separate blots.

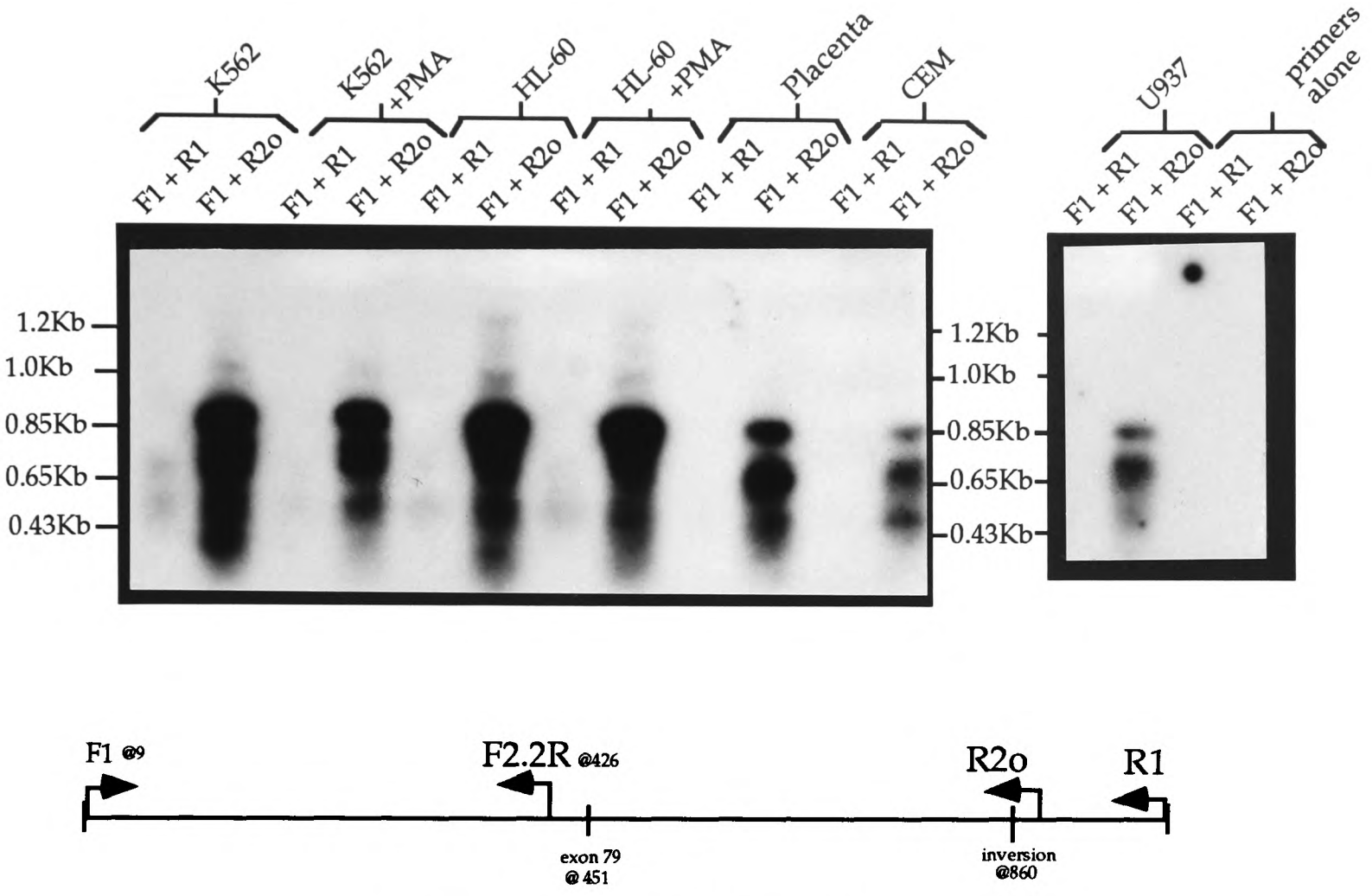
2) F1+R1 were used to attempt to reproduce the complete 1Kb ap-4 cDNA. From this set of primers, only one product of about 1.2Kb from HL-60 RNA has been produced which successfully hybridised to other primers from regions internal to the product in a Southern blot (Figure 6.15C), although an apparent apo-4 product was reproduced from the HPB-ALL cDNA library (Figure 6.14) and other cDNA libraries using standard PCR (data not shown). A summary of the PCR products obtained with and without Southern blotting is provided in Table 6.1. The only potentially spliced RNA products obtained with these primers were 0.29 and 0.43Kb from K562 and HL-60 RNA (Figure 6.14) and one 0.6Kb one which did not appear to contain the inversion from placental cDNA (Figure 6.15C). Other F1+R1 potential splice products from RNA which hybridised to the F2 oligo included those from HepG2 (.1Kb), K562 (0.2 and 0.1Kb), and HL-60, CEM and U937 (0.1Kb) all (Figure 6.15C).

3) F2+R1 were used to reproduce the inversion. These primers predicted an approximately 300 bp product and were used to test whether the inversion could be reproduced in cell lines. In two separate experiments, the inversion was reproduced first in brain, THP1, KG-1 and second in K562, HL-60 and placental DNA and a very faint amount in HepG2 (Figure 6.15A). In the first RT-PCR reaction with dystrophin-negative patient samples, the 0.3Kb full-length inversion product alone was visibly reproduced in Human brain, THPI and KG-1 RNA with F2+R1 but did not appear in the negative control samples with primers alone, patient NJ or LB RNA (Figure 6.14B). Significantly, the inversion was reproduced in genomic DNA. The lack of a full-length reproduced apo-4 product and no primer signal in the original gel indicates that the products are not due to apo-4 plasmid contamination. The inversion was consistently reproduced in genomic DNA and 1st round cDNA in a number of future RT-PCR experiments.

Of the candidates sequenced following cloning into Bluescript SK+ (Stratagene), brain cDNA shows highly conserved homology at the peptide level to the inverted sequence (Table 6.2). This product was from the group of three in the initial RT-PCR which successfully hybridised to probes internal to the PCR product and showed no products in the control lanes or with primers alone (Figure 6.15A).

These findings suggest that the 5' 876 bp of apo-4, previously considered "intronic" can be obtained from RNA, and that the overlapping inversion sequence can be reproduced. The overlapping clones suggest that the full-length clone could be obtained with the right combination of primers or nested PCR. All of the techniques with RNA depend on high quality RNA and high abundance as well as the correct PCR strategy or hybridisation conditions, all of

which can be highly variable. It was thus important to consider the expression of apo-4 *in vitro* and through *in vivo* expression and staining studies.



The Apo-dystrophin cDNA*

*RT-PCR primers and Southern Blotting probe are indicated. F2.2R was used to probe the blot.

Figure 6.14. The 5' 859 base pairs of apo-4 are reproduced in several cell lines with RT-PCR using a probe internal to the primers. Upper Panels: K562 and K562 + PMA (F1+R2o) show three main signals at 0.85, 0.65 and 0.29Kb which are generally reflected in all the samples. These smaller bands may reflect splice products from the original gene. All templates are cDNA made from RNA to test for the presence or absence of the inverted region. No hybridisation is shown for p Blot was exposed 5 days at -80°C.

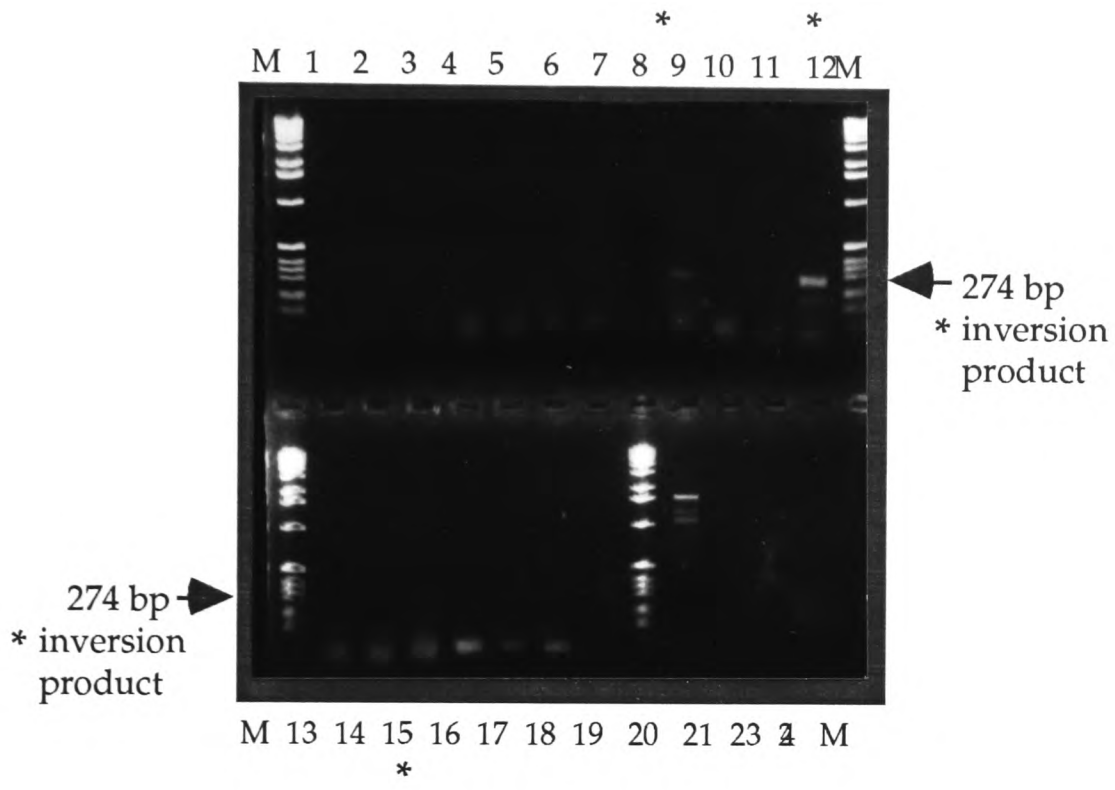
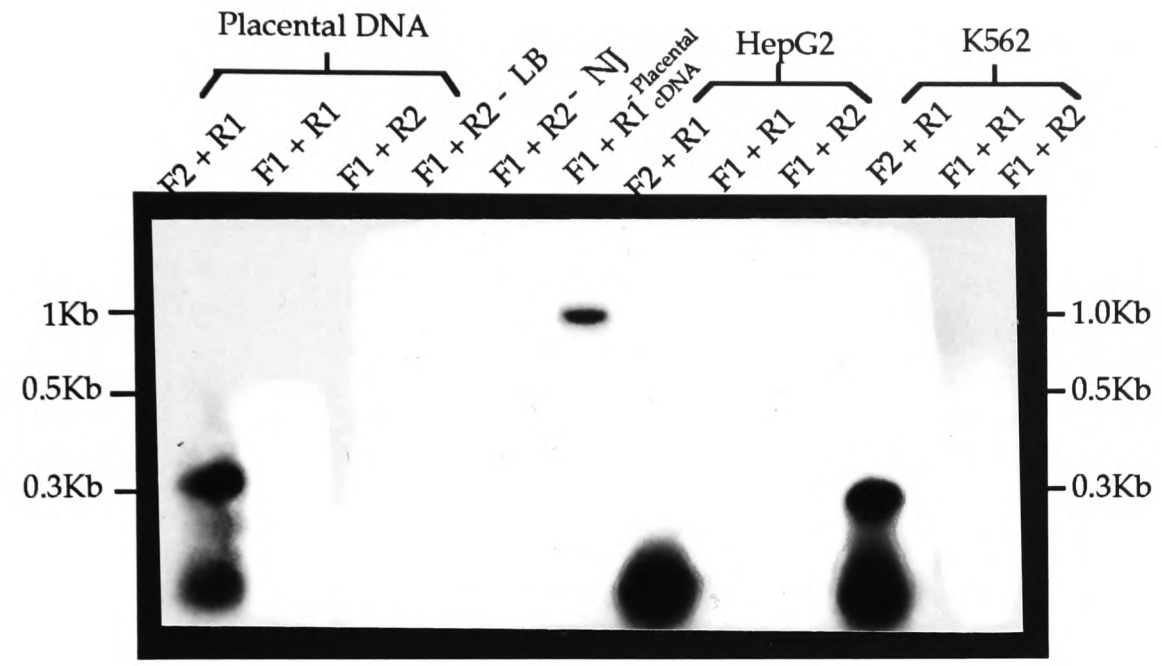
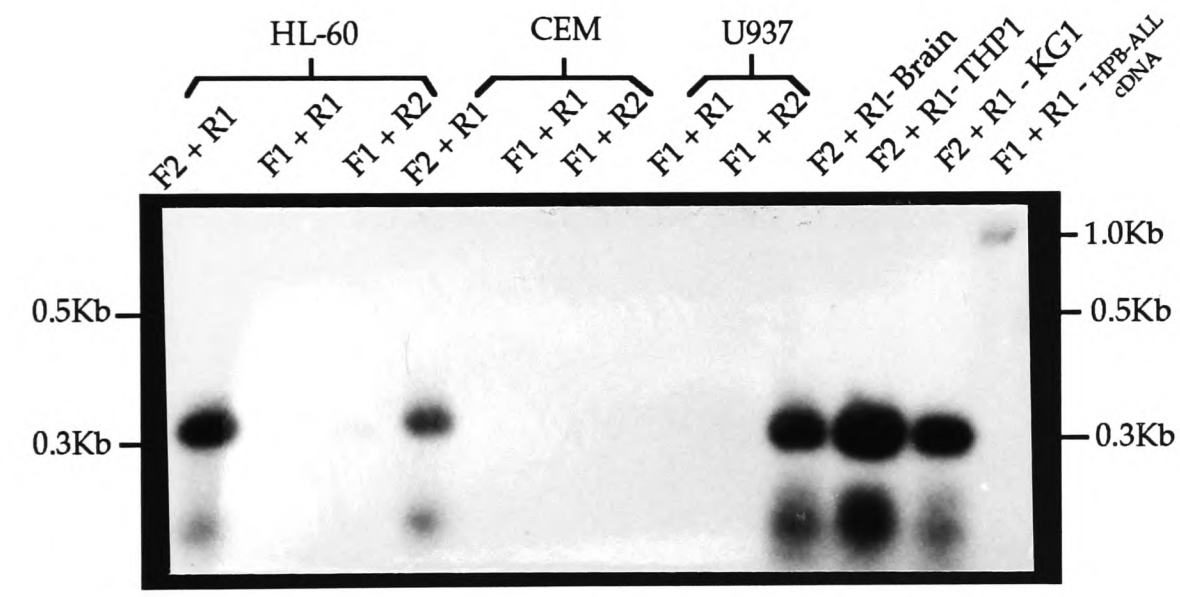
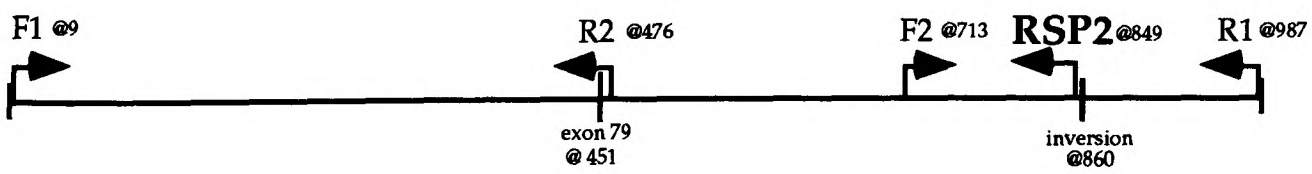


Figure 6.15 RT-PCR performed in the absence of apo-4 shows a reproduced inversion and no primer products. A. RT-PCR on 1st round cDNA with 30 min. rather than 30 sec. annealing shows inversion products. 1. Lymphoblastoma dystrophin (D) negative (-) **patient LB** (F1+R1) 2. LB (F1+R2) 3. LB (F2+R1) 4. D-**patient NJ** (F1+R1) 5. NJ (F1+R2) 6. NJ (F2+R1) 7. **Adult brain** (B) (F1+R1) 8. B (F1+R2) 9. B (F2+R1) 10. **THP1** (F1+R1) 11. THP1 (F1+R2) 12. THP1 (F2+R1) 13. **KG-1** (F1+R1) 14. KG-1 (F1+R2) 15. KG-1 (F2+R1) 16. **primers alone** (PA) (F1+R1) 17. PA (F1+R2) 18. PA (F2+R1) 19. space 20. markers 21. Placenta cDNA library (F1+R1) obtained from previous PCR reaction is shown for a size control.



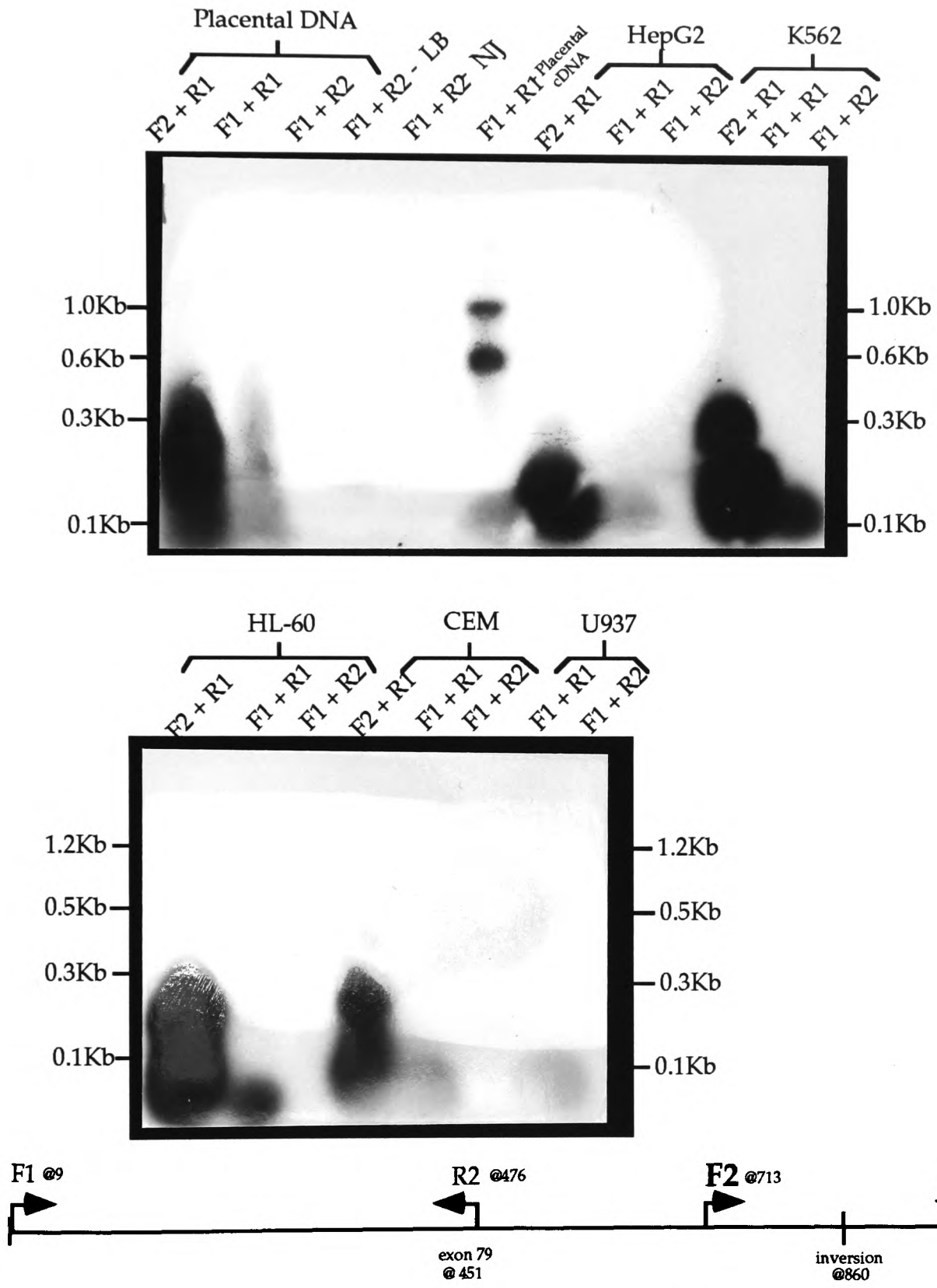




The Apo-dystrophin cDNA*

*RT-PCR primers and Southern Blotting probe (RSP2) are indicated by the arrows.

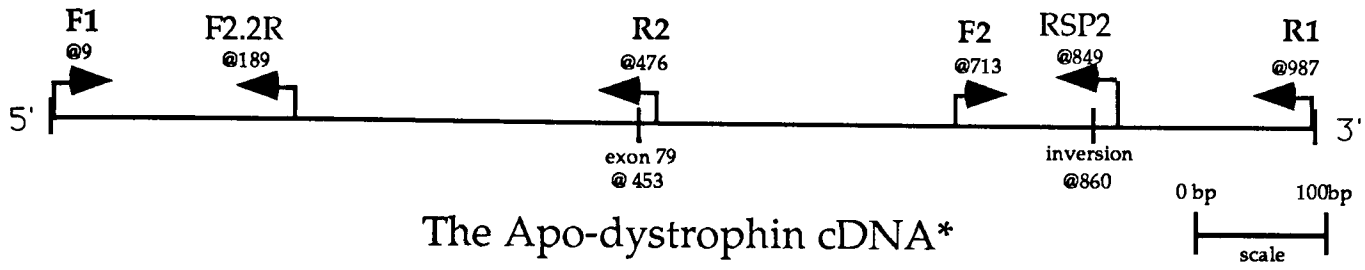
Figure 6.15B. A probe just upstream of the inversion strongly hybridises to all PCR products potentially containing the inversion and the full-length product from HPB-ALL cDNA. Top Panel: A 0.3 Kb fragment is shown for F2 + R1 for (left to right) placental DNA and K562. A faint signal is shown for HepG2. A very small product of about 100 bp also hybridises in placental DNA, HepG2 and K562, which may be primer dimers. **Bottom panel:** A 0.3 Kb fragment in the F2+R1 samples hybidises in HL-60, CEM, Brain, THP1 and KG-1 (final three from a previous run). The F1+R1 fragment from HPB-ALL cDNA also hybridises at 1.0 Kb showing that it carries the region just upstream of the inversion. Blot was exposed 3 hrs. at -80°C.



The Apo-dystrophin cDNA*

*RT-PCR primers are indicated with the arrows. F2 was used as the Southern Blotting probe.

Figure 6.15C. The F2 probe hybridises to a 1.2 Kb PCR product in HL-60 and Placental cDNA which approximates the size of the apo-dystrophin cDNA. Upper Panel: Placental DNA shows a broad signal from 0.1-0.5Kb rather than the expected 0.3Kb for F2+R1 and 1Kb for F1+R1. As expected, no signal is shown to patient DNA (NJ and LB). Placental cDNA shows a potential splice product at 0.6Kb and the expected product at 1Kb and weakly at 1.2Kb in F1+R1. HepG2 strongly hybridises at 0.1 Kb and to a faint band at 0.3Kb in F2+R1. K562 shows a strong signal from 0.1-0.3Kb for F2+R1 and a strong signal at 0.1Kb for F1+R1. **Lower Panel:** HL-60 and CEM show a broad band from 0.1-0.5Kb for F2+R1. HL-60 shows a signal at 1.2Kb and a signal at 0.1Kb which is also in CEM and U937 F1+R1. The blot was exposed for 2 hrs. at -80°C.



The Apo-dystrophin cDNA*

*RT-PCR primers (bold) and Southern Blotting probes are indicated by the arrows.

- Oligo Probe - cDNA position - series. order probed		F2.2R @188 3.0	F2 @713 2.1	RSP2 @849 1.2
RT-PCR sample origin (order follows blots 6.15A-D)	Approximate Hybridised Fragment Size (Kb)			
Placental DNA F2+R1	0.3 0.2 (r) 0.1 (r)	not tested (NT)	++ ++ ++	+ + +
Placental DNA F1+R1	1.0 0.6 0.3 (r) 0.2 (r) 0.1	NT	- - + + +	- - - - -
Placental DNA F1+R2	0.7 0.5 0.4 0.2 0.1	NT	- - - - -	- - - - -
Patient LB ** RNA F1+R2	0.5 0.4 0.1	NT	- - -	- - -
Patient NJ ** RNA F1+R2	0.5 0.4 0.3 0.1	NT	- - - -	- - - -
Placental cDNA F1+R1 (+ control)	1.2 1.0 0.6 (s)	NT	(+) + +	- + (+)
HepG2 RNA F2+R1	0.3 0.1 (s)	NT	(+) ++	- ++
HepG2 RNA F1+R1	1.0 0.1 (s)	NT	- (+)	- -
HepG2 RNA F1+R2	0.5 0.1 (s)	NT	- -	- -
K562 RNA F2+R1	0.3 0.2 (s) 0.1 (s)	NT	++ ++ ++	+ - ++
K562 RNA F1+R1	1.0 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	- (+) (+) - -	- - - + ++	- - - - -
K562 RNA F1+R2	0.5 0.1 (s)	NT	- -	- -
K562 RNA F1+R2o	0.9 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	+ ++ ++ ++ +	NT	NT
K562+PMA RNA F1+R2o	0.9 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	(+) + + + -	NT	NT
HL-60 RNA F2+R1	0.3 0.1 (s)	NT	++ ++	+ (+)
HL-60 RNA F1+R1	1.2 0.29 (s) 0.1 (s)	- (+) -	+ - ++	- - -

Oligo Probe	Fragment Size	F2.2R	F2	RSP2
HL-60 RNA F1+R2	0.5 0.3 (s)	NT	- -	- (+)
HL-60 RNA F1+R2o	1.2 0.9 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	(+) (+) ++ + + (+)	NT	NT
HL-60+PMA RNA, F1+R1	1.0 0.2 (s)	- (+)	NT	NT
HL-60+PMA RNA F1+R2o	1.2 0.9 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	(+) (+) ++ + + (+)	NT	NT
CEM RNA F2+R1	0.3 0.1 (s)	NT	++ ++	- (-)
CEM RNA F1+R1	1.0 0.1 (s)	-	- (+)	-
CEM RNA F1+R2	0.5 0.1 (s)	NT	- -	- -
CEM RNA F1+R2o	0.9 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	- + + + -	NT	NT
Placental RNA F1+R1	1.0	-	NT	NT
Placental RNA F1+R2o	0.9 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	- + + + -	NT	NT
U937 RNA F1+R1	1.0 0.1 (s)	- -	- (+)	- -
U937 RNA F1+R2	0.5 0.1 (s)	NT	- -	-
U937 RNA F1+R2o	0.9 0.7 (s) 0.4 (s) 0.2 (s) 0.1 (s)	- + + + -	NT	NT
Brain RNA** F2+R1	0.3 0.1 (s)	NT	NT	+ (+)
THP1 RNA** F2+R1	0.3 0.1 (s)	NT	NT	+ (+)
KG-1 RNA ** F2+R1	0.3 0.1 (s)	NT	NT	+ (+)
HPB-ALL cDNA, F1+R1	1.0	NT	NT	+
Primers Alone (- control)	< 1.2	-	NT	NT

Table 6.1. RT-PCR provides evidence of potential splice products. * Expected RT-PCR fragment sizes, which correspond to the apo-dystrophin-4 cDNA sizes, are: **F2+R1** - 274 bp (~0.3Kb); **F1+R1** - 979 bp (~1.0Kb); **F1+R2** - 468 bp (~0.5 Kb); **F1+R2o** - 868 bp (0.9Kb) bp. All cDNA was made directly from RNA before subjecting it to PCR unless otherwise indicated ("cDNA" means that the sample was taken from a cDNA library and DNA means that the sample was genomic DNA); ++ = strong signal; + = medium signal; **(+)** = weak signal; - = no signal. **(r)** = potential rearrangements; **(s)** = potential splice product; products less than 0.1 Kb are typically considered "primer dimers". ** primers alone were used in these RT-PCR reactions, and gave a negative result, but were not reloaded on the Southern due to lack of room on the gel. In the same reaction, Patients LB and NJ gave negative results for F1+R1, F1+R2 and F2+R1.

6.5.4 Subcloning reveals close homology with apo-4

From the initial RT-PCR, several products were successfully subcloned into the pBluescript SK$^+$ (pBS) vector (Table 6.2). Based on these cloning results,

Tissue	Primers	Predicted Product Size	Product size	Product Sequenced	Alignment with apo cDNA	Splicing?
Human Brain (a)	F2+R1	274bp	~274bp	274bp	96.9%	no
Human Brain (c)	F2+R1	"	~274bp	195bp	95.8%	no
U937 cDNA	F1+R2	468bp	~450bp	238bp	97.9%	no
HPB-ALL cDNA	F1+R1	979bp	~950bp	139bp	67.2% (50% at peptide)	unlikely

Table 6.2. Subcloned RT-PCR products in pBluescript SK$^+$. Several potential splice products were obtained, with those sequenced shown above. In brain, two products were obtained sharing close but not exact homology with the original apo-4 cDNA, as well as one 5' product (8-11) and one cDNA full-length product which was not an exact duplication of the apo-4 cDNA from an HPB-ALL cDNA library.

three important features emerge. First, it is apparently possible to duplicate the inverted region at 860 bp in brain cDNA, which suggests that the inversion found in placenta may also be present in brain tissue and a number of cell lines. Second, none of the above subclones shows evidence of splicing as they match the expected size. Third, although the full-length transcript was apparently reproduced from the HPB-ALL cDNA library, upon sequencing it demonstrated a very low nucleotide homology to apo-dystrophin-4 and thus is not a fully reproduced transcript, although some homology was shown at the peptide level. Finally, the 5' 875 bp of apo-4, including 15 bp of the inversion appears to be completely reproduced from freshly isolated RNA in placenta, K562, the version of HL-60, CEM and U937.

6.6 Inversion breakpoints are identified on genomic DNA

The apo-dystrophin-4 cDNA appears to be a piece of genomic DNA which has undergone rearrangement but not RNA cleavage, polyadenylation or splicing. The rearrangement could be due to signal sequences surrounding the 136 bp region in genomic DNA to trigger it, possibly as a transposable element, to insert itself at the AT-rich stretch which becomes position 860 bp in apo-4. As will be shown, *in vitro*, the absence of the inverted region eliminates the production of apo-4 proteins, suggesting that it may function as a downstream enhancer or post-transcriptional modifier to activate a normally dormant exon (Figure 3.18). In other words, DNA which was previously considered to be intronic could act as an exon in the presence of this rearrangement. Such is the case with some inverted sequences in V-D-J joining in immunoglobulin rearrangement, where an inversion occurs proximally to an enhancer which then activates transcription. In that case, the enhancer is active only in B lymphocytes as suggested in a model for transcription (Lewis et al. 1985). A precedent has also been shown for inversions in DNA rearrangement to occur via recombination signal sequences (RSS'S) in transgenic neonatal brain development with a distribution and affinity for older neural centres (Matsuoka et al. 1991). A proposed site of apo-4 recombination to produce the inversion is imprecise relative to V-D-J joining, as with brain rearrangement. Imprecise joining of Ig elements to the oncogene c-*myc* in Burkitt's lymphoma has also been demonstrated (Gerondakis 1984). A comparison of apo-4 cDNA and genomic DNA with known regions of conservation of the recombination signal sequences (Figure 6.16A) reveals that similar sequences are present in the 3'UTR of dystrophin suggesting a role for them in rearrangement at these sequences (Figure 6.16B). Significantly, at the upstream inversion, the A at 862 which is found in apo-4 was not included in the alignment with the dystrophin 3' UTR to preserve the homology for 15 bp 3' to

the inversion breakpoint (Figure 6.16A). If the A were included the homology would be significantly lower in this region. This A appears to be an "orphan" nucleotide which was randomly inserted upon the formation of the inversion. However, if the inversion resulted from the insertion of an 8 bp direct repeat as suggested (Figure 6.16C) the A was inserted to accomodate the final T in the direct repeat which is not in the unrearranged sequence but is in the inverted sequence. The insertion of the A could have contributed to the inversion breakpoint and cemented the decision to include the inverted sequence in the gene rather than the usual genomic sequence.

A.

12/23bp spacer

CACAGTG----------ACAAAAACC

heptamer nonamer

B.

```
                                  inversion breakpoint1
        11640       11650       11660    |   11670       11680
            *           *           *    |       *           *
dystrophin  T TTATAACAGT TATAAAGAAA GA^TTGTAAAC TAAAGTGTGC
            A AATATTGTCA ATATTTCTTT CT^AACATTTG ATTTCACACG
                                       a
apo-4 cDNA   840         850           |       870
[  138 ]    T TTATAACAGT TATAAAGAAA GA^TTaTAAAg gAAAaaGaaa>
            ^ ^^^^^^^^^^ ^^^^^^^^^^ ^^ ^^v^^^^v v^^^vv^vvv
dystrophin  T TTATAACAGT TATAAAGAAA GA^TTGTAAAC TAAAGTGTGC

-------------------------------------------------------------------
                   11690       11700       11710       11720       11730
                      *           *           *           *           *
dystrophin  TTTATAAAAA AAAGTTGTTT ATAAAAACCC CTAAAAACAA AACAAACACA
            AAATATTTTT TTTCAACAAA TATTTTTGGG GATTTTTGTT TTGTTTGTGT
apo-4 cDNA   880    890         900       910         920         930
[  138 ]    aTaAaAtggA cAAGTgGTga ATgtgAACtC aggtgtgCAc AAttAtCAgg>
            v^v^v^vvv^ v^^^^v^^vv ^^vvv^^^v^ vvvvvvv^^v ^^vv^v^^vv
dystrophin  TTTATAAAAA AAAGTTGTTT ATAAAAACCC CTAAAAACAA AACAAACACA

-------------------------------------------------------------------
                   11740       11750
                      *           *
dystrophin  CACACACACA CATACACACA
            GTGTGTGTGT GTATGTGTGT
apo-4 cDNA    940           950
[  138 ]    aACAC-CcCA -AaAC-CAaA>
            v^^^^ ^v^^  ^v^^ ^^v^
dystrophin  CACACACACA CATACACACA
```

Figure 6.16A-B. Recombination signal sequences are contained around the upstream inversion breakpoint in genomic DNA. A. The nonamer/heptamer motif found in immunoglobulin sequences surrounding rearrangement breakpoints. Underlining shows regions of 87-100% conservation. **B.** Sequence begins at 11662 in the 3' UTR of exon 79 of dystrophin cDNA (Koenig et al. 1987) or 861 on apo-4. The "a" at 862 in apo-4 was ejected from the alignment by the computer because it decreased the homology. This A appears to be an "orphan" nucleotide" which was inserted to allow a downstream direct repeat to integrate itself into the sequence at the underlining (11,654) similar to the orphan sequence found immediately 5' of the inverted region in the γ globin gene, a causative factor in globin rearrangement (Jones et al. 1981). Unrearranged genomic DNA contains two sets of recombination signal sequences immediately downstream of the inversion breakpoint in genomic DNA. RSS's are in boldface type and spacers underlined. The second RSS is in a CA repeat, commonly seen in *Alu* repeats, starting at 11727.

While not all bases in the RSS's are conserved in the dystrophin genomic DNA, those with high conservation are, and the spacing is accurate (Figure 6.16B). Near the point of the apo-4 inversion (5'-3') two sets of RSS's are found with the second contained in a CA repeat on unrearranged genomic DNA beginning at 11,727. It is of interest that the inversion point occurs 66 bp upstream of the start of an "imperfect" (CA)7(TA)(CA)7 repeat in the genomic sequence. The CA microsatellite, often part of the *Alu* repeat, represents one of the most abundant families of interspersed repetetive DNA in eukaryotic genomes (Hamada 1982; Jeang and Hayward 1983; Miesfeld et al. 1981) and has characteristically been used as an often polymorphic marker for genomic mapping between patients and/or species (Weber 1990). This study shows that a CA repeat could contain signal sequences used in recombination, although its distance from the inversion breakpoint implies that it plays a secondary role in producing the inversion. Repetitive DNA is known to contribute to hairpin loop formation, however, which may have contributed to the formation of the inversion.

6.6.1 Downstream RSS's flank an inversion breakpoint

The 136 bp of the inversion was located in unrearranged Genomic DNA. RSS's were found in the forward and reverse strands of DNA at the upstream breakpoint of the inversion as well as one in the reverse strand near the downstream breakpoint, both of which could have been utilised during rearrangement (Figure 6.16B).

```
                                                          inversion breakpoint2
                  13130       13140       13150       13160     |   13170
                    *           *           *           *       |       *
dystrophin  AATTAGCTTT TGGAGAGTGG GTTTTGTCCA TTATTAATAA  TTAATTAATT
            TTAATCGAAA ACCTCTCACC CAAAACAGGT AATAATTATT  AATTAATTAA

                                                                 990
apo-4       ____________________________________________________<AATTAATT
                                                                  ^^^^^^^^
dystrophin                                                        AATTAATT

-------------------------------------------------------------------------
                  13180       13190       13200       13210       13220
                    *           *           *           *           *
dystrophin   AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTTGGG
             TTGTAGTTTG TGCCGAAGAG TACGATAAAG ATGGAGTGAA ACCAAAACCC

                   980         970         960         950         940
apo-4       <AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTTGGG
             ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^
dystrophin   AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTTGGG

-------------------------------------------------------------------------
                  13230       13240       13250       13260       13270
                    *           *           *           *           *
dystrophin  GTGTTCCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC
            CACAAGGACT ATTAACACGT GTGGACTCAA GTGTCGAAGT GGTGAACAGG

                   930         920         910         900         890
apo-4       <GTGTTCCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC
             ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^
dystrophin  GTGTTCCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC

-------------------------------------------------------------------------
                  13280       13290       13300       13310       13320
                    *           *           *           *           *
dystrophin  ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTTT CCTTCATAAT
            TAACGCAATA AAAGAAAAAG GAAATATTAA GAAAGAAAAA GGAAGTATTA
                                          |
                                inversion breakpoint3
                   880         870         860         850         840
apo-4       <ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTaT aacTgtTAta
             ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^v^ vvv^vv^^vv
dystrophin  ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTTT CCTTCATAAT
```

Figure 6.16B Recombination Signal Sequences and direct repeats are found on either side of the uninverted 136 bp sequence in genomic DNA RSS's exist on the forward and reverse strands at the upstream junction at 13,163 bp and one with poor homology which overlaps the downstream junction of the inversion breakpoint at 13,299bp (inclusive) in dystrophin by 1 bp. Potential nonamers/heptamers are in boldface type. A perfect 10 bp overlap is highlighted between apo-4 and dystrophin at 13300-13309, containing a direct repeat and forms a perfect palindrome with its 8 bp inverted repeat 5′ at 11,654. In bold are a perfect 12 bp inverted repeat upstream of inversion breakpoint2 and a 6 bp inverted repeat upstream of breakpoint3 relative to the lower strands.

6.6.2 Potential mechanisms to produce the inversion

Immediately 5' to the point of inversion in the 3' end of apo-4 is an 8 bp direct repeat, AAAGAAAG 5'-3' at 851-858 bp which forms a palindrome with its reverse and complement CTTTCTTT in dystrophin cDNA at 13000-13307 bp at the second inversion breakpoint in the 3'UTR of dystrophin (Figure 6.16B). It would be a 16 bp palindrome if the upstream point of inversion were ligated to the original sequence 1.62 Kb downstream. Direct repeats have been identified as signal points to produce inversions in transposable elements (Temin 1964) and histones (Vitelli and Weinberg 1983) and result in increased crossing-over (Dornfeld and Livingston 1992) while palindromes are recognised points for protein recognition in DNA. The inversion could also have resulted from the formation of a hairpin loop structure or triggered insertion by pairing of the inverted repeats and the formation of a palindrome at the point of inversion during an overlap of the arms of genomic DNA in recombination (Figure 6.16C).

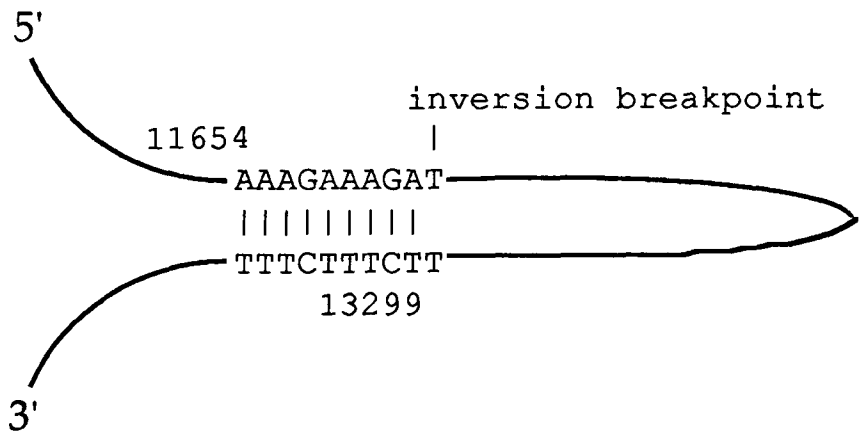


Figure 6.16C. A proposed mechanism of inversion formation due to direct repeats. The above structure proposes that during recombination, the dystrophin gene could loop out and rearrange "illegitimately" upstream at direct repeats found to flank the inversion in apo-4. To accommodate the mismatched T at 13,300, an A appears to be inserted into the upstream sequence upon matching of the direct repeats, potentially contributing to the formation of the inversion breakpoint.

A stem loop would form between the 3' end of the top strand and the 5' end of the lower strand to form the inversion seen in apo-4. Recognition of the repeats could then influence upstream factors following rearrangement. In histone genes, a relationship between breakpoints and gene coding regions has been found in which normal repeat structures invert to appear in apposition to each other, specifically TATAA and CAAT motifs (Vitelli and Weinberg, 1983). Examples of these sequences have also been identified around the apo-4 inversion (Figure 6.17)

```
                          inversion @ 860
    TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
841 ---------+---------+---------+---------+---------+---------+ 900
    ATTGTCAATATTTCTTTCTTAATATTTCCTTTTTCTTTTATTGCGTTACCTGTTCACCAC
```

Figure 6.17. Signal Sequences and Promoter elements could trigger an inversion. "TATAA" and "CAAT" boxes have been found immediatedly upstream of the inversion which have also been found in a rearrangement of histone genes.

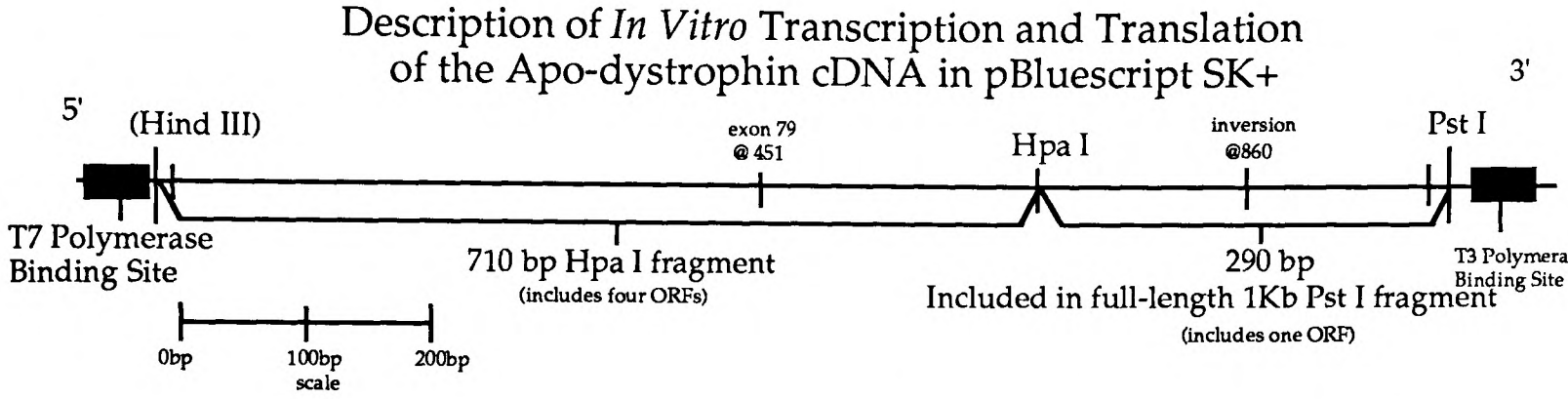
These observations also suggest that nucleotide sequence homology may permit alignment of the genes and the subsequent rearrangement having the unusual outcome of producing an inversion, and that the interaction of the sequences may depend upon signalling motifs within regions with transcription potential (Jennings et al. 1985). RSS's, direct repeats and histone motifs could all contribute to the production of an inverted sequence, although direct repeats form the most precise alignments at the point of inversion. The 3' UTR may play a role in upstream functions (reviewed in (Jackson 1993)) and we thus attempted to determine whether the inversion could influence the production of a transcript from apo-4 by studying its protein expression.

6.7. The inversion is necessary to produce proteins *in vitro*

To test whether apo-4 cDNA could synthesise a protein *in vitro* apo-4 cDNA in pBS SK$^+$ was linearised with Hind III and Pst I to produce a full-length template and HindIII and HpaI to produce a truncated 710 bp template lacking the 290 bp containing the inversion. In vitro transcription products were produced by incubating each template with T7 polymerase which bound to the endogenous T7 promoter present in pBS SK$^+$. To produce translated products from these transcripts, both Wheat Germ Extracts (WGE) and Rabbit Reticulocyte Lysates (RRL) were incubated with the transcripts. The full-length apo-4 transcripts in RRL yielded ^{35}S labelled translation products at 40 Kd and 50 Kd with a more weakly labelled species at 30 Kd and a faint band at 40 Kd from WGE, while no proteins were produced from truncated apo-4 templates (Figure 6.18).

These translation products established several important points. First, apo-4 was not a pseudogene, as these are defined by an inability to produce a transcript or a protein as shown in with the repetitive genes in actin and tubulin (Firtel 1981). Second, the suppression of the stop codons in the apo-4 gene was most likely due to readthrough rather than splicing, as this was unlikely to happen in a cell-free system, although it could not be completely ruled out (personal communication, Promega). Third, stop-codon translational readthrough was aided by the presence of the inversion as a downstream enhancing element and/or secondary structure which allowed the nonsense codons to be bypassed and amino acids made in their place, possibly by uncharacterised mammalian suppressor tRNAs. The third base of any codon is the least important in determining the amino acid produced according to the "wobble hypothesis" and suppressor tRNAs are often caused by changes in the modification of bases in the anticodon. Thus, the amino acid sequence produced cannot be entirely predicted from the relevant triplet

sequences but may be influenced by other features of the molecule (Murgola et al. 1983; Swanson et al. 1988). According to the "wobble hypothesis" an A at the end of a codon cannot have a unique meaning because U must also recognise G. Of the 23 stop codons in the apo-4 reading frame presented, 22 end in an A, which could give rise to cysteine or tryptophan and one ends in a G, which with a modified or "wobbly" tRNA, could produce a tyrosine rather than a nonsense codon.



Linearize plasmid with either HpaI (truncated) or Pst I (full length). Gene Clean and incubate with T7 polymerase and dNTPs to produce RNA in vitro.

↓

Incubate RNA with Wheat Germ Extract or Rabbit Reticulocyte Lysates to produce *in vitro* translation

↓

Separate translation products by SDS-PAGE. Fix, Amplify and Dry Gel. Perform Autoradiography

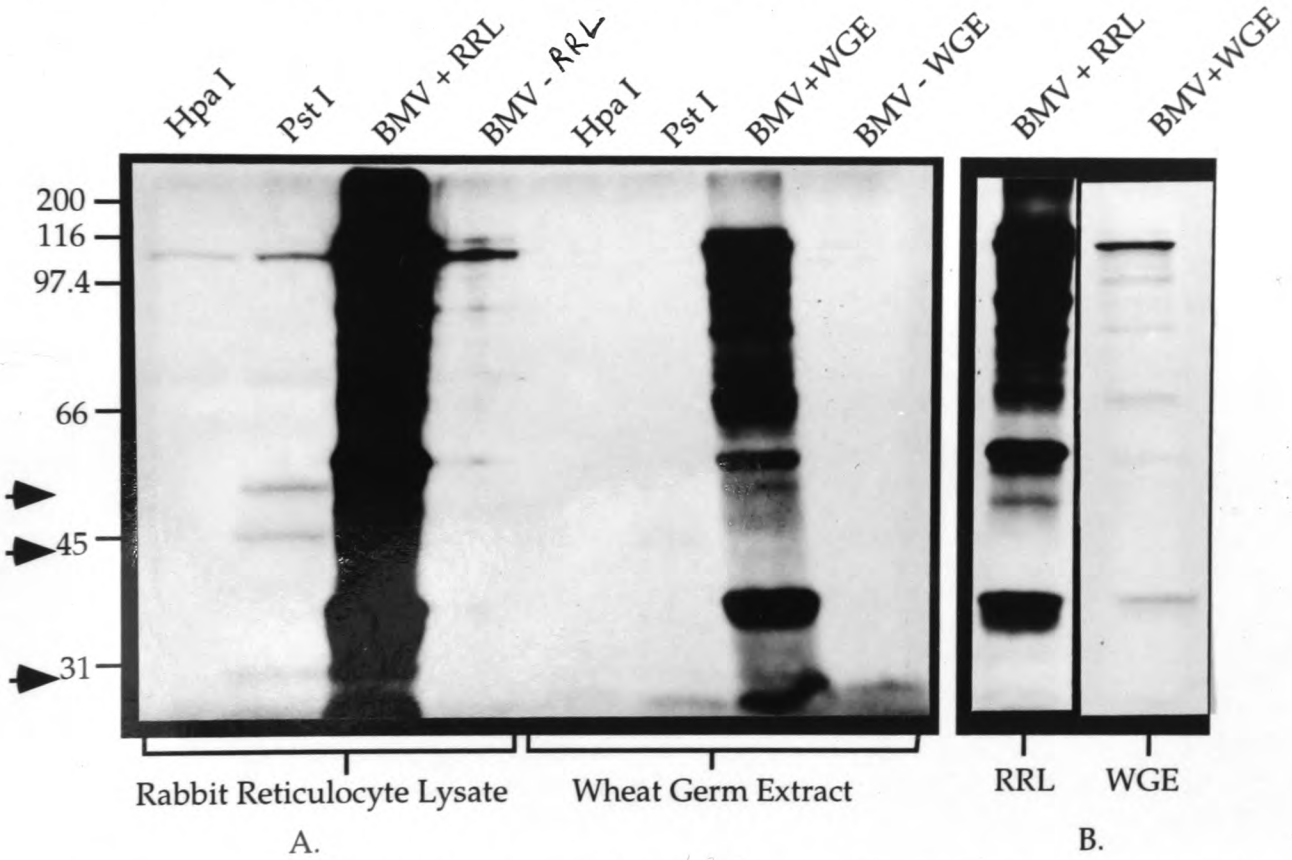


Figure 6.18. *In Vitro* Transcription and Translation (IVT&T) demonstrates that the full-length apo-dystrophin transcript produces proteins of 40 Kd and 50 Kd under reducing conditions. A. Reduced gel (left to right). Rabbit Reticulocyte Lysate (RRL) (Lanes 1-4) - Hpa I shows no bands; Pst I shows strong signals at 40 Kd (2Kd smaller than on a non-reduced gel) and 50 Kd (1 Kd smaller than on a NR gel) and 25Kd; the positive control, Bovine Mosaic Virus (BMV) + RRL does not show bands produced in Pst I (RRL) in a less exposed version shown in (**B.**). **Wheat Germ Extracts (WGE)** (Lanes 5-8) - Only Pst I shows a faint broad bands at 34Kb and 44Kb, 6Kb smaller than those in RRL which may reflect differences in glycosylation or an incomplete translation reaction. BMV + WGE shows no band at 33 and a very broad band from 39-46Kd which is distinct from those in the Pst I lane when compared with the shorter exposure of BMV + WGE in (**B.**). BMV - WGE shows no bands in the same position as Pst I. Reduced samples had 5% β-ME added before SDS-PAGE. The 10% gel was incubated in Amplify for 30 minutes and exposed at -80°C for 18 hrs and then for five days. at -80°C.

The presence of the inversion may have facilitated translation by functioning as a cis-acting sequence to suppress the stop codons found in the apo-4 sequence. Such a phenomenon has been shown in the pol gene of the Moloney murine leukemia virus (M-MuLV) in which translational readthrough is mediated by an RNA sequence of at least 50 base pairs located 3' to the *gag* UAG termination codon, containing a short purine-rich sequence adjacent to an amber suppression codon, highly conserved among different retroviruses (Honigman et al. 1991). To test the ability of translation products to be specifically recognised *in vivo*, peptide antisera were designed against a putative N-terminus and two other regions of apo-4 to analyse the expression of the gene in COS transfectants and K562.

6.8 Peptide antisera designed against putative antigenic regions

Three peptides were designed from the apo-4 sequence for the production of polyclonal rabbit antisera to analyse apo-4 cell-surface expression. Based on the full apo-4 sequence these were: P1: (aa 30-42) - MYPIMEYSCSDRN, a putative apo-4 N-terminus; P2: (aa 53-62 in frame 1 predicted to splice onto exon 79 (DTM in dystrophin) at aa 151-153 in frame 2) - YIYIGNLNVADTM); P3: (aa 133-150 DDLGRAMESLVSVMTDEE -- at the beginning of exon 79 but in the apo-4 frame, which also appears in Dp71 (Lederfein 1992)). P1 was of primary importance in analysing the N-terminus of the predicted protein. P1 and P2 were primarily used in these studies, as they were able to precipitate specific polypeptides. All antibodies were affinity purified, although the crude antisera retained specific activity longer than the purified version. P1 displayed a superior antigenic profile based on its greater number of peptide turns and hydrophilicity in a MacVector comparison of the three antisera. P3 displayed

one turn in its C-terminal 5 aa, but low specificity in ELISAs. P2 was only hydrophilic in the region of exon 79 in a proposed splice. No significant homologies were revealed in a FASTA homology search against the P1 peptide, showing only 46% complete homology with the cytomegalovirus hypothetical protein and the large tegument protein of the herpes simplex virus. 30-38% homology was shown for the IL-1β convertase precursor and the homeobox protein HOX-C4 (Pearson and Lipman 1988).

6.8.1 Antisera precipitate specific protein products

To test the unpurified (crude) peptide antisera raised against two regions of the ORF at +88 on the cDNA, COS cells transfected with the full length cDNA were pulsed on day two post transfection in -Met media with ^{35}S and lysed in 1% NP-40. Immunoprecipitation revealed heavily labelled bands at about 50 Kd for anti-P1c and about 53Kd for anti-P2c which were distinct from the respective prebleed sera. Weaker bands were appeared at 30 and 97Kd for both P1 and P2. A faint band was immunoprecipitated with Fc-CD33 at 50Kd in addition to a 28Kd band common to P1, P2 and Fc-CD33 (Figure 6.19).

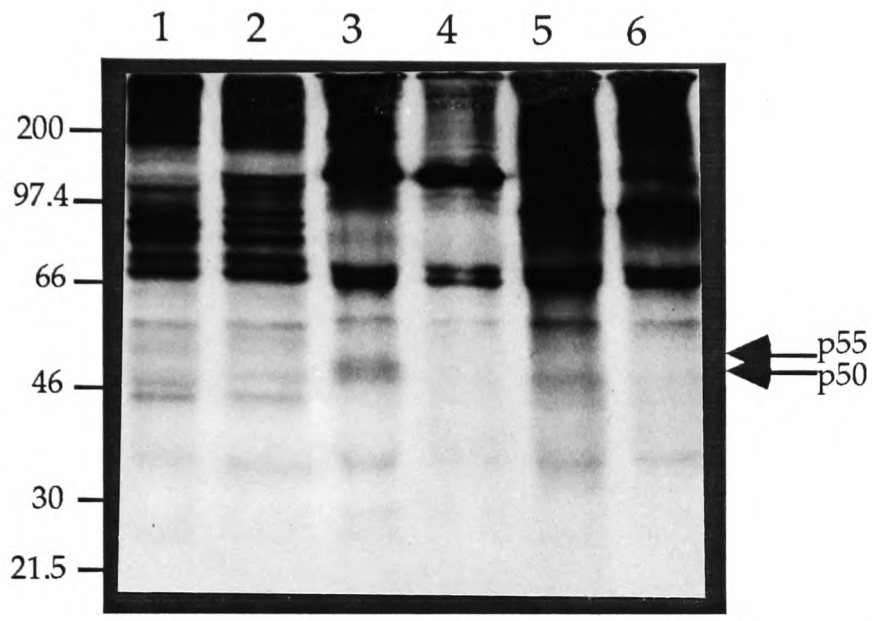


Figure 6.19. Anti-apo-4 antisera precipitates unique proteins in metabolically labelled apo-4 transfectants which may be immunoprecipitated from Fc-CD33. Fc-CD33 precipitates faint bands at 50-55Kd (Lane 1-2). Rabbit-anti-P2c antisera (Lane 3) precipitates a band at 55Kd not shown in the anti-P2c pre-bleed (Lane 4). A band 50Kd is shown by rabbit-anti-P1c antisera (Lane 5) not found in the P1c pre-bleed (Lane 6). This assay demonstrates that anti-apo-4 antisera are capable of precipitating at least two proteins from transfected apo-4.

6.8.2 Surface labelling demonstrates at least four species

Intact COS cells transfected with the apo-4 gene were then surface biotinylated either 1) following EDTA lifting 2) in situ for 20min at 25℃ followed by labelling at 4℃ or 3) in situ at 4℃ without EDTA lifting. Cells were then lysed in 1% NP-40 and immunoprecipitated with the P1 or P2 antisera to examine protein product size. Labelling in situ at 4℃ was found to produce the largest number of unique immunoprecipitated proteins. Apo-4 P1 polyclonal antibodies precipitate prominent bands at p45 and p95, and a minor band at p22 (Figure 6.20, Lane 10). P2 antibodies precipitate major bands at p50 and p97, and minor bands at p26 and p22 (Lane 12). On a longer exposure, p45 (Lane 5) and p97 (Lane 7) appear from cells labelled at 25℃ while p95 (Lane 5) and p97 (Lane 7) appear to be

temperature-independent. Only p95 (Lanes 1) and p97 (Lane 3) bind independently of divalent cations while the others require labelling at 4°C in the absence of EDTA to achieve efficient labelling.

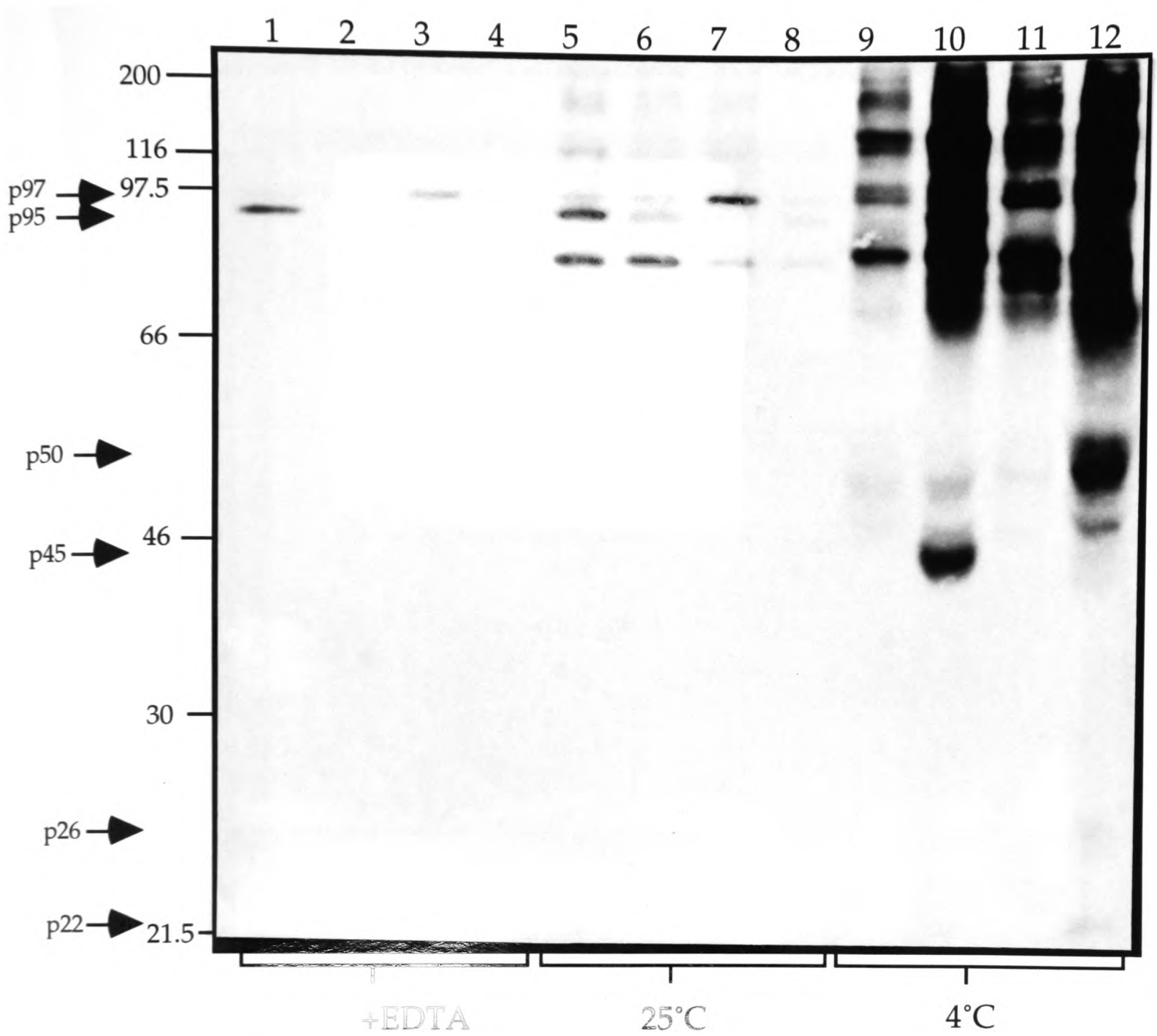


Figure 6.20. Anti-Apo-4 P1 polyclonal antibodies precipitate prominent bands at 45Kd and 95Kd and anti-apo-4 P2 antibodies precipitate major bands at 50Kd and 97Kd from surface biotinylated transfectants. Apo-4 transfectants were first EDTA lifted (Lanes 1-4) or surface labelled in situ at 25°C (Lanes 5-8) or 4°C (Lanes 9-12). Crude anti-apo-4 P1c pre-bleed sera - Lanes 2, 6 and 9; anti-P1c antisera - Lanes 1, 5 and 10; Anti-apo-4 P2c pre-bleed sera - Lanes 4, 8 and 11; anti-P2c antisera - Lanes 3, 7 and 12. Only p95 and p100 bind independently of divalent cations while the rest require labelling at 4°C in the absence of EDTA to achieve efficient labelling. On a longer exposure, anti-P1 showed a minor band at 22Kd and anti-P2 showed minor bands at 26Kd and 22Kd.

The sizes of the protein products obtained in IVT&T were then compared with those precipitated with the P1 and P2 polyclonal antibodies using COS transfectants metabolically labelled with ^{35}S and surface labelled with biotin (Table 6.4). Only proteins distinct from the control are listed. ^{35}S labelled

proteins were shown to run about 5Kd higher than biotinylated transfectants, possibly due to the interaction of label with the protein or a difference between types of gel used.

IVT&T-RRL-R	IVT&T RRL-NR	IVT&T WGE-R	IVT&T WGE-NR	P1-R COS ^{35}S	P2-R COS ^{35}S	Fc-CD33-R COS ^{35}S	P1-R biotin	P2-R biotin	Fc-CD33 ^{125}I COS
Pst I	Pst I	Pst I	Pst I				95Kd	97Kd	100
50Kd	52Kd	44Kd-f	45Kd-f	50Kd	55Kd	50/55Kd	45Kd	50Kd	66-f
40Kd	42Kd	32Kd	34Kd	30Kd	30Kd	30Kd?		26Kd-f	50-f
25Kd	27Kd	22Kd-f	24Kd				22Kd-f	22Kd-f	45-f
Hpa-1	Hpa-1	Hpa-1	Hpa-1						
44Kd-f	45Kd-f	45Kd-f	45Kd-f						
		32Kd	37Kd						

Table 6.4 Size Comparison of proteins obtained from *in vitro* translation and immunoprecipitation. The table shows that proteins obtained with different labelling systems are similar in weight. IVT&T=*in vitro* transcription & translation; RRL=rabbit reticulocyte lysates; WGE=wheat germ extracts; R=reduced; NR=nonreduced; f=faintly labelled band; P1=protein immunoprecipitated from COS cell transfectants with purified polyclonal antisera P1, P2 and P3; biotin=COS cells surface labelled with biotin.

6.8.3 Apo-4 antibody and Fc-CD33 precipitate a 50 Kd band

Crude anti-P1 (P1c), P2 (P2c) and P3 (P3c) antisera were purified against affinity columns coupled to the appropriate peptide and subjected to immunoprecipitation and immunohistochemistry to compare reactivity. In an initial assay with purified P1 (P1p), P2 (P2p) and P3 (P3p) and Fc-CD33 on K562, a differential reactivity was shown (Figure 6.21)

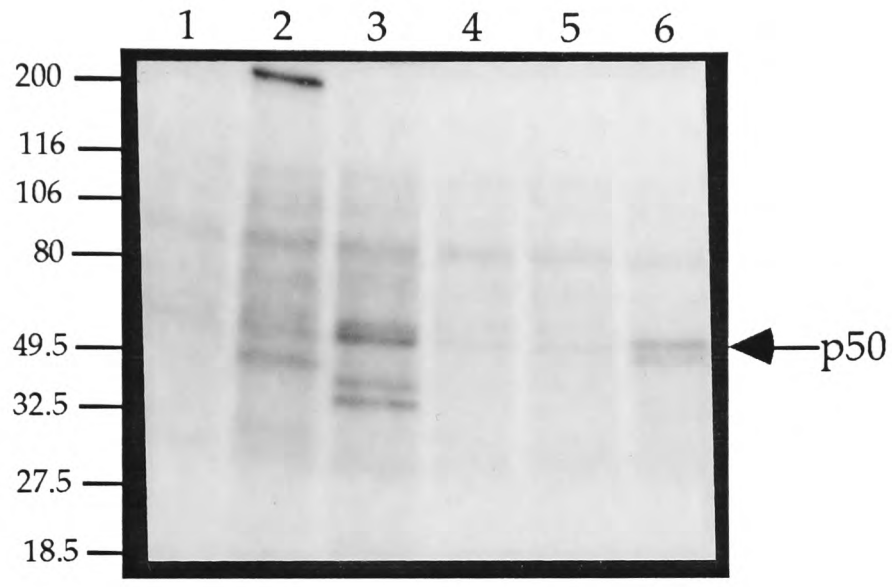


Figure 6.21. Purified P1 is able to precipitate proteins in the same region as Fc-CD33 from K562. ^{35}S labelled K562 were immunoprecipitated with purified apo-4 antisera and Fc-CD33 using stringent washing conditions. Mouse Ig (Lane 1) and purified anti-CD22 antisera (Lane 2) do not show the 50/55Kd doublet precipitated by anti-apo-4 P1 (Lane 3). A 50Kd band is faintly shown by anti-apo4 P2 (Lane 4) and anti-apo-4 P3 (Lane 5) and more strongly by Fc-CD33 (Lane 6) as shown at the arrow. Proteins were run on a 5-20% gradient gel for SDS-PAGE.

Following immunoprecipitation using stringent "high-salt/low-salt" washing conditions, purified P1 (P1p) demonstrated four distinct bands at about 38, 39, 50 and 55Kd in K562 (Figure 6.21, Lane 2). p50 co-migrated with Fc-CD33 alone (Lane 6) and was absent from the controls (Lanes 1 and 2) and P2p and P3p (Lanes 4 and 5), demonstrating that P1p and Fc-CD33 appeared precipitate a band of the same weight.

6.8.4 Fc-CD33 precipitates a 50Kd band from K562 and TFs

Fc-CD33 and purified apo-4 antisera were coupled to beads in an attempt to increase their avidity in immunoprecipitations. In this assay, the ability of crude anti-apo-4 P1 (P1c) and purified anti-apo-4 P1 (P1p) antisera to precipitate specific proteins was compared using stringent washing conditions (Figure 6.22). In apo-4 transfectants (TF) P1c was shown to precipitate a pronounced band at 50Kd (Lane 14) not seen with P1p (Lane 6) or with P1c in sham TF (Lane 17) demonstrating its specificity. Fc-CD33 also weakly precipitated a band at 50Kd in apo-4 TF (Lane 12) and a stronger signal in K562 (Lane 4) weakly shown with P1p (Lane 2) which was partially blocked with P1 peptide (Lane 3) and not shown with anti-CD22 antisera (Lane 1). P3p showed a faint band at 50Kd and a stronger one at 68Kd (Lane 10) which were blocked by peptide (Lane 11), reinforcing the readthrough hypothesis by demonstrating that P3p could detect the 50Kd band, likely in the same frame as P1 (Lane 14). P2p showed faint bands at 55 and 68Kd which were mostly blocked by peptide (Lane 8). This gel shows that anti-apo-4 P1c antisera can precipitate a protein at 50Kd in COS transfectants which is in the same region as one precipitated by Fc-CD33, distinct from SHAM transfectants. P1p is less efficient than P1c in precipitating this protein and P1c was thus used to detect p50 in future immunoprecipitation experiments. It was suspected that protein purification may have decreased the stability of purified P1c.

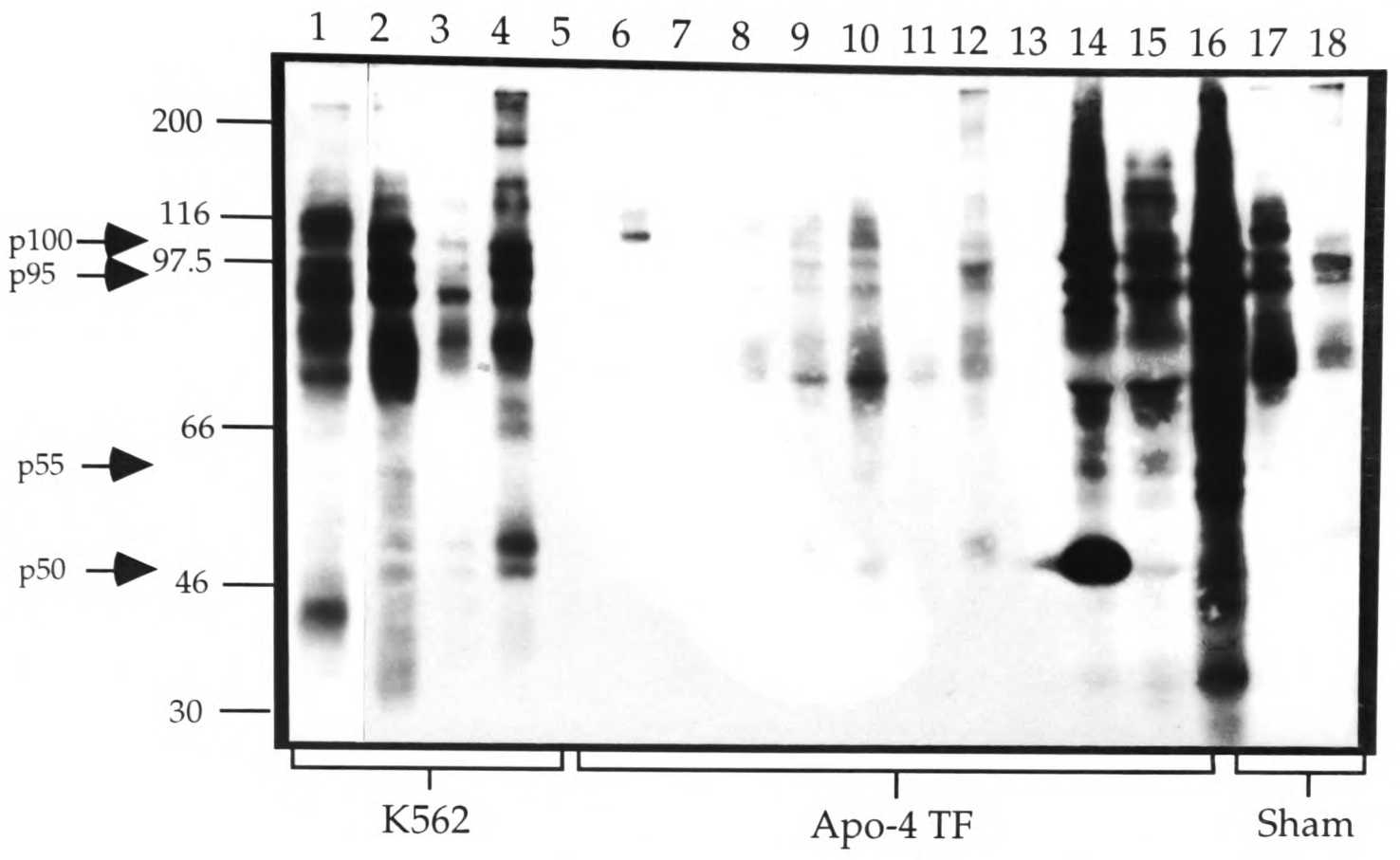


Figure 6.22. Fc-CD33 precipitates a band at 50 Kd from K562 and apo-4 transfectants. The band at 50Kd is compared with anti-P1, P2, and P3 crude and purified antisera in NHS-biotin labelled K562 (Lane 1-5), and apo-4 (Lane 6-16) and sham transfectants (Lane 17-18). In **K562,** a band at 48/50 was weakly precipitated by P1p (Lane 2), more strongly by Fc-CD33 (Lane 4) and was partially blocked by the addition of P1 peptide as were bands at 95 and 105Kd (Lane 3). Denatured p48/50 from K562 P1c immunoprecipitates was not recognised by Fc-CD33 (Lane 5) p48/50 was not precipitated by purified anti-CD22 antisera (22p) (Lane 1). In **apo-4 transfectants,** P1c precipitated a pronounced band at an average weight of 50Kd (Lane 14), faintly seen in P2c which precipitated a higher band at 55Kd (Lane 15) which comigrates with the 55Kd band seen with P1c (Lane 14). P3c only precipitated a distinct band at 32Kd (Lane 16). Fc-CD33 also precipitated a band at 50Kd (Lane 12) which was not seen in sham transfectants (Lane 18). P1p was unable to precipitate a protein at 50Kd in either apo-4 transfectants (Lane 6) or sham transfectants (Lane 17); P1p peptide blocked the band at 100Kd (Lane 7). P2p (Lane 9) showed a weak band at 50Kd which was blocked by peptide (Lane 8). P3p precipitated a faint band at 50Kd (Lane 10). All P3p bands were blocked by peptide (Lane 11). Fc-CD33 does not reprecipitate proteins from P1p precipitates (Lane 13). This assay shows that purified antibodies do not precipitate the predominant 50Kd apo-4 transcript as well as the crude antisera, that Fc-CD33 cannot recognise denatured proteins precipitated with P1 and that Fc-CD33 can recognise p50 in COS transfectants.

6.8.5 p50 can be specifically blocked by peptide

To test whether p50 could be specifically immunoprecipitated and blocked by peptide, apo-4 and CD22 transfectants (TF) and K562 cells were surface-labelled with NHS-biotin and immunoprecipitated with anti-apo-4 P1c to reveal bands at p50 and 95-100Kd in apo-4 TFs (Figure 6.23, Lane 8) which could be blocked by P1 peptide (Lane 6). CD22 TFs were unable to precipitate this band (Lane 9). A band at p50 was precipitated in K562 with P1c (Lane 5) and Fc-CD33 (Lane 3) and partially blocked by peptide (Lane 4) which may be due to the polyclonal nature of the antisera. P1p was only able to specifically precipitate a band at 68Kd in K562 (Lane 2) which was completely blocked by P1 peptide (Lane 1).

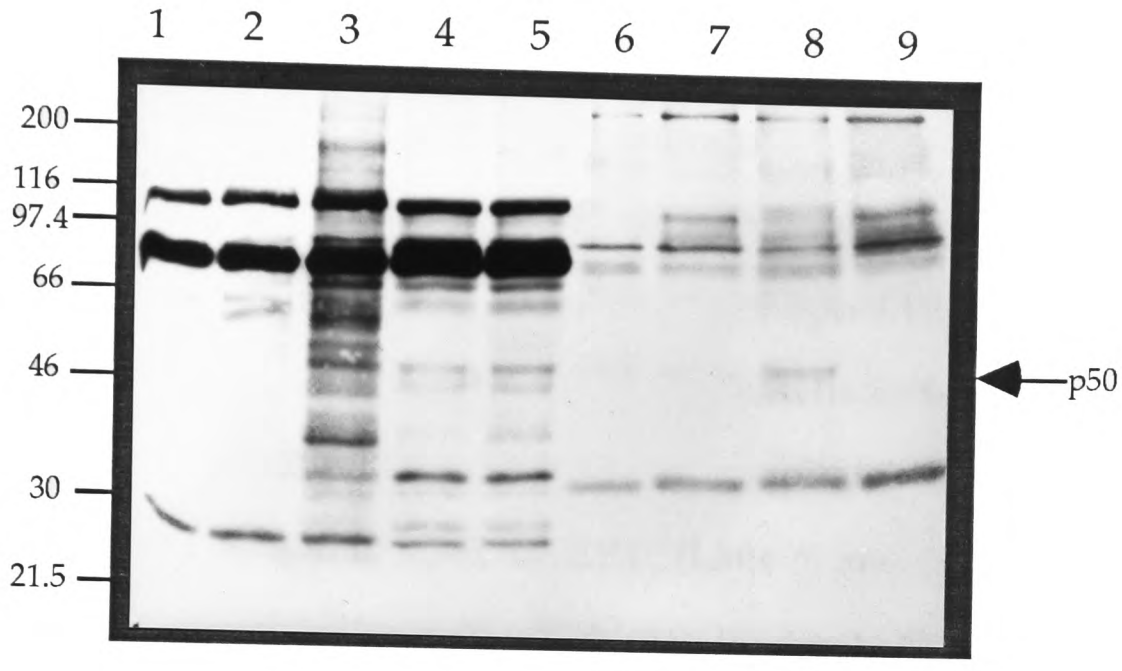


Figure 6.23. Anti-P1 antisera specifically precipitates a band at p50 in apo-4 transfectants which can be blocked by peptide. NHS-biotin labelled K562 and apo-4 and CD22 transfectants were immunoprecipitated with crude P1 (P1c) or purified (P1p) antisera in the presence and absence of P1 peptide and washed using stringent conditions and ECL detection. p50 was immunoprecipitated with P1c in apo-4 transfectants (Lane 8) and blocked with P1 peptide (Lane 6) but was not precipitated by CD22 without P1 peptide (Lane 9) or with peptide (Lane 7). P1c also precipitated a band at 50Kd in K562 (Lane 5) which was partially blocked by peptide (Lane 4), which also appeared in Fc-CD33 immunoprecipitates (Lane 3). P1p precipitated a band at about 58Kd which was completely blocked by peptide. P1p precipitated a doublet at 62/64Kd (Lane 2) which was completely blocked by peptide (Lane 1).

This gel reinforces the finding that p50 is specific to apo-4 transfectants, that it can be blocked by peptide and that a surface-labelled band of the same size can be precipitated by P1c and Fc-CD33 in K562, although it is only partially blocked by peptide.

6.8.6 Apo-4 appears to exist as a cell surface heterodimer

To determine whether apo-4 was disulfide linked to itself or to another protein, a peptide blocking experiment was performed under reducing and non-reducing conditions (Figure 6.23.1). Under non-reducing conditions, p50 was shown exist as a single band at about 62Kd (Lane 1). Under reducing conditions, bands of 50Kd and 56-58Kd (Lane 3) were blocked by peptide (Lane 4) indicating that apo-4 may exist as a heterodimer on the cell surface.

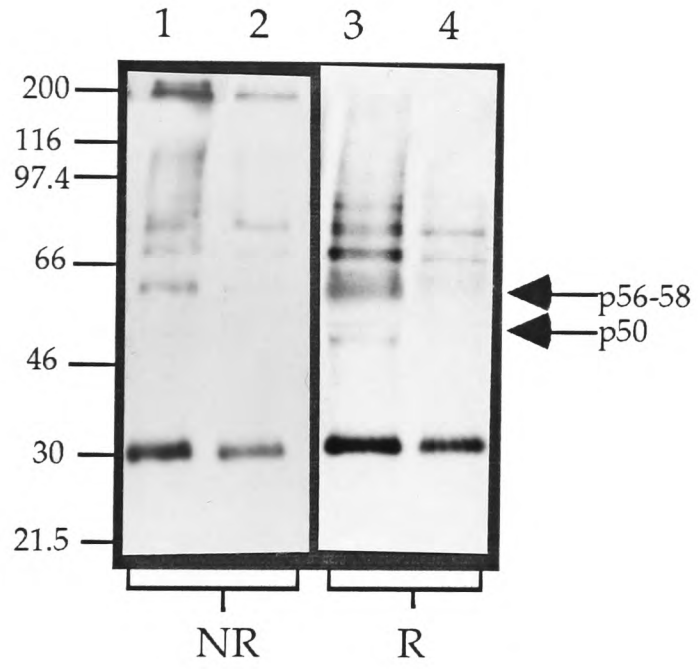


Figure 6.23.1 Immunoprecipitation of apo-4 under nonreduced and reduced conditions shows possible linkage with a separate subunit. Apo-4 transfectants were labelled with NHS-biotin and immunoprecipitated with the anti-apo 4 P1 antisera (Lanes 1 and 3) and the antisera blocked with peptide (Lanes 2 and 4) and subjected to SDS-PAGE under nonreducing (Lanes 1 and 2) and reducing (Lanes 3 and 4) conditions. Under non-reducing conditions, apo-4 runs at about 62 Kd and a high molecular weight species is blocked at 106-108 Kd (Lane 1). Under reducing conditions, two bands are blocked by peptide at 50Kd and 56-58Kd indicating that apo-4 may exist as a heterodimer on the cell surface.

6.9 Predicted STADEN and GeneID splice sites

The presence of stop codons in the P1 ORF identified in apo-4 prompted a search for potential splice sites in the cDNA. Potential donor and acceptor sites appeared throughout the apo-dystrophin-4 cDNA. The GeneID program predicted 32 donor and 8 acceptor sites (above 72% probability with known acceptor sites) (Guigo et al. 1992). The Staden program predicted 28 splice donor and 28 acceptor sites within the default range and 39 more outside the default range (Staden 1990). The equation $2^{(\text{donors} + \text{acceptors})-2}$ with the maximum number of both is $2^{(31+67)-2}$ or 2^{96} or 7.9^{28} possible combinations of splice donors and acceptors in the apo-dystrophin-4 cDNA. This increases the probability that the regions, upstream and downstream of exon 79, previously thought to be only intronic or untranslated could contain exons expressed under certain circumstances, such as when the apo-4 cDNA is placed downstream of a strong promoter or when an inversion or other enhancing element is present downstream. The exons and introns predicted here, according to given splice sites, are named in relation to known introns and exons, specifically, exon 79. An attempt was made to predict reasonably spliced mature cDNAs while conserving the proposed upstream exon, 78.3, and using the splice donor and acceptor sites generated from GeneID and Staden to achieve the longest ORF possible.

A splice product "H2" is shown (Figure 6.24) which includes the proposed N-terminal exon 78.3. Other proposed splice products appear in Appendix II. Based on the amino acid sequence of the spliced product, all potential splice products predict a Type II transmembrane (TM) protein using the TopPred 1.2 program (Claros and von Heijne 1994). With Type II proteins, the

transmembrane domain acts as the signal peptide and the N-terminal region usually remains in the cytoplasm, except in the case of some cell-surface molecules with multiple TM domains. If apo-dystrophin-4 coded for a Type I protein using the first available methionine, a signal peptide could be cleaved off to leave a 14aa extracellular domain, assuming the suppression of the first stop codon. TopPred has predicted that the same 20 amino acid region in the putative exon 78.3 is a "certain" transmembrane domain in all proposed splice products. This finding lends weight to the suggestion that exon 78.3 could act as a functional exon to produce a transmembrane protein either through readthrough of stop codons or splicing in the COS cell system and potentially *in vivo*. The proposed splice products are listed with their proposed donor and acceptor sites. Predicted proteins are presented with the transmembrane domain(s) underlined.

A. H2 starting at second methionine - 321 bp, predicted weight = 17.4Kd + 1 N-glycosylation site + 20.4 Kd.

B. Splice sites for peptide
MYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFTTPDSRMVFIIFIQQRGLDSK
SLQEINLYFCEGFYTSMQLYKKVIRKLHKITQWTRTPQNQSEVEIA 107

C.

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@88 bp	78.3	@74-180	106 bp	79.1	@181-529	349 bp
	79.1	@530-654	125 bp	79.4	@655-720	66 bp
	79.4	@721-769	49 bp	79.55	@770-875	105 bp
	79.55	@876-893	18 bp	79.75	@894-932	39 bp
	79.85	@933- 966	33 bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 12- 32 1.8833

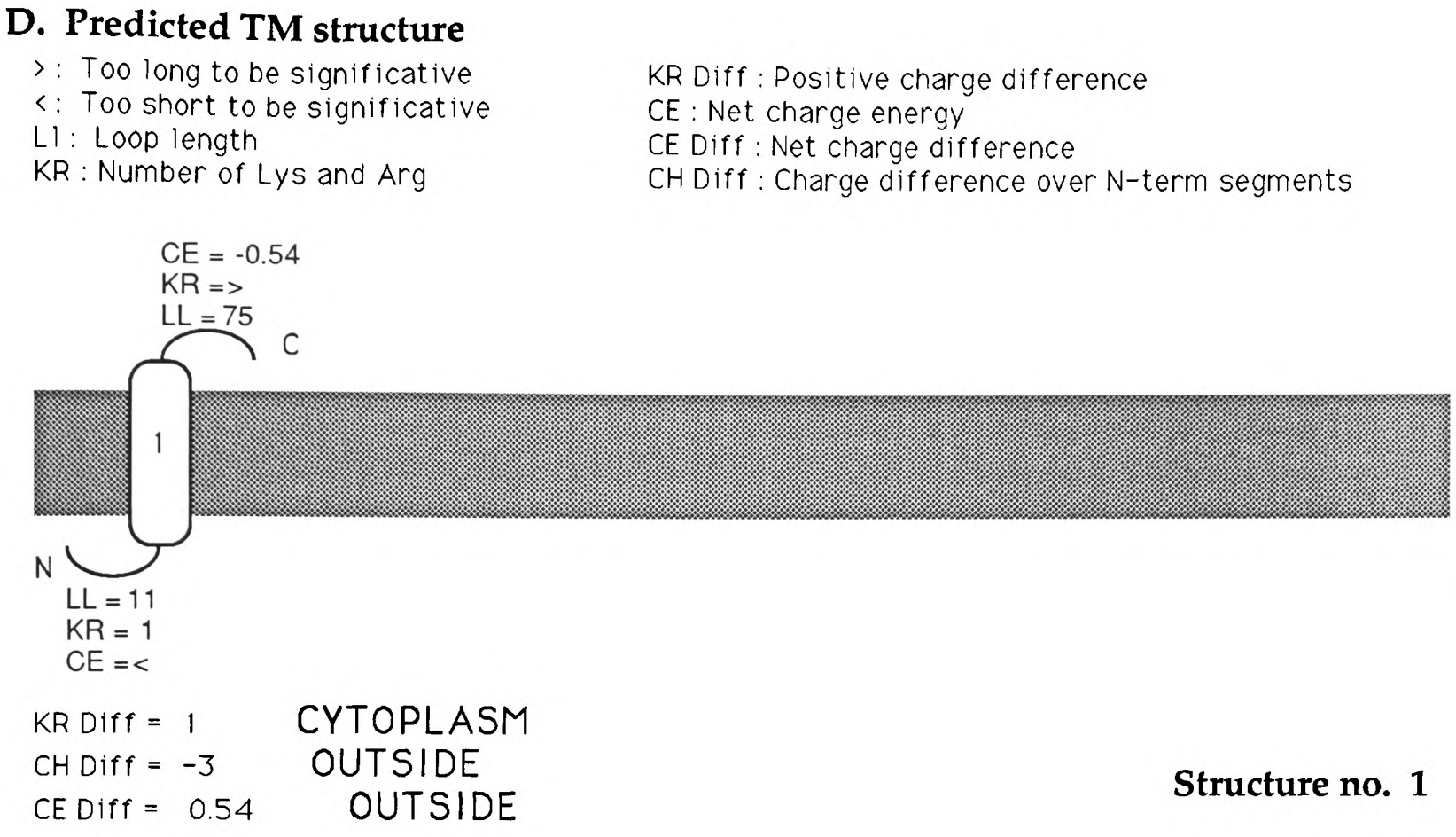


Figure 6.24. Potential apo-4 splice product H2. A. A model of potential splicing of apo-4 using available splice sites. **B.** Peptide structure with TM underlined; **C.** potential "new" exons identified; and **D.** TM domain structure predicted with TopPred showing a cytoplasmic N-terminus. Total spliced length, 321 bp, 107amino acids = 16Kd + 1 N-glycosylation site=3Kd; proposed weight = 19Kd.

H2 represents the longest splice product yet obtainable. These findings argue that, if any, only the 22Kd band produced from *in vitro* transcription and translation could have been the product of splicing or premature truncation of the translated product due to the presence of internal stop codons. It is highly unlikely that the 45-50Kd bands could have been produced from a spliced apo-4 cDNA.

6.9.1 Predicted structure of full-length apo-4 protein

The full-length apo-4 gene was able to produce non-glycosylated proteins *in vitro* at average weights of 50, 40 and 25Kd from a 997 bp apo-4 gene plus 45 bp to the PstI site (Figure 6.18). If all 1042 base pairs were utilized, a maximum 52 Kd product should have been produced. 50 Kd was near this, thus raising the possibility that the stop codons in the sequence were not being recognised, possibly from a suppression element in the inversion or in the sequence immediately upstream, considering that no protein was produced in the absence of the inversion. The UTR 3' of the dystrophin full-length cDNA comprises 2.677Kb. Based on the precedents indicating that elements within the 3'UTR are capable of suppressing stop codons (Honigman et al. 1991), Type I and Type II proteins were predicted from a proposed gene in which all of the thymidines in stop codons were changed to adenines to allow for readthrough, which happens when RNA is "edited" by adenosine deaminase (Bass and Weintraub 1987) In addition, the AATTAA site at 989 bp may have been utilised as a cryptic polyadenylation site, as recognition of this sequence has been reported at least 2% of the time in those cases studied (Hames and Glover 1988). The "RNA edited" 997 bp of the apo-4 gene was translated and demonstrated a continuous open reading frame. The peptide sequence was then subjected to a MacVector search for putative start, signal peptide cleavage and glycosylation sites which appear below along with predicted protein sizes. Assuming readthrough, nine potential initiating methionines are identified in apo-4 (Table 6.3). Staden-predicted probability based on Kozak sequence are listed in parentheses, with 93% at the M at +34 (+100nt) being the highest. Type I potential protein sizes (methionines with cleaved signal peptides between 17-27 amino acids were included; 20-23 amino acids is the average length, but known signal peptides range from 14-45 amino acids (Barclay 1993)). Type II proteins do not cleave off

the signal peptide thus allowing for a longer predicted protein product. The largest potential apo-4 protein product is a Type II 97 Kd protein which would require the 660 bp of the CDM8 vector sequence to its internal polyadenylation site and utilisation of the first available methionine (Table 6.3B).

A. Type I proteins predicted from the 997 bp apo-4 gene

Methionine start site	**Signal Peptidase Cleavage Site (+length of Signal Peptide)**	**Cleaved Peptide length (aas) (+N-terminus length)**	**Position & Length of 1st TM domain**	**Predicted protein Weight (no. aa x 150)**
@9 **(.76)** +	@25 (17 aas)	307 (15)	@40-60	**46.0Kd (+2N=52Kd)**
@9 (.76) (+)	@39 (31 aas)	293 (1)	@40-60	43.9Kd(+2N=49.9Kd)
@30 (.77)	@48 (19 aas)	284 (5)	@40-60	**42.6Kd**(+2N=48.6Kd)
@34 (.93)	@48 (15 aas)	284 (5)	@40-60	42.6Kd(+2N=48.6Kd)
@67 (.74)	@93 (27 aas)	239 (27)	@101-121	35.8Kd(+2N=41.8Kd)
@120 (.77)	@142 (23 aas)	188 (74)	@217-237	28.2Kd(+1N=31.2Kd)

B. Type II proteins predicted from the apo-4 gene

Methionine start site position (+GeneID probability)	amino acids to 1st predicted TMD = Nterminus	Position & Length of 1st TM domain	Pred. Pep. length (aas)	Predicted protein Weight (no. aa x 150)	+Extra Weight from vector to *Pst I (IVT&T) (+15 aas=2.25Kd) or **AATAAA/CDM8
@9(.76)$_{(+CDM8)}$+	31	@40-60	547	82.1Kd(+5N=**97Kd**)	**97.00Kd
@9 (.76)+	31	@40-60	324	48.6Kd(+2N=**54.6Kd**)	*50.85Kd(56.85K)
@30 (.77)+	10	@40-60	302	45.3Kd(+2N=51.3Kd)	*47.55Kd(53.55K)
@34 (.93)+	6	@40-60	298	44.7Kd(+2N=**50.7Kd**)	*46.95Kd(52.95K)
@67 (.74)	34	@101-121	265	39.7Kd(+2N=45.7Kd)	*41.95Kd(47.95K)
@110 (.77)	0	@110-120	222	33.3Kd(+1N=36.3Kd)	*35.55Kd(41.55K)
@120 (.77)	12	@132-152	212	31.8Kd(+1N=34.8Kd)	*34.05Kd(40.05K)
@160 (.84)	57	@217-237	172	25.8Kd(+1N=**28.8Kd**)	*28.05Kd(34.05K)
@168 (.83)	49	@217-237	164	24.6Kd(+1N=27.6Kd)	*26.85Kd(32.85K)
@175 (.77)	42	@217-237	157	23.5Kd(+1N=26.5Kd)	*25.75Kd(31.75K)
@184(<.67)	33	@217-237	148	22.2Kd(+1N=25.2Kd)	*24.45Kd(30.45K)

Table 6.3A-B. Type I (A) and Type II (B) apo-4 proteins predicted with nonsense suppression. All T's in stop codons were changed to A's in the existing apo-4 gene, as happens with RNA editing with adenosine deaminase. Predicted amino acid sequence was analysed using MacVector for starting methionines, N-glycosylation sites and predicted weight. The Staden predicted probability of M being utilised based on Kozak sequences appears in parentheses after each M. Product weights, with N-glycosylation added, which are close to products obtained from *in vivo* labelling are highlighted. All weights are based on the assumption that the AATTAA site at 989 in apo-4 is used for polyadenylation unless otherwise indicated, as the original transcripts produced *in vitro* could not have used polyadenylation sites in the vector. "+" indicates the only products which would allow full or (+) partial recognition by the anti-apo-4 P1 antisera.

In theory, many proteins could be produced from an apo-4 gene in which the stop codons are suppressed. With the predicted N-glycosylation sites along this region, the protein predicted from this gene would be 97Kd (Table 6.3B); without the CDM8 sequence it would be about 50.85 Kd with a 31 aa N-terminus and will be termed apo-4F, the predicted structure of which appears in Appendix IV. As previously shown, a 97Kd protein could be precipitated from COS transfectants with an anti-apo-4 P1 polyclonal antisera in the presence and absence of EDTA (Figure 6.20). Alternatively, if the AATAAA sequence at position 632 in apo-4 were utilised, starting at M@34, a 210aa Type II protein is predicted with two N-glycosylation sites and a weight of 32.5Kd. In addition, two histone methylation sites have been identified in the inversion as shown below.

inversion
RK^NYKGKRK

These sites could provide an additional explanation for the ability of the cDNA to produce protein in the absence of a conventional Poly A site. In histones, these sites are utilised instead of polyadenylation to give stability to the RNA message prior to producing a protein. This analysis demonstrates that the N-terminus of apo-4 could appear on the surface of a transfectant as calculated from the amino acid sequence.

6.9.2 TopPred analysis predicts five TM domains

The apo-4 readthrough product beginning at the second available methionine (apo-4S) was then subjected to TopPred to identify putative transmembrane (TM) domains using Kyte-Doolittle analysis. Apo-4S predicts an 11 aa N-terminus and five putative transmembrane domains (Figure 6.25A-C). Domains 1 and 3 are certain and 2, 4 and 5 are putative. The sequence obtained when stop codons are suppressed shows a 303 amino acid structure with an N-glycosylation site at +12

(NQS) which could be utilised if the N-terminus were extracellular. Although the structure predicted using all five TM domains calculates an N-terminus on the outside of the cell (Figure 6.25B), the predicted transmembrane domain structure to which the program defaulted shows a cytoplasmic N-terminus but predicts that it is probably on the outside of the cell (Figure 6.25C). An extracellular N-terminus with multiple TM domains would share a similar channel-like structure with the C5aR, f-Met-Leu-Phe-R, and IL-8 receptors (Barclay 1993).

A readthrough apo-4S product using the second available methionine

A. The Apo-4S peptide sequence

P1 Begin TM_1(R)
+30 | P2

MYPIMEYSCSD RNLVLIYGIL LI**YIYIGNLN V**A**R**HFSM**K**TP VA**R**SNIKLIL 80

TNNV**K**WLH**KK** GFASSWKLVK NQTLLCTPSM QLLCCLHPEM GNDFPNGKET 130

P3

E**R**CYLSAPFV KSVFLSLCFP GHNVGSLFHM A**DDLGRAMES LVSVMTDEE**G **180**

AE**K**MFYNS**R**F PHGFYNIHTT KRIRQ**K**EFTR NKSIFL**R**RVV VLYCRFQ**K**FL 230

SLLLFC**K**QWQ VLHVYAIVQK SYKKTTCKIL IAK**K**LAISLY GTHFGLFKNL 280

KQL**K**RKNYKG KRK**K**RNGQVV KL**R**TQVCTII RNTPKPK**R**GR NSMRSRV**R**C**K** 330

LI 332 (302aa in predicted polypeptide)

B. Candidate membrane-spanning segments:

Certain 1 41-61 1.9073
Putative 2 101-121 0.8052
Certain 3 132-152 1.2552
Putative 4 217-237 1.1833
Putative 5 254-274 0.9240

Transmembrane segments included in structure No. 8: 1 2 3 4 5

Loop lengths: 11 39 10 64 16 58; K+R profile: 1 2 5 (9 >22)

K+R difference: -23: -> Orientation: **N-out**

Charge-difference over N-terminal Membr. segs. (±15 residues): -4
-> Orientation: **N-out**

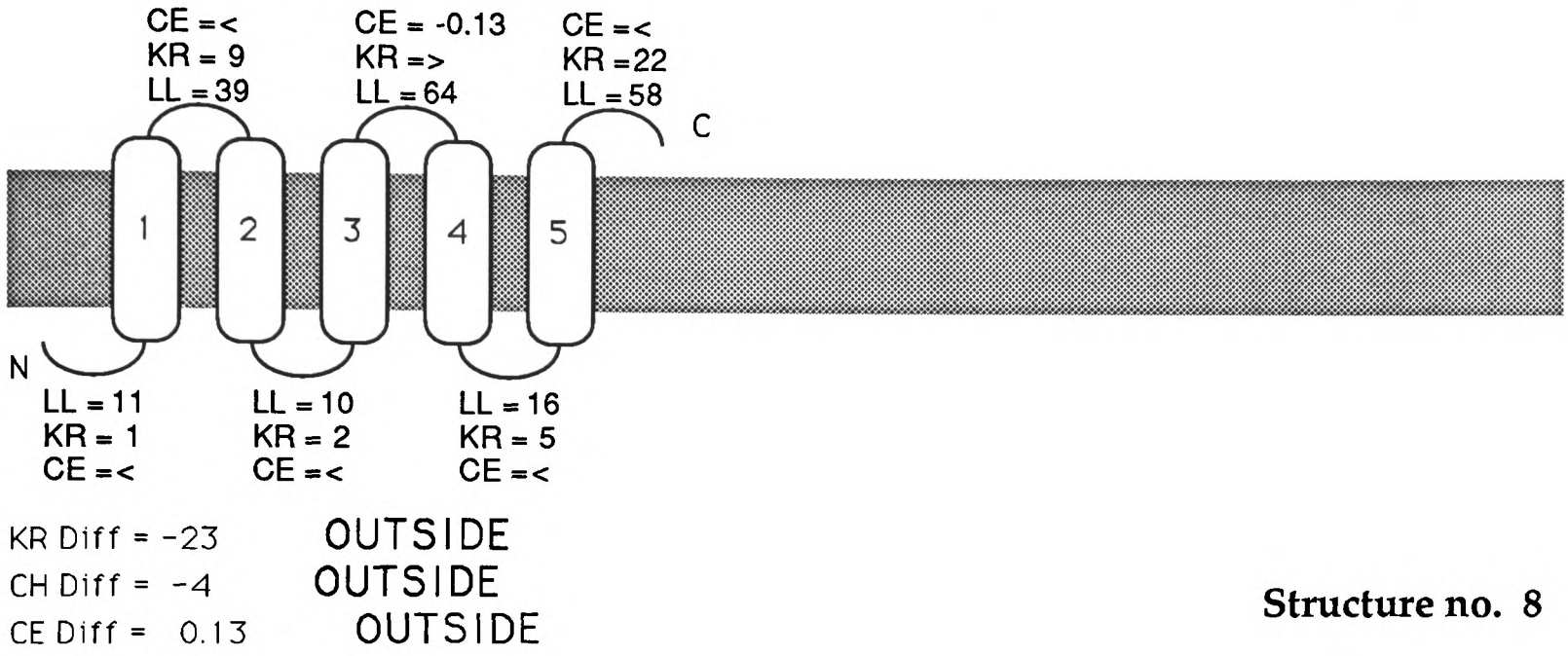
CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 <

CYT-EXT difference: 0.13: -> Orientation: **N-out**

C.

> : Too long to be significative
< : Too short to be significative
Ll : Loop length
KR : Number of Lys and Arg

KR Diff : Positive charge difference
CE : Net charge energy
CE Diff : Net charge difference
CH Diff : Charge difference over N-term segments



Structure no. 8

Figure 6.25A-C. Five transmembrane domains are predicted for the full-length apo-4S readthrough product. A. The amino acid sequence shows the predicted TM sequence as underlined and the regions against which P1, P2 and P3 peptide antisera were designed in bold and underlined beginning with R at +64. Amino acids in which T was changed to A are in bold alone. Structure begins at the "weak" M (+30) included in the P1 peptide antisera followed by the "strong" M at +34. **B.** A description of the predicted TM domains with those described as "certain" and "putative"; the position of the N-terminus as "outside" is relative to a prediction including all five TM domains, although most predicted TMs which could be utilised predict an "outside" N-terminus. **C.** The TopPred model of the readthrough apo-4S product which would utilise the second available methionine at +30 aa. The model shows a cytoplasmic N-terminus, although it is calculated to be on the outside of the cell below the model. Appendix IV shows the structure predicted beginning at the first M at +8 which would allow for two N-glycosylation sites on an extracellular N-terminus.

A TopPred product beginning at the first available M at +9 aa, apo-4F (full-length) appears in Appendix IV and also predicts two certain and four possible TM domains, although the 31 aa N-terminus is predicted to be cytoplasmic in a putative four-domain structure. An apo-4 product beginning at the M at +34 shares an equivalent predicted structure with apo-4F with an extracellular N-terminus. Whether the N-terminus is extra or intracellular, extracellular recognition could occur via the extracellular portions of the apo-4 protein and/or the N-terminus (Appendix IV). Two other N-glycosylation sites were identified in the full-length apo-4 readthrough product at 93 (NQT), 203 (NKS). In a five TM domain structure, the final two could only be utilised on Apo-4S or F if the N-terminus were intracellular and the first two could only be utilised on Apo-4F if the N-terminus were extracellular. If only the two certain TM domains were utilised and the N-terminus were on the outside, the first two and the last site could be used. N-glycosylation would account for a size discrepancy between an *in vitro* synthesised protein and one immunoprecipitated from transfectants and cell lines. Peptide sequencing of both major products from transfectants would be the only way to confirm the protein sequence.

6.10 Apo-4 P1 antibodies suggest specific staining in tissues

Staining was performed in collaboration with Helen Turley, Department of Cellular Science, John Radcliffe Hospital on frozen sections from cerebellum, cerebral cortex, muscle, tonsil, and spleen and cytospins of K562. Sections and cells were probed with crude and purified anti-P1, P2 and, in some cases, P3 antibodies using the biotin-avidin system to attempt to detect apo-4 transcripts in these tissues. To assess their binding capability in situ, anti-apo-4 antibodies and

Fc-CD33 were used to probe apo-4 and control transfectants at 4℃ and the signal was detected with fluorescent antibodies.

In cerebral cortex (Figure 6.26I, A-D) anti-P1c (C) and anti-P2c (D) showed positive staining on microglial cells and vessels compared with pre-bleed controls (P1c-A, P2c-D). Purified antisera was then tested on cerebral cortex (Figure 6.26I, E-H) and cerebellum (Figure 6.26I-L). Cerebral cortex retained microglial staining but lost most vessel staining with P1p (F) seen more clearly with a 40x magnification (G), while P2p showed some microglial and no vessel staining (H). In response to tissue damage, microglia transform into large amoeboid phagocytic cells and are thus considered to be members of the macrophage-monocyte defence system (Wheater et al. 1987), suggesting a potential connection between the brain and lymphocytes. In cerebellum, a marked contrast was shown between the absence of staining with no antibody (I), and some Purkinje cell and granular staining with P1p (J) less so with P2p (K) and much less with P3p (L), similar to the control. In cerebellum, each granule synapses with several hundred Purkinje cells to coordinate movement (Wheater et al. 1987).

Purified P1, P2 and P3 antibodies were then tested on Muscle (Figure 6.26II, A-C), Tonsil (Figure 6.26II, D-I) and Spleen (Figure 6.26II, J-L) on which P2p and P3p were also tested. In muscle, P1p (B) gave stronger staining than P2p (C) which appeared to give no staining. Crude and purified antisera were tested in tonsil in which P1c (F) showed some basal layer lymphocyte and pronounced basal epithelial staining compared with no antibody (D) and the P1c pre-bleed (E). P1p showed a similar but less intense pattern of staining (G) while P2p (H) and P3p (I) showed fainter staining on vessels and epithelium. In spleen, no

staining was shown without antibody (J) while P1p (K) and P2p(L) showed some red pulp but no white pulp staining of lymphocytes.

In K562 cytospins (Figure 6.26III, A-H) P1c (B) and P2c (D) showed apparent intracellular staining, although P1c stained throughout the cytoplasm and nucleus, excepting the nucleolus, and P2c stained only around the nucleus. The staining could have been extracellular, as the cell surface covers the nucleus and could give what appears to be nuclear and cytoplasmic staining. P1c gave the most pronounced staining, but both gave intense staining relative to their respective pre-bleeds, P1c (A) and P2c (C). Purified antisera gave less intense staining, although again it appeared to be intracellular for P1p (F), P2p (G) and less so for P3p (H), although all three were more pronounced than the negative control (E). In a FACS analysis comparing the ability of anti-apo-4 P1c to stain cell lines, K562 showed 5.64% positive staining followed by 4.31% on the leukemic T cell line CEM and 1.84% on the leukemic T cell line HSB2 (data not shown). Protein searches of the peptide sequence used to generate the antisera revealed no matches and some distant homologues with an identity of 30% indicating that the antisera does not recognise any known proteins. The ability of anti-apo-4 P1 to stain cortical glial and cerebellar granule and Purkinje cells is very similar to the staining pattern shown for brain and Purkinje dystrophin and apo-dystrophins 1 and 2 (reviewed in Ahn and Kunkel 1993; Blake 1994) indicating that apo-4 may be involved in movement.

6.10.1 An extracellular N-terminus is stained

Crude and purified Anti-apo-4 and anti-CD22 antisera were compared in Apo-4 and CD22 transfectants stained in situ with signal detected by FITC-labelled

secondary antibodies (Figure 6.27I). A comparison with phase-contrast images allowed an assessment of transfection and staining efficiency. Apo-4 transfectants show superior staining with anti-P1c antisera (A-B) compared with anti-P1p (E-F) and negligible staining with anti-CD22p (I-J). CD22 transfectants show strong but inferior staining with anti-CD22c (C-D) compared with strong staining with anti-CD22p (G-H) and negligible staining with anti-P1p (K-L).

Phase contrast microscopy and immunofluorescent microscopy were again used to compare anti-apo-4 staining in transfectants (Figure 6.27II). Anti-apo-4 P1c gave no staining on sham transfectants (A-B) but showed highly specific staining on apo-4 transfectants (E-F) compared with anti-CD22p antisera which gave no staining (E-F) and anti-apo-4 P1p which gave slightly less staining (G-H). The apo-4 staining in figures 6.27IB and 6.27IIF demonstrate that apo-4 staining clearly occurs at the cell surface indicating that the sequence recognised by anti-apo-4 P1c is extra-cellular and arguing that either the apo-4S and/or apo-4F models predict an extra-cellular N-terminus in COS cell transfectants.

6.10.2 Fc-CD33 shows low-level apo-4 staining

Apo-4 transfectants were examined for Fc-CD33 staining (Figure 6.27III). Fc-CD33 showed no staining on sham transfectants (A-B) and faint, although visible, staining on apo-4 transfectants confirming that Fc-CD33 was capable of binding apo-4 in a COS cell transfectant (C-D). CD33 polyclonal antibody staining is shown on CD33 transfectants (E-F).

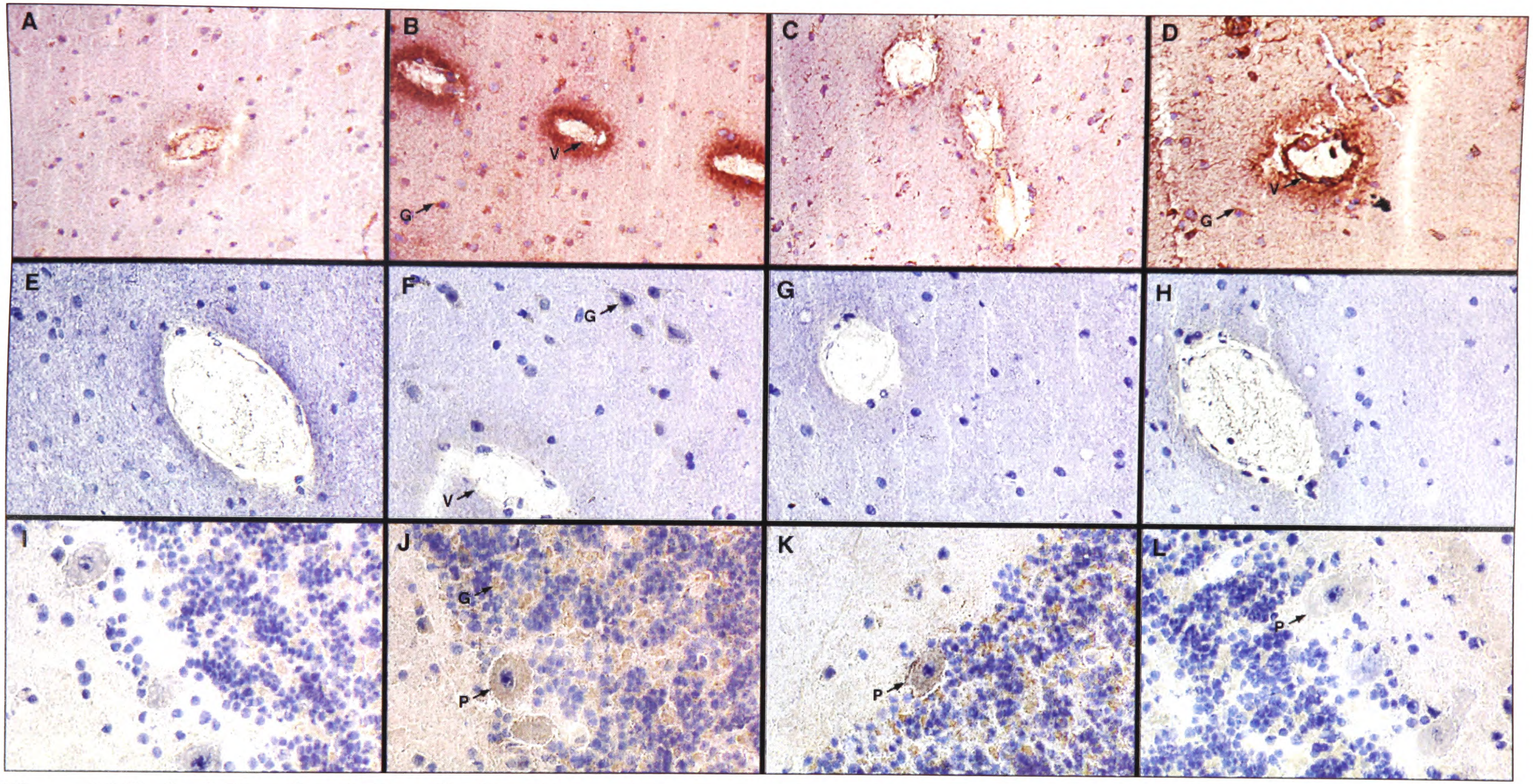


Figure 6.26I. Anti-P1 and anti-P2 antibodies show staining on vessels and glial cells in brain. Frozen sections from cerebral cortex (A-H) and cerebellum (I-L) were probed with crude and purified P1 and P2 antisera. P1c (B) and P2c (D) showed pronounced staining of vessels and glial cells over P1-pb (A) and P2-pb (C). Purified P1p (F-G) retained staining of glial cells but showed reduced staining of vessels, as did P2p (H) as compared with no antibody (E). In cerebellum, P1p (J) and P2p (K) showed staining of Purkinje cells more pronounced in P1p than in P3p (L) or no antibody (I); all three anti-apo-4 antisera showed staining of the granule cell layer "G". Images are a 40X magnification.

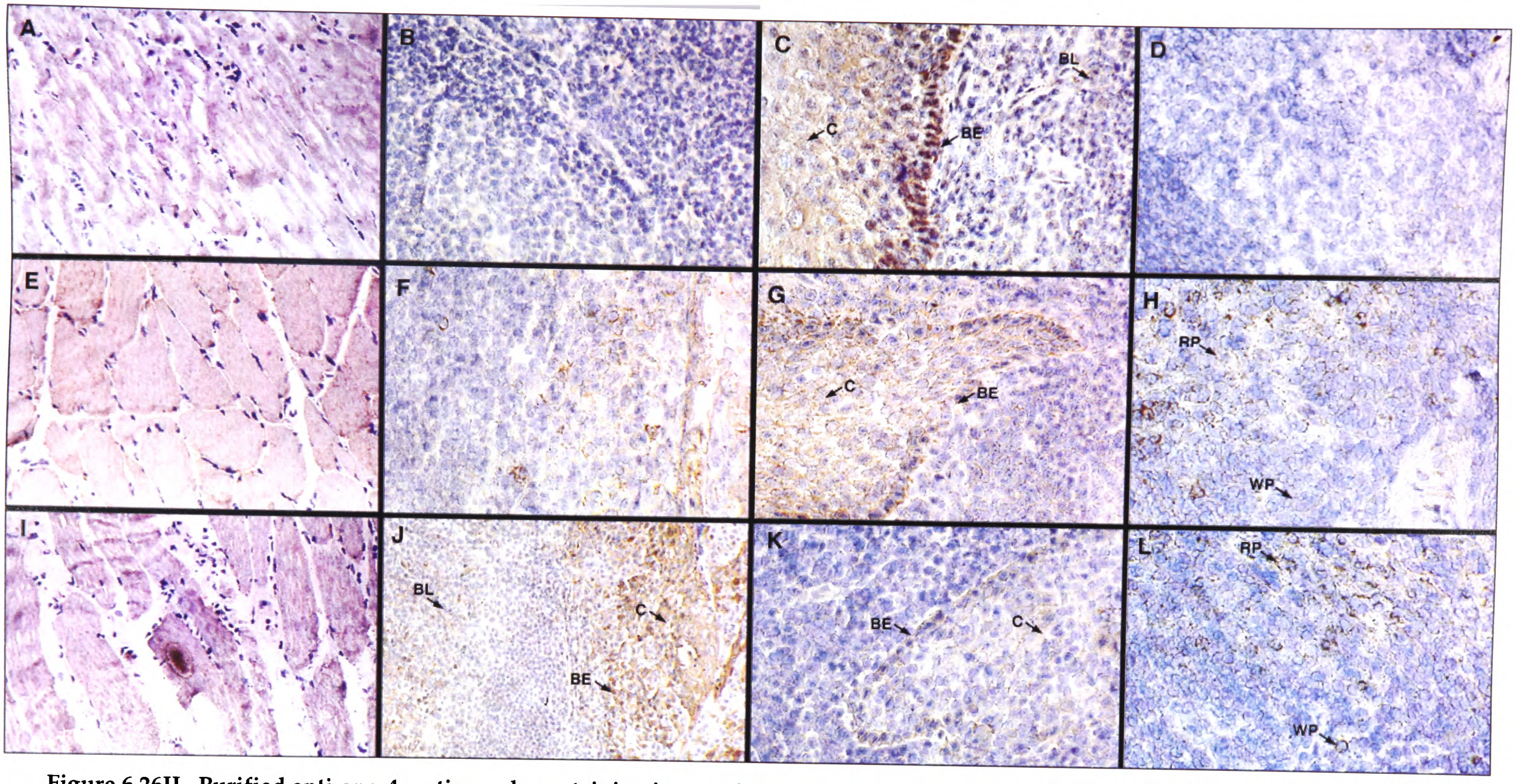


Figure 6.26II. Purified anti-apo-4 antisera show staining in muscle, tonsil and spleen. In muscle, P1p (E) shows some muscle cell staining as compared with P2p (I) or no antibody (A). Crude and purified antisera were tested in tonsil in which P1c (J) showed some lymphocyte staining in the basal layer "BL" and crypt epithelial "C" staining compared with no antibody (B) and the P1c pre-bleed (F). P1p showed similar but fainter staining (C) while P2p (G) and P3p (K) also showed weaker staining on basal epithelium "BE" and BL . In spleen, no staining was shown without antibody (D) while P1p (H) and P2p (L) showed some red pulp "RP" but no white pulp "WP" staining. Images are a 40X magnification.

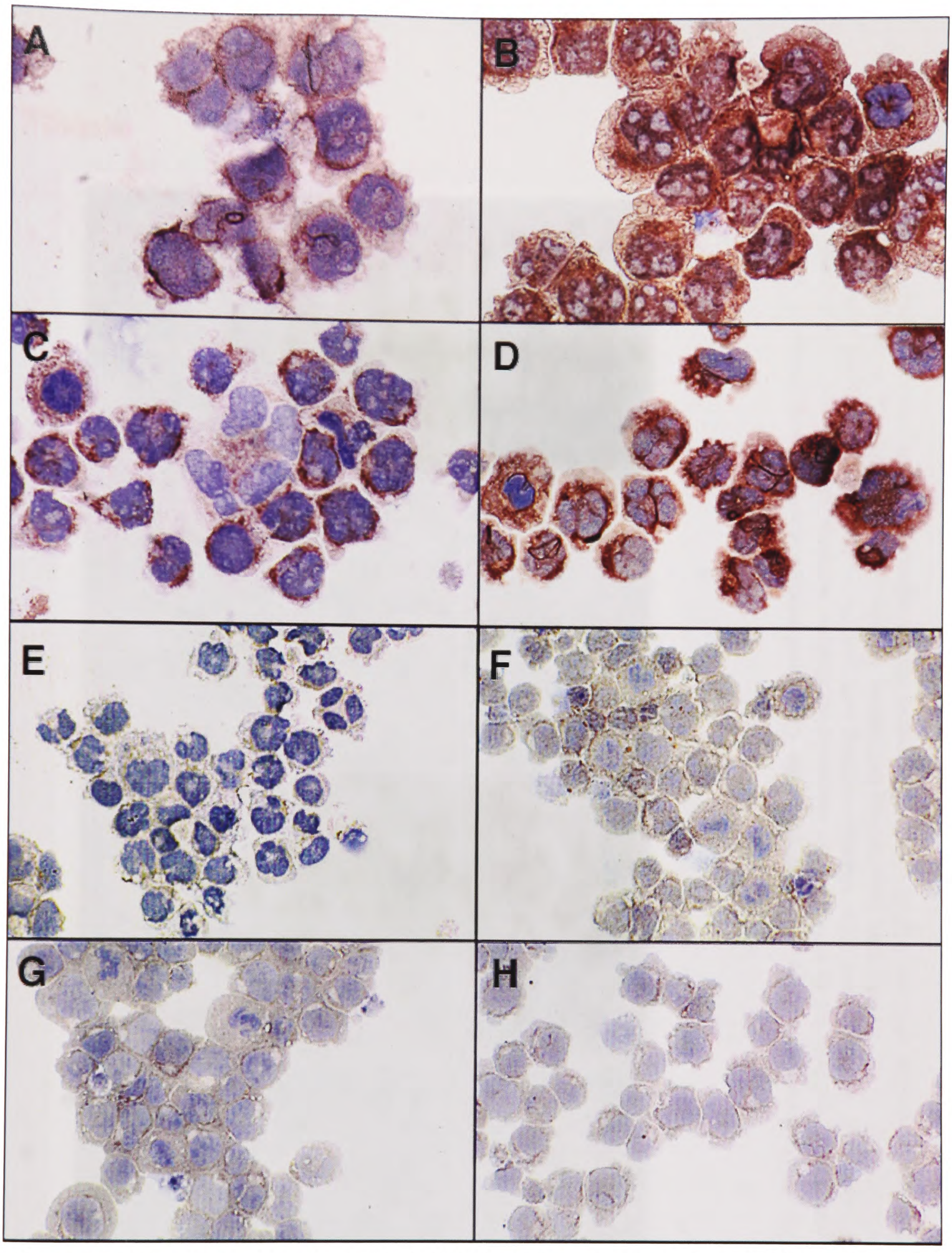


Figure 6.26III, A-H. Crude and purified antisera show intracellular staining in K562. In K562 cytospins, anti-apo-4 P1c (B) showed apparent nuclear but not nucleolar staining as shown by the white regions within the nucleus; the unstained nucleus may be mitotic. Staining of the cell surface could give the appearance of intracellular staining, suggested by the staining throughout the cytoplasm. Anti-apo-4 P2c (D) shows no nuclear and some apparent cytoplasmic staining concentrated around the nucleus. Both gave stronger staining than their respective pre-bleeds, P1c-pb (A) and P2c-pb (C). Purified antisera gave less intense staining than crude antisera. As in B, the staining with P1p (F) appeared to be across the cell surface and/or within the cytoplasm and possibly in the nucleus. P2p (G) gave only apparent cytoplasmic staining as did P3p (H) more weakly, although all three were more pronounced than a control with no antibody (E). Images are a 40X magnification.

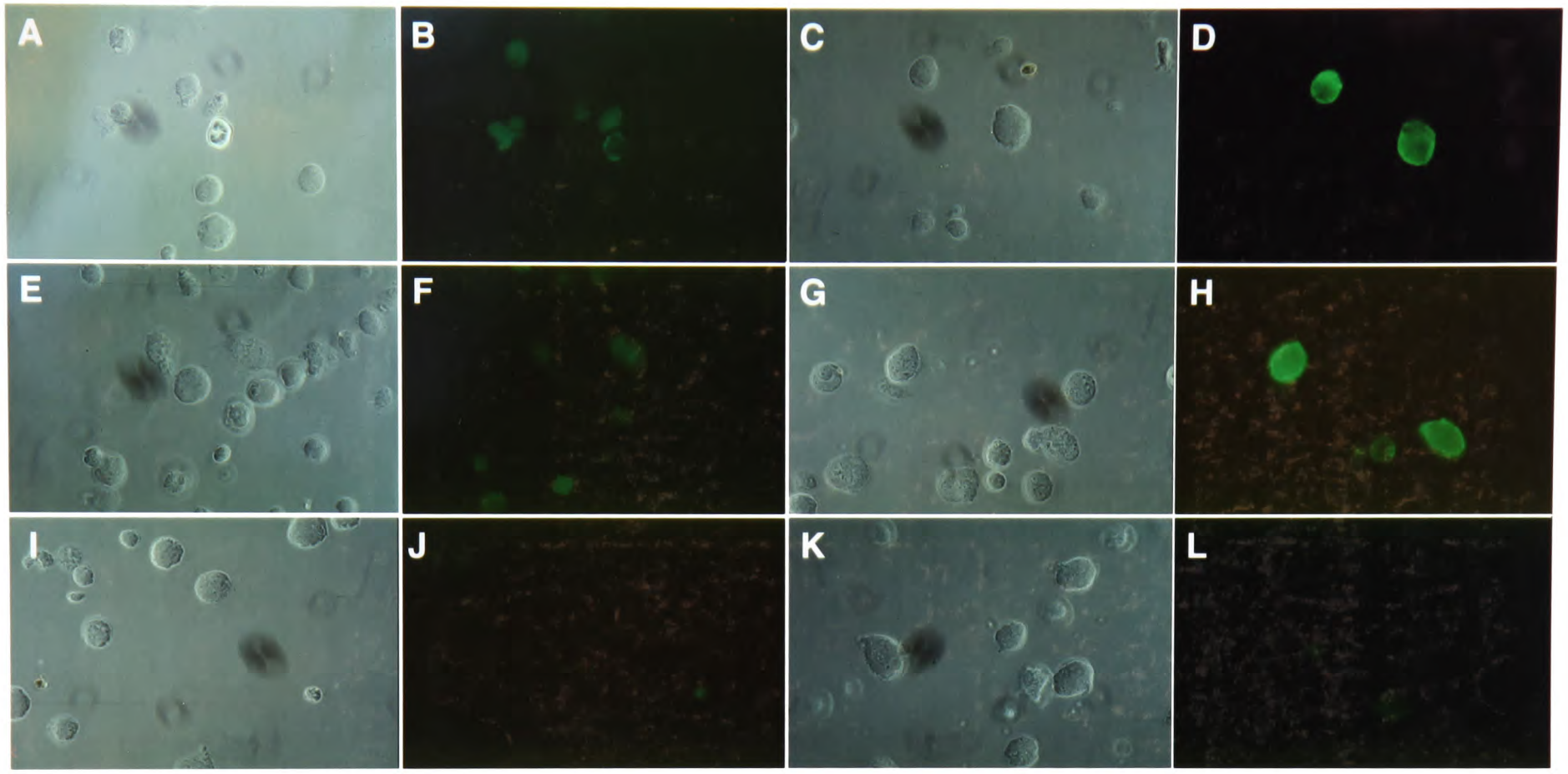


Figure 6.27I. Anti-apo-4 P1 antibodies specifically stain transfectants. Crude and purified anti-CD22 and anti-apo-4 antisera were compared in Apo-4 and CD22 transfectants and signal detected by FITC-labelled secondary antibodies. Samples are shown with phase-contrast images to assess transfection and staining efficiency. Apo-4 transfectants show superior staining with anti-P1c antisera (A-B) compared with anti-P1p (E-F) and negligible staining with anti-CD22p (I-J). CD22 transfectants show strong but inferior staining with anti-CD22c (C-D) compared with strong staining with anti-CD22p (G-H) and negligible staining with anti-P1p (K-L). All fluorescent images represent eight second exposures.

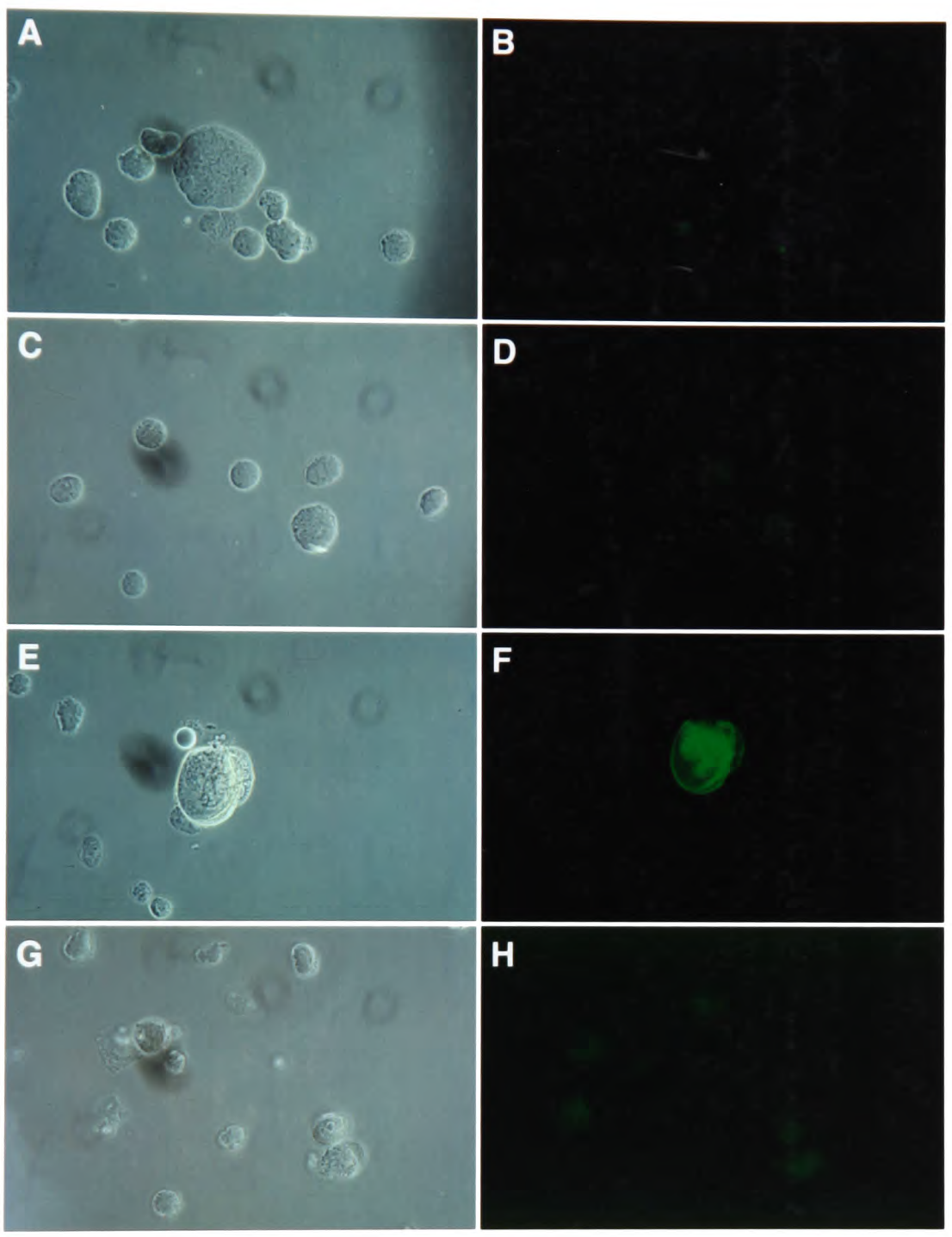


Figure 6.27II. Anti-apo-4 P1 antisera show pronounced staining on apo-4 transfectants compared with sham transfectants. Phase contrast microscopy and immunofluorescent microscopy was used to compare anti-apo-4 staining in transfectants. Anti-apo-4 P1c gave no staining on sham transfectants (A-B) but showed highly specific staining on apo-4 transfectants (E-F) compared with anti-CD22p antisera which gave no staining (C-D) and anti-apo-4 P1p which gave slightly less staining (G-H).

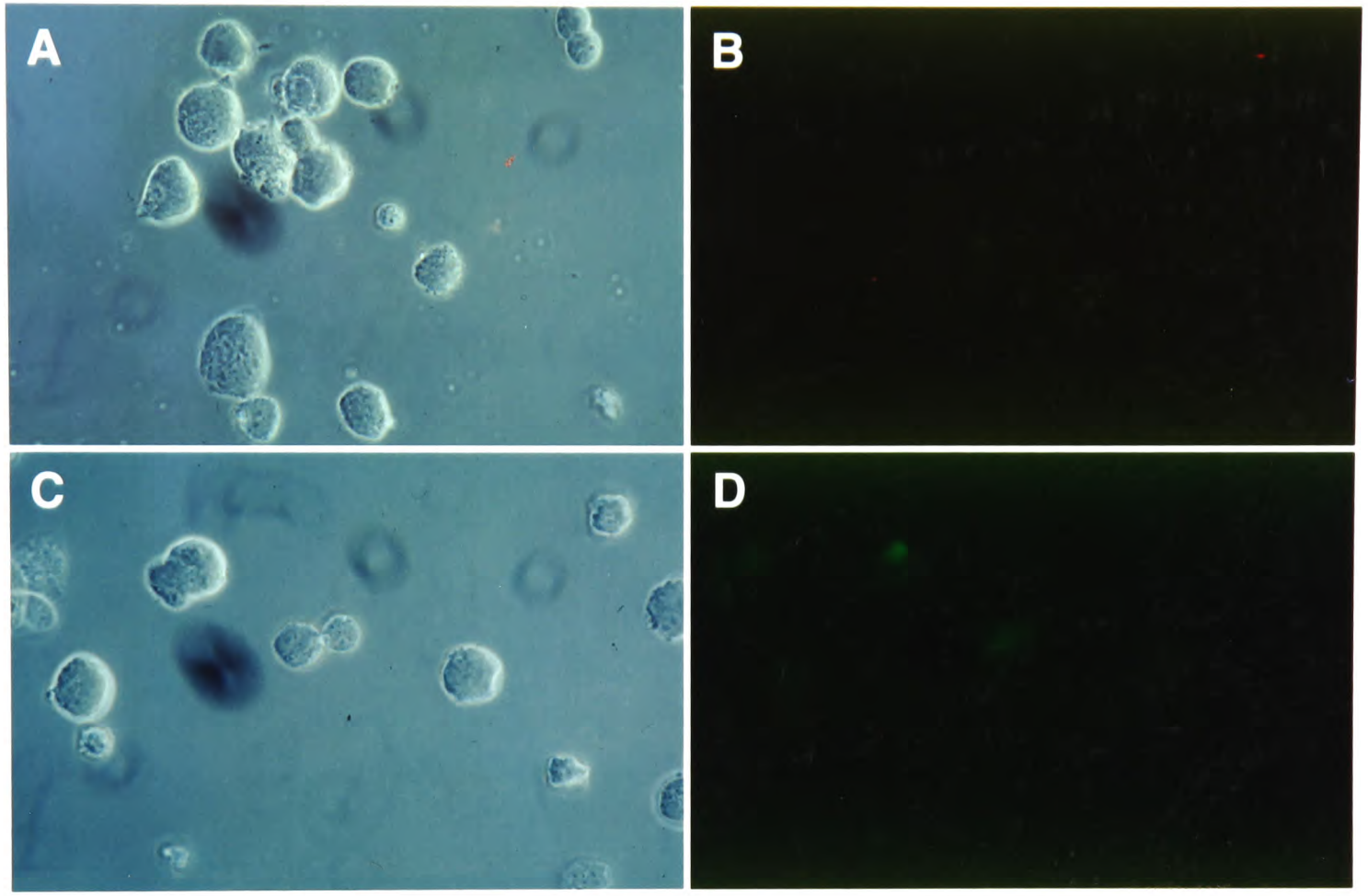


Figure 6.27III. Fc-CD33 shows some staining on apo-4 transfectants. Fc-CD33 shows no staining on sham transfectants (A-B) and some staining on apo-4 transfectants (C-D).

6.11 Discussion

Binding studies with Fc-CD33 on iodinated cells revealed that it was capable of binding several ligands, at, on average 46Kd, 54Kd, and 61Kd on K562, and an additional ligand from 97.4-110Kd on U937 and possibly HL-60. Transfected candidate cDNAs isolated from panning bound between 2-10%, on average in FACS analyses and sequencing of three of these candidates revealed that they were ICAM-1 (p95-110Kd) and a novel transcript with homology to the 3′ end of dystrophin, termed here as apo-dystrophin-4 which demonstrated low-affinity binding to Fc-CD33 in iodination with proteins of 50/66Kd and more strongly at 97-100Kd, with a slight variation in protein product weight between assays. Attempts to repeat the ICAM-1 binding with metabolic labelling were unsuccessful however antibody blocking studies on various cell lines are necessary to assess its ability to bind Fc-CD33 *in vivo*. Isolating multiple ligands which require coexpression to achieve binding could not be achieved from a number of attempts with the low-avidity and low-affinity binding properties of at least five Fc-adhesins tested. We thus attempted to characterise apo-4 as a low-affinity ligand.

6.11.1 Apo-4 expression and Fc-CD33 binding

Apo-dystrophin-4 cDNA was isolated in the second round of panning from a placental cDNA library with Fc-CD33. Protein expression studies and in situ staining on apo-4 transfectants suggest that it can bind CD33 with low-affinity,

but only under optimal transfection conditions with freshly isolated Fc-protein. A full-length readthrough product from apo-4 with a 7, 9 or 31 amino acid N-terminus predicts two to four N-glycosylation sites available for intercellular binding. CD33 has recently been shown to bind N-linked sugars exclusively (Sgroi 1996) and the proposed structure of a full-length product of apo-4 displays two potential N-glycosylation sites, while ICAM-1 possesses five N-glycosylation sites compatible with CD33 binding .

6.11.2 The inversion could result from several factors

The structure of the 1Kb apo-4 clone showed a 136 bp inversion at the 3' end which could result from the insertion of a downstream direct repeat into a compatible upstream location during recombination in genomic DNA (Figure 6.16C). The 3′ 274 bp including the inversion was reproduced in RT-PCR products obtained from placental DNA (Sigma) and in first-round cDNA from the cell lines K562, HL-60, U937, CEM, THP1, KG-1, brain and a small amount from HepG2. The partial conservation of recombination signal sequences (RSS's) or repeats near the inversion in genomic DNA may contribute to an explanation for the reproducibility of the results in leukemic cell lines. In dystrophin, the 3Kb of 3' UTR is conserved to 95% homology in chicken but has no known function. A novel regulatory role for 3' untranslated regions (UTR's) in growth and differentiation has recently been demonstrated in differentiated skeletal muscle cells (Rastinejad and Blau 1993). It was suggested that the 3' UTRs of certain differentiation-specific RNAs can function as trans-acting regulators in a feedback loop which inhibits cell division and promotes differentiation.

Transcripts from the 3'UTR of alpha tropomysin have also been shown to act as tumor suppressors (Rastinejad et al. 1993). Apo-4 appears to be the product of a functional illegitimate recombination event in the 3' UTR of dystrophin and may carry out a related function.

6.11.3 A potential role for RAG-1 and RAG-2 in the inversion

The recombinase activating genes (RAG-1 and RAG-2) genes have been shown to be essential to produce rearrangement as recently reviewed (Gellert 1992). The inversion products obtained in T cell lines may suggest a role for the RAG-1 and RAG-2 genes. Although only RAG-1 is present in murine brain and both are normally found only in lymphocytes, it is possible that myeloid cells contain low levels of RAG-1 and RAG-2 in association with other factors important in rearrangement. Other known factors include the TdT enzyme which generates non-templated insertions, some endonucleases which cut at or near signals, two heptamer-binding proteins, a nonamer-binding protein, and a protein (RBP-Jk) that binds to 23-spacer signals have been found in murine lymphoid extracts. The inversion was originally isolated from placental DNA which may suggest a role for a factor with a broader distribution than lymphoid cells. A suggested pathway of rearrangement, diagrammed below, may have resulted from a multifactorial contribution from RSS's, and/or indirect and direct repeats aiding the action of the inversion as a small transposon-like element.

Likely Pathway of apodystrophin transcript to protein

Genomic rearrangement of 3' end of transcript during replication

↓

Pre-spliced RNA transcripts made from rearranged genomic DNA

↓

cDNA library made with the CDM8 vector capturing rearranged and unspliced apodystrophin transcript

↓

COS cells transfected with rearranged DNA allow for the expression of one or more protein products which are captured during the panning process.

↓

Sequencing reveals that the first 860 bp of apo-4 cDNA aligns with genomic DNA, although several smaller products are obtainable with PCR, indicating that at least one of the protein products from the gene may result from splicing.

6.11.4 Transcription of apo-4 does not require a TATA box

A transcription initiation site was mapped at -70 bp upstream of the first met in apo-4. Although the longest ORF and potential splice product only predicted a protein of 21.5Kd, *in vitro* translation produced products at 22, 40 and 50 Kd, which depended on the presence of the inversion to be produced. This suggests that the inversion controls the 5' end of the protein, possibly as a downstream enhancer, consistent with previous reports that the 3' UTR can regulate the 5' end of the gene (Jackson 1993). Anti-apo-4-P1c antisera were able to precipitate a major product at p48-50 Kd and one at p95 and P2c precipitated products at p50 and p97 from COS cell transfectants (Figure 6.20) suggesting that stop codons in the sequence were not recognised. Known mechanisms of "readthrough" include RNA editing (T to A or A to G) by adenosine deaminase (dsRAD), which usually only occurs at a few RNA sites, or suppression of stop codons by signalling motifs in the sequence or by modification of the tRNAs reading the sequence. It is also possible that the hydrophobic regions identified in the apo-4 gene could act to bind DNA and potentially regulate transcription as occurs in the DNA binding regions found in *c-myc* to which the DNA binding protein MAX binds (Blackwood and Eisenman 1991).

6.11.5 Proposed explanations for stop-codon "readthrough"

This is the first known case of a dystrophin transcript being expressed on the surface of leukemic and potentially other cells. How can protein(s) be obtained

using this system if stop codons appear in the sequence between the beginning of the cDNA and the start of the ORF of interest? Anthony Monaco originally described the "reading-frame hypothesis" which stated that DMD would result from a mutation which shifted the reading frame in dystrophin, thus prematurely truncating the protein (Monaco et al., 1988). This hypothesis has been shown to be accurate in 92% of DMD cases (Koenig et al., 1989). However, recent work on DMD patients shows that despite nonsense mutations which should result in the complete absence of dystrophin, a small proportion of "revertant" dystrophin is present in muscle fibres. "Exon skipping" has been proposed to account for this, as the exons flanking the mutation are skipped in the revertant mRNA (Nicholson et al., 1992; Sherratt et al., 1993). Exon skipping resulting in frame restoration has also been observed in illegitimate (ectopic) dystrophin transcripts isolated from lymphocytes and fibroblasts (Chelly et al. 1991; Roberts et al. 1991). Frameshift mutations which should delete dystrophin also demonstrate 12-15% dystrophin levels and ribosomal frameshifting and reinitiation are two mechanisms proposed to account for the correction (Gangopadhyay 1992). The dystrophin gene appears to use a variety of mechanisms to correct mutations to preserve some form of this life-saving protein.

Four additional mechanisms are also suggested: 1) splicing of the RNA following transcription to eliminate stop codons 2) "readthrough" of the stop codons by the translational apparatus or by the transcriptional apparatus which allows the stop codons to be bypassed, 3) RNA editing , and 4) a combination of events which accounts for transcripts of varying lengths, such as readthrough of transcripts initiated at different methionines. Even one splicing event would require the excision of intronic sequence of at least 20-100 bp which would reduce the weight

of the longest possible product from 1-5Kd. Assuming a 100 bp intron per splice, each splice would represent a loss of 5Kd in weight from the longest potential product. The longest product which could be produced from the available potential splice sites in apo-4 was 19.5Kd making splicing an unlikely event. Using the first available methionine at 9aa in the given sequence, if all stop codons were eliminated, would produce an ORF of 323 amino acids if polyadenylation occured at the 3'end of apo-4. That the *in vitro* products are 40 and 50Kd implies that they occured through either readthrough of stop codons, RNA editing, or some combination of the two. Readthrough is the most likely since levels of adenosine deaminase are unlikely to be high enough in an *in vitro* system to produce RNA editing. RNA editing cannot be completely ruled out without further experimentation, however, and has been shown to account for differentially polyadenylated forms of apolipoprotein which enable it to be directed to different tissues (Davidson 1993) and has been used therapeutically to correct an upstream nonsense mutation (Woolf et al. 1995).

Several theories have been advanced to explain the mechanism of "readthrough" mainly based on the "wobble hypothesis" which allows for flexibility in reading the third base by tRNA (Atkins et al. 1991; Tuohy et al. 1992). Indeed, a codon ending in A does not necessarily code for any specific amino acids given the potential wobble of tRNA and/or the modification of the tRNA reading it. Readthrough could also result from the strong T7 promoter or some other element interacting with the RNA polymerase to allow transcription to continue. This is more common in phage and may be related to the homology apo-4 shares with viruses found in protein homology searches of the 5' end of the gene. In mice, a 60 bp sequence at the extreme 3' end of the sequence can control protein production (Braun 1990). In apo-4 all or part of the 136bp sequence at the 3' end

could control protein production so that all stop codons are suppressed by a proposed mechanism which will be addressed in Chapter 7. The AATTAA sequence at the extreme 3' end of apo-4 appears to serve as a cryptic polyadenylation site, as it is in the right place to produce p50 from the "strong" met at +34 aa. In addition to the novel proteins characterised in this study, the inversion itself likely contains powerful clues about 3' motifs capable of regulating protein production which are worthy of further study, principally by placing the inversion sequence downstream of other DNA with stop codons which terminate translation. The products of apo-4 appear to be regulated despite readthrough and it is important to study further the effects of apo-4 *in vivo*.

The observations reported in this chapter have been provoking in that the explanation for the production of both the apo-4 transcript and its translation products demand the involvement of rare or novel mechanisms. It is also appreciated that the techniques employed are working at the limits of sensitivity and resolution in some cases. Validation of the entire genomic rearrangement predicted to underly the production of the apo-4 transcript has yet to be obtained. Wide acceptance for the model presented may await complete sequencing of the genomic region containing the rearrangement and of the corresponding protein products of apo-4. Until then, the results obtained can be taken to represent an intriguing combination of findings suggesting the existence of a potentially highly significant phenomenon.

Chapter 7

Concluding Discussion

7.0 The limits and advantages of Fc-adhesins

The underlying theme of this work is that both CD33 and CD34 and their ligands appear to play a role both in differentiation and leukemia and that these processes appear to be linked, particularly in the case of CD33. The use of recombinant soluble Fc-CD33 and Fc-CD34 has allowed these molecules to be studied in isolation in several different systems *in vitro*. The main advantage of these reagents is that they can be produced relatively quickly, purified in large quantities and used in a similar fashion to antibodies with many standard techniques. Unlike antibodies, however, they bind with a low affinity and usually require presentation to target ligand in a "cluster" or in higher quantities than antibodies to achieve a similar level of binding, as shown with adhesion assays (Simmons 1993; Watt et al. 1993) and, in these studies, with flow cytometry analyses. An additional potential limitation is the level of O-linked glycosylation on the ECD of the recombinant molecule. It is questionable whether COS cells possess the enzyme to allow correct O-glycosylation of the 30 potential sites on CD34, despite a correctly sized molecule on an SDS-PAGE gel. Even if the sites were occupied, they may be in a conformation unrelated to that *in vivo*. The same problem could apply to molecules from cDNA libraries expressed on the cell surface. Both of these problems present an inherent limitation to the panning process which became particularly evident with CD34. N-linkages produced in COS cells appear to be accurate, however, which allowed for a more reliable, yet still low affinity panning probe for ligand. In addition, Rg and pIg1 recombinant proteins appeared to bind proteins with different affinities, despite the nearly identical nature of the vector. Although these differences may be relatively minor, they could bind different low-affinity ligands, particularly important in studies *in vitro*. These limitations necessitated using a number of other

approaches to characterise potential CD33 and CD34 ligands. This work attempted to shed light on the ligand binding behaviour of these important cell-surface antigens.

7.1 Chapter 3: Fc-CD33 Binding Increases with Differentiation

Fc-CD33 was a particularly useful reagent in elucidating the ligand binding behaviour of CD33 along the differentiation pathway in a series of FACS analyses and in characterising the size of potential CD33 ligands via cell-surface iodination. In FACS analyses, the highest percent positive fluorescence in gated was shown for promonocytic U937, followed closely by promelocytic HL-60 and then erythroleukemic HEL and K562. In ungated cells, however, HL-60 followed by HEL and K562 were shown to predominate (Barber 1993). This is in contrast to data obtained some time later using Fc-CD33 in adhesion assays which showed little binding for K562 obtained from the same source but also showed the highest percentage of binding for U937 and general CD33 binding to sialic acid (Freeman 1995). This was unexpected considering the high levels of sialic acid on erythroleukemic cells, suggesting that the adhesion assays should be repeated with K562 from several sources and used at less than six passages, considering that those originally used in the adhesion assay had undergone many more passages than the six recommended (personal communication, Cell Culture Lab, ICRF Central Services, Herts, UK). In FACS analyses, the binding to K562 and U937 was shown to be concentration dependent and the earlier high affinity binding shown to U937 suggested that it may possess an additional high-affinity ligand not or only weakly expressed on K562. An explanation for the earlier high-affinity CD33 binding to U937 was found in the iodination assays which

demonstrated that Fc-CD33 was capable of binding to increasing numbers of subunits as myeloid differentiation progressed. In promyelociytic KG-1b, Fc-CD33 precipitated a 58Kd subunit. In K562, a middle stage of differentiation, Fc-CD33 precipitated 54Kd, 58Kd and 61Kd subunits. In promonocytic U937 binding was also pronounced and a new subunit appeared from 97-100Kd in addition to those at 54 and 69Kd. HL-60 also displayed a broad band at 97-100Kd in addition to one at 54-61Kd and one at 23Kd. These findings suggested that the 97-100Kd band could be a high-affinity ligand for Fc-CD33 and that Fc-CD33 binding was heterophilic. Only one of these ligands could be isolated from the panning process at a time, however and would almost certainly require cotransfection to achieve maximal binding.

7.2 Chapter 4: CD33 and ligand kinase associations

The first evidence of the existence of a group of transducing retroviruses came with the discovery of extra genetic sequence, termed*v-src* , found in the Rous sarcoma virus which was not found in nontransforming retroviruses but was identified in normal cellular genes (*c-src*) (Temin 1964). This genetic material became known as the first protooncogene (from the Greek *onkos,* meaning a bulk or mass) as it provided evidence of a normal gene which could become transformed following the integration of a retrovirus near the protooncogene, resulting in a much higher level of transcription and oncogene products capable of phosphorylating substrates within the cell (Purchio 1978) which are now known as the *src*-like kinases. The $p60^{src}$ protein, shown to be expressed at increasing levels along the differentiation pathway represented by myeloid leukemic cell lines, is thought to be a constituent of normal cells (Gee et al. 1986). The *src*-like kinases are shown

for the first time in this study to associate with CD33 and at least some of its ligands. Phosphoproteins associate with CD33 at an increasing level as differentiation progresses, with the most pronounced phosphorylation shown in U937 followed by KG-1b which, it is suggested, actually represents a more differentiated cell or perhaps demonstrates an early elevation of CD33 expression followed by a decrease in KG-1a and recovery in phosphorylation in the more differentiated U937. An unexpectedly low level of phosphorylation was shown in K562 and a drift in the expression of CD33 in this cell line cannot be ruled out. The pronounced phosphorylation in U937 was shown to result from a complex of *src*-like kinases, including *c-fgr, fyn, hck, lyn* which associate with the cytoplasmic tail of CD33. This is not unexpected considering that YXXL/I/V motifs were identified in CD33 which have been shown to act as docking sites for the *src*-homology 2 (SH2) domains phosphorylated in*src*-like kinases to allow them to activate/inactivate the target receptor (Hatada et al. 1995). The phosphorylation of a 67Kd band obtained under novel serum conditions suggests that CD33 is phosphorylated directly, but further experimental evidence is needed to confirm this. The correlation of pronounced phosphorylation in U937 with the highest level of ligand binding suggests that phosphorylation could play a role in ligand binding by exposing the CD33 ligand binding site by removing it from cis-sialic acid interactions or simply increasing the expression level of CD33 on the cell surface to allow it to "cluster" and bring it into contact with ligand. Elevated expression could be sufficient to disengage CD33 from neighboring sialic acids which would normally mask its binding site. It will be important to determine the levels of phosphorylation associated with CD33 in normal monocytes, macrophage and dendritic cells and compare it with phosphorylation in transformed cell lines to assess a potential transforming role for phosphorylation. A number of kinase-specific therapies are currently

under development, and the cytoplasmic tail of CD33 may provide a useful target for such a therapy to inhibit potentially oncogenic kinase activity.

A reciprocal phosphorylation mechanism was suggested by *in vitro* kinase assays performed on Fc-CD33 immunoprecipitates. An unexpected pattern emerged which was often the reciprocal of that obtained with CD33 immunoprecipitates. Where there was little CD33-associated phosphorylation in KG-1a and K562, there was pronounced Fc-CD33 ligand-associated phosphorylation. Similarly, where there was pronounced CD33-associated phosphorylation in U937, there was comparatively less phosphorylation associated with CD33 ligands. Importantly, Fc-CD33 ligands appear to associate with the *src*-like kinase *lck* identified in K562 (Rouer et al. 1994) which was previously thought to exist only in T-cells. Soluble recombinant Fc-CD33 offered a unique tool to visualise the associated phosphorylation differences between CD33 and its ligand and suggests that a cloning strategy based on degenerate primers and YXXL motifs might allow for identification of a ligand for CD33. A more straightforward approach would involve affinity purifying CD33 ligands from U937, HL-60 and/or K562 lysates precleared of high-affinity Fc-receptors over an Fc-CD33 coupled affinity column. An important correlate of the CD33-associated phosphorylation in U937 is the high-level of ligand binding shown for Fc-CD33 on U937 in FACS analyses. It will be important to assess whether phosphorylation decreases as differentiation progresses. If so this would lend weight to the argument that CD33-associated phosphorylation plays an important role in CD33 ligand binding.

7.3 Chapter 5: Fc-CD34 binds to unique ligands on HUVEC

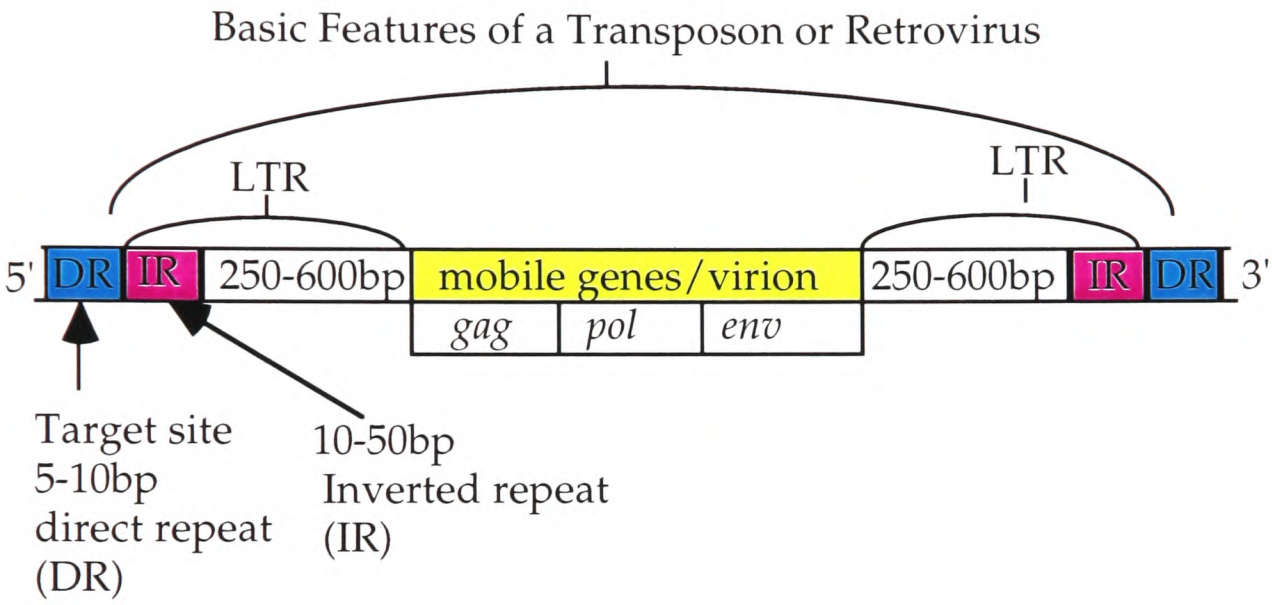
Fc-CD34 demonstrated binding to HUVEC in the presence of growth factors and specific precipitation of potential ligands on surface-labelled HUVEC at 66Kd and 110Kd which appear to be divalent-cation dependent. These ligands are distinct in from L-Selectin, which is 74Kd on lymphocytes and 95Kd on neutrophils. Fc-CD34 binding was determined to be heterophilic through *in vivo* metabolic labelling. The Fc-CD34 ligand also precipitated unique phosphoproteins at 70Kd and 100Kd from primary HUVEC which were primarily tyrosine and threonine phosphorylated. The evidence for a second ligand for CD34 corroborates previous work which suggests that anti-L-selectin antibodies are not able to completely block CD34 binding in L-selectin expressing cells (Oxley 1994). The existence of a separate ligand for CD34 on endothelial cells suggests that it possesses a discrete function in stem cells from that in High Endothelial Venules on which it was shown to bind L-selectin. It will be important to identify this ligand either through peptide sequencing or antibody blocking with known endothelial markers to further elucidate the function of CD34 in endothelium. One of the major limitations to this system is the need to prepare primary cells with each experiment. The development of an endothelial cell line which could maintain the expression of CD34 would prevent both the difficulty of preparation and the risk of blood-borne and bacterial contamination inherent in the preparation of primary cells.

7.4 Chapter 6: Characterisation of an apo-dystrophin cDNA

Through the panning process, a putative low-affinity ligand for CD33 was isolated and termed apo-dystrophin-4. The apo-4 cDNA contains an inversion at its 3′ end which appears neccessary for the production of its two major protein products. In chronic and acute leukemic cells, and less commonly in lymphoma, both chromosomal and genomic DNA translocations and inversions often give rise to unique "fusion proteins" (Rabbitts 1991). The chromosomal rearrangement of 11q23 leads to the production of an active myeloid-lymphoid leukemia (MLL) gene at the point of fusion of the translocation consisting of the 5′ segment of MLL and a 3′ segment of the other gene (Rowley 1995). A similar fusion product, *bcr-abl,* is produced with the fusion of the *bcr* element with the abelson leukemia virus element in the chromosome 9:22 translocation often found in Chronic Myelogenous Leukemia (CML) cells (reviewed in (Varmus 1984). The triggers for an aberrant rearrangement of these unstable genetic elements are not known but repetive elements in the DNA, such as *Alu* repeats or aberrant rearrangement of the type found in immunoglobulin are thought to be among them (reviewed in Bishop 1983). Retroviral insertion is another common mechanism of gene activation of protooncogenes as described for *v-src* and *c-src*. Insertion of the Rous sarcoma virus next to normal genes brings about acute transformation due to high transcription of the activated oncogene. Similarly, retroviral elements already present in the DNA can rearrange themselves next to protooncogenes elsewhere and transform them into oncogenes. The DNA intermediates which occur during the replication of retroviruses are structurally and, as has been proposed, perhaps functionally related to transposable elements (Majors et al. 1981) which were first described in maize (McClintock 1956) (Figure 7.1A). A similar

mechanism of activation is proposed for apo-dystrophin-4 gene which appears to insert an inverted sequence containing the basic hallmarks of a retrovirus or transposable element (Figure 7.1B), into a specific target site in the dystrophin gene prior to splicing and most likely during gene rearrangement.

A.



B.

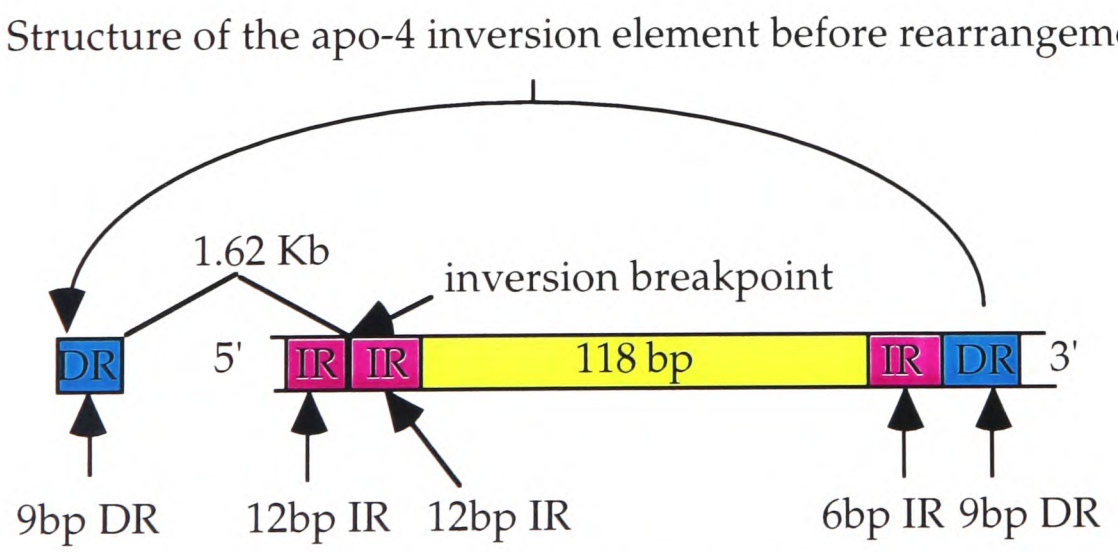


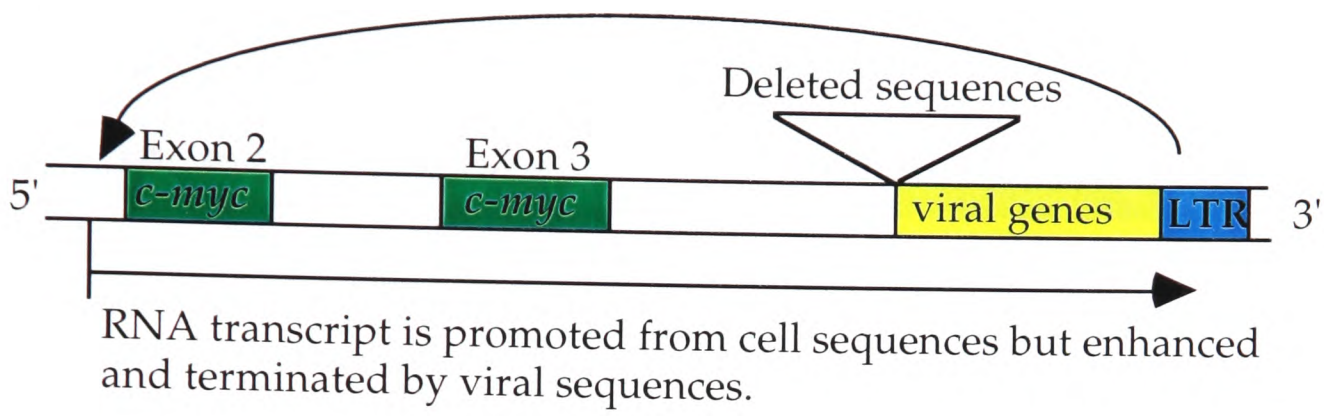
Figure 7.1 The apo-4 inversion shares several features with transposable elements. A. The 5′ and 3′ terminal inverted repeats characterise retroviruses and transposable elements as well as direct repeats at one end and at the point of insertion. **B.** The apo-4 pre-inverted sequence has the basic structure of a retrovirus with the exception of a single direct repeat at the 3′ end, which probably accounts for the upstream deletion in apo-4 rather than a complete transposition, and a short 118 bp sequence LTR-like sequence containing some repeats. "A" is adapted from (Temin 1964) and based on the common elements found in retroviruses and transposable elements (Majors et al. 1981).

A well studied example of retroviral transposition occurs with the insertion of the avian leukemia virus next to *c-myc* by virtue of the two LTRs which allow insertion upstream or downstream of the *myc* gene (Payne et al. 1982) (Figure 7.2A). *c-myc* is also implicated in Burkitt's Lymphoma (Croce 1984), BL-Acute T-cell Leukemia (reviewed in Rabbitts 1991). Insertion of the retrovirus upstream in the correct 5′-3′ orientation is referred to as "promoter insertion" while insertion upstream in the 3′-5′ orientation upstream or the 5′-3′ direction downstream results in "enhancer insertion" by virtue of the ability of the 3′ LTR to act as an enhancer in either orientation (Payne et al. 1982) (Figure 7.2A). Of primary importance to this study is the third model in which the 5′ LTR is deleted and the 3′ LTR acts as an transcription enhancer of the 5′ end of independent genes which are not transcribed in its absence, as illustrated below (Figure 7.2A). Retroviruses have the ability to ignore the stop codons present in their own genomes as is shown by the production of the readthrough protein *gag-pol* rather than the individual proteins *gag* and *pol* by the action of a glutamyl-tRNA which allows the stop codon to be bypassed and a single protein to be generated. It is possible that the same process could function to suppress the stop codons in apo-4 to allow translation to proceed.

Following rearrangement of the dystrophin 3′ sequence as shown above, it is proposed that the inversion could act as a downstream *myc*-like transcriptional enhancer (Figure 7.2B). In a similar fashion, the 72 bp SV40 enhancer can increase transcription in either orientation at distances of up to several Kb upstream or downstream of the start site (Banerji 1981) (Figure 7.2A). The inversion may not only direct upstream transcription by acting as an enhancer to recruit transcriptional activators, but may recruit some of

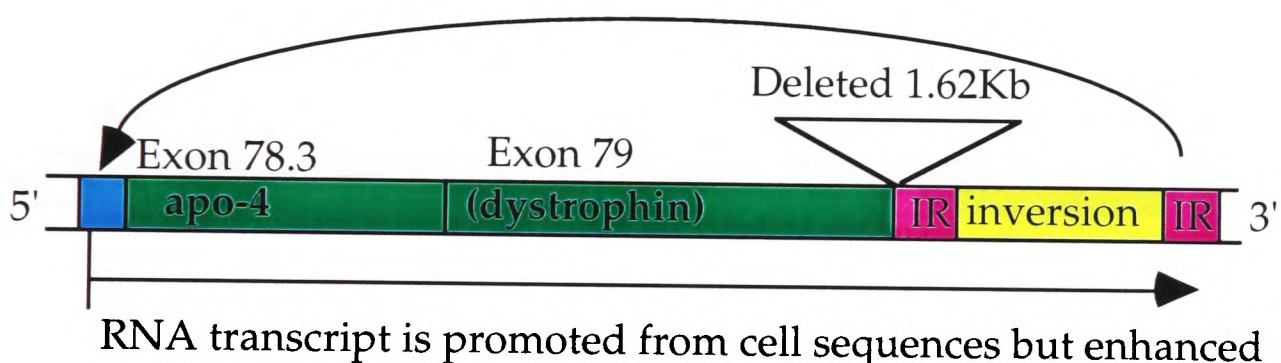
these DNA binding proteins to its hydrophobic regions to direct the production of translation products.

A.



RNA transcript is promoted from cell sequences but enhanced and terminated by viral sequences.

B.



RNA transcript is promoted from cell sequences but enhanced and terminated by inversion sequences which may also activate suppressor tRNAs or reverse transcriptase activity to prevent the recognition of stop codons. Inverted repeats (IR) are present at both ends of the inversion, as they are in retroviruses and transposable elements.

Figure 7.2 A proposed mechanism of activation of the apo-4 gene based on the retroviral insertion model of *c-myc*. The apo-4 inversion in vitro appears to act as an enhancer closely resembling a model previously outlined for the action of the c-*myc* oncogene with a retroviral *enhancer insertion*. This is illustrated by the removal of the inversion in apo-4 which produces a complete loss of translation in the gene. The upper diagram is adapted from (Darnell 1986) based on retroviral insertion behaviour previously described (Payne et al. 1982).

If this mechanism is indeed active, the 5′ 453 bp of the apo-4 gene may represent an activated oncogene which is normally dormant and/or spliced

out of the dystrophin gene. This may explain why the inversion was readily reproducible in several leukemic cell lines. It may have an independent function in normal cells, however, and such a possibility cannot be ruled out. The apparent retroviral or transposable behavior of the inversion may have larger implications. It may help explain why nucleotide and protein sequence searches of the 5′ 453 bp of the apo-4 gene often yielded significant homology only to viruses. It is possible that part or all of the 3′ end of dystrophin represents an ancient form of viral material which has integrated itself into the genome to become a permanent part of the gene, as suggested by identifying the inversion in genomic DNA. Indeed, the *Alu* repeat found downstream of the inversion and spliced out in apo-4 may form part of an enhancer suppressor as was recently shown with the probable *in vivo* formation of an *Alu* cruciform, thought to silence the activity of a 3' CD8 enhancer (Hanke et al., 1995). *Alu* repeats are thought to be over 30 million years old (Britten 1994) and originally retrotransposons (Weiner et al., 1986). Although apo-4 appears to have the unusual property that all of its internal stop codons are not recognised, only two major and a possible third minor transcript are produced from the gene indicating that there is some element of active transcriptional and translational control. In bacteriophage λ transcription can continue by virtue of changed secondary structure under conditions of low tryptophan; if a 2-3 stem-loop forms rather than a 3-4 stem-loop transcription will continue (Yanofsky 1981). Secondary structure may also contribute to the readthrough of apo-4 stop codons. All of these activities may be controlled by the inversion which could enhance the production of normal and mutated dystrophin. If the inversion does act as a bi-directional enhancer it may help explain the exon-skipping observed with other dystrophin mutations (Nicholson et al. 1992; Sherratt et al. 1993). To test this,

it will be important to put the sequence next to other genes and examine its ability to control transcription and/or translation.

To summarise the findings on apo-4 it is important to look at the evidence for its expression in COS cells and in cell lines, most of which is known from K562. This study suggests that in a COS cell, apo-4 produces major bands at 55Kd and 50Kd which can be blocked by peptide (Figures 6.22-6.23). p50 has an extracellular N-terminus as shown by fluorescent staining and it is likely that the p55 does as well, although fluorescent staining will need to be done to prove this. It is also possible for Fc-CD33 to bind p50 *in vitro*, given optimal transfection conditions. In two separate immunoprecipitation experiments on labelled K562, Fc-CD33 appeared to bind a 50Kd doublet in the same region as purified anti-apo-4 P1p antisera, which was almost completely blocked in one experiment (Figure 6.22) and partially blocked in another (Figure 6.23). In the first experiment, P1p failed to precipitate the 50Kd band in transfectants although it could precipitate the bands in K562. Crude anti-apo-4 could precipitate a band in both transfectants and K562, but it was more difficult to block the crude antisera (Figure 6.23). Immunohistochemistry showed that crude antisera (P1c) also gave pronounced staining on K562 which was distinct from that shown with P2c, which is not unexpected considering that P2c recognises sequence in the putative transmembrane domain (Figure 6.26III). In addition, FACS analysis on K562 using anti-apo P1c as a probe produced a significant shift of the curve above the pre-bleed control (data not shown). Although the possibility of cross-reactivity cannot be excluded with a polyclonal antisera on the cell lines, the evidence suggests that apo-4 is specifically expressed there, particularly the staining pattern in brain which is in the same areas as other apo-dystrophin proteins. Peptide sequencing of the large amounts of protein products obtainable from transfectants would allow

a definitive answer to the peptide sequence obtained from the transfected apo-4 gene, particularly which amino-acids were produced from "wobble" or readthrough. Making monoclonal antibodies against these products would also allow for greater specificity in both the immunoprecipitation and staining work to confirm the results presented here. It might be useful to ligate the 5′ end of apo-4 into the pIG1 vector and raise monoclonal antibodies against the soluble protein although if the N-terminus were only 7-9 amino acids it may restrict this avenue.

Whether or not the full-length apo-4 product can be detected in RNA from tissues has not been completely confirmed in these studies. The first 876 bp of the gene were reproducible as was the inversion in a number of RT-PCR experiments attempted, but these "overlapping clones" were not reproduced as a full-length transcript with the primers designed near the extreme ends of the gene. The finding that the inversion appears to exist in genomic DNA lends weight to the evidence that the transcript is in normal genomic DNA. Parts of other rare transcripts have been amplified by using a "nested PCR" technique in which primers at the putative ends of a clone were used on the first PCR, and the product was reamplified using another set of primers immediately inside the original ones. In this way, two overlapping clones were produced after 60 rounds of PCR (Lejeune 1994). On examining the primers used, however, it may be that the F1 primer was unstable with 1st round cDNA when used with R1 due to the stretch of eight As at its 5′ end. That the 1st 876 bp and the inversion could be produced from RNA suggests that the entire gene could be produced. A reasonable next step would be to try to reproduce the entire gene by "walking" back to the R1 primer with existing antisense oligonucleotides 3′ of the inversion in combination with primers immediately 5′ of the putative starting methionines.

The primer extension data produced two compelling pieces of evidence. First, it identified a putative transcription start site which is in the standard location predicted for use by the second or third available methionines. Second, it coincided with the start of an upstream exon identified by the GRAIL search. This exon is in the same frame as apo-4 and shares the same hydrophobic character seen in the 5′ hydrophobic promoter element GCN4 in yeast which acts as a DNA binding protein to stimulate enhancer activity and is dependent on its hydrophobic residues to do so (Drysdale 1995) as is the DNA binding protein tax which binds the basic portion of the HTLV-1 genome (Baranger et al. 1995) and the MAX protein in its binding to c-*myc* (Blackwood and Eisenman 1991). The existence of this sequence as a continuous transcript from the apo-4 gene suggests that apo-4 may have resulted from a truncated transcript which actually possesses a hydrophobic 5′UTR which could also act as a DNA binding protein to regulate transcription. It may work in concert with the other hydrophobic regions seen in apo-4 to regulate transcription and/or translation. The apo-4 gene will provide many exciting avenues for future work and it is suggested that it may be most useful as a new marker for leukemia therapy.

In future work it would be informative to cotransfect the ICAM-1 and apo-4 transcripts to determine whether enhanced CD33 ligand binding could be produced. Binding could be confirmed by blocking studies with anti-ICAM-1 antibodies. Although several attempts failed to reproduce Fc-CD33 binding to ICAM-1 in immunoprecipitation studies, *in vivo* studies using blocking antibodies against both Fc-CD33 and anti-ICAM-1 antibodies on cell lines were not attempted and these may provide some evidence of binding. It is hoped that the results presented in this thesis will offer useful insights both into basic biological processes and potential therapeutic advances in leukemia.

Chapter 8

Bibliography

Bibliography

Ahn, A. H., and Kunkel, L. (1993). The structural and functional diversity of dystrophin. *Nature Genetics*, **3**: 283-291.

Alexander, E. L., and Sanders, S. K. (1977). F(ab')2 reagents are not required if goat, rather than rabbit, antibodies are used to detect human surface immunoglobulin. *J Immunol*, **119**: 1084-1088.

Anderson, S. J., Chou, H. S., and Loh, D. Y. (1988). A conserved sequence in the T-cell receptor beta-chain promoter region. *Proc. Natl. Acad. Sci. USA*, **85**: 3551-3554.

Andrews, R. G., Singer, J.W., and Bernstein, I.D. (1989). Precursors of Colony-Forming Cells in Humans Can Be Distinguished From Colony-Forming Cells By Expression of the CD33 and CD34 Antigens and Light Scatter Properties. *J. Exp. Med.*, **169**: 1721-1731.

Andrews, R. G., Torok-Storb, B., and Bernstein, I. D. (1983). Myeloid-associated differentiation antigens on stem cells and their progeny identified by monoclonal antibodies. *Blood*, **62**: 124-132.

Andrews, R. G., Takahashi, M., Segal, G. M., Powell, J. S., Bernstein, I. D., and Singer, J. W. (1986). The L4F3 antigen is expressed by unipotent and multipotent colony-forming cells but not by their precursors. *Blood*, **68**: 1030-1035.

Appelbaum, F. R., Matthews, D.C., Eary, J.F., Badger, C.C., Kellogg, M., Press, O.W., Martin, P.J., Fisher, D.R., Nelp, W.B., and Thomas, E.D. (1992). The use of radiolabelled anti-CD33 antibody to augment marrow irradiation prior to marrow transplantation for acute myelogenous leukemia. *Transplantation*, **54**: 829-833.

Aruffo, A., Stamenkovic, I., Melnick, M., Underhill, C.B., and Seed, B. (1990). CD44 Is the Principal Cell Surface Receptor for Hyaluronate. *Cell*, **61**: 1303-1313.

Aruffo, A., and Seed, B. (1987). Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Nat'l. Acad. Sci. USA.*, **84**: 8573-8577.

Ashman, L. K., Cambareri, A.C., Bik To, L., Levinsky, R.J., and Juttner, C.A. (1991). Expression of the YB5.B8 Antigen (c-*kit* proto-oncogene product) in Normal Human Bone Marrow. *Blood*, **78**: 30-37.

Atkins, J. F., Weiss, R. B., Thompson, S., and Gesteland, R. F. (1991). Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frame shifts and hops. *Annu. Rev. Genet.*, **25**: 201-228.

Ausubel, F. M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1993). In Current Protocols in Molecular Biology, (New York: John Wiley & Sons, Inc.).

Bain, B. J. (1990). *Leukemia diagnosis: A Guide to the FAB Classification*. 116. Gower Medical Publishing.

Bambrick, L. L., and Braun, P. E. (1991). Phosphorylation of myelin-associated glycoprotein in cultured oligodendrocytes. *Dev. Neurosci.*, **13**: 412-416.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β–globin gene is enhanced by remote SV40 DNA sequences. *Cell*, **27**: 299-308.

Baranger, A. M., Palmer, C. R., Hamm, M. K., Giebler, H. A., Brauweiler, A., Nyborg, J. K., and Schepartz, A. (1995). Mechanism of DNA-binding enhancement by the human T-cell leukaemia virus transactivator Tax. *Nature*, **376**: 606-608.

Barber, E. K., Dasgupta, J.D., Shlossman, S.F., Trevillyan, J.M., and Rudd, C.E. (1989). The CD4 and CD8 antigens are coupled to a protein tyrosine-kinase ($p56^{lck}$) that phosphorylates the CD3 complex. *Proc. Natl. Acad. Sci. USA.*, **6**: 3277-3281.

Barber, E. K., and Simmons, D.L. (1993). Functional-analysis of the early hematopoietic surface-markers CD33 and CD34. *J. Cell. Biochem.*, **17a**: 346.

Barclay, A. N., Birkeland, M.L., Brown, M.L., Beyers, A.D., Davis, S.J., Somoza, C., and Williams, A.F. (1993). The Leucocyte Antigen Facts Book. 424. Academic Press Limited.

Bass, B. L., and Weintraub, H. (1987). A developmentally regulated activity that unwinds RNA duplexes. *Cell*, **48**: 607-613.

Baumhueter, S., Singer, M.S., Henzel, W., Hemmerich, S., Renz, M., Rosen, S.D., and Lasky, L.A. (1993). Binding of L-Selectin to the Vascular Sialomucin CD34. *Science*, **262**: 436-438.

Berendt, A. R., McDowall, A., Craig, A. G., Bates, P. A., Sternberg, M. J., Marsh, K., Newbold, C. I., and Hogg, N. (1992). The binding site on ICAM-1 for Plasmodium falciparum-infected erythrocytes overlaps, but is distinct from, the LFA-1-binding site. *Cell*, **68**: 71-81.

Bevilacqua, M. P., Stengelin, S, Gimbrone, M.J. and Seed, B. (1989). Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complement regulatory proteins and lectins. *Science.*, **243**: 1160-1165.

Bishop, J. M. (1983). Cellular oncogenes and retroviruses. *Annu. Rev. Biochem.*, **52**: 301-354.

Blackwood, E. M., and Eisenman, R. N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science*, **251**: 1211-1217.

Blake, D. J., Tinsley, J.M., and Davies, K.E. (1994). The emerging family of dystrophin-related proteins. *Tr. in Cell Biol.*, **4**: 19-23.

Boyle, W. J., Van der Geer, P., and Hunter, T. (1991). Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Methods Enzymol.*, **201**: 110-149.

Braesch-Anderson, S., and Stamenkovic, I. (1994). Sialylation of the B-Lymphocyte Molecule CD22 by α2,6-Sialyltransferase Is Implicated in the Regulation of CD22-mediated Adhesion. *J. Biol. Chem.*, **269**: 11783-11786.

Braun, R. E. (1990). Temporal translational regulation of the protamine 1 gene during mouse spermatogenesis. *Enzyme*, **44**: 120-128.

Britten, R.J. (1994). Evidence that most human *Alu* sequences were inserted in a process that ceased about 30 million years ago. *Proc. Nat'l. Acad. Sci.* **91**: 6148-6150.

Brown, S. C., and Lucy, J.A. (1993). Dystrophin as a Mechanochemical Transducer in Skeletal Muscle. *BioEssays*, **15**: 413-419.

Bruggemann, M., Winter, G., Waldmann, H., and Neuberger, M.S. (1989). The Immunogenicity of Chimeric Antibodies. *J. Exp. Med.*, **170**: 2153-2157.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.*, **212**: 563-78.

Burton, D. R., Jefferis, R., Partridge, L. J., and Woof, J. M. (1988). Molecular recognition of antibody (IgG) by cellular Fc receptor (FcRI). *Mol. Immunol*, **25**: 1175-1181.

Byers, T. J., Licov, H.G.W. and Kunkel, L.M. (1993). An alternative dystrophin transcript specific to peripheral nerve. *Nature Genet.*, 4: 77-81.

Capon, E. J., Chamow, S.M., Mordenti, J., Marsters, S.A., Gregory, T., Mitsuya, H., Byrn, R.A., Lucas, C., Wurm, F.M., Groopman, J.E., Broder, S., and Smith, D.H. (1989). Designing CD4 immunoadhesins for AIDS therapy. *Nature*, **337**: 525-531.

Caron, P. C., Schwartz, M.A., Co, M.S., Queen, C., Finn, R.D., Graham, M.C., Divgi, C.R., Larson, S.M., and Scheinberg, D.A. (1994). Murine and humanized constructs of monoclonal-antibody m195 (anti-CD33) for the therapy of acute myelogenous leukemia. *Cancer*, **73**: 1049-1056.

Chelly, J., Gilgencrantz, H., Lambert, M., Hamard, G., Chafey, P., Recan, D., Katz, P., de la Chapelle, A., Koenig, M., Ginjaar, I. B., et al. (1990). Effect of dystrophin gene deletions on mRNA levels and processing in Duchenne and Becker muscular dystrophies. Cell **63**: 1239-1248.

Chelly, J., Kaplan, J. C., Maire, P., Gautron, S., and Kahn, A. (1988). Transcription of the dystrophin gene in human muscle and non-muscle tissue. *Nature*, **333**: 858-860.

Chelly, J., Montarras, D., Pinset, C., Berwald, N. Y., Kaplan, J. C., and Kahn, A. (1990). Quantitative estimation of minor mRNAs by cDNA-polymerase chain reaction. Application to dystrophin mRNA in cultured myogenic and brain cells. *Eur. J. Biochem*, **187**: 691-698.

Chirgwin, J. M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*, **18**: 5294 - 5299.

Church, G.M., and Gilbert, W. (1984). Genomic Sequencing. *Proc. Natl. Acad. Sci. USA*, **8**: 1991-1995.

Civin, C. I. (1990). Human Monomyeloid Cell Membrane Antigens. *Exp. Hematol.*, **18**: 461-467.

Claros, M. G., and von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *CABIOS*, **10**: 685-686.

Cockerill, G. W., Meyer, G., Noack, L., Vadas, M. A. and Gamble, J. R. (1994). Characterization of a spontaneously transformed human endothelial cell line. *Lab Invest.*, **71**: 497-509.

Cole, S. R., Ashman, L. K., and Ey, P. L. (1987). Biotinylation: An Alternative to Radioiodination for the Identification of Cell Surface Antigens in Immunoprecipitates. *Mol. Immunol.*, **24**: 699-705.

Collins, S. J., Gallo, R.C., and Gallagher, R.E. (1977). Continuous growth and differentiation of human myeloid leukaemic cells in suspension culture. *Nature*, **270**: 347-349.

Croce, C. M., Erikson, J., Ar-rushdi, A., Aden, D. and Nishikura, K. (1984). Translocated *c-myc* oncogene of Burkitt lymphoma is transcribed in plasma cells and repressed in lymphoblastoid cells. *Proc. Natl. Acad. Sci. USA*, **81**: 3170-3174.

Crocker, P. R., Mucklow, S., Bouckson, V., McWilliam, A., Willis, A.D., Gordon, S., Milon, G., Kelm, S., and Bradfield, P. (1994). Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 imunoglobulin-like domains. *EMBO J.*, **13**: 4490-4503.

Crocker, P. R., Kelm, S., Hartnell, A., Freeman, S., Nath, D., Vinson, M., and , and Mucklow, S. (1996). Sialoadhesin and related cellular recognition molecules of the immunoglobulin superfamily. *Biochem. Soc. Trans.* **24**: 150-156.

Darnell, J., Lodish, H., and Baltimore, D. (1986). Molecular Cell Biology. 1192. Scientific American Books.

Davidson, N. O. (1993). Apolipoprotein B mRNA editing: a key controlling element targeting fats to proper tissue. *Ann. Med.*, **25**: 539-543.

Degliantoni, G., Murphy, M., Kobayashi, M., Francis, M. K., Perussia, B., and Trinchieri, G. (1985). Natural killer (NK) cell-derived hematopoietic colony-inhibiting activity and NK cytotoxic factor. Relationship with tumor necrosis factor and synergism with immune interferon. *J. Exp Med*, **162**: 1512-1530.

Dinarello, C. A. (1985). An update on human interleukin-1: from molecular biology to clinical relevance. *J. Clin. Immunol*, **5**: 287-297.

Doody, G. M., Justement, L. B., Delibrias, C. C., Matthews, R. J., Lin, J., Thomas, M. L., and Fearon, D. T. (1995). A role in B cell activation for CD22 and the protein tyrosine phosphatase SHP. *Science*, **269**: 242-244.

Dornfeld, K. J., and Livingston, D. M. (1992). Plasmid recombination in a rad52 mutant of Saccharomyces cerevisiae. *Genetics*, **131**: 261-76.

Downward, J., Parker, P., and Waterfield, M. D. (1984). Autophosphorylation sites on the epidermal growth factor receptor. *Nature*, **311**: 483-485.

Drysdale, C. M., Duenas, E., Jackson, B.M., Reusser, U., Braus, G.H., and Hinnebusch, A.G. (1995). The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Mol. Cell Biol.*, **15**: 1220-1233.

Dulac, C., Tropak, M. B., Cameron, C. P., Rossier, J., Marshak, D. R., Roder, J., and Le, D. N. (1992). Molecular characterization of the Schwann cell myelin protein, SMP: structural similarities within the immunoglobulin superfamily. *Neuron*, **8**: 323-334.

Dynan, W. S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell*, **35**: 79-87.

Eck, M. J., Atwell, S. K., Shoelson, S. E., and Harrison, S. C. (1994). Structure of the regulatory domains of the Src-family tyrosine kinase Lck. *Nature*, **368**: 764-769.

Eliceiri, B., Labella, T., Hagino, Y., Srivastava, A., Schlessinger, D., Pilia, G., Palmieri, G., and D'Urso, M. (1991). Stable integration and expression in mouse cells of yeast artificial chromosomes harboring human genes. *Proc Natl Acad Sci U S A*, **88**: 2179-2183.

Elices, M. J., and Goldstein, I. J. (1990). Initiation of poly-N-acetyllactosamine chain biosynthesis occurs preferentially on complex multiantennary asparagine-linked oligosaccharides. *Carbohydr. Res*, **203**: 109-118.

Ellison, J. P., Berson, B.J., and Hood, L.E. (1982). The nucleotide sequence of a human immunoglobulin Cγ1 gene. *Nuc. Acids Res.*, **10**: 4071-4079.

Emerson, S. G., Reilly, P., and Cone, R. E. (1979). Differential radiolabelling of lymphocyte membrane alloantigens and immunoglobulins: variation of H2O2 concentration during lactoperoxidase catalyzed cell surface radio-iodination. *J. Immunogenet*, **6**: 87-97.

Emery, A. E. H. (1987). Duchenne Muscular Dystrophy. *(Oxford Monographs in Medical Genetics)*, 99-102. Oxford University Press.

Erikstein, B. K., Funderud, S., Beiske, K., Aas, E. A., De, L. D. C., Blomhoff, H. K., and Smeland, E. B. (1992). Cell cycle-dependent regulation of CDw75 (beta-galactoside alpha-2,6-sialyltransferase) on human B lymphocytes. *Eur. J. Immunol*, **22**: 1149-1155.

Ervasti, J. M., and Campbell, K. P. (1991). Membrane organization of the dystrophin-glycoprotein complex. *Cell*, **66**: 1121-1131.

Fackler, M. J., Civin, C. I., Sutherland, D. R., Baker, M. A., and May, W. S. (1990). Activated Protein Kinase C Directly Phosphorylates the CD34 Antigen on Hematopoietic Cells. *J. Biol. Chem.*, **265**: 11056-11061.

Fang, F., Ahmad, S., Lei, J., Klecker, R.W., Trepel, J.B., Smithgall, T.E., and Glazer, R.I. (1993). Effect of the Mutation of Tyrosine 713 in p93c-*fes* on Its Catalytic Activity and Ability To Promote Myeloid Differentiation in K562 Cells. *Biochemistry*, **32**: 6995-7001.

Fawcett, J. (1994a). Molecular Aspects of Angiogenesis and Metastasis (Oxford).

Fawcett, J., Holness, C.L., and Simmons, D.L. (1994b). Molecular Tools for the Analysis of Cell Adhesion. *Cell Adhesion and Communication*, **2**: 275-285.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry*, **132**: 6-13.

Firtel, R. A. (1981). Multigene families encoding actin and tubulin. *Cell*, **24**, 6-7.

Freeman, S. D., Kelm S., Barber, E.K., and Crocker, P.R. (1995). Characterisation of CD33 as a new member of the sialoadhesin family of cell adhesion molecules. *Blood*, **85**: 2005-2012.

Furie, M. B., and McHugh, D.D. (1989). Migration of neutrophils across endothelial monolayers is stimulated by treatment of the monolayers with interleukin-1 or tumor necrosis factor-alpha. *J. Immunol.*, **143**: 3309-3317.

Furley, A. J., Reeves, B.R., Mizutani, S., Altass, L.J., Watt, S.M., Jacob, M.C., van den Elsen, P., Terhorst, C., Greaves, M.F. (1986). Divergent molecular phenotypes of KG1 and KG1a myeloid cell lines. *Blood*, **68**: 1101-1107.

Gangopadhyay, S.B., Sherratt, T.G., Heckmatt, J.Z., Dubowitz, V., Miller, G., Shokeir, M., Ray, P.N, Strong, P.N. and Worton, R.G. (1992). Dystrophin in frameshift deletion patients with Becker muscular dystrophy. *Am. J. Hum. Genet.* **51:** 562-570.

Garvin, A. M., Pawar, S., Marth, J. D., and Perlmutter, R. M. (1988). Structure of the murine lck gene and its rearrangement in a murine lymphoma cell line. *Mol. Cell Biol.*, **8**: 3058-3064.

Gee, C. E., Griffin, J., Sastre, L., Miller, L. J., Springer, T. A., Piwnica, W. H., and Roberts, T. M. (1986). Differentiation of myeloid cells is accompanied by increased levels of pp60*c-src* protein and kinase activity. *Proc. Natl. Acad. Sci. USA*, **83**: 5131-5135.

Gellert, M. (1992). Molecular analysis of V(D)J recombination. *Annu. Rev. Genet.*, **26**: 425-446.

Genetics Computer Group, (G. C. G.) (1991). Program Manual for the GCG Package, Version 7. Genetics Computer Group.

Gerondakis, S., Cory, S., and Adams, J.M. (1984). Translocation of the *myc* cellular oncogene to the immunoglobulin heavy chain locus in murine plasmacytomas is an imprecise reciprocal exchange. *Cell*, **36**: 973-982.

Gimbrone, M. A., Cotran, R. S., and Folkman, J. (1974). Human vascular endothelial cells in culture: Growth and DNA synthesis. *J. Cell Biol.*, **60**: 673 - 674.

Giordano, S., Di, R. M., Cirillo, D., Naldini, L., Chiado'Piat, L., and Comoglio, P. M. (1987). Proteins phosphorylated on tyrosine as markers of human tumor cell lines. *Int. J. Cancer*, **39**: 482-487.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. *Cell*, **23**: 175-182.

Gorecki, D. C., Monaco, A. P., Derry, J. M., Walker, A. P., Barnard, E. A., and Barnard, P. J. (1992). Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. *Hum. Mol. Genet*, **1**: 505-10.

Gorman, C. M., Rigby, P. W., and Lane, D. P. (1985). Negative regulation of viral enhancers in undifferentiated embryonic stem cells. *Cell*, **42**: 519-26.

Gregersen, J. P., Mehdi, S., Gelderblom, H., and Zettlmeissl, G. (1990). A CD4:Immunoglobulin fusion protein with antiviral effects against HIV. *Arch. Virol.*, **111**: 29-43.

Griffin, J. D., and, Lowenberg B. (1986). Clonogenic cells in acute myeloblastic leukemia. *Blood*, **68**: 1185-1189.

Griffin, J. D., Linch, D., Sabbath, K., Larcom, P., and Schlossman, S. F. (1984). A monoclonal antibody reactive with normal and leukemic human myeloid progenitor cells. *Leuk. Res*, **8**: 521-534.

Guigo, R., Knudsen, S., Drake N., and Smith, T. (1992). Prediction of Gene Structure. *J.Mol.Biol.*, **226**: 141-157.

Hamada, H., and Kakunaga, T. (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature*, **298**: 396-398.

Hames, B. D., and Glover, D.M. (1988). Transcription and Splicing. In Frontiers in Molecular Biology, (Oxford: IRL Press), pp. 212.

Hanahan, D. (1983). Studies on Transformation of *Escherichia Coli* with Plasmids. *J. Mol. Biol.*, **100**: 557-580.

Hanasaki, K., Varki, A. and Powell, L.D. (1995). CD22-mediated Cell Adhesion to Cytokine-activated Human Endothelial Cells. *J. Biol. Chem.*, **270**: 7533-7542.

Handgretinger, R., Schafer, H.-J., Faur, F., Frank, D., Ottenlinger, C., Buhring, H.-J., and Niethammer, D. (1993). Expression of an early myeopoietic antigen (CD33 on a subset of human umbilical cord blood-derived natural killer cells. *Immunol. Letts.*, **37**: 223-228.

Hanke, J.H., Hambor, J.E., and Kavathas, P. (1995). Repetitive *Alu* Elements form a Cruciform Structure that Regulates the Function of the Human CD8α T Cell-specific Enhancer. *J. Mol. Biol.*, **246**: 63-73.

Harlow, E., and Lane, D. (1988). Antibodies: a Laboratory Manual. 723. Cold Spring Harbor Laboratory.

Hatada, M. H., Lu, X., Laird, E. R., Green, J., Morgenstern, J. P., Lou, M., Marr, C. S., Phillips, T. B., Ram, M. K., Theriault, K., Zoller, M. J., and Karas, J. L. (1995). Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. *Nature*, **377**: 32-38.

Hauer, C., Getty, R., and Tykocinski, M. (1989). Epstein-Barr virus episome-based promoter function in human myeloid cells. *Nucleic Acids Res.*, **5**: 1989-2003.

Hay, R., Caputo J., Chen T.R., Macy M., McClintock P., and Reid Y. (1992). American Type Culture Collection (ATCC) Catalogue of Cell Lines & Hybridomas, (Maryland).

Hesketh, R. (1995). The Oncogene Facts Book. In FactsBook Series, (London: Academic Press.), pp. 370.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol*, **26**: 365-369.

Hnatowich, D. J., Virzi, F., and Rusckowski, M. (1987). Investigations of avidin and biotin for imaging applications. *J. Nucl. Med.*, **28**: 1294-1302.

Hollenbaugh, D., Grosmaire, L. S., Kullas, C. D., Chalupny, N. J., Braesch, A. S., Noelle, R. J., Stamenkovic, I., Ledbetter, J. A., and Aruffo, A. (1992). The human T cell antigen gp39, a member of the TNF gene family, is a ligand for the CD40 receptor: expression of a soluble form of gp39 with B cell co-stimulatory activity. *EMBO J.*, **11**: 4313-4321.

Holness, C., and Simmons, D. L. (1993). Molecular Cloning of CD68, a Human Macrophage Marker Related to Lysosomal Glycoproteins. *Blood*, **81**: 1607-1613.

Honigman, A., Wolf, D., Yaish, S., Falk, H., and Panet, A. (1991). cis Acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. *Virology*, **183**: 313-319.

Hugnot, J. P., Gilgenkrantz, H., Vincent, N., Chafey, P., Morris, G.E., Monaco, A.P., Berwald-Netter, Y., Koulakoff, A., Kaplan, J.C., Kahn, A., and Chelly, J. (1992). Novel products of the dystrophin gene: A distal transcript initiated from a unique alternative first exon encoding a 75 kDa protein widely distributed in non-muscle tissues. *Proc. Natl. Acad. Sci. USA*, **89**: 7506-7510.

Huhn, R. D., Posner, M. R., Rayter, S. I., Foulkes, J. G., and Frackelton, A. J. (1987). Cell lines and peripheral blood leukocytes derived from individuals with chronic myelogenous leukemia display virtually identical proteins phosphorylated on tyrosine residues. *Proc .Natl. Acad. Sci. USA*, **84**: 4408-4412.

Imai, Y., Lasky, L. A. and Rosen, S. D. (1993). Sulphation requirement for GlyCAM-1, an endothelial ligand for L-selectin. *Nature.*, **361**: 555-557.

Jackson, D. G., Buckley, J., and Bell, J. I. (1992). Multiple Variants of the Human Lymphocyte Homing Receptor CD44 Generated by Insertions at a Single Site in the Extracellular Domain. *J. Biol. Chem.*, **267**: 4732-4739.

Jackson, R. J. (1993). Cytoplasmic regulation of mRNA function: the importance of the 3′ untranslated region. *Cell*, **74**: 9-14.

Jaffe, E. A., Nachman, R. L., Becker, C. G., and Minick, C. R. (1973). Culture of human endothelial cells derived from umbilical veins - identification by morphological and immunological criteria. *J. Clin. Inves.*, **52**: 2745-2756.

Jaramillo, M. L., Afar, D. E., Almazan, G., and Bell, J. C. (1994). Identification of tyrosine 620 as the major phosphorylation site of myelin-associated glycoprotein and its implication in interacting with signaling molecules. *J. Biol. Chem*, **269**: 27240-27245.

Jeang, K. T., and Hayward, G. S. (1983). A cytomegalovirus DNA sequence containing tracts of tandemly repeated CA dinucleotides hybridizes to highly repetitive dispersed elements in mammalian cell genomes. *Mol. Cell. Biol.*, **3**: 1389-1391.

Jennings, M. W., Jones, R. W., Wood, W. G., and Weatherall, D. J. (1985). Analysis of an inversion within the human beta globin gene cluster. *Nucleic Acids Res..*, **13**: 2897-906.

Jones, R. W., Old, J. M., Trent, R. J., Clegg, J. B., and Weatherall, D. J. (1981). Major rearrangement in the human beta-globin gene cluster. *Nature*, **291**: 39-44.

Kamata, T., Wright, R., and Takada, Y. (1995). Critical threonine and aspartic acid residues within the I domains of beta 2 integrins for interactions with intercellular adhesion molecule 1 (ICAM-1) and C3bi. *J. Biol. Chem*, **270**: 12531-12535.

Kawada, H., Ichikawa, Y., Kobayashi, N., Fukuda, R., Yonekura, S., and Nagao, T. (1993). Acquisition of CD13 and CD33 expression at relapse on acute myeloid-leukemia cells with an unusual phenotype - MPO+CD13-CD33-. *Internal Med.*, **32**: 733-736.

Kelm, S. (1994b). Modifications of cell surface sialic acids modulate cell adhesion mediated by sialoadhesin and CD22. *Glycoconj. J.*, **11**: 576-585.

Kelm, S., Pelz, A., Schauer, R., Filbin, M. Tang, S., Bellard, M.E.D., Schnaar, R., Mahoney, J., Hartnell, A., Bradfield, P., and Crocker, P. (1994a). Sialoadhesin, MAG, and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. *Current Biology*, **4**: 965-972.

Kienhuis, C. B., Heuvel, J. J., Ross, H. A., Swinkels, L. M., Foekens, J. A., and Benraad, T. J. (1991). Six methods for direct radioiodination of mouse epidermal growth factor compared: effect of nonequivalence in binding behavior between labeled and unlabeled ligand. *Clin Chem.*, **37**: 1749-1755.

Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, J., and Playfair, J. H. (1987). Expression and rescuing of a cloned human tumour necrosis factor gene using an EBV-based shuttle cosmid vector. *Embo J*, **6**: 355-361.

Kirchhoff, F., and Hofer, H. W. (1993). Myelin-associated glycoprotein is phosphorylated by protein-kinase C. *J. Neurosci. Res.*, **36**: 368-381.

Kloetzer, W., Kurzrock, R., Smith, L., Talpaz, M., Spiller, M., Gutterman, J., and Arlinghaus, R. (1985). The human cellular *abl* gene product in the chronic myelogenous leukemia cell line K562 has an associated tyrosine protein kinase activity. *Virology*, **140**: 230-238.

Koeffler, H. P. (1983). Induction of Differentiation of Human Acute Myelogenous Leukemia Cells: Therapeutic Implications. *Blood*, **62**: 709-721.

Koeffler, H. P., Billing, R., Lusis, A.J., Sparkes, R., and Golde, D.W. (1980). An Undifferentiated Variant Derived From the Human Acute Myelogenous Leukemia Cell Line (KG-1). *Blood*, **56**: 265-273.

Koeffler, P. H., and Golde, D. W. (1978). Acute Myelogenous Leukemia: A Human Cell Line Responsive to Colony-Stimulating Activity. *Science*, **200**: 1153-1154.

Koenig, M., Beggs, A.H., Moyer, M., Scherpf., S., Heindrich, K., Bettecken, T., Meng, G., et al., (1989). The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. *Am. J. Hum. Genet.*, **45**: 498-506.

Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C., and Kunkel, L. M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*, **50**: 509-517.

Koenig, M., Monaco, A. P., and Kunkel, L. M. (1988). The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell*, **53**: 219-226.

Koller, U., and Peschel, C.H. (1989). Cluster report: CD33. In Leucocyte Typing IV: White Cell Differentiation Antigens, (Oxford: Oxford University Press), pp. 812-813.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*, **15**: 8125-8148.

Kudo, A., and Melchers, F. (1987). A second gene, VpreB in the lambda 5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. *EMBO J*, **6**: 2267-2272.

Kudo, A., Pravtcheva, D., Sakaguchi, N., Ruddle, F. H., and Melchers, F. (1987). Localization of the murine lambda 5 gene on chromosome 16. *Genomics*, **1**: 277-279.

Kunkel, L. M., Lalande, M., Monaco, A. P., Flint, A., Middlesworth, W., and Latt, S. A. (1985). Construction of a human X-chromosome-enriched phage library which facilitates analysis of specific loci. *Gene*, **33**: 251-258.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**: 680-685.

Landau, N. R., St, J. T., Weissman, I. L., Wolf, S. C., Silverstone, A. E., and Baltimore, D. (1984). Cloning of terminal transferase cDNA by antibody screening. *Proc. Natl. Acad. Sci. USA*, **81**: 5836-5840.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA*, **88**: 4123-4127.

Lederfein, D., Levy, Z., Augier, N., Mornet, D., Morris, G., Fuchs, O., Yaffe, D., and Nudel U. (1992). A 71-kilodalton protein is a major product of the Duchenne muscular dystrophy gene in brain and other nonmuscle tissues. *Proc. Natl. Acad. Sci. USA*, **89**: 5346-5350.

Lehmann, J. M., Riethmuller, G., and Johnson, J.P. (1989). MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily. *Proc. Natl. Acad. Sci. USA.*, **86**: 9891-9895.

Lejeune, S., Huguet, E.L., Hamby, A., Poulsom, R. and Haris, A. (1994). *Wnt5a* Cloning, Expression, and Up-regulation in Human Primary Breast Cancers. *Clin. Cancer Res.*, **1**: 215-222.

Lewis, S., Gifford, A., and Baltimore, D. (1985). DNA elements are asymmetrically joined during the site-specific recombination of kappa immunoglobulin genes. *Science*, **228**: 677-85.

Lidov, H. G., Byers, T.J., Watkins, S.C. and Kunkel, L.M. (1990). Localization of dystrophin to postsynaptic regions of central nervous system cortical neurons. *Nature*, **348**: 725-728.

Lozzio, C. B., and Lozzio, B.B. (1975). Human Chronic Myelogenous Leukemia Cell-Line With Positive Philadelphia Chromosome. *Blood*, **45**: 321-334.

Majdic, O. (1989). Cluster report: CD64. In Leucocyte Typing IV: White Cell Differentiation Antigens, (Oxford: Oxford University Press), pp. 835-836.

Majors, J. E., Swanstrom, R., DeLorbe, W. J., Payne, G. S., Hughes, S. H., Ortiz, S., Quintrell, N., Bishop, J. M., and Varmus, H. E. (1981). DNA intermediates in the replication of retroviruses are structurally (and perhaps functionally) related to transposable elements. *Cold Spring Harb. Symp. Quant. Biol,*, **2**: 731-738.

Marth, J. D., Peet, R., Krebs, E.G., and Perlmutter, R.M. (1985). A lymphocyte-specific protein-tyrosine kinase gene is rearranged and overexpressed in the murine T cell lymphoma LSTRA. *Cell*, **80**: 80-87.

Martin, P. J. and Papayannopoulou, T. (1982). A New Human Erythroleukemia Cell Line with Spontaneous and Induced Globin Expression. *Science.*, **216**: 1233-1235.

Matsuoka, M., Nagawa, F., Okazaki, K., Kingsbury, L., Yoshida, K., Muller, U., Larue, D. T., Winer, J. A., and Sakano, H. (1991). Detection of somatic DNA recombination in the transgenic mouse brain. *Science*, **254**: 81-86.

Matutes, E., Rodriguez, B., Polli, N., Tavares de Castro, J., Parreira, A., Andrews, C., Griffin, J. D., Tindle, R. W., and Catovsky, D. (1985). Characterization of myeloid leukemias with monoclonal antibodies 3C5 and MY9. *Hematol Oncol*, **3**: 179-86.

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harbor Symposium*, **21**: 197-216.

McCourt, P. A. G., Ek, B., Forsberg, N., and Gustafson, S. (1994). Intercellular Adhesion Molecule-1 Is a Cell Surface Receptor for Hyaluronan. *J. Biol. Chem.*, **269**: 30081-30084.

Melton, D. A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984). Efficient in vitro synthesis of biologically active RNA land RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucl. Acids Res.*, **12**: 7035-7056.

Miesfeld, R., Krystal, M., and Arnheim, N. (1981). A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ- and β- globin genes. *Nucleic Acids Res.*, **9**: 5931-5947.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., Kunkel, L.M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics*, **2**: 90-95.

Monaco, A. P., Walker, A. P., Millwood, I., Larin, Z., and Lehrach, H. (1992). A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene. *Genomics*, **12**: 465-473.

Moore, K. L., Eaton, S. F., Lyons, D. E., Lichenstein, H. S., Cummings, R. D. and McEver, R. P. (1994). The P-selectin glycoprotein ligand from human neutrophils displays sialylated, fucosylated, O-linked poly-N-acetyllactosamine. *J . Biol. Chem.*, **269**: 23318-22327.

Murgola, E. J., Prather, N. E., Mims, B. H., Pagel, F. T., and Hijazi, K. A. (1983). Anticodon shift in tRNA: a novel mechanism in missense and nonsense suppression. *Proc. Natl. Acad. Sci. USA*, **80**: 4936-4939.

Muta, T., Kurosaki, T., Misulovin, Z., Sanchez, M., Nussenzweig, M. C., and Ravetch, J. V. (1994). A 13-amino-acid motif in the cytoplasmic domain of Fc gamma RIIB modulates B-cell receptor signalling [published erratum appears in Nature 1994 May 26;369(6478):340]. *Nature*, **368**: 70-73.

Nakamura, Y., Noma, M., Kidokoro, M., Kobayashi, N., Takei, M., Kurashima, S., Mukaiyama, T., and Kato, S. (1994). Expression of CD33 antigen on normal human activated T-lymphocytes. *Blood.*, **83**: 1442-1443.

Naldini, L., Stacchini, A., Cirillo, D. M., Aglietta, M., Gavosto, F., and Comoglio, P. M. (1986). Phosphotyrosine antibodies identify the p210c-abl tyrosine kinase and proteins phosphorylated on tyrosine in human chronic myelogenous leukemia cells. *Mol. Cell Biol.*, **6**: 1803-1811.

Neame, P. B., Soamboonsrup, P., Browman, G.P., Meyer, R.M., Benger, A., Wilson, W.E., Walker, I.R., Saeed, N., and McBride, J.A. (1986). Classifying acute leukemia by immunophenotyping: A combined FAB-immunologic classification of AML. *Blood*, **68**: 1355-1362.

Neuberger, M. S., Williams, G.T., Mitchell, E.B., Jouhal, S.S., Flanagan, J.G., and Rabbitts, T.H. (1985). A hapten-specific chimaeric IgE antibody with human physiological effector function. *Nature*, **314**: 268-270.

Nicholson, L.V., Bushby, K.M., Johnson, M.A., den Dunnen, J.T., Ginjaar, I.B., van Ommen, G.J. (1992). Predicted and observed sizes of dystrophin in some patients with gene deletions that disrupt the open reading frame. *J. Med. Genet.* **29:** 892-896.

Norgard, S. K., Varki, N. M. and Varki, A. (1993). Calcium-dependent heparin-like ligands for L-selectin in nonlymphoid endothelial cells. *Science.*, **261**: 480-483.

O'Connor, R., Torigoe, T., Reed, J. C., and Santoli, D. (1992). Phenotypic changes induced by interleukin-2 (IL-2) and IL-3 in an immature T-lymphocytic leukemia are associated with regulated expression of IL-2 receptor beta chain and of protein tyrosine kinases LCK and LYN. *Blood*, **80**: 1017-1025.

Osborn, L., Hessian, C., Tizard, R., Vasallo, C., Lubowskyj, S., Chi-Rosso, G., and Lobb, R. (1989). Direct expression cloning of VCAM 1, a cytokine induced endothelial protein that binds to lymphocytes. *Cell.*, **59**: 1203-1211.

Oxley, S. M., and Sackstein, R. (1994). Detection of an L-selectin ligand on a hematopoietic progenitor cell line. *Blood*, **84**: 3299-3306.

Pani, G., Kozlowski, M., Cambier, J.C., Mills, G.B., and Siminovitch, K.A. (1995). Identification of the tyrosine phosphatase PTP1C as a B cell antigen receptor-associated protein involved in the regulation of B cell signalling. *J. Exp. Med.*, **181**: 2077-2084.

Patel, V. P., and Lodish, H. F. (1987). A Fibronectin Matrix Is Required for Differentiation of Murine Erythroleukemia Cells into Reticulocytes. *J. Cell Biol.*, **105**: 3107-3118.

Patel, V. P., and Lodish, H. F. (1984). Loss af adhesion of murine erythroleukemia cells to fibronectin during erythroid differentiation. *Science*, **224**: 996-998.

Payne, G. S., Bishop, J. M., and Varmus, H. E. (1982). Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. *Nature*, **295**: 209-214.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, **85**: 2444-2448.

Peiper, S. C., Ashmun, R.A., and Look, T.A. (1988). Molecular Cloning, Expression, and Chromosomal Localization of a Human Gene Encoding the CD33 Myeloid differentiation Antigen. *Blood.*, **72**: 314-321.

Peiper, S. C., Leboeur, R.D., Hughes, C.B., Prasthofer, E.F., Borowitz, M.J., Dewutter-Dambuyant, C., Katz, D.R., Walker, W.S., Ashmun, R.A., and Look, A.T. (1989). Report on the CD33 cluster Workshop: biochemical and genetic characterization of gp67. In Leucocyte Typing IV: White Cell Differentiation Antigens, (Oxford: Oxford University Press), pp. 814-816.

Peters, J. H., Gieseler, R., Thiele, B., and Steinbach, F. (1996). Dendritic cells: from ontogenetic orphans to myelomonocytic descendants. *Immunol. Today*, **17**: 273-278.

Piller, V., Piller, F. and Fukuda, M. (1989). Phosphorylation of the major leukocyte surface sialoglycoprotein, leukosialin, is increased by phorbol 12-myristate 13-acetate. *J. Biol. Chem.*, **264**: 18824-18831.

Powell, L. D., and Varki, A. (1995). I-type lectins. *J. Biol. Chem.*, **270**: 14243-14246.

Prasad, K. V., Janssen, O., Kapeller, R., Raab, M., Cantley, L. C., and Rudd, C. E. (1993). Src-homology 3 domain of protein kinase p59fyn mediates binding to phosphatidylinositol 3-kinase in T cells. *Proc. Natl. Acad. Sci. USA*, **90**: 7366-7370.

Purchio, A. F., Erikson, E., Brugge, J.S., and Erikson, R.L. (1978). Identification of a polypeptide encoded by the avian sarcoma virus *src* gene. *Proc. Nat'l. Acac. Sci. USA*, **75**: 1567-1571.

Rabbitts, T. H. (1991). Translocations, Master Genes, and Differences between the Origins of Acute and Chronic Leukemias. *Cell*, **67**: 641-644.

Rapaport, D., Fuchs, O., Nudel, U., and Yaffe, D. (1992). Expression of the Duchenne muscular dystrophy gene products in embryonic stem cells and their differentiated derivatives. *J. Biol. Chem.*, **267**: 21289-21292.

Rastinejad, F., and Blau, H. M. (1993). Genetic complementation reveals a novel regulatory role for 3′ untranslated regions in growth and differentiation. *Cell*, **72**: 903-917.

Rastinejad, F., Conboy, M. J., Rando, T. A., and Blau, H. M. (1993). Tumor suppression by RNA from the 3′ untranslated region of alpha-tropomyosin. *Cell*, **75**: 1107-1117.

Rearden, A., Taetle, R., Elmajian, D.A., Majda, J. and Baird, S.M. (1985). Glycophorin A on Normal and Leukemia Cells Detected by Monoblonal Antibodies, Including a New Monoclonal Antibody Reactive with Glycophorins A and B. *Mol. Immunol.*, **22**: 369-378.

Reth, M. (1989). Antigen receptor tail clue [letter]. *Nature*, **338**: 383-384.

Richardson, J. M., Morla, A. O., and Wang, J. Y. (1987). Reduction in protein tyrosine phosphorylation during differentiation of human leukemia cell line K-562. *Cancer Res*, **47**: 4066-4070.

Ritz, J., Takvorian, T., Sallan, S.E., Fredman, A.S., Anderson, K., Coral, F., Schlossman, S.F., and Nadler, L.M. (1987). In vitro applications of monoclonal antibodies for bone marrow transplantation. In Leucocyte Typing III, Proceedings of the Third International Workshop on Human Leukocyte Differentiation Antigens, (Oxford: Oxford University Press), pp. 938.

Roberts, R. G., Barby, T. F. M., Manners, E., Bobrow, M., and Bentley, D. R. (1991). Direct detection of dystrophin gene rearrangements by analysis of dystrophin mRNA in peripheral blood lymphocytes. *Am. J. Hum. Genet.*, **49**: 298-310.

Robertson, M. J., Soiffer, R.J., Freedman, A.S., Rabinowe, S.L., Anderson, K.C., Ervin, T.J., Murray, C., Dear, K., Griffin, J.D., Nadler, L.M., and Ritz, J. (1992). Human Bone Marrow Depleted of CD33-Positive Cells Mediates Delayed but Durable Reconstitution of Hematopoiesis: Clinical Trial of MY9 Monoclonal Antibody-Purged Autografts for the Treatment of Acute Myeloid Leukemia. *Blood*, **79**: 2229-2236.

Rosenstein, Y., Park, J. K., Hahn, W. C., Rosen, F. S., Bierer, B. E. and Burakoff, S. J. (1991). CD43, a molecule defective in Wiskott-Aldrich syndrome, binds ICAM-1. *Nature.*, **354**: 233-235.

Rouer, E., Van Huynh, T., Lavareda de Souza, S., Lang, M.-C., Fischer, S. and Benarous, R. (1989). Structure of the human *lck* gene: differences in genomic organisation within src-related genes affect only N-terminal exons. *Gene*, **84**: 105-113.

Rouer, E., Dreyfus, F., Melle, J., and Benarous, R. (1994). Pattern of expression of five alternative transcripts of the *lck* gene in different hematopoietic malignancies: correlation of the level of *lck* messenger RNA I B with the immature phenotype of the malignancy. *Cell Growth Differ.*, **5**: 659-666.

Rowley, J. D. (1995). Chromosome translocations: good genes gone wrong. *Pathol. Biol. Paris*, **43**: 197-201.

Rudd, C. E., Trevillyan, J.M., Dasgupta, J.D., Wong, L.L. and Schlossman, S.F. (1988). CD4 is phosphoylated by the protein tyrosine-kinase ($p56^{lck}$). *Proc. Natl. Acad. Sci. USA*, **85**: 5190-5194.

Rudd, C. E., Barber, E. K., Burgess, K. E., Hahn, J. Y., Odysseos, A. D., Sy, M. S., and Schlossman, S. F. (1991). Molecular analysis of the interaction of p56*lck* with the CD4 and CD8 antigens. *Adv. Exp. Med. Biol.*, **292**: 85-96.

Rudd, C. E., Janssen, O., Prasad, K. V., Raab, M., da Silva, A., Telfer, J. C., and Yamamoto, M. (1993). src-related protein tyrosine kinases and their surface receptors. *Biochim. Biophys. Acta*, **1155**: 239-266.

Rusten, L. S., Smeland, E. B., Jacobsen, F. W. Lien, E., Lesslauer, W., Loetscher, H., Dubois, C. M., and Jacobsen, S. E. W. (1994). Tumor-necrosis-factor-alpha inhibits stem-cell factor-induced proliferation of human bone-marrow progenitor cells in-vitro - role of p55 and p75 tumor-necrosis-factor receptors. *J. Clin. Invest.*, **94**: 165-172.

Rutherford, T., Clegg., J.B., Higgs, D.R., Jones, R.W., Thompson, J. and Weatherall, D.J. (1981). Embryonic erythroid differentiation in the human leukemic cell line K562. *Proc. Natl. Acad. Sci. USA*, **78**: 348-352.

Sabbath, K. D., and Griffin, J.D. (1986). Differentiation-associated states of clonogenic cells in acute myeloblastic leukemia using monoclonal antibodies. In Leukocyte Typing II, Proceedings of the Second International Workshop on Human Leukocyte Differentiation Antigens, (Berlin: Springer-Verlag), p. 261.

Saksela, K., Cheng, G., and Baltimore, D. (1995). Proline-rich (PXXP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of Nef$^+$ viruses but not for down-regulation of CD4. *EMBO J.*, **14**: 484-491.

Salzer, J. L., Holmes, W. P., and Colman, D. R. (1987). The amino acid sequences of the myelin-associated glycoproteins: homology to the immunoglobulin gene superfamily. *J. Cell Biol.*, **104**: 957-965.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. (N. Ford), Cold Spring Harbor Laboratory Press.

Sandri, G. R., Goldrin, A. L., Levine, M., and Glorioso, J. (1983). High-efficiency transfer of DNA into eukaryotic cells by protoplast fusion. *Methods in Enzymology*, **101**: 402 - 411.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat'l. Acad. Sci. USA*, **74**: 5463 - 5467.

Schwarzinger, I., Valent, P., Koller, U., Marosi, C., Schneider, B., Haas, O., Knapp, W., Lechner, K., and Bettelheim, P. (1990). Prognostic significance of surface marker expression on blasts of patients with de novo acute myeloblastic leukemia. *J. Clin. Oncol.*, **8**: 423-430.

Seed, B. (1987a). An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor CD2. *Nature.*, **329**: 840-842.

Seed, B., and Aruffo, A. (1987b). Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. USA.*, **84**: 3365-3369.

Seligmann, M., Mihaesco, E., Preud'homme, J.-L., Danon, F., and Brouet, J.-C. (1979). Heavy Chain Diseases: Current Findings and Concepts. *Immunol. Rev.*, **48**: 145-167.

Sgroi, D., Varki, A. Braesh.-Anderson, S. and Stamenkovic, I. (1993). CD22, a B Cell-specific Immunoglobulin Superfamily Member, Is a Sialic Acid-binding Lectin. *J. Biol. Chem.*, **268**: 7011-7018.

Sgroi, D., Nocks, A., and Stamenkovic, I. (1996). A Single N-linked Glycosylation Site Is Implicated in the Regulation of Ligand Recognition by the I-type Lectins CD22 and CD33. *J. Biol. Chem.*, **271**: 18803-18809.

Sheen, J. Y., and Seed, B. (1988). Electrolyte gradient gels for DNA sequencing. *Biotechniques*, **6**: 942-944.

Sherratt, T.G., Vulliamy, T., Dubowitz, V., Sewry, C.A., Strong, P.N. (1993). Exon skipping and translation in patients with frameshift deletions in the dystrophin gene. *Am. J. Hum. Genet.* **53:** 1007-1015.

Simmons, D., Makgoba, M.W. and Seed, B. (1988a). ICAM, an adhesion ligand of LFA-1, is homologous to the neural cell adhesion molecule NCAM. *Nature.*, **331**: 624-627.

Simmons, D., and Seed., B. (1988b). Isolation of a cDNA encoding CD33, a differentiation antigen of myeloid progenitor cells. *J. Immunol.*, **141**: 2797-2800.

Simmons, D. L. (1993). Cloning cell surface molecules by transient expression in mammalian cells. In *Cellular Interactions in Development--A Practical Approach*, (Oxford: IRL Press), pp. 93-128.

Simmons, D. L., Satterthwaite, A.B., Tenen, D.G., and Seed, B. (1992). Molecular Cloning of a cDNA Encoding CD34, a Sialomucin of Human Hematopoietic Stem Cells. *J. Immunol.*, **148**: 267-271.

Simmons, D. L., and Needham, L. (1991). Cloning cell surface molecules using monoclonal antibodies. In *Vascular endothelium: Interactions with circulating cells.*, (Oxford: Elsevier Science Publishers B.V.), pp. 3-29.

Smale, S. T., and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell*, **57**: 103-13.

Smith, L. J., Curtis, J.E., Messner, H.A., Senn, J.S., Furthmayr, H., and McCulloch, E.A. (1983). Lineage infidelity in acute leukemia. *Blood*, **61**: 1138-1145.

Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., and Cantley, L. (1993). SH2 domains recognize specific phosphopeptide sequences. *Cell*, **72**: 767-778.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98**: 503-517.

Staden, R. (1990). Searching for patterns in protein and nucleic acid sequences. *Methods in Enzymology*, **183**: 193-211.

Stamenkovic, I., Sgroi, D., Aruffo, A., Sy, M. S. and Anderson, T. (1991a). The B lymphocyte adhesion molecule CD22 interacts with leukocyte common antigen CD45RO on T cells and alpha 2-6 sialyltransferase, CD75, on B cells [see comments]. *Cell.*, **66**: 1133-1144.

Stamenkovic, I., Amiot, M., Pesando, J. M., and Seed, B. (1989). A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. *Cell*, **56**: 1057-1062.

Stamenkovic, I., Aruffo, A., Amiot, M., and Seed, B. (1991b). The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. *EMBO J*, **10**: 343-348.

Stamenkovic, I., and Seed, B. (1990). The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion. *Nature*, **345**: 74-77.

Staunton, D., Dustin, M., and Springer, T. (1989). Functional cloning of ICAM-2: a cell adhesion-ligand for LFA-1 homologous to ICAM-1. *Nature.*, **339**: 61-64.

Sundstrom, C., and Nilsson, K. (1976). Establishment and Characterisation of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer*, **17**: 565-577.

Sussman, D. J., and Milman, G. (1984). Short-term, high-efficiency expression of transfected DNA. *Mol. Cell. Biol.*, **4**: 1641-1643.

Sutherland, D. R., Marsh, J. C., Davidson, J., Baker, M. A., Keating, A. and Mellors. A. (1992). Differential sensitivity of CD34 epitopes to cleavage by Pasteurella haemolytica glycoprotease: implications for purification of CD34-positive progenitor cells. *Exp. Hematol.*, **20**: 590-599.

Appendix

Apo-4 Primers and Structural Information

Appendix I

Additional Oligonucleotide primers used for apo-dystrophin-4 southern blotting and sequencing

FORWARD

GTT CGT TAA TAC AAG TAG	F2.3(@28)
GCC AAG GTG GAA AAG ATG	F2.2(@73)
CCA GTA GCC TGA TCC AAC	F3.2(@208)
GGC TTC ATT AAT AAG	F3.1(@257)
GGC AAA GAA ACA GAG TG	F4.2(@379)
CAG GAC ACA ATG TAG GA	F4.1(@449)
GTT ATA AAG AAA GAA TTA TAA AG	FJ^{n}(@846)
GAA AAT AAC GCA ATG GAC	F5.1(@875)

REVERSE

GAT GGG ATA CAT CTT TTC C	R6.1(@99)
CAA GCT ACA TTC AGG TTC CC	F2.2R(@188)
GGA CTC CAT CGC TCT GCC	R4.1(@510)
GAC TTA GAA ACT ACT G	R3.4(@694)
ATA GAC GTG TAA AAC CTG C	R2.1(@735)
AAC TGT TAT AAA TTT TTA	RSP2(@848)
CTT TTT CCT TTA TAA TTC TTT C	$R2.3_{O}$(@875)

Appendix II

An Additional Splice Product Predicted From The Apo-4 Gene

A second potential theoretical splice product which retains exon 78.3 is shown below.

A. H2 p1-124 spliced product =351bp, 117 amino acids + 10 from vector + 1 N-glycosylation site; predicted weight = 21.9Kd

B. Peptide Generated
MFVNTTKVEKMYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFTTPD
SRMVFIIFIQQRGLDSKSLQEINLYFCEGFYTSMQLYKKVIRKLHKITQWTRTPQ
NQSEVEIA 117

C.

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@26bp	78.1	@16-41	26bp	78.3	@42-74	33bp
	78.3	@75-181	106bp	79.1	@182-530	349bp
	79.1	@531-655	125bp	79.4	@656-721	66bp
	79.4	@722-770	49bp	79.55	@771-876	105bp
	79.55	@877-894	18bp	79.75	@895-933	39bp
	79.85	@934- 967	33bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 22- 42 1.8833

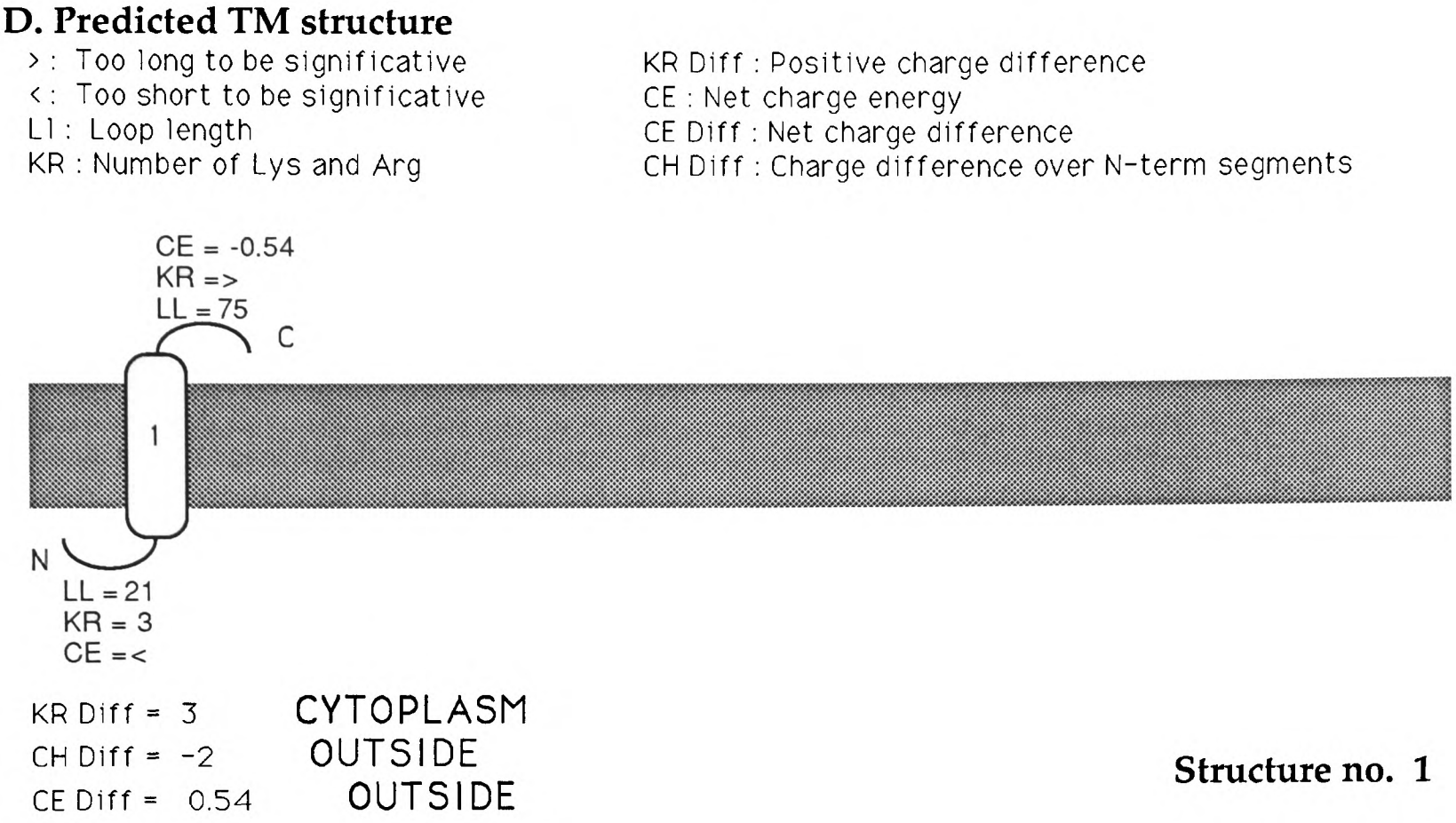


Figure 6.28. Potential apo-4 splice product p1-124. A. A model of potential splicing of apo-4 using available splice sites. **B.** Peptide structure with TM underlined; **C.** potential "new" exons identified; and **D.** TM domain structure predicted with TopPred showing a cytoplasmic N-terminus. Total spliced length, 468 bp, 156 amino acids + 10 amino acids in vector, + 1 N-glycosylation site; proposed weight = 29.5Kd.

Appendix III

Nucleic Acid Subsequence Sites Identified In Apo-4

Motif	Position	Significance
CpG	-7, (+28, +106)	DNA methylation site
CAAT	-132, (+127, +131)	Binding of CAAT factors
TATAAT (5/6)	-120, -114, (+10)	TFIID Binding site
TATA	-154	Binds RNA polymerase II and TFIID
CCATTCA	-162, -131	Cap Site I
TATCAGT	+12, (+25)	Cap Site II
TGGCTGCAAGCCCAA (10/14)	-57, (+41)	Binds CTF/NF-I protein
GTGATGG	-140, -4, +11, +32	Eucaryotic Transcription Initiation Site

Appendix IV

Top Pred predicts 4-5 transmembrane domains for a full-length apo-4F product in which all the stop codons are suppressed.

A. Protein sequence and position of predicted TM domains

```
                                        Begin TM1 (R)
                          P1                 |              P2
MFVNTSREKV INQSLIAKVE KMYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHFSMK60

TPVARSNIKL ILTNNVKWLH KKGFASSWKL VKNQTLLCTP SMQLLCCLHP EMGNDFPNGK 120
                                        P3
ETERCYLSAP FVKSVFLSLC FPGHNVGSLF HMADDLGRAM ESLVSVMTDE EGAEKMFYNS180

RFPHGFYNIH TTKRIRQKEF TRNKSIFLRR VVVLYCRFQK FLSLLLFCKQ WQVLHVYAIV 240

QKSYKKTTCK ILIAKKLAIS LYGTHFGLFK NLKQLKRKNY KGKRKKRNGQ VVKLRTQVCT 300

IIRNTPKPKR GRNSMRSRVR CKLI 324
Hydrophobicity Scale KD
```

B. Apo-4F : Candidate membrane-spanning segments:

Certain	1	33- 53	1.9073
Putative	2	93- 113	0.8052
Certain	3	124- 144	1.2552
Putative	4	209- 229	1.1833
Putative	5	246- 266	0.9240

I. Transmembrane segments included in structure 8: **1 2 3 4 5;** Loop lengths: 32 39 10 64 16 58

K+R difference: -19; -> Orientation: **N-out;** Charge-difference over N-terminal Membr. segs.

(±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < < < < -0.13 <

CYT-EXT difference: 0.13
-> Orientation: **N-out**

II. Transmembrane segments included in structure 7: **1 3 4 5;** Loop lengths: 32 70 64 16 58

K+R profile: 5 > 22 > 5; K+R difference: 22 -> Orientation: **N-in**

Charge-difference over N-terminal Membr. segs. (±15 residues): -3; -> Orientation: **N-out**

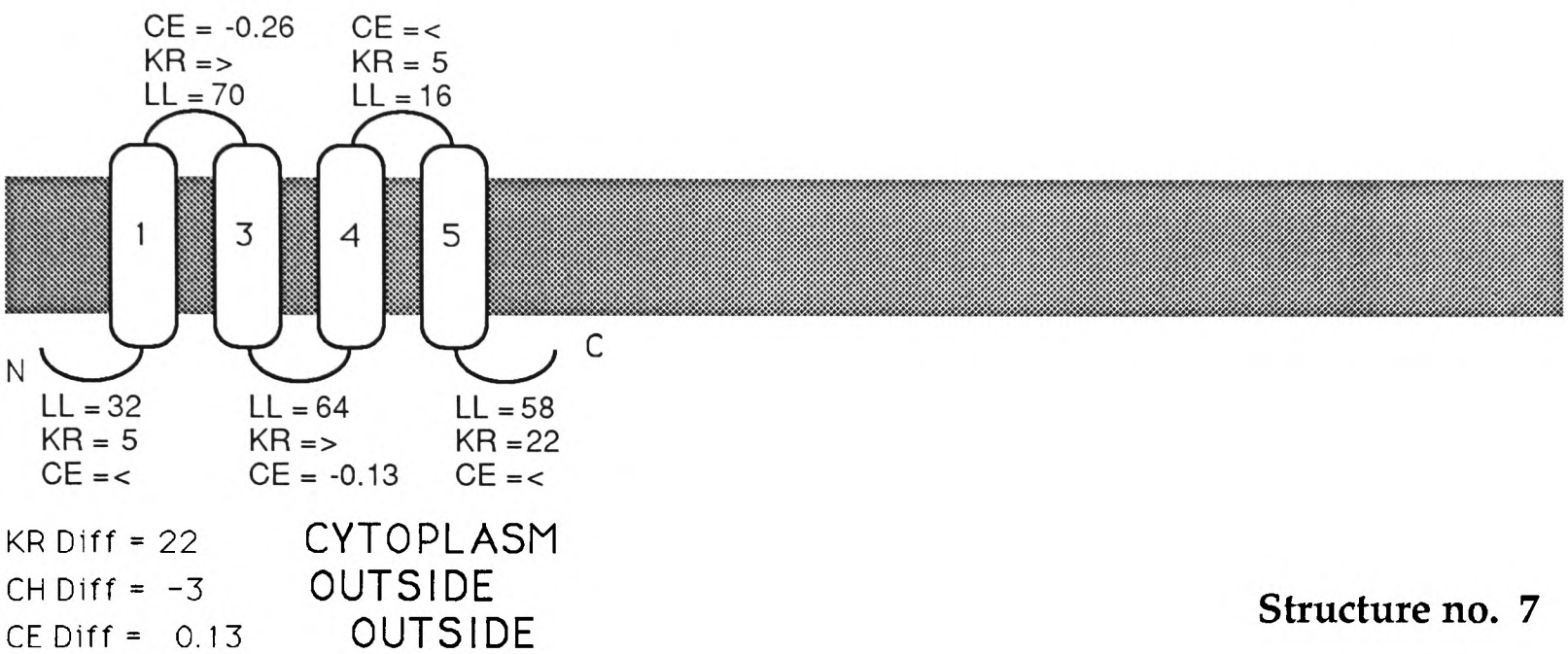
CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 < -0.26 <

CYT-EXT difference: 0.13; -> Orientation: **N-out**

C. TopPred predicts a cytoplasmic N-terminus for four TM domains

> : Too long to be significative
< : Too short to be significative
Ll : Loop length
KR : Number of Lys and Arg

KR Diff : Positive charge difference
CE : Net charge energy
CE Diff : Net charge difference
CH Diff : Charge difference over N-term segments



Structure no. 7

Figure 6.29A-C. Five transmembrane domains are predicted for the full-length apo-4F readthrough product. The amino acid sequence begins at the first available M at 9 aa, underlining the predicted TM domains and the regions against which anti-apo-4 P1, P2 and P3 peptide antisera were designed in boldface type and underlined beginning with R at +64. Amino acids in which T was changed to A are in boldface alone. **B.** A description of the predicted TM domains with those described as "certain" and "putative"; I. the N-terminus is predicted as "outside" relative to a prediction including all five TM domains; II. most predicted combinations of TMs which could be utilised predict an "outside" N-terminus, although the structure to which the program defaulted predicted a cytoplasmic N-terminus, as shown in **C.** The model shows a cytoplasmic N-terminus although the algorithm predicts an external one.