MUTATION ANALYSIS IN HAEMOPHILIA A AND INVESTIGATION OF A KERATINOCYTE BASED GENE THERAPY FOR HAEMOPHILIA B

SANJAY I. BIDICHANDANI

Thesis submitted to
The Faculty of Medicine
University of Glasgow
For the degree of Doctor of Philosophy

DECLARATION

I certify that this thesis does not contain material previously published or written by any other person except where referred to in the text and that the results in this thesis have not been submitted for any other degree or diploma.

SANJAY I. BIDICHANDANI

To Smita

ACKNOWLEDGEMENTS

I would like to thank the following people,

Professor Connor for his patient and expert guidance throughout.
Dr. George Lanyon for his supervision and helpful discussions.
Dr. Rosemary Akhurst for her supervision and continuous support.
All the clinicians who organised the collection of samples, especially Dr. Caroline Shiach and Dr. Gordon Lowe, and the patients for their participation in this study.

My appreciation also extends to everyone in the Dev Gen, CMG and PGR labs.

I would also like to thank Glasgow University for a Postgraduate scholarship.

I would like to specially thank Carol for her friendship and for all her help.
My deepest thanks to my family for their love and support throughout.
And finally I would like to thank my wife Smita for always being there.

TABLE OF CONTENTS

LIST OF FIGURES

Following page

LIST OF TABLES

Following page

ABBREVIATIONS

A	adenine
a.a.	amino acid
ASA	allele-specific amplification
ATP	2'-adenosine 5'-triphosphate
bp	base pair
C	cytosine
°C	degrees centigrade
CCM	chemical cleavage of mismatches
Ci	Curie (3.7×10^{10} Becquerel)
cpm	counts per minute
cm	centimetre
cDNA	complementary DNA
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytosine 5'-triphosphate
ddNTP	dideoxyribonucleoside triphosphate
DEPC	diethyl pyrocarbonate
DGGE	denaturing gradient gel elctrophoresis
dGTP	2'-deoxyguanosine 5'-triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
ELISA	enzyme linked immunosorbent assay
G	guanine
g	gram
h	hour
HA	heteroduplex analysis
kb	kilobase
kD	kilodalton
kg	kilogram
M	molar
min	minute
MOPS	3' (N-morpholino) propanesulfonic acid
mg	milligram
ml	millilitre
µg	microgram
µl	microlitre
mM	millimolar
mRNA	messenger RNA
ng	nanogram
nm	nanometre
OD	optical density
PCR	polymerase chain reaction
pm	picomole
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNase	ribonuclease

rpm	revolutions per minute
RT-PCR	reverse transcription PCR
SDS	sodium dodecyl sulphate
SOE	splicing by overlap extension
SSC	standard saline citrate
SSCP	single strand conformational polymorphism
SV40	simian virus 40
T	thymidine
TBE	tris borate ethylenediamine tetra acetic acid
TE	tris ethylenediamine tetra acetic acid
TEMED	N,N,N,N-tetramethylethylenediamine
Tris	tris (hydroxymethyl) aminomethane
UTR	untranslated region
UV	ultraviolet

ABSTRACT

Haemophilia A and B are X-linked recessive bleeding disorders caused by a deficiency of the procoagulant function of clotting factors VIII and IX in the intrinsic coagulation pathway, respectively. The two main objectives of this project were: (I) to carry out a comprehensive search for mutations in the factor VIII gene of patients with haemophilia A and (II) to investigate the possibility of keratinocyte-specific expression of factor IX as a potential method for somatic cell gene therapy for haemophilia B.

Using a combination of RT-PCR of ectopic factor VIII transcripts and DNA-PCR of individual exons, the entire essential sequence of the factor VIII gene was screened for mutations using chemical cleavage of mismatches (CCM) and single strand conformation analysis (SSCP). A total of 11 different mutations were characterised. These include 7 single base substitutions, 3 deletions and 1 gross mRNA abnormality due to the intron 22-mediated inversion. The 7 single base substitutions included 5 missense (D56E, V162M, G701D, A1834T and R1869I), 1 nonsense (R-5X) and 1 splice donor site mutation in intron 6 (787+3A→G). The 3 deletions included a single base deletion in exon 17 (5697delC), a five base pair deletion in exon 4 (519delTACCT) and a gross deletion of exon 16. Six mutations are predicted to result in a truncated or shortened factor VIII polypeptide. The five missense mutations were evaluated for pathogenicity by analysing the mutated amino acid positions for evolutionary conservation and their involvement with functionally important sequence motifs. Analysis of the sequence in the immediate vicinity of the various mutations revealed sequence features that may have been involved in the mechanism of mutagenesis.

To investigate the role of keratinocytes as a target cell type for somatic gene therapy of haemophilia B, a number of factor IX (FIX) gene expression vectors were constructed in order to achieve keratinocyte-specific expression from the bovine counterpart of the human cytokeratin 10 (K10) gene promoter. To optimise the efficiency of recombinant

factor IX expression by the basic K10-FIX-II construct, two sequence modifications were performed. A favourable sequence context was created surrounding the translation initiation codon (K10-FIX-G4) and synthetic oligonucleotides were used to lengthen the 5' leader sequence thus closely resembling the endogenous K10 gene (K10RB-FIX and K10RB-FIX-G4).

Since primary keratinocytes do not express K10 in monolayer cultures, a spontaneously transformed, non-tumorigenic, human keratinocyte cell line (HaCaT) was used to assess the absolute and relative factor IX expression by the various gene constructs. It is known that 5-10% of HaCaT cells in confluent cultures express K10 at low levels. Stably transfected HaCaT cell lines (pools and individual clones) were generated with each gene expression construct. Expression of factor IX mRNA was detected in every transfected keratinocyte cell line using RT-PCR analysis. Northern blot analysis of total cellular RNA extracted from all pools of transfected clones (each representing 125-750 clones) detected a 2.5 kb recombinant factor IX transcript. Factor IX protein that was γ-carboxylated and secreted into conditioned medium, was detected using a double antibody (A5/A7) ELISA. A maximum expression of 35 ng / 10^6 cells / 24 hours was achieved by the K10-FIX-G4 construct. The three constructs with potentially favourable sequence manipulations, expressed significantly higher amounts of factor IX than the basic K10-FIX-II construct. As only a few HaCaT cells express K10, the observed level of factor IX expression can be considered comparable to that achieved by other investigators who used strong viral promoters in similar experiments. These results indicate the suitability of using cytokeratin gene promoters to target the expression of biologically active factor IX in keratinocytes, for somatic gene therapy of haemophilia B.

SUMMARY

Haemophilia A (classical haemophilia) and haemophilia B (Christmas disease) are clinically indistinguishable, X-linked recessive bleeding disorders, caused by deleterious mutations in the genes for clotting factors VIII and IX, respectively. These two proteins are essential components of the intrinsic coagulation cascade. The genes for factors VIII and IX were cloned over a decade ago and in the past five years, a number of disease-causing mutations have been reported and some progress been made towards correcting these diseases by gene therapy. The overall objectives of this project were: (I) to carry out a comprehensive search for mutations in the factor VIII gene of patients with haemophilia A and (II) to investigate the possibility of keratinocyte-specific expression of factor IX as a potential method for somatic cell gene therapy for haemophilia B.

In haemophilia A, analysis of a sufficiently large number of naturally occurring mutations in the factor VIII gene would help to further elucidate the functionally important domains of the large and complex factor VIII protein. In an attempt to identify mutations in the factor VIII gene, two strategies were employed in this study. The choice of the strategy was mainly governed by the availability of RNA from the patient. When RNA was available, the method involving a combination of RT-PCR of ectopic transcripts and DNA-PCR to amplify the entire essential sequence of the factor VIII gene was used (Naylor et al., 1991, 1993a). Thirteen randomly selected unrelated haemophilia A patients were analysed by this method (group 1). Large sequence alterations detected following the initial PCR amplification, were completely characterised by direct sequencing. Chemical mismatch cleavage (CCM) analysis was used to screen and accurately localise small sequence alterations, which were then fully characterised by direct sequencing. Twelve unrelated patients from whom only DNA samples were available, were also selected (at random) for mutation analysis (group 2). Exons 4, 7, 8, 11, 12 and 16 were screened using single strand conformation polymorphism (SSCP) analysis. Exons 7, 11, 12 and 16 in particular were selected because of the disproportionate harbouring of mutations in these small exons (Tuddenham et al., 1991).

Using the above two strategies, a total of 4 gross mRNA abnormalities, 13 positive CCM and 2 positive SSCP results were detected while screening for mutations within the factor VIII gene. All mutations were tentatively mapped to individual exons of the factor VIII gene. A total of 11 different and potentially pathogenic mutations were fully characterised. These include 7 single base substitutions, 3 deletions and 1 gross mRNA abnormality due to the intron 22-mediated inversion (Naylor et al., 1992, 1993b; Lakich et al., 1993). The 7 single base substitutions included 5 missense (D56E, V162M, G701D, A1834T and R1869I), 1 nonsense (R-5X) and 1 splice donor site mutation in intron 6 (787+3A→G). The 3 deletions included a single base deletion in exon 17 (5697delC), a five base pair deletion in exon 4 (519delTACCT) and a gross deletion of exon 16. Eight of these are novel changes affecting the factor VIII gene.

Correlation of the genotype with the phenotype was carried out for each of the potentially pathogenic mutations. A total of 6 mutations are predicted to result in either a truncation or internal loss of amino acids from the factor VIII polypeptide. All these mutations, except the intron 6 splice donor mutation which resulted in the skipping of exons 5 and 6, resulted in a severe phenotype. Naylor et al. (1993a) reported a DNA deletion of the same two exons in a severe haemophilia patient. The moderate phenotype observed in the two patients with the splice mutation is probably due to a small fraction of normally spliced factor VIII mRNA detected. Four of the five missense mutations were the only change detected on screening the entire essential sequence of the factor VIII gene (A1834T was one of two CCM results in the same patient). These are therefore likely to be the cause of the severe (G701D, R1869I), moderate (D56E) and mild (V162M) haemophilia A in these patients. The pathogenicity of the missense mutations was assessed by analysing the affected residues for evolutionary sequence conservation and for involvement with known or predicted functional domains, as catalogued in the "PROSITE dictionary of protein sites and patterns" (Release 10.1; using the computer program "MOTIFS"). The "A" domains of human and murine factor VIII, factor V and ceruloplasmin were aligned using a computer program, "PileUp" (all missense mutations mapped to the "A" domains of factor VIII). Three missense mutations affected highly conserved residues (D56E and G701D occurred in 11/12 and V162M in 9/12 "A"

domains analysed). Analysis with "MOTIFS" showed that G701D involved a multicopper oxidase I signature sequence and V162M, a casein kinase II phosphorylation site.

Two of the twelve patients in whom a potentially pathogenic mutation was detected, developed inhibitors to therapeutically administered factor VIII (the inhibitor status in three patients was not known). One was a patient with the frame-shifting single base deletion in exon 17, predicted to cause a severe truncation of the factor VIII polypeptide. Interestingly, the other was a patient with a missense mutation (G701D; 9.4 Bethesda units).

The intron 6 splice donor mutation (787+3A→G) was detected in two of the thirteen patients in group 1 (a four generation family history suggested they were unrelated, but the mothers of both patients were shown to be carriers and a haplotype analysis was unable to exclude identity by descent). The two single base substitutions, V162M and R-5X, have each been reported twice before. Only one of the seven single base substitutions detected, occurred in a hypermutable CpG dinucleotide (R-5X). Of the eight severe haemophilia A cases in whom a complete ascertainment of mutations was carried out (group 1), only one showed the characteristic intron 22 mRNA abnormality. The latter two findings suggest that this small group of patients was not a representative sample of the population. The distribution of mutations among the various exons was not totally random. Exons 4, 16, 17 and 21 were seen to be affected by two mutations each and three mutations were found to affect exon 14.

The sequence in the immediate vicinity of the various mutations was analysed for potential mechanisms of their origin (and recurrence in some cases). The five base pair deletion in exon 4 (519delTACCT) was shown to be concordant with the model of slipped mispairing at DNA replication, proposed by Krawczak and Cooper (1991). The nonsense mutation, R-5X, occurred at a hypermutable CpG dinucleotide and this was the third independent report of this mutation. This study is also the third independent report of the V162M missense mutation and sequence analysis surrounding this non-CpG site mutation revealed a tetranucleotide motif "TGGA" immediately flanking the G→A at position 541. This is a putative arrest site (during DNA replication) for DNA polymerase α and is reported to be a mutation hotspot for both single base substitutions and small

deletions (Krawczak and Cooper, 1991; Cooper and Krawczak, 1993). In addition, the trinucleotide topoisomerase I cleavage site consensus sequence "CTT", was also noted in the immediate vicinity. The D56E mutation affects the same codon that harbours a known polymorphic missense mutation (D56V). A tetranucleotide direct repeat "TCAC" was seen immediately flanking the base change responsible for D56E (and overlapping the base change causing D56V) that may have mediated a four-base slippage, producing the misincorporation error(s). The trinucleotide topoisomerase I cleavage site sequence "CTT", was also noted in the immediate vicinity of the D56E mutation site. The R-5X and A1834 mutation sites were both flanked on either side by a "CTT" motif. Another such sequence was also observed on the opposite strand as the R1869I mutation, which occurred within a direct dinucleotide "GA" repeat sequence.

The two strategies used for mutation detection were evaluated for their efficacy. Thirteen cleavage products were obtained after a complete screen of the factor VIII gene in nine patients (i.e., those in group 1 who did not show a grossly abnormal segment), with >1 seen in three of them. It is thought that a near total ascertainment was achieved using this strategy. The efficiency of SSCP analysis of selected exons (all amplified as PCR products of <400 bp) was estimated to be <70%.

Haemophilia is a candidate disease for somatic gene therapy. Many groups have used strong viral promoters to achieve ectopic expression of factor IX in various cell types. Most groups attempting an *ex vivo* approach, however, achieved low and transient expression after transplantation of the genetically modified cells *in vivo*. It is now clear that the choice of cell type and promoter is crucial in determining the long-term success of expression *in vivo*. In general, it was shown that the use of endogenous cellular promoters are capable of directing sustained heterologous expression *in vivo*, albeit at least presently, only at low levels (Scharfmann et al., 1991; Dai et al., 1992).

Keratinocytes are attractive target cells for somatic gene therapy. Gerrard et al. (1993) have recently demonstrated the ability of keratinocytes to successfully carry out the necessary post-translational modifications of factor IX and actively secrete it into the systemic circulation. But the use of a viral promoter *in vivo* achieved only a transient effect. In the present project, an attempt was made to achieve sustained and high level

expression of factor IX, by using a strong cytokeratin gene promoter to direct expression in a keratinocyte-specific fashion. The bovine counterpart of the differentiation-specific human cytokeratin 10 (K10) gene promoter was used for this purpose. The ability of this promoter to direct keratinocyte-specific gene expression has previously been demonstrated (Bailleul et al., 1990).

The aim of this project was to create a variety of K10 promoter-driven human factor IX cDNA-based gene constructs and test their relative efficiencies in an *in vitro* keratinocyte model. In collaboration with Dr. Y. Alexander, the efficiency of a construct was also tested *in vivo* in transgenic mice. A basic construct (K10-FIX-II) was first created by linking the human factor IX cDNA to the K10 promoter (in an expression vector that also contained an intron and polyadenylation signal sequence from the SV40 viral genome). Three sequence modifications were then introduced into the basic construct, in order to theoretically improve factor IX expression. These changes included: (i) site-directed mutagenesis to create a favourable sequence context surrounding the "AUG" initiation codon (K10-FIX-G4) to improve the efficiency of translation initiation (Kozak, 1987, 1991a). This necessitated the induction of a missense (probably conservative) mutation in the second codon of factor IX (Q-45V); (ii) the use of synthetic oligonucleotides to repair the 5' untranslated region (5'UTR), which involved the replacement of the existing polylinker with the original K10 cap site and 5'UTR (K10RB-FIX and K10RB-FIX-G4), effectively increasing its length and reducing the GC content (Kozak, 1991b,c,d, 1994); (iii) "splicing by overlap extension" (Horton et al., 1989) to replace the SV40 viral sequence at the 3' end of the gene construct with the 3' end of the K10 gene. All sequence modifications were confirmed by direct sequencing.

Primary keratinocytes do not express keratin 10 (K10) in conventional monolayer cultures. The HaCaT cell line, which is a spontaneously transformed (non-tumorigenic) keratinocyte cell line expresses a wide range of keratins *in vitro*, including K10 (Ryle et al., 1989). This cell line was therefore used as the *in vitro* model to test the relative efficiencies of the various K10-FIX gene constructs. Liposome-mediated transfections were performed using the basic construct (K10-FIX-I and II) and those gene constructs with the favourable translation initiation consensus (K10-FIX-G4), the repaired 5'UTR

(K10RB-FIX) and with both these changes (K10RB-FIX-G4). The efficiency of transfection was estimated using a *lacZ* reporter construct. Transfected cells were selected for the co-transfected neomycin resistance gene (pSV2Neo) and both, pools and individually selected stably transfected clones were analysed for expression of factor IX. All transfected pools and clones showed expression of factor IX mRNA on RT-PCR. All transfected pools showed the presence of a ~2.5 kb recombinant factor IX mRNA by Northern blot analysis. Secreted factor IX was estimated in conditioned tissue culture medium using a double antibody ("sandwich") ELISA. The "A7" antibody used in the assay detects only that fraction of factor IX that is γ-carboxylated at its N-terminal glutamic acid residues (Smith et al., 1986; Smith, 1988) and was used as an indirect estimate of post-translationally modified and potentially biologically active factor IX (vitamin K was added in the tissue culture medium as it is an essential cofactor in the carboxylation). All pools and clones of transfected HaCaT cells expressed factor IX above background levels.

One of the basic constructs (K10-FIX-I) was shown to have a missense mutation (S136T) and since it had already been used for the generation of transgenic mice, *in vitro* gene expression studies were also carried out using this construct. Both the basic constructs i.e., the one with the normal factor IX sequence and the one with the S136T mutation expressed similar levels of factor IX, suggesting that the S136T mutation may be a conservative change. K10-FIX-G4 was the highest expressing gene construct and secreted factor IX at a level of 35 (26-45) ng/10^6 cells/24 hours and expressed significantly more than both the basic constructs (*t* test showed $p<0.01$ for K10-FIX-II and $p<0.02$ for K10-FIX-I). This indicated that Q-45V is probably a conservative change. The other manipulated constructs, K10RB-FIX and K10RB-FIX-G4 also expressed significantly higher levels than the basic construct. Surprisingly, K10RB-FIX-G4, which had the repaired 5'UTR in addition to the change in K10-FIX-G4, did not express higher levels of factor IX than the latter construct. Retrospectively, an analysis of the repaired 5'UTR sequence (with a computer program, "STEMLOOP") predicted the presence of a detrimental hairpin structure close to the cap site (Kozak, 1991b, 1994). Interestingly, Northern analysis revealed that all three manipulated constructs showed more intense

recombinant factor IX bands than both the basic constructs. Since the sequence modifications were aimed at increasing the translational efficiency of the basic construct (seen from the factor IX protein levels), it is proposed that the discrepancy in the levels of factor IX mRNA represent some post-transcriptional protection, offered either by the process or by the product of translation.

HaCaT cells express very low levels of K10 (only at confluence) and *en face* immunofluoresence indicates that this is due to only a few cells showing expression (the signals from which become more intense as cell density increases, Ryle et al., 1989). Thus even after the selection of transfected cells, it would be expected that only a few cells would express factor IX. Even though the *in vitro* factor IX expression is 10-15 fold less than that obtained using primary keratinocytes transduced with a retroviral construct containing factor IX cDNA driven by a strong viral promoter, followed by G418 selection (Gerrard et al., 1993), deficiencies in the *in vitro* model used in this project indicate that the expression levels may be comparable. K10-FIX-I was also used to generate transgenic mice (created by Dr. Y. Alexander). Screening of the mice for transgene integration and measurement of plasma factor IX levels were performed as part of this project, to assess the efficacy of the K10-FIX construct *in vivo*. The mice produced enough γ-carboxylated factor IX to maintain an average level of 28 (22-34) ng/ml in their plasma, as measured by the same ELISA using the "A7" antibody. This level is similar to that achieved by repeated administrations of retrovirally transduced myoblasts (containing a recombinant factor IX construct and a cellular muscle creatine kinase enhancer element), in nude mice (Dai et al., 1992).

Keratinocytes have been shown to be effective target cells for somatic gene therapy of haemophilia B (Gerrard et al., 1993). The results obtained in this project (including the data gathered in collaboration with Dr. Y. Alexander) substantiate this finding, and in addition, indicate the suitability of using tissue-specific cellular promoters to achieve sustained and high level expression of factor IX for this purpose.

CHAPTER: 1

INTRODUCTION

1.0 INTRODUCTION

1.1 HAEMOPHILIA A and B

1.1.1 INTRODUCTION

Genetic defects or deficiencies in blood coagulation proteins are associated with a bleeding disorder known as haemophilia. Deficiency or defects in clotting factor VIII result in haemophilia A (classical haemophilia) which accounts for about 85% of all haemophilias. Deficiency or defects in clotting factor IX result in haemophilia B (Christmas disease) which comprises 10-12% of this hereditary disorder. Defects in other coagulation proteins are rare, but deficiency states have been recognised for most of the blood clotting proteins (Furie and Furie, 1988). Haemophilias A and B are both X-linked recessive disorders (McKusick catalogue no. 306700 and 306900, respectively) with an incidence of about 1:5,000 (Hoyer, 1994) and 1:30,000 (Connor et al., 1985) males, respectively. The two disorders are clinically indistinguishable and can only be separately diagnosed by specific assays for their respective cofactors. The clinical hallmarks include haemorrhage that is either spontaneous or follows minor trauma, and mainly involves large joints and deep muscles.

Factor VIII and IX are both essential components of the intrinsic coagulation cascade. Factor IX is a vitamin K-dependent serine-protease that, in its activated form (factor IXa), serves to proteolytically activate factor X. Activated factor VIII (factor VIIIa) acts as a cofactor to increase the V_{max} of this reaction by at least 4 orders of magnitude (in the presence of negatively charged phospholipids and calcium ions; van Dieijen et al., 1981). Factor VIII is stabilised in the circulation (in plasma) by non-covalent complex formation with a large multimeric plasma protein, von Willebrand factor (vWF; Hoyer, 1981). The genes responsible for clotting factors VIII (Gitschier et al., 1984) and IX (Anson et al., 1984; Yoshitake et al., 1985) have been cloned and completely characterised.

1.1.2 CLINICAL FEATURES & COMPLICATIONS

Haemophilia A and B are both X-linked recessive disorders of bleeding, and therefore almost invariably affect males (although some females have been reported to have clinical disease due to non-random X-inactivation; Nisen et al., 1986; Nisen and Waber, 1989; Giannelli et al., 1993). Haemophilia A and B are clinically heterogeneous disorders in which the severity of disease correlates well with the biochemical phenotype. Normal levels of factor VIII and factor IX coagulant activity (factor VIII:C and IX:C) range between 50-150 IU/dl. Levels of <2, 2-5 and 5-30 IU/dl result in severe, moderate and mild clinical disease, respectively.

Severe haemophilia is characterised by spontaneous and recurrent haemarthroses into knees, elbows, ankles, shoulder and hip joints. Bleeds usually start at the age of 6 months when the baby becomes mobile. The bleeding may be as frequent as 2-4 times per month. Untreated haemarthroses result in persistent severe pain and due to their recurrent nature, eventually lead to loss of cartilage and secondary osteoarthrosis; the haemophilic arthropathy. Deep muscle haematomas and frank haematuria are also observed. Spontaneous intracranial bleeds can occur and in adults can be severe and often fatal. This was the most common cause of death before the AIDS epidemic which started in the early eighties (Hoyer, 1994).

Moderate haemophilia usually leads to bleeding in response to minor trauma and is usually only seen 1-5 times a year. Mild haemophiliacs on the other hand only bleed after relatively severe trauma and because of this, diagnosis may be delayed until adult life. An interesting subgroup of haemophilia B patients, the Leyden variant (Veltkamp et al., 1970), have a severe phenotype during childhood, but gradually improve with eventual conversion to mild disease after puberty.

Before effective replacement therapy was introduced at the end of the 1960s, haemophilia caused severe disability and death at a young age. A spectacular improvement in life expectancy has resulted from the introduction of specific replacement therapy (Larrson, 1985; Smit et al., 1989; Jones and Ratnoff, 1991).

1.1.3 CURRENT MODALITIES OF THERAPY

In the UK, clinical management of haemophilia is carried out in specialised comprehensive haemophilia centres. Treatment is often initiated following a bleeding episode. The main objective being to increase and maintain the plasma levels of the deficient cofactor. Presently, this is best achieved by infusions of treated factor VIII (reviewed by Brettler and Levine, 1989) or factor IX concentrates. Due to the short half-lives of factors VIII and IX in circulation, (10 and 24 hours, respectively) repeated administrations are required. Mild haemophiliacs require intravenous therapy only if they have a haemarthrosis or if they are to undergo a surgical procedure. Those with more severe haemophilia may be candidates for home therapy as they are aware of the onset of haemorrhage before there is any clinical evidence. In Sweden, severe haemophiliacs have been treated prophylactically for the past twenty-five years with infusions given thrice a week. This has effectively converted their disease to that of moderate severity (Nilsson et al., 1992).

However, since concentrates are usually prepared from plasma pools derived from 2000-30,000 individuals, they form a potential source for the transmission of blood borne diseases. Infectious complications from transfusion-transmitted viruses began to be noted in haemophiliacs in the late 1970s and are still a major concern. Many newer types of cofactor concentrates are now produced, both to eliminate infectious complications and to increase purity. Even though the infectivity of the currently available concentrates (heat / chemically treated; Brettler and Levine, 1989) is considerably reduced, the sizeable number of patients already infected in the past, continue to have associated complications. Contamination of the plasma pool by human immunodeficiency virus (HIV) has led to many haemophiliacs being infected, and a survey performed a few years ago showed 1201 UK haemophiliacs to be HIV positive, with 85 cases of AIDS and 47 deaths (Darby et al., 1989). AIDS is now the most common cause of death in severe haemophilia A patients. Virtually every individual who received untreated cofactor preparations developed hepatitis B or non-A, non-B. Consequently, many haemophilia patients have, and will develop chronic progressive liver disease (Hay et al., 1985). Various trials have suggested that the viruses responsible for hepatitis are more difficult

to inactivate (Brettler and Levine, 1989), however, the "super dry heat" method for inactivation used in the UK, has been reported to be more successful (Study Group of the UK Haemophilia Centre Directors, 1988).

To eliminate the potential complications of plasma-derived products, two biotechnology companies have invested efforts in the development of recombinant factor VIII synthesised in mammalian cell cultures. Clinical trials have shown these products to be as effective as concentrates and with no immediate side effects (White et al., 1989; Lusher et al., 1993). These products are not yet being widely used and it is thought that costs may be prohibitive (Hoyer, 1994).

One of the most difficult problems in the clinical management of haemophilia (especially haemophilia A) is the development of inhibitory antibodies to the exogenously administered cofactor, making them ineffective. This complication is seen to affect about 20% of haemophilia A patients (Ljung et al., 1992; Lorenzo et al., 1992). Patients whose haemophilia is due to gross deletions or nonsense mutations are more likely to develop inhibitors although certain missense mutations have also resulted in an inhibitor phenotype (Tuddenham et al., 1991).

Clearly, due to the reactive nature of treatment initiation, the need for repeated injections and the associated complications, the current form of therapy is sub optimal.

1.2 HAEMOSTASIS AND THE ROLE OF CLOTTING FACTORS VIII AND IX

Haemostasis is maintained by a complex sequence of interactions between endothelial cells, platelets and a series of plasma proteins (Davie and Ratnoff, 1964; MacFarlane, 1964). Trauma to the vessel wall results in damage of the endothelial lining and exposure of the subendothelial matrix. vWF plays an active role in binding and subsequent accumulation of platelets at the site of injury (Tuddenham, 1989; Handin and Wagner, 1989). This platelet plug forms the first line of action in the prevention of blood loss. The intrinsic and extrinsic coagulation cascades are initiated by the exposure of the subendothelial negatively-charged matrix, and the release of tissue thromboplastin from the damaged cells in the vicinity, respectively. Sequential proteolytic reactions result from

the activation of these coagulation pathways, which converge at the activation of factor X (figure 1).

Both, factor VIII and IX are part of the intrinsic coagulation cascade. The coagulation cascade can be divided into three main stages namely, the *contact* phase, *procoagulant* and *anticoagulant* phases. There are two main classes of proteins that take part in the coagulation cascade. These are the serine proteases and the protein cofactors. One of each, forms an enzyme complex on a suitable surface (in the presence of calcium ions), which forms the functional units of the coagulation cascade (Mann et al., 1988). The serine protease in each enzyme complex, in the presence of the active cofactor (which has no proteolytic function), proteolytically activates another zymogen to protease, which in turn forms another complex downstream in the cascade. As seen in figure 1, in the activation of factor X (inactive zymogen) to Xa (active protease), activated factor IX (IXa) is the serine protease whose proteolytic function is catalysed by active cofactor VIII (VIIIa). In the presence of negatively charged phospholipids and calcium ions the proteolytic cleavage of factor X by factor IXa is increased by at least four orders of magnitude by factor VIIIa (van Dieijen et al., 1981; Mann et al., 1990).

The intrinsic coagulation cascade is initiated by the activation of factor XII, which at the end of the *contact* phase results in activated factor XI (XIa). Factor XIa activates factor IX, which is the first enzyme complex after the *contact* phase. The extrinsic pathway on the other hand first activates factor VII (VIIa). Both, VIIa and IXa serve to activate factor X (Xa). The intrinsic and extrinsic cascades follow a common pathway after this step.

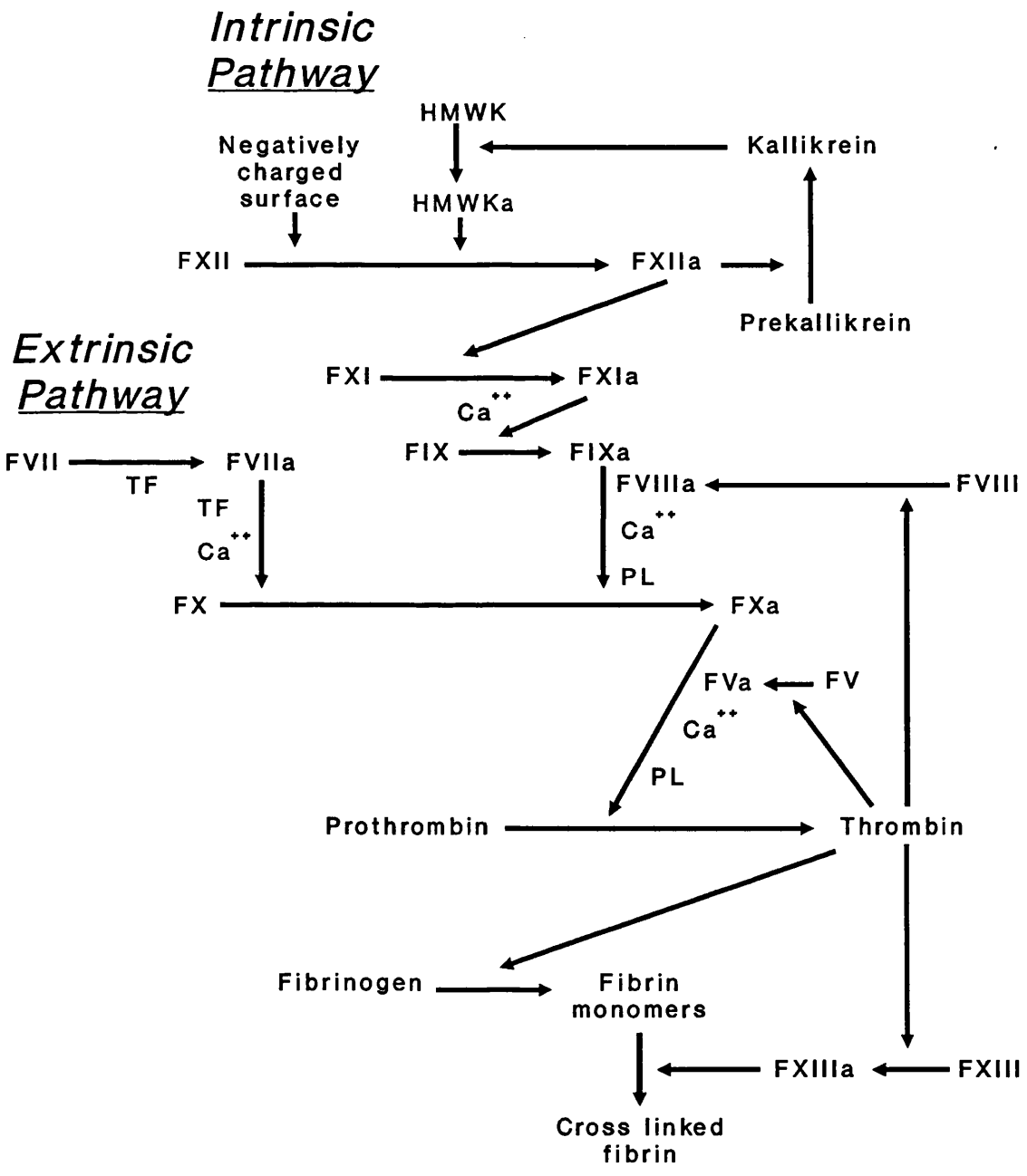
Factor Xa converts prothrombin to thrombin, which in turn converts fibrinogen to insoluble fibrin, which gets polymerised to form a fibrin network in the presence of XIIIa (which was also activated by thrombin). Thrombin also activates platelets, factor V and VIII, thus playing an important role in the amplification of the cascade reaction.

Reactions within the cascade are further regulated via a variety of feedback loops. The *anticoagulant* pathway is activated to prevent the excessive formation of thrombin. This is done by plasma proteins, antithrombin III and heparin cofactor II, as well as thrombomodulin on the endothelial cell surface. Eventual dissolution of the fibrin clot is

Figure 1: The coagulation cascade

Glycoprotein components of the intrinsic pathway include factors XII, XI, IX, VIII, V, prothrombin and fibrinogen. Glycoprotein components of the extrinsic pathway, initiated by the action of the tissue factor located on cell surfaces include, factors VII, X, II and fibrinogen. The intrinsic and extrinsic cascade reactions are shown, which culminate in the conversion of fibrinogen to fibrin and the formation of a fibrin clot. Where indicated reactions are Ca^{+2} dependent. Numbered factors are abbreviated FXI, FXII, etc. Other abbreviations include HMWK=high molecular weight kininogen, TF=tissue factor, PL=phospholipid.

The coagulation cascade



promoted by plasmin which is broken down from plasminogen by tissue plasminogen activator.

Thus the enzyme complexes which form the functional units of the clotting cascade, not only cause an amplification of the reaction but also serve to localise it to the site of vascular injury due to the membrane binding domain of the vitamin K-dependent serine-proteases (Mann et al., 1988). The formation of the complexes also provide numerous opportunities for the modulation of the reaction rate via the various positive and negative feedback loops.

The three main phases are also distinct in their pathophysiological consequences on the coagulation system, in the event of genetic or acquired defects in their protein components. Whilst a functional deficiency in any of the factors of the *procoagulant* phase (e.g. factor VIII and IX) causes a bleeding disorder warranting protein replacement therapy, a similar deficiency in factors of the *contact* phase results in little, if any, effects. This is probably due to the capacity of the activated platelets to convert factor XI to XIa, thus activating the intrinsic cascade (bypassing the factor XII activation and initiation). A deficiency of the factors in the *anticoagulant* phase (antithrombin III, proteins C and S) predisposes to venous thrombosis. A deficiency of platelet number or function results in a characteristic bleeding profile involving mainly mucocutaneous surfaces.

1.3 CLOTTING FACTORS VIII AND IX

1.3.1 THE FACTOR VIII GENE

Cloning of the factor VIII gene

The factor VIII gene was independently cloned by two groups, one at Genentech (Gitschier et al., 1984) and the other at the Genetics Institute (Toole et al., 1984). Both groups used the same basic method to characterise factor VIII genomic clones, followed by the cloning and sequencing of the entire factor VIII cDNA (Wood et al., 1984; Toole et al., 1984). Both groups also transfected the full length cDNA into mammalian cells and

demonstrated the successful expression of active factor VIII protein (Toole et al., 1984; Wood et al., 1984).

The method employed by both groups involved the use of partial amino acid sequence of human (Vehar et al., 1984) or porcine (Toole et al., 1984) highly purified factor VIII protein (human, Rotblat et al., 1983; porcine, Fass et al., 1982) to synthesise a cocktail of oligonucleotide probes for the screening of genomic DNA libraries. Genomic DNA libraries were screened instead of cDNA libraries because it was not known which cells expressed factor VIII and due to the low concentration of plasma factor VIII (100-200 ng/ml), a low mRNA abundance was anticipated. Both groups then used probes isolated from the genomic DNA library to screen a series of cell lines by Northern blotting, in order to identify cell types expressing factor VIII. Both groups identified the liver as the source of human factor VIII, but the Genentech group used an expressing T-cell hybridoma cell line cDNA library to isolate the factor VIII cDNA.

To confirm that they had indeed isolated the gene and cDNA for factor VIII, both groups constructed full length cDNA clones and linked them to viral promoters. These constructs were introduced into mammalian cell lines (Genentech: hamster kidney cells; Genetics Institute: monkey kidney cells), following which a human factor VIII:C-like activity was secreted into conditioned media. This was quantified using a sensitive assay based on the activation of factor X and hydrolysis of a chromogenic substrate. The factor VIII:C activity was found to be 1% and 7% of normal plasma concentrations by the groups at the Genetics Institute and at Genentech, respectively. On affinity purification this was shown to partially correct the clotting time of plasma from a haemophiliac.

The clotting activity was shown to be increased by thrombin (a physiological activator of factor VIII) and decreased by specific anti-factor VIII antibodies. The Genentech group also demonstrated specific and reversible binding to vWF on an affinity column, which was inhibited by protein C (an inactivator of factor VIII).

The structure of the factor VIII gene and cDNA

The factor VIII gene is relatively large and has 26 exons and 25 introns that span 186 kb of genomic DNA (Gitschier et al., 1984). The gene maps to Xq28 i.e., at the telomere of the long arm of the X chromosome (Purrello et al., 1985; Tantravahi et al.,

1986). It is estimated to constitute up to 0.1% of the entire length of the X chromosome (Gitschier et al., 1984).

The 26 exons of the factor VIII gene, range from 69 to 3,106 bp (median size of 164 bp). All exons are fairly normal in size, except for exons 5, 20, 14 and 26. Exons 5 and 20, at 69 and 72 bp respectively, are about half the average size of vertebrate exons (137 bp; Hawkins, 1988). Exon 14 is 3,106 bp long and codes for the central connecting region or "B" domain of the factor VIII polypeptide (section 1.3.2). Exon 26, which is the last exon, is 1,958 bp long, of which 1,805 bp constitutes the 3' untranslated sequence of the factor VIII gene.

The 25 introns of the factor VIII gene range from 0.2 kb to 32.4 kb in length. All the splice donor and acceptor sites conform to the GT..AG rule (Breathnach and Chambon, 1981). Intron 22, which is 32.4 kb in length, is the largest intron in the factor VIII gene. It contains an unmethylated GC rich sequence (CpG island) of the type known to be associated with 5' regions of vertebrate genes (Bird, 1986). This island appears to serve as a bi-directional promoter for two genes emanating from it, referred to as factor VIII-associated genes A and B (*F8A* and *F8B*). *F8A* is an intronless gene that is completely nested within intron 22 of the factor VIII gene (Levinson et al., 1990). It is transcribed in a direction opposite to that of the factor VIII gene, and produces a transcript that is 1,739 nucleotides long. This was the first example of an intron-embedded gene to be identified in a vertebrate genome. Two additional copies of this gene have been identified approximately 500 kb upstream of the factor VIII gene i.e., further telomeric on the X chromosome (Patterson et al., 1989; Levinson et al., 1990; Freije et al., 1992; Lakich et al., 1993). This intron 22-embedded gene, along with the two telomeric copies, has recently been implicated in the causation of about half the cases of severe haemophilia A (Lakich et al., 1993; Naylor et al., 1993a,b; section 1.7.2). *F8B* in contrast, is transcribed in the same direction as the factor VIII gene. Its first exon is contained within intron 22, but it uses exons 23 - 26 of the factor VIII gene as its own exons 2 - 5 (Levinson et al., 1992). It produces a transcript which is formed by the splicing of its exon 1 with exon 23 of the factor VIII gene and utilises the polyadenylation

signal of the factor VIII gene. In contrast to the factor VIII gene, both, *F8A* and *F8A* are ubiquitiously transcribed.

The factor VIII cDNA is approximately 9 kb long and has an open reading frame that codes for a polypeptide of 2,351 amino acids, the first 19 of which comprise the hydrophobic signal sequence. Both, RNase protection and Northern blotting revealed that the liver was the main site of factor VIII expression. RNase protection assays revealed that the mRNA start site was 170 (less frequently 172) nucleotides upstream of the translation start site. The poly (A) tail is attached 1,805 nucleotides following the translation termination codon. A consensus polyadenylation signal, "AATAAA" is located 19 nucleotides upstream from the poly (A) tail.

A consensus promoter element, "GATAAA", is seen at position -30 from the mRNA start site (200 nucleotides before the translation initiation site) and is thought to be the equivalent of the TATA-box seen in many tissue-specific genes (Breathnach and Chambon, 1981). No "CAAT" sequence was observed around position -70 from the mRNA start site.

1.3.2 THE FACTOR VIII PROTEIN

Factor VIII is a high molecular weight glycoprotein (360 kDa; Rotblat et al., 1985) which is proteolytically activated by thrombin and factor Xa. In plasma, factor VIII exists as a heavy chain extending from 90 to 200 kDa in a metal ion association with a light chain of 80 kDa (Vehar et al., 1984; Rotblat et al., 1985; Andersson et al., 1986). This complex is stabilised by a non-covalent association with another plasma glycoprotein, von Willebrand factor (vWF; Hoyer, 1981).

For many years the characterisation of factor VIII was complicated by its low concentration in plasma (200 ng/ml or 1/2,000,000 the molar concentration of serum albumin), its heterogeneity in size and its extreme sensitivity to proteolysis. Small amounts of highly purified factor VIII were prepared from porcine and human plasmas in the early 1980s (Fass et al., 1982; Fulcher and Zimmerman et al., 1982; Rotblat et al., 1983). As described before, this information was used to clone the factor VIII gene and to sequence the entire factor VIII cDNA. This work allowed the determination of the

primary structure of factor VIII. Expression of factor VIII in mammalian cells using recombinant DNA techniques (Kaufman et al., 1988, 1989), coupled with the use of site-directed mutagenesis (Toole et al., 1986; Pittman and Kaufman, 1988; Pittman et al., 1992a,b) and classical protein biochemistry experiments (Fulcher et al., 1984; Rotblat et al., 1985; Eaton et al., 1986; O'Brien et al., 1992) have led to an increased understanding of the structure-function correlation of this large and complex protein. Furthermore, the delineation of naturally occurring mutations in patients with haemophilia A (Tuddenham et al., 1991) have also contributed to our understanding of some aspects of the functioning of factor VIII (Arai et al., 1989a; O'Brien et al., 1990; McGinniss et al., 1993). Electron microscopy has also added some information regarding the structural organisation of factor VIII (Mosesson et al., 1990; Fowler et al., 1990). These results suggest that the heavy-chain subunits (A1 and A2) and the light-chain subunit (A3-C1-C2) are closely associated in a small globular core (10-12 nm in diameter) and the B-domain forms a 25 nm tail-like structure. Analysis of factor VIIIa revealed only the small globular core, consistent with the thrombin-mediated cleavage of the tail-like B-domain appendage (a diagrammatic representation is presented by Hoyer, 1994).

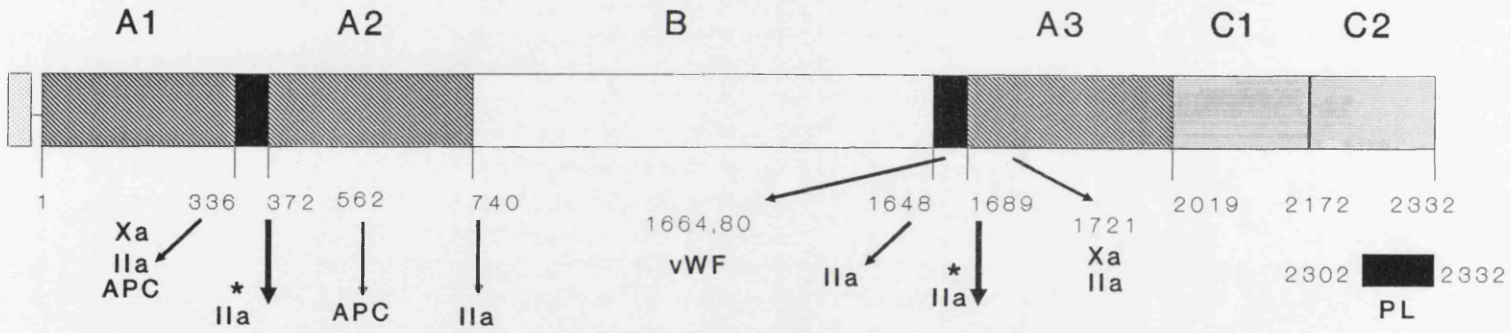
Computer-aided analysis of the factor VIII amino acid sequence revealed two types of internal homologies (Vehar et al., 1984; figure 2). The first consists of a triplicated segment called the "A" domains (A1-A3) and the second consists of a duplicated segment called the "C" domains (C1, C2). The three "A" domains consist of about 330 residues each, and are found at the following positions, A1: 1-336, A2: 372-740 and A3: 1689-2019, of the factor VIII polypeptide. The A2 and A3 domains are separated by a large region of 983 amino acids which is called the "B" domain. The "C" domains consist of 150 residues each, and are found at the following positions, C1: 2020-2172 and C2: 2173-2332, of the factor VIII polypeptide.

The three "A" domains share about 30% sequence identity with each other and the two "C" domains about 40% (Vehar et al., 1984). Factor V, which is another plasma glycoprotein cofactor in the coagulation cascade, also shows a similar structure with three repeated "A" domains (with a large connecting region separating A2 and A3) and two repeated "C" domains (Kane and Davie, 1986; Kane et al., 1987; Jenny et al., 1987).

Figure 2: Structure of Factor VIII

The figure shows a diagrammatic representation of the factor VIII polypeptide. The triplicated A domains (hatched boxes), the duplicated C domains (dotted boxes) and the central B domain represented by an open box are indicated. The positions of cleavage by thrombin (IIa), activated factor X (Xa) and activated protein C (APC) are shown. The sites of thrombin cleavage essential for activation of factor VIII (372, 1689) are marked by an asterisk. The sites of binding with von Willebrand factor (vWF) and phospholipid (PL) are indicated. The amino acid numbering is according to Wood et al., 1984.

Structure of factor VIII



The "A" and "C" domains of factor VIII and V share approximately 40% and 35% sequence identity, respectively (Kane and Davie, 1988). Ceruloplasmin, a major copper binding protein in plasma, also has three repeated "A" domains, but unlike factor VIII and V, lacks the duplicated "C" domains and the large connecting region separating the A2 and A3 domains (Takahashi et al., 1984; Kochinsky et al., 1986). The three "A" domains of ceruloplasmin share about 30% sequence identity with those of factor VIII (Vehar et al., 1984) and V (Kane and Davie, 1988). In addition, the "C" domains of factor VIII and V share about 20% sequence identity with the first 150 residues of the slime mold lectin discoidin I (Kane and Davie, 1988; Poole et al., 1981) and about 40% with the more recently described mouse milk fat globule membrane protein (Stubbs et al., 1990). The large interconnecting regions ("B" domain) between the A2 and A3 domains of factor VIII and V show no sequence similarity (Kane and Davie, 1988).

There are 23 cysteine residues in the mature factor VIII polypeptide. Most of these are clustered in the "A" and "C" domains and occupy very similar positions (Vehar et al., 1984). This suggests that the structures of both repeated domains of factor VIII reflect conserved disulphide bonding arrangements.

The exceptionally large "B" domain of factor VIII is delimited by residues 740 and 1648. Although these are sites for proteolytic cleavage by thrombin, site-directed mutagenesis has revealed that they are not important for factor VIII activity (Pittman and Kaufman, 1988). This region is extremely rich in potential asparagine-linked glycosylation sites, containing 20 of the 25 sites found in factor VIII. Comparison of the deduced amino acid sequence of porcine and human factor VIII shows a striking divergence between the "B" domains, whereas the region of the A2 and A3 domains immediately flanking them, revealed 80-85% homology (Toole et al., 1986). Since porcine factor VIII can be used to treat haemophilia A, it was speculated that the "B" domain may be dispensable for factor VIII procoagulant activity. This was shown to be the case by site-directed mutagenesis (Toole et al., 1986; Eaton et al., 1986). The entire "B" domain deleted from the full length cDNA in an expression construct used to transfect mammalian cells, expressed factor VIII with procoagulant function, thrombin cleavage products and an activation profile comparable to the wild type molecule. This is the

rationale for using factor VIII cDNA constructs devoid of the large "B" domain, in current haemophilia A gene therapy experiments (Hoeben et al., 1990, 1992; Lynch et al., 1993).

Treatment of factor VIII with thrombin causes a rapid activation followed by a subsequent first order decay of its procoagulant activity. The activation coincides with proteolysis of both the heavy and light chains. The heavy chain is first cleaved at position 740 which releases the "B" domain. Then two further thrombin cleavages occur at positions 372 and 1689 resulting in the heavy chain splitting into 50 and 43 kDa fragments and the light chain being cleaved into a 73 kDa fragment. By site-directed mutagenesis at these residues it was shown that the cleavages at 372 and 1689 are in fact crucial for activation (Pittman and Kaufman, 1988; figure 2). The codons for the arginine residues at 372 and 1689 have both been affected by recurrent mutations in unrelated haemophilia A patients due to the hypermutable CpG dinucleotides at these sites (Pattinson et al., 1990a; Tuddenham et al., 1991). Hill-Eubanks and Lollar (1990) showed that cleavage at position 372 was the rate-limiting step in the activation of factor VIII. Inactivation of factor VIII is associated with proteolytic cleavage by activated protein C at positions 336 and 562 (Fulcher et al., 1984; Fay et al., 1991; Walker and Fay, 1992). There is also evidence to indicate that the inactivation of factor VIII takes place as a result of spontaneous dissociation of the low-solubility A2 subunit (Lollar and Parker, 1990).

The factor VIII molecule is known to interact with vWF, factors Xa and IXa, phospholipid, calcium ions and perhaps with copper ions. The residues for the binding of vWF were shown to map in the acidic region between positions 1649 and 1689. Deletional mutagenesis revealed the consequent loss of the normal stabilising effect of vWF *in vitro* (Pittman and Kaufman, 1989). The sulphated tyrosine residues at 1664 and 1680 were mutated to phenylalanine (a conservative change that would block sulphation at these residues), which showed the same effect. Further, monoclonal antibodies that recognise the aspartic acid and tyrosine residues at 1679 and 1680 inhibited the interaction with vWF (Foster et al., 1988). Thus residues at 1664 and 1680 are likely to be responsible for vWF binding of the factor VIII molecule.

The C2 domain has been shown to be the site of phospholipid binding to factor VIII. Antibodies to the factor VIII light chain that recognise epitopes between residues 2178-2332 (Arai et al., 1989b) and peptides spanning residues 2302-2332 (Foster et al., 1990), were shown to inhibit the binding of phosphatidylserine to factor VIII.

The homology of factor VIII with ceruloplasmin suggests the possible involvement of factor VIII with copper or other metal ions. The residues implicated for copper binding in ceruloplasmin are conserved in the A1 and A3 domains of factor VIII, which may be involved in some metal ion binding function (Vehar et al., 1984).

1.3.3 THE FACTOR IX GENE

Cloning of the factor IX gene

By 1982, 20% of the human (DiScipio et al., 1978) and 100% of the bovine (Katayama et al., 1979) factor IX amino acid sequence had been determined. Two groups, used this information and screened human cDNA libraries using a mixture of synthetic oligonucleotides (Kurachi and Davie, 1982; Choo et al., 1982). Similar methods were used by other groups to isolate human factor IX cDNA clones (Jaye et al., 1983; Anson et al., 1984; Jagadeeswaran et al., 1984). Kurachi and Davie (1982) also used another probe for this purpose, where they obtained a single stranded DNA probe prepared from enriched mRNA for baboon factor IX. This was obtained by injecting a baboon with goat anti-human factor IX antibodies, followed by extraction of the enriched poly (A) RNA from its liver.

Following the isolation of the factor IX cDNA, these clones were used to isolate genomic DNA clones containing the factor IX gene. Anson et al. (1984) thus reported the characterisation and structure of the factor IX gene. Yoshitake et al. (1985) reported the sequence of the entire 33.5 kb factor IX gene.

Structure of the factor IX gene and cDNA

The entire factor IX gene has been sequenced (Yoshitake et al., 1985). The gene spans 33.5 kb of genomic DNA at Xq27.1. It has eight exons (I-VIII) and seven introns (A-G) (Anson et al., 1984; Yoshitake et al., 1985). The exons range in size from 25 bp (exon III, which codes for only 8 amino acids), to 1,935 bp (exon VIII, which encodes

182 amino acids and forms the 1,390 nucleotide 3' untranslated sequence). The introns range in size from 188 bp (intron B) to 9,473 bp (intron F). All splice sites conform to the GT..AG rule (Breathnach and Chambon, 1981). The seven introns in the factor IX gene are located essentially in the same position as those of the gene for human protein C, and the first three are in positions identical to those in the human prothrombin gene (Yoshitake et al., 1985).

There are several potential "TATA" sequences upstream from the mRNA start site (Anson et al., 1984; Yoshitake et al., 1985; Salier et al., 1990). The "TGTA" observed at position -27 may serve as the "TATA box" sequence. The "G" in this sequence is unusual but not unknown (Breathnach and Chambon, 1981). No "CCAAT" box was observed upstream of this box. The polyadenylation signal sequence "AATAAA" was observed 16 nucleotides before the addition of the poly (A) tail (1,367 nucleotides downstream of the "TAA" stop codon; Yoshitake et al., 1985).

The factor IX cDNA is 2,802 bp long. It has a 29 nucleotide 5' non-coding sequence and a 3' non-coding sequence of 1,390 nucleotides. The translation initiation codon is at position 30 and the termination codon is just after position 1412. The codons -46 to -1 encode the precursor region, where -46 to -21 codes for the hydrophobic signal peptide and -20 to -1 the hydrophilic propeptide. The mature peptide is encoded by codons 1 - 415 (Anson et al., 1984).

1.3.4 THE FACTOR IX PROTEIN

Factor IX is a vitamin K-dependent, serine protease, that circulates in the plasma as a 415 amino acid zymogen. In the intrinsic coagulation cascade, its activated form proteolytically activates factor X in the presence of factor VIIIa, calcium ions and negatively charged phospholipids. It has a domain structure very similar to that of other vitamin K-dependent proteins in the coagulation pathway, namely, prothrombin, factors VII and X, protein C and protein S (reviewed in Furie and Furie, 1988).

The eight exons (I-VIII) of the factor IX gene code for six major domains of factor IX. Exon I (-46 to -17) codes for the hydrophobic signal peptide which targets the protein for secretion from the hepatocyte into the blood stream. Exons II (-17 to 37) and

III (38 to 47) encode the propeptide and the *gla* domain. The latter domain has the 12 N-terminal glutamyl residues (the first 11 encoded by exon II and the last one by exon III) that are modified to γ-carboxyglutamyl residues by vitamin K-dependent carboxylase (Vermeer, 1990). This post-translational modification occurs in the endoplasmic reticulum of the hepatocyte and is essential for this domain to bind calcium ions and negatively charged phospholipid surfaces. Exon IV (47 to 85) encodes the first (type B) epidermal growth factor-like domain which is homologous to epidermal growth factor (EGF) and has conserved carboxylate residues, including a β-hydroxyaspartate at amino acid position 64 (Stenflo, 1991). This domain is involved in calcium ion binding (Handford et al., 1990) and may also bind factor VIII. Exon V (85 to 128) codes for the second (type A) epidermal growth factor-like domain whose function is not known. Exon VI (128 to 195) codes for the activation domain within which factorIX a is cleaved twice (at positions 145 and 180), thus activating factor IX. Exons VII (196 to 234) and VIII (234 to 415) encode the serine protease or catalytic domain which is responsible for the proteolytic activation of factor X to Xa. This region is homologous to other well studied serine proteases or chymotrypsin-like enzymes, and it is thought likely that the conserved residues, histidine-221 (encoded by exon VII), aspartate-269 and serine-365 (both encoded by exon VIII), all participate in the classical catalytic mechanism.

Factor IX is initially synthesised as a precursor molecule, some 40 residues longer at its N-terminus than the 415 amino acid peptide found circulating in the plasma. The sequential removal of the hydrophobic signal peptide and propeptide occur in the hepatocyte prior to secretion (Bristol et al., 1993). In addition to the γ-carboxylation of the N-terminal glutamyl residues (Vermeer, 1990) and the partial β-hydroxylation of aspartate at 64 (Stenflo, 1991), N-linked carbohydrate side chains are added at residues 157 and 167 and an O-linked carbohydrate at serine-53 and 61 (Hase et al., 1990; Nishimura et al., 1992).

1.4 BIOCHEMICAL STUDIES

Screening tests of the coagulation cascade include the activated partial thromboplastin time (APTT) for the intrinsic system, prothrombin time for the extrinsic

system and fibrinogen levels and thrombin time for the dysfibrinogenaemias. Tests used in the initial investigation of bleeding disorders include, APTT, prothrombin time, thrombin clotting time, bleeding time and platelet count. If an abnormality in any of these tests is observed, then the appropriate coagulation factors are assayed. In mild deficiencies, screening tests may not reveal the defect and a strong clinical suspicion may warrant a coagulation factor assay.

The activities of factor VIII and IX can be measured by functional or immunological assays. The former activity is termed factor VIII:C and IX:C, respectively, and can be measured by clotting and chromogenic assays. The immunological activity, termed factor VIII:Ag and IX:Ag, respectively, are measured using polyclonal or monoclonal antibodies. Individuals with haemophilia A have a good correlation between the functional and antigenic assays, and factor VIII levels are either not detectable or are comparably reduced in most cases. In a few (~5%), however, a functionally defective protein is produced and the antigenic levels are considerably higher than the functional assay would suggest. The latter are termed cross-reacting-material positive (CRM^+) haemophilia. A significantly higher proportion of haemophilia B (>30%) patients fall in this category (Bertina and Veltkamp, 1978; Kasper et al., 1977). Since these mutations affect mainly the function of an otherwise stable clotting factor molecule, delineation of these help in correlation with functionally important residues / domains of the protein (Tuddenham et al., 1991; McGinniss et al., 1993).

Although the factor VIII:C and IX:C levels vary widely among normal males (50-150 IU/dl), there is good correlation between levels of coagulant activity and clinical severity. Levels of <2, 2-5 and >5 IU/dl represent severe, moderate and mild clinical phenotypes, respectively.

Patients with von Willebrand's disease also have factor VIII deficiency. The vast majority of these patients can be distinguished from haemophilia A by the bleeding profile that mainly involves bruising and gastro-intestinal bleeding, von Willebrand factor antigen level and ristocetin cofactor activity. Rarely, however, a variant form of von Willebrand's disease, where the defect lies in the binding of factor VIII by vWF (Nishino et al., 1989; Mazurier et al., 1990; Gaucher et al., 1991), could be confused with haemophilia A. This

condition mimics haemophilia A on standard laboratory tests and it is necessary to measure factor VIII-vWF binding to establish the diagnosis.

A particular problem arises when it comes to diagnosing female carriers of haemophilia A or B. Although a female carrier is heterozygous for the deleterious mutation at the factor VIII or IX locus and should theoretically have half the normal level of the concerned clotting factor, due to random X-inactivation this is not always the case. In the case of haemophilia A, the measurement of a ratio of factor VIII to vWF improves the accuracy of carrier prediction (Klein et al., 1977), but even this may result in misclassification of up to 30% of haemophilia A carriers (Antonarakis et al., 1985). The problem is even greater in haemophilia B, where 30-50% mutations result in a CRM^+ phenotype. In this case, measurement of both, factor IX:C and factor IX antigen levels as a ratio helps to predict the carrier status. The reliance on only biochemical parameters for carrier assignment is not entirely satisfactory, and presently, the diagnosis of female carriers of haemophilia A and B is primarily performed by DNA-based methods.

1.5 GENETIC PREDICTION

Due to the fact that the factor VIII and IX genes are affected by a heterogeneous group of mutations, of which the vast majority include single base alterations, no uniform direct mutation detection strategy is feasible for the genetic prediction of either of these conditions. Haemophilia A and B were among the first common genetic disorders for which linked markers became available to help with carrier assignment and prenatal diagnosis (Gitschier et al., 1985a,b; Antonarakis et al., 1985; Winship et al., 1984; Giannelli et al., 1984). Several intragenic and extragenic markers have since been described for this purpose. These were initially described as restriction fragment length polymorphisms (RFLPs) detected by Southern blotting and hybridisation with appropriate radiolabelled DNA probes. Although most of these RFLPs are still being used, they have now been replaced by polymerase chain reaction (PCR) based detection methods (Kogan et al., 1987; Kogan and Gitschier, 1990). Genetic prediction using the battery of polymorphic markers can predict the carrier status of up to 70% of female relatives (of

familial cases) at risk. In addition to these di-allelic polymorphisms, a different class of PCR-based, highly polymorphic dinucleotide repeat elements have been identified, which greatly increase the feasibility and accuracy of genetic prediction (Lalloz et al., 1991).

Gene tracking using linked markers has certain distinct disadvantages. A family segregation analysis entails the collection of samples from a number of individuals, some of whom may be either dead, otherwise unavailable or even unwilling to comply. In order to be able to differentiate between the two X chromosomes in a carrier female, the individual has to be heterozygous (informative) for these markers. Even though a large number of RFLPs have been identified in and around the factor VIII and IX genes, it is not always possible to get informative markers (although this is now changing due to the availability of highly polymorphic repeat elements). Another problem is the reduction in the useful information from multiple markers because of a high percentage of linkage disequilibrium amongst the markers (Connor et al., 1986; Kogan and Gitschier, 1990). For the large factor VIII gene, recombination between the causative mutation and the linked marker may necessitate the use of flanking markers. A third of haemophilia cases represent new mutations (true sporadic cases) and therefore it is not possible to use this system, which essentially relies on the assumption that one of the X chromosomes of the mother of the affected male bears the causative mutation. This is further complicated by the possibility of gonadal mosaicism. Interpretation of the segregation analysis relies on the accuracy of the pedigree information obtained. This is clearly not always the case, as was exemplified by the finding of a relatively high level of non-paternity among haemophilia B families in the West of Scotland (Bidichandani, 1991).

1.6 MUTATION ANALYSIS

1.6.1 IMPORTANCE OF MUTATION DETECTION

In the field of human genetics, detection of naturally occurring germline/somatic mutations are used to determine whether a candidate gene is causally related to a phenotype or to identify new alleles at a known locus. The latter involves the detection of

novel and recurrent mutations in previously characterised genes. While the former is essential to confirm the causative role of a particular gene in the pathogenesis of human disease, the latter expands on the biological properties of the gene/protein and is useful for genetic diagnosis. Analysis of mutations at a previously characterised locus allow the determination of: (i) mutational spectrum, for example: ~65% of patients with Duchenne/Becker muscular dystrophy (DMD/BMD) have large deletions (den Dunnen et al., 1989), >80% of haemophilia B patients have point mutations (Giannelli et al., 1993) (ii) mutational distribution and the identification of hotspots, for example: deletion hotspots in the dystrophin gene (Forrest et al., 1987, 1988; Wapenaar et al., 1988) and the clustering of mutations affecting the first nucleotide-binding fold of the CFTR gene (exon 11) in cystic fibrosis patients (Cutting et al., 1990), (iii) identification of functional domains in proteins, for example: the GAP-related domain in the neurofibromatosis type 1 gene (Li et al., 1992), (iv) population genetics, for example: the geographic variation in the frequency of the ΔF508 mutation in Europe (European Working Group on CF Genetics, 1990), (v) identification of novel mechanisms of mutagenesis, for example: trinucleotide repeat expansions in fragile X syndrome (Fu et al., 1991; Verkerk et al., 1991; Oberle et al., 1991; Yu et al., 1991) and myotonic dystrophy (Harley et al., 1992; Buxton et al., 1992) and (vi) the role of somatic mutations, for example: in the *ras* group of oncogenes in human malignancies (Bos, 1989).

At the start of the present project, the molecular pathology was not known in about 50% of severe haemophilia A patients (Higuchi et al., 1991a,b). This meant that the knowledge of the spectrum and distribution of factor VIII mutations was probably incomplete. Although various site-directed mutagenesis studies and other *in vitro* studies have contributed to an understanding of the structure-function relationship of the factor VIII protein, this knowledge is rather limited (reviewed by Pittman and Kaufman, 1989). It is hoped that with careful cataloguing and analysis of many naturally occurring factor VIII mutations, a better understanding of the functions of the various domains of the factor VIII protein might be possible (Tuddenham et al., 1991).

1.6.2 METHODOLOGY OF MUTATION DETECTION

In general, the methods for the detection of mutations can be subdivided into those that are designed to predominantly detect small or large sequence alterations. The detection of mutations has been revolutionised with the introduction of PCR technology (Saiki et al., 1988) and presently, the detection of small sequence alterations is almost exclusively performed by PCR-based methods (Grompe, 1993). The methods for the detection of large mutations include the use of Southern blotting (Southern, 1975), pulse field gel electrophoresis (PFGE; Smith et al., 1988) and Northern blotting (Alwine et al., 1977).

The strategy used for the detection of mutations greatly depends on the known (or expected) spectrum of mutations affecting a particular gene. This usually means tailoring the strategy to suit a particular disease. For example in DMD/BMD, where the majority of cases have gross deletions (or duplications), either Southern blotting or PFGE is used as an initial screen. As the dystrophin gene is on the X-chromosome, hemizygous males are detected with relative ease. With the identification of deletion hotspots and the optimisation (and commercial availability) of a multiplex PCR strategy (Chamberlain et al., 1988), it is now possible to rapidly detect the deletion in ~60% of cases. In a disease like Charcot-Marie-Tooth disease type 1A where >90% cases are due to a large DNA duplication involving a 1.5 Mb monomer unit (Lupski et al., 1991), PFGE is used as the first diagnostic test before using other PCR-based methods to detect smaller sequence alterations (Roa et al., 1993).

Both, genomic DNA and mRNA can be used as templates for the PCR-based mutation detection strategies. The advantage of using mRNA (reverse transcribed to cDNA) as the template are, that large non-coding regions are omitted thus permitting the analysis of a potentially smaller target region (with fewer PCRs) and the rapid detection of gross deletions or splicing errors. Until recently, this was not possible for genes expressed in specific tissues (example dystrophin). Chelly et al. (1988, 1989) and Sarkar and Sommer (1989) demonstrated that all genes were expressed in all cell types, albeit at low levels. This made it possible to exploit the extreme sensitivity and specificity of PCR to analyse cDNAs of large tissue-specific genes for the detection of disease causing

mutations (Roberts et al., 1990; Berg et al., 1990; Naylor et al., 1991). For small genes with fewer exons, it is possible to achieve the amplification of the entire essential sequence using a few DNA-based PCRs (example for the β-globin gene).

Mutations (especially small sequence alterations) are ultimately characterised by nucleotide sequencing. It is now possible to directly sequence PCR products without the necessity for prior cloning (Gyllensten, 1989). In order to avoid having to sequence the entire amplified template, several mutation screening strategies have been developed and used with varying degrees of success (Grompe, 1993). The various strategies have varying capacities for mutation detection and localisation. Some genes with relatively short coding sequences have been directly sequenced in order to detect mutations (porphobilinogen deaminase cDNA in acute intermittent porphyria, Mgone et al., 1992). The most commonly used mutation screening methods include, (i) analysis of single strand conformation polymorphisms (SSCP; Orita et al., 1989), (ii) heteroduplex analysis (HA; White et al., 1992), (iii) denaturing gradient gel electrophoresis (DGGE; Fischer and Lerman, 1983; Myers et al., 1985a), (iv) chemical cleavage of mismatches (CCM; Cotton et al., 1988) and (v) direct sequencing (Gyllensten, 1989).

SSCP is based on the principle that the electrophoretic mobility of a single stranded DNA molecule is a function of both its length and its nucleotide sequence. Double stranded PCR products when denatured and electrophoresed on a non-denaturing polyacrylamide gel (under conditions that enhance the formation of conformations in single strands), show altered mobilities caused by differences as small as single base pair substitutions. The detectability is greatly increased in smaller sized PCR products, being ~79-95% in <200 bp segments, and rapidly falls to <50% when the size is >400 bp (reviewed by Hayashi, 1991 and Grompe, 1993). HA is based on the principle that the heteroduplex formed between wild-type and mutant DNA segments has a different secondary structure from either wild-type or mutant homoduplexes, and would be detected as altered electrophoretic mobilities on non-denaturing gels. White et al. (1992) reported a higher detectability rate than SSCP, but since the same PCR product can be used for both methods, it is possible to potentially complement each technique. Furthermore, novel gel matrices have been developed (Hydrolink™ and MDE™; Hoefer)

in order to increase the sensitivity of SSCP and HA. DGGE is based on the principle that when double stranded DNA is electrophoresed through a gradient of denaturant, it migrates up to a certain point where the strands melt and no further migration occurs (Fischer and Lerman, 1983). Since the melting behaviour of a double stranded DNA molecule is the function of its nucleotide sequence (melting domains), a single base substitution in a heteroduplex can be detected as an altered mobility on such a denaturing gradient gel. Incorporation of GC-clamps (at the 5' ends) increases the sensitivity of this method to over 90%. This method however requires the use of specialist equipment, longer primers with GC-clamps and does not accurately localise the mutation in a given DNA segment.

CCM is a mutation detection method that has the potential for detecting and accurately localising all mutations in a given PCR product. Unlike SSCP, HA and DGGE, it is not dependent on the size of the DNA fragment being analysed (up to 1.7 kb, Grompe, 1993) or the electrophoretic conditions. This method however is lengthy, complicated and makes use of toxic chemicals. CCM is based on the chemical reactivity of mismatched "C" or "T" bases in DNA-DNA hybrids, that make them susceptible to cleavage with piperidine (Cotton et al., 1988). This method is not only extremely sensitive but also aids in the accurate localisation of the mutation by estimation of the size of the cleavage product. Prior to the description of this method, other less sensitive methods had been described that were based on a similar principle. Cleavage of the single stranded mismatch was attempted using S1 nuclease (in DNA-DNA hybrids; Shenk et al., 1975), RNase A (in RNA-DNA hybrids; Myers et al., 1985b) and reactivity with carbodiimide (in DNA-DNA hybrids; Novack et al., 1986).

Finally, all positive mutation screening results have to be completely characterised by nucleotide sequencing. Direct sequencing of PCR products has also been used as the primary mutation screening method to detect mutations in parts of large genes (Higuchi et al., 1990) or by sequencing entire coding sequences (Mgone et al., 1992). This is however very labour intensive and perhaps even unnecessary due to the availability of various rapid mutation screening methods. In the direct sequencing of PCR products, the quality of sequence was shown to be superior when a preliminary asymmetric PCR was

performed in order to generate single stranded PCR products (Gyllensten and Ehrlich, 1988; Gyllensten, 1989; Mgone et al., 1992).

Mutations that have been detected by the above methods may need to be screened for the purpose of familial segregation or population studies. This can be easily achieved by one of the following methods: (i) restriction enzyme digestion (if a site has been created or abolished), (ii) allele-specific oligonucleotide discriminant hybridisation (ASO; Connor et al., 1983), (iii) allele-specific PCR amplifications (ASA/ARMS; Wu et al., 1989; Sarkar et al., 1990) or (iv) the same screening method that was used to detect it originally (example SSCP or HA).

1.7 DETECTION OF MUTATIONS IN THE FACTOR VIII GENE

1.7.1 DIFFICULTIES ASSOCIATED WITH DETECTION OF MUTATIONS IN THE FACTOR VIII GENE

The main factors complicating the detection of mutations in the factor VIII gene are: (i) a mutational spectrum that largely comprises point mutations, (ii) the high *de novo* mutation rate with the consequence that most unrelated patients have unique mutations, (iii) the large size of the gene and (iv) the hepatocyte-specific expression of factor VIII.

Haemophilia A, with an incidence of 1 in 5000 males, is a relatively common monogenic disease affecting man. Unlike other common genetic diseases like cystic fibrosis (CF), Charcot-Marie-Tooth disease type 1A (CMT1A), fragile X syndrome and Duchenne/Becker muscular dystrophy (DMD/BMD) that are mainly caused by one common mutation or type of mutation, a heterogenous group of mutations have been reported in the factor VIII gene of haemophilia A patients (Tuddenham et al., 1991). Recently a novel mechanism of mutagenesis was discovered that implicated the sequence of *F8A*, a gene embedded in intron 22 of the factor VIII gene, in the causation of half the severe cases of haemophilia A (Lakich et al., 1993; Naylor et al., 1993b). However, the vast majority of patients, perhaps over 70%, have a variety of single base substitutions. Due to the high *de novo* mutation rate affecting the factor VIII gene, the vast majority of

these are unique mutations (a few recurrences have nevertheless been reported, Tuddenham et al., 1991). The fact that unique single base substitutions (or small sequence alterations) have to be identified in the vast majority of haemophilia A patients, is the main reason complicating the analysis of factor VIII mutations.

Furthermore, the factor VIII gene is very large, comprising 26 exons, scattered over 186 kb of genomic DNA (Gitschier et al., 1984). The factor VIII gene produces a 9 kb, hepatocyte-specific transcript, that codes for a single polypeptide of 2332 amino acids preceded by a 19 amino acid hydrophobic leader sequence (Wood et al., 1984; Toole et al., 1984; Vehar et al., 1984). For this reason, until recently, most groups have only screened limited sections of the factor VIII gene (Gitschier et al., 1985c; Higuchi et al., 1990; Pattinson et al., 1990a; Traystman et al., 1990). Recently however, at least four groups have screened the entire essential sequence of the factor VIII gene i.e., including the complete coding sequence, splice sites, putative promoter and polyadenylation / cleavage sequences (Higuchi et al., 1991a,b; Naylor et al., 1991, 1993a; Diamond et al., 1992; Lin et al., 1993).

The factor VIII mRNA is expressed predominantly in hepatocytes and represents about 5% of the size of the factor VIII gene (Gitschier et al., 1984; Wood et al., 1984; Toole et al., 1984). There are several advantages in searching for mutations within the factor VIII mRNA. These include: (i) screening of a relatively shorter target region is required (due to the exclusion of large intronic sequences), (ii) screening of only the coding sequence increases the probability for most mutations to be pathogenic and (iii) errors of splicing and large gene alterations are readily detectable. Until the discovery of ectopic / illegitimate transcription, it was thought that liver biopsies would be required to allow such an analysis. Chelly et al. (1988) discovered spliced dystrophin transcripts (normally expressed in muscle and brain) in cultured fibroblasts, lymphoblasts and Hep G2 hepatoma cells. They and others later reported (Chelly et al., 1989; Sarkar and Sommer, 1989) the detection of spliced forms of a variety of tissue-specific mRNAs in various cell types i.e., concluding that any gene was transcribed (albeit at low levels) in any cell type. Chelly et al. (1988) estimated that these illegitimate or ectopic transcripts were expressed at a level of less than 1 in every 500 to 1000 non-specific cells. The

discovery of this amazing phenomenon obviated the need to obtain cells that predominantly express a tissue-specific gene. In fact recently, RT-PCR of various large tissue-specific mRNAs have been used as templates for various mutation screening strategies (Berg et al., 1990; Chelly et al., 1991; Chalkey and Harris, 1991; Roberts et al., 1991, 1992; Naylor et al., 1991, 1993a; Fonknechten et al., 1992). In order to facilitate a comprehensive search for mutations, the method described by Naylor et al. (1991, 1993a) i.e., using a combination of RT-PCR of ectopic transcripts and DNA-PCR to amplify the entire essential sequence of the factor VIII gene, was employed in this project.

1.7.2 MUTATIONAL SPECTRUM IN HAEMOPHILIA A

Haemophilia A is the most common severe inherited disease of blood coagulation, affecting approximately 1 in 5,000 males. It is caused by a heterogeneous group of mutations, predominantly point mutations, affecting the factor VIII gene.

By analysing the factor VIII coding sequence, adjacent intron sequences, putative promoter and polyadenylation signal sequences in DNA, Higuchi et al. (1991a,b) aimed to define the causative mutation in every haemophilia A patient analysed. Mutations were found in almost all the mild and moderate cases, but were identified in only about half of the severe cases. This led to speculation that causative mutations in these cases might lie outside the regions analysed, for example, in other expression regulatory elements, within introns or even at another (tightly linked) locus. Subsequently, Naylor et al. (1992, 1993a) made a discovery using RT-PCR of ectopic factor VIII mRNA transcripts, that it was not possible to perform an amplification across the junction of exons 22 and 23 in 40% of their severe cases, whereas, amplification on either side yielded products of the expected size. Naylor et al. (1992, 1993a) concluded that changes in intron 22 were responsible for half the cases of severe haemophilia A.

Intron 22 is the largest intron in the factor VIII gene and has two genes emanating from a CpG island within it (Levinson et al., 1990, 1992). *F8A*, which is fully embedded in intron 22 and transcribed in a direction opposite to that of factor VIII, has two additional copies approximately 500 kb upstream of the factor VIII gene (Patterson et al., 1989; Levinson et al., 1990; Freije et al., 1992; Lakich et al., 1993). It was subsequently

shown by two groups (Lakich et al., 1993; Naylor et al., 1993b) that the mutation affecting half the severe cases of haemophilia A was due to a homologous recombination between the intron 22 embedded copy of *F8A* and one of the two additional copies upstream (telomeric on the X chromosome) of the factor VIII gene. This crossover results in a large inversion, dividing the factor VIII gene into two parts, with an intact promoter and exons 1-22 greatly separated from, and in opposite orientation to exons 23-26. At least six groups have now shown that this mutation is responsible for up to 45% of all severe cases or approximately 20% of all haemophilia A patients (Lakich et al., 1993; Naylor et al., 1993a; Goodeve et al., 1994; Collins et al., 1994; Ljung, 1994; Tizzano et al., 1994). It was observed that, of the two homologues of *F8A*, recombination with the distal copy was more common than with the proximal one (Collins et al., 1994; Ljung, 1994). A rare third variant pattern was seen on Southern blot analysis, that indicated the presence of additional copies of *F8A* (Collins et al., 1994 observed it in 1 of their 40 patients with this inversion; Ljung, 1994 observed it in 1 of their 19 inversion cases; Rossiter et al., 1994). Furthermore, Rossiter et al. (1994) tested and confirmed their hypothesis that this inversion was possible only in male meiosis, probably because the lone X chromosome does not have a homologue to pair with.

The rest of the severe haemophilia A cases are caused by a variety of mutations including, minor and major gene deletions or insertions and point mutations that include nonsense, missense and those causing errors of splicing (Tuddenham et al., 1991). The frequency of large deletions of the factor VIII gene determined by Southern blotting in a recent study of pooled data was 2.5% (34 out of 1386 haemophilia A patients; Millar et al., 1990). The frequency of small deletions / insertions (ranging between 1 - 23 bp) was determined to be about 7% of all haemophilia A patients, in a comprehensive search for mutations by Higuchi et al. (1991a,b). Over 60 partial deletions have been incorporated in the database of all mutations affecting the factor VIII gene (Tuddenham et al., 1991). All but 3 of these result in a severe phenotype. Three in-frame deletions, two of exon 22 and one involving exons 23 and 24 resulted in moderately severe haemophilia A. No evidence of hotspots have been identified for the large deletions. Two of the small deletions described in the database (involving 2 and 4 bp) occurred in the same amino acid position

340-341 and involved the postulated deletion hotspot consensus sequence of Krawczak and Cooper (1991). Antonarakis and Kazazian (1988) reported in two patients with severe haemophilia A, insertions of partial *LINE*-1 (L1) repeat elements. Both insertions were in exon 14, involved the 3' portions of the L1 sequence (including the poly (A) tract) and occurred within 12-13 nucleotide duplication sites.

Mild and moderate haemophilia A on the other hand, almost always results from a wide variety of missense mutations that presumably affect regions important for factor VIII function and/or stability. In general all nonsense mutations result in severe disease and all point mutations causing mild and moderate phenotypes are of the missense type (Tuddenham et al., 1991).

An excess of point mutations are reported to affect exons 8, 14, 18, 23, 24 and 26. This is likely to be due to the large size of exons 14 and 26 and due to the fact that mutations at CpG sites in exons 18, 23, 24 and 26 were specifically looked for using restriction digestion (Gitschier et al., 1985c). Evidence from groups searching the entire factor VIII coding sequence and from the database of haemophilia A mutations, suggest that there may be a disproportionate harbouring of mutations in the small exons 7, 11, 12 and 16 (Tuddenham et al., 1991).

38% of point mutations reported in the haemophilia A mutation database are at hypermutable CpG dinucleotides resulting in CG→TG/CA substitutions. This is consistent with deamination of the methylated cytosine (5-methyl cytosine) to thymine, which is not corrected by the DNA repair mechanisms and results in either of the above changes depending on its occurrence on the sense or antisense strand (Cooper and Youssouffian, 1988; Cooper and Krawczak, 1990). The true frequency is probably lower, perhaps 32% as gathered from Higuchi et al. (1991a,b), because some of the point mutations reported in the database have been detected by the deliberate screening for CpG dinucleotide mutations (Gitschier et al., 1985c, 1988; Pattinson et al., 1990a). There are a total of 71 CpG dinucleotides in the factor VIII cDNA and recurrent mutation in at least 16 such sites has been reported in unrelated haemophilia A patients (Tuddenham et al., 1991). A similar directed search for nonsense mutations at 12 CpG sites was

conducted by Reiner and Thompson (1992). Recurrence at non-CpG sites is rarely seen and when seen, identity by descent may account for the majority of these.

About half of all haemophilia A patients have no detectable factor VIII protein levels (termed Cross Reacting Material Negative, CRM^-) and ~5% have an excess of factor VIII antigen over functional activity due to the presence of a dysfunctional factor VIII molecule (termed Cross Reacting Material Positive, CRM^+). About 45% have reduced factor VIII levels comparable to the reductions in factor VIII coagulant activity (CRM reduced). Mutations resulting in the CRM^+ or CRM reduced phenotype are considered to be those that affect sites critical for factor VIII function. Indeed CRM^+ haemophilia A has been caused by recurrent mutations in the two arginine residues at 372 and 1689 (Gitschier et al., 1988; Arai et al., 1989a, 1990; Shima et al., 1989; Higuchi et al., 1990; Pattinson et al., 1990a,b; Schwaab et al., 1991), which are thrombin cleavage sites for the proteolytic activation of factor VIII. This phenotype was also seen with a missense mutation affecting the putative binding site for von Willebrand factor (Y1680F; Higuchi et al., 1990). In this respect, it is important to identify those mutations resulting in a CRM^+ phenotype in order to delineate those residues critically important for factor VIII function. In one such study, 11 of the 26 mutations resulting in a CRM^+ or reduced phenotype were found to affect the A2 domain of factor VIII (McGinniss et al., 1993).

Inhibitor development is an important complication associated with factor VIII replacement therapy (seen in ~20% of those who receive exogenous factor VIII). Deletion of the factor VIII gene is associated with a five-fold higher risk of this complication. Inhibitor development has been reported for 12 patients with point mutations described in the database of mutations causing haemophilia A (Tuddenham et al., 1991). Ten of these are in severe haemophiliacs with nonsense mutations. Goodeve et al. (1994) reported that none of their 10 patients with the intron 22-mediated inversion mutation developed inhibitors, whereas 7 of 12 without the inversion did. Two subsequent studies showed this to be atypical and thus ruled out any such correlation (Ljung, 1994; Tizzano et al., 1994).

1.8 GENE THERAPY

1.8.1 DEFINITIONS

Gene therapy is defined as "the transfer of genetic material into the cells of an organism to treat disease" (Miller, 1990). Although theoretically it is possible to transfer genetic material into both somatic (somatic gene therapy) and germ cells (germline gene therapy), for socio-ethical (and medical) reasons only somatic cells are presently being targeted. This has led to some investigators defining gene therapy as "the delivery of a functional gene for expression in somatic tissues with the intent to cure disease" (Mitani and Caskey, 1993). A document published by the Center for Biologics Evaluation and Research at the Food and Drug Administration (FDA) in the USA (1991), used a similar definition but extended the application of somatic gene therapy for prophylactic and diagnostic uses as well. So far there has not been much success with homologous recombination (Smithies et al., 1985) to correct a defective gene by physically replacing it i.e., "replacement therapy". At present, therefore, gene therapy can be regarded as "gene supplementation" since it involves the addition of corrective genes to the host cell genome (Akhurst, 1989). Consequently, most work has focused on the treatment of recessive single gene disorders.

Somatic gene therapy can be delivered by either an *ex vivo* or *in vivo* approach. In the *ex vivo* approach, cells are first harvested from an *in vivo* site and then expanded and manipulated *in vitro*. This is then followed by re-implantation of the genetically modified cells. The type of cells used for the *ex vivo* approach can be described as either autologous, allogenic or xenogenic living cells. Manipulation of cells *ex vivo*, may involve propagation, expansion, selection, and pharmacological or other biological modifications. Genetically modified cells can be returned by infusion, injection or surgical implantation. Advantages of the *ex vivo* approach are: (i) gene-transfer efficiency is generally high and a wide variety of transfer methods can be used, (ii) the transduced cells can be enriched if the vector has a selectable marker and the (iii) successful transduction and expression can be assessed before re-implantation. The *in vivo* approach involves the direct transfer of genes to cells residing in the body. Problems associated with this method include: (i)

requirement for a means of targeting the desired cell type or achieving tissue-specific expression, (ii) sufficient numbers of cells should be transduced, (iii) association with a low efficiency of stable gene transfer necessitates repeated treatment and (iv) problems with host immune response.

Proposals for gene therapy made by research groups are reviewed by various governmental bodies. In the USA various guidelines for human gene transfer have been adopted by the Recombinant DNA Advisory Committee (RAC; 1990) and the FDA (1991). The principle concerns include the safety of the procedure (for both the patient and the general public) and the anticipated benefit in comparison to the potential risk of the procedure. Since the first approved gene marking trial in 1989 (Rosenberg et al., 1990) and the first gene therapy trial in 1990 (Culver et al., 1991), many more clinical trials have been approved (Anderson, 1992; Miller, 1992). It is also now clear that even though the initially intended targets for this form of therapy were largely single gene disorders, acquired diseases such as malignancies and AIDS have recently become the focus of much attention (Anderson, 1992; Miller, 1992; Mulligan, 1993). On the whole, somatic gene therapy is now generally viewed as a viable and ethical therapeutic option for the treatment of serious disease (Anderson, 1992).

1.8.2 GENE DELIVERY SYSTEMS

Highly efficient gene transfer is essential for gene therapy. Several technologies have been considered to be applicable for human gene transfer. Although, the most successful and commonly used method includes the use of viral vectors, many non-viral methods are also promising.

Viral vectors

The use of viruses as gene transfer vehicles arose from the observation that viruses are extremely efficient at introducing genes into their host cells, and that they might therefore be useful as vectors for introducing genes in order to correct genetic diseases in man. It was proposed that if the deleterious features of these viruses were eliminated, they might be able to safely deliver a desirable gene to correct cellular defects caused by genetic disorders (Anderson, 1984; Freidmann, 1989).

Retroviral vectors are the best characterised viral vectors for human gene transfer and were the first to be employed in human gene therapy clinical trials (Culver et al., 1991; Anderson, 1992). Retroviral vectors transduce dividing cells in culture with an efficiency approaching 100% and stably integrate into the host cell genome without causing gross rearrangements. Although these vectors are ideal for *ex vivo* applications, several features of the transfer method may limit their applicability, particularly with regard to *in vivo* therapies. First, retroviral entry into cells is absolutely dependent on the existence of appropriate viral receptor on the target cell. Although packaging cell lines that generate vectors with a wide host range have been developed (Miller, 1990), problems have still been encountered in the transduction of certain cell types. Secondly, replication of the target cell is a prerequisite for proviral integration to occur. It is for this reason that it is necessary to perform partial hepatectomies in order to achieve a reasonable efficiency of hepatocyte transduction (Kay et al., 1993). Thirdly, the retroviral particle is relatively labile in comparison to other viruses and is rapidly inactivated *in vivo* in primates (Mulligan, 1993). In addition, questions regarding the safety of the procedure were raised after three monkeys recently developed malignant T-cell lymphomas after the transplantation of bone marrow that had been retrovirally transduced (due to contamination with helper virus; Miller, 1992).

Adenoviral vectors are capable of transducing a wide variety of cells and may be particularly attractive when considering gene transfer into non-dividing cells *in vivo*. (e.g. airway epithelium: Rosenfeld et al., 1991; 1992; cardiac and skeletal muscle: Stratford-Perricaudet et al., 1992; Ragot et al., 1993; central nervous system: Davidson et al., 1993; Akli et al., 1993; Bajocchi et al., 1993; hepatocytes: Smith et al., 1993). They are structurally stable and can be prepared at high titre (10^9-10^{11} plaque forming units per ml, as opposed to 10^6 for retroviral vectors) making them suitable for *in vivo* delivery. In addition, US military recruits vaccinated with wild type adenoviruses showed no side effects, demonstrating their safety for human use (Mitani and Caskey, 1993). Adenoviral vectors however, do not result in stable integration into the host cell genome. This feature may be advantageous from the point of view that insertional mutagenesis would not occur, but makes it necessary to perform repeated deliveries. Gene transfer with

adenoviral vectors *in vivo* has been complicated by the immune response that develops against the administered adenoviral particles, which also neutralises the effect of any further administrations (Smith et al., 1993).

Vectors derived from the non-pathogenic human adeno-associated virus (AAV) are also promising tools for human gene transfer. They specifically integrate into a region in chromosome 19. They however have a limited substitution capacity of ~5 kb and although not pathogenic themselves, they require co-infection with helper adenoviruses or other herpes viruses which may contaminate AAV stocks. Herpes simplex virus type 1 (HSV-1) vectors have been found to be useful for gene delivery to post-mitotic neurons. In common with adenoviruses, they too are not integrated into the host cell genome. Recently, the Sendai virus (Haemagglutinating virus of Japan [HVJ]; Kaneda et al., 1989) has been used for efficient *in vivo* delivery of genes (Dzau et al., 1993). In this method foreign DNA is complexed with liposomes, a nuclear protein and the viral protein coat of HVJ.

Non-viral delivery systems

The efficiency with which intact viruses introduce genes into cells led to the search to identify synthetic systems that could emulate viral delivery while avoiding viral infectivity. Cationic liposomes (cytofectins) have been developed that are much more efficient than the conventional liposome preparations which had much lower encapsulation and transfection efficiencies (Felgner and Ringold, 1989; Nabel and Felgner, 1993). Cationic-lipid-mediated gene transfer has been shown to be effective as an *in vivo* delivery system, by direct administration to the lung using aerosols or intubation (Yoshimura et al., 1992; Stribling et al., 1992).

In addition, direct application to the skin surface (Burns et al., 1991), intramuscular injections (Wolff et al., 1990), high energy microprojectile bombardment (Williams et al., 1991) and catheter-based delivery (Nabel et al., 1990; 1992) of genetic material have all been shown to be effective as *in vivo* delivery systems. The polylysine-ligand conjugate system (Wu et al., 1987; Cotten et al., 1992) has also shown promise as an alternate to viral methods for *in vivo* gene delivery. Transferrin-polylysine to target transferrin receptors (Wagner et al., 1992) and asialoglycoprotein-DNA conjugates to

target asialoglycoprotein receptors (Wu et al., 1991) have been shown to achieve hepatocyte-specific delivery.

1.8.3 CANDIDATE DISEASES AND TARGET TISSUES

Research in gene therapy has focused on a variety of diseases, most of which include recessive single-gene disorders that can be corrected by the addition of a functional gene. In cases like the clotting factor deficiencies resulting in the haemophilias, almost any somatic cell can be targeted, as long as the biologically-active gene product can enter the systemic circulation. On the other hand, correction of various diseases require the corrective gene product to be expressed in a tissue-specific manner, for example, β-globin in haematopoietic cells (in β-thalassaemia) and dystrophin in skeletal muscle cells (in Duchenne muscular dystrophy). Acquired diseases such as cancer, AIDS, certain cardiovascular and neurological disorders have also been considered as suitable candidates for somatic gene therapy.

In general, for the *ex vivo* strategy to be effective, it is necessary to select cells that can be easily obtained (biopsied) and manipulated *in vitro*. The *in vivo* approach is particularly attractive for those organs or tissues that cannot be easily accessed, as in the case of neurological or cardiovascular disorders. The various tissues that have been considered as suitable targets for somatic gene therapy include, haematopoietic stem cells, lymphocytes, hepatocytes, airway epithelial cells, fibroblasts, myoblasts, endothelial cells and keratinocytes.

Preclinical studies in the case of those diseases where the gene product needs to be delivered to the systemic circulation have generally been more successful than when a specific cell type has to be targeted. In this case the choice of target cells is governed by certain practical considerations: (i) the ease of isolation, (ii) feasibility of culture and genetic manipulation *in vitro*, (iii) half-life of the cells *in vivo*, (iv) capacity of the cells to express and actively secrete the particular protein, including the capacity to carryout post-translational modifications necessary for biological activity and (iv) capacity to make contact with the general circulation. More recently researchers have started using primary cells as opposed to established cell lines. Although a wide variety of cells have been used,

most progress has been made in the transplantation of transduced fibroblasts, myoblasts and keratinocytes (Mulligan, 1993). They are all readily explanted, grown in culture, genetically modified and transplanted. Fibroblasts can be returned to the body by subdermal injections. Considerable experience has been gained in grafting keratinocytes to heal epidermis in human burns patients (Gallico et al., 1984; Compton et al., 1989). Myoblasts injected directly into skeletal muscles fuse with existing muscle cells and continue to express the transferred gene (Dhawan et al., 1991; Dai et al., 1992). Each of these cell types have been induced to produce a variety of gene products following transduction with retroviral vectors. Fibroblasts have been made to express adenosine deaminase (ADA), glucocerebrosidase, purine nucleoside phosphorylase, low density lipoprotein receptors (LDL-R), factor VIII and factor IX. Myoblasts and keratinocytes have both been induced to produce factor IX (Dai et al., 1992; Gerrard et al., 1993) and growth hormone (Dhawan et al., 1991; Morgan et al., 1987; Teumer et al., 1990).

Success with somatic gene therapy when specific cell types have to be targeted has been less frequent. The most successful example was the transduction of autologous T-lymphocytes with retroviral vectors expressing the human adenosine deaminase (ADA) gene (Culver et al., 1991; Anderson, 1992). Many groups have reported transfer and long-term expression of genes into mouse haematopoietic stem cells (reviewed in Miller, 1990), but similar success has not been achieved in larger animals, like monkey or dog (Miller, 1990; 1992). The main problems in the transduction of haematopoietic stem cells are their sparse population, primarily quiescent nature and the unavailability of an effective assay to identify and quantify them. Moreover, transduction of haematopoietic stem cells with retroviral vectors expressing the β-globin gene did not yield high levels of expression (Dzierzak et al., 1988; Bender et al., 1989, Bodine et al., 1989), and has been one of the main reasons for the delay in progress towards gene therapy for β thalassaemia (which was the focus for most of the initial research in gene therapy; Anderson, 1992). Diseases such as hyperammonaemia due to ornithine transcarbamylase (OTC) deficiency, phenylketonuria and disorders of cholesterol metabolism are due to mutations that affect hepatic function and are considered to be candidate diseases for somatic gene therapy. Hepatocytes are natural targets for the correction of these disorders. Hepatocytes cannot

be extensively manipulated *in vitro* because they undergo only a few cell divisions and transduction efficiencies with retroviral vectors is rather low (Chowdhury et al., 1991). Both *ex vivo* and *in vivo* methods have been investigated. Obviously the *in vivo* methods would be more appropriate for clinical application. Both, adenoviral (Jaffe et al., 1992; Smith et al., 1993) and retroviral (Ferry et al., 1991, Kay et al., 1993) vectors have been shown to be efficient methods for direct gene delivery to the liver. Retroviral vectors require that the hepatocytes be actively proliferating, which has been achieved by partial hepatectomies (Kay et al., 1993). Adenoviral vectors do not integrate in the genome and their effect is therefore short-lived and re-infection does not result in a similar beneficial effect as with the first infection (Smith et al., 1993). Asialo-glycoprotein receptors have been exploited for *in vivo* delivery to the liver (Wu et al., 1991). Herpes simplex virus (type 1) vectors (HSV-1) have been used to effect *in vivo* gene delivery to the central nervous system (Breakfield and DeLuca, 1991; Friedmann and Jinnah, 1993). Respiratory epithelium has been targeted *in vivo* using direct instillation of adenoviral vectors (Rosenfeld et al., 1991, 1992) and by lipofection (Yoshimura et al., 1992).

More recently, acquired diseases have been considered as suitable targets for somatic gene therapy. Several modalities are being considered for the treatment of malignant disease and AIDS (Mulligan, 1993). Preclinical experiments for therapy of cardiovascular diseases (e.g. familial hypercholesterolaemia, restenosis after angioplasty) has been investigated by the *ex vivo* (genetically modified endothelial or smooth muscle cells; Wilson et al., 1989; Nabel et al., 1989; Dichek et al., 1989) and *in vivo* methods using various viral and nonviral methods (Nabel et al., 1992; Leclerc et al., 1992), including direct intracardiac injections (Lin et al., 1990).

1.9 GENE THERAPY FOR HAEMOPHILIA B

1.9.1 WHY GENE THERAPY FOR HAEMOPHILIA?

The main argument that may be forwarded in favour of developing gene therapy for haemophilia is the inadequacy of, and complications associated with, the current

therapeutic regimen. A major consideration for this disease is also the feasibility of a gene therapeutic strategy.

Currently, treatment is initiated only once bleeding starts and this "reactive strategy" results in considerable morbidity due to recurrent spontaneous bleeds. Due to the short half-life of the cofactors in circulation, repeated administrations are necessary. The treatment which depends on replacement of the deficient or defective clotting factor is not without complications. Although the spread of HIV and the hepatitis B virus have been greatly reduced by appropriate screening programmes and heat treatment, cofactor concentrates still transmit hepatitis non A-non B (C) infection (Brettler and Levine, 1989). This is likely to result in the development of chronic progressive liver disease in a number of these patients (Hay et al., 1985). The cost of the management of bleeding episodes is estimated to be between $50,000 to $100,000 per year, per severe haemophiliac (in the USA). This does not include the cost of treatment of associated complications and the loss of productivity caused by the loss of working hours from both physical and psychological effects.

The haemophilias are relatively easy diseases to treat by gene therapy, from a technological point of view. The factor VIII and IX genes have both been cloned and reasonably well studied. Clotting factor levels as low as 5 to 10% of normal can prevent spontaneous bleeds and protect against minor trauma, resulting in a considerable improvement in the quality of life. Moreover, the clotting factor levels do not have to be precisely regulated, unlike other proteins e.g. insulin. As long as a delivering tissue can express the active form of the deficient clotting factor and secrete it into the plasma, no special tissue-specific expression is required, unlike for certain other proteins e.g. dystrophin expression in skeletal muscle.

In summary, the current treatment is sub-optimal, expensive and associated with unacceptable complications and the delivery of both factor VIII and IX by gene therapeutic strategies can be considered as a feasible option.

Haemophilia A is at least five times more prevalent than haemophilia B and the normal human plasma level of factor VIII is 25-fold less than that of factor IX (200 ng/ml and 5 μg/ml, respectively). For these reasons it would seem that haemophilia A is a better

disease target for a gene therapy strategy. However, due to the large size of the factor VIII coding sequence compared to that of factor IX (7 kb and 1.4 kb, respectively) and the detection of negative sequence elements in the factor VIII cDNA, that are responsible for reduced RNA accumulation (Lynch et al., 1993), factor VIII has proved to be a more difficult protein to express. Several groups have attempted and achieved efficient expression of factor IX by targeting various cell types (see next section). On the other hand, only a few groups have tried expressing factor VIII. They have mainly used retroviral vectors and some have used factor VIII cDNA clones with a deletion of the large segment coding for the central B domain (~2.5 kb; with no known procoagulant function). All attempts have resulted in low levels of factor VIII expression (Israel and Kaufman, 1990; Hoeben et al., 1992; Lynch et al., 1993).

1.9.2 GENE THERAPY FOR HAEMOPHILIA B

Many research groups have been attempting to develop gene therapy protocols for haemophilia B with the aim to prevent bleeding by providing a constant, prophylactic blood level of factor IX. So far, research efforts have mainly focused on *ex vivo* gene transfer approaches. A number of primary cell types have been targeted for somatic gene therapy of haemophilia B. These include fibroblasts (St. Louis and Verma, 1988; Palmer et al., 1989, Axelrod et al., 1990; Scharfmann et al., 1991), myoblasts (Yao and Kurachi, 1992, Dai et al., 1992), endothelial cells (Yao et al., 1991), hepatocytes (Kay et al., 1993) and keratinocytes (Gerrard et al., 1993). All of the above groups used recombinant retroviral vectors to transduce the various cell types and demonstrated high levels of expression *in vitro*. Most of them, however, experienced very low and transient levels of factor IX expression *in vivo*. Scharfmann et al. (1991) demonstrated that this was due to the use of viral promoters and that the use of a house keeping gene (dihydrofolate reductase) promoter resulted in prolonged, but low levels of expression. The same group (Dai et al., 1992) later demonstrated sustained expression of factor IX (over 6 months) from retrovirally transduced primary myoblasts that were transplanted into athymic mice, due to the addition of the (207 bp) mouse muscle creatine kinase (MCK) enhancer element upstream from the original (579 bp) cytomegalovirus (CMV) promoter/enhancer

(Axelrod et al., 1990) driving the expression of the factor IX cDNA. Kay et al. (1993) transduced canine hepatocytes *in vivo*, with retroviral vectors containing the factor IX cDNA being expressed by the viral LTR (long terminal repeats) itself. On intraportal delivery (following a partial hepatectomy), hepatocytes were transduced and sustained, but low level expression was observed in the canine haemophilia B model targeted (see discussion for choice of promoter and cell type, section 1.10.2).

Smith et al. (1993) used adenoviral vectors expressing factor IX, using an *in vivo* approach and achieved high level but transient expression. They showed that even though there was very efficient hepatocyte transduction, development of an immune response resulted in transient expression.

Furthermore, a canine model for haemophilia B exists (Brinkhous et al., 1973). It was shown that affected animals had a missense mutation (G379E; corresponding to position 381 of human factor IX) at a highly conserved residue in the factor IX catalytic domain (Evans et al., 1989). This mutation (unexpectedly) resulted in a complete lack of circulating factor IX in the affected animals. This canine model is being used by at least two groups in the USA for preclinical trials of somatic gene therapy for haemophilia B (Savio Woo, Baylor College of Medicine; Inder Verma, Salk Institute).

1.10 KERATINOCYTE-BASED GENE THERAPY

1.10.1 KERATINOCYTES AS A TARGET FOR SOMATIC CELL GENE THERAPY

The keratinocyte is an attractive target cell to provide somatic gene therapy for haemophilia. Keratinocytes are easily obtained from small skin biopsies and readily expanded in culture (Rheinwald and Green, 1975). It is possible to expand a 1 cm^2 biopsy to 1 m^2 of epithelium within about 3 weeks (Morgan et al., 1987). Cultured epithelial cells can be detached from the surface of the tissue culture dish as an intact sheet and grafted onto a suitable dermal bed *in vivo*. This method has been extensively used to regenerate the epidermis in patients with burns (Gallico et al., 1984). Due to the ease of

their culture, keratinocytes are suitable target cells for genetic manipulation *in vitro*. These properties make keratinocytes ideal for targeting by an *ex vivo* method and also due to their easy access, DNA could be applied directly to the surface of the skin as a method for *in vivo* delivery. Easy access also facilitates the monitoring of the graft, including a subsequent diagnostic biopsy to assess transgene integrity and expression *in vivo*. The graft can be easily removed if there is an undesirable response. If an inducible promoter were used, for example with a steroid-responsive element, it may be possible to selectively induce expression by topical steroid applications (Vogel, 1993).

There are however certain characteristics of keratinocytes that have to be addressed, namely (a) the continuous loss of the genetically modified cells by normal desquamation, (b) the avascular nature of epidermal keratinocytes and (c) the ability of keratinocytes to perform the necessary post-translational modifications required to result in the active form of the ectopically synthesised protein (factor IX).

Keratinocytes have the ability for self-renewal throughout life due to the proliferative potential of their stem cell population. Targeting of the stem cells would therefore serve to continuously replenish the genetically-modified cells being lost by desquamation. The isolation / enrichment of stem cells has been an overriding goal for somatic gene therapeutic strategies in general. For long it has proved difficult to identify the presence of keratinocyte stem cells *in vivo* and *in vitro*, although recently Jones and Watt (1993) have reported some progress in enriching and identifying them in culture systems. They found that keratinocytes with characteristics of stem cells could be isolated from cultured human epidermis on the basis of high surface expression of β_1 integrins and rapid adhesion to extracellular matrix proteins. This method is especially useful to achieve separation of the authentic stem cells (~10% of basal keratinocytes) from "transit amplifying cells" (equivalent to the committed progenitor cells of haematopoietic tissue; ~50% of basal keratinocytes; [Potten and Morris, 1988]) up to greater than 90% purity. This has the potential to greatly facilitate the long-term persistence of genetically-modified keratinocytes *in vivo*.

Keratinocytes in the epidermis are not vascularised, and proteins synthesised in these cells will have to traverse the dermo-epidermal basement membrane to successfully

enter the systemic circulation. Several factors would lead us to believe that ectopically synthesised factor IX in keratinocytes would traverse the basement membrane. Fenjves et al. (1989) showed that apolipoprotein E (ApoE) (normally) synthesised by a human epidermal graft on an athymic mouse was detected in the systemic circulation. This was detected for as long as the graft was present (12 weeks) and disappeared when the graft was removed. ApoE has 299 amino acids and a molecular weight of 30 kDa. Milstone and Edelson (1988) have reviewed various studies indicating the active secretory nature of keratinocytes and have shown them to be responsible for the secretion of lymphokines, hormones and growth factors for systemic delivery. Gerrard et al. (1993) recently showed that keratinocytes were able to (at least partially) post-translationally modify and secrete biologically active factor IX.

1.10.2 VIRAL Vs CELLULAR GENE PROMOTERS

An important consideration for any somatic cell gene therapeutic strategy is the choice of promoter in driving expression of the foreign gene. Besides being biologically safe, it is also important to achieve long-term, efficient and in some cases tissue-specific expression. Most groups attempting somatic gene therapy have made use of viral LTRs (long terminal repeats) and heterologous viral promoters to achieve efficient expression. Notably, most groups attempting to achieve high level expression of factor IX have used either recombinant retroviral vectors (St. Louis and Verma, 1988; Palmer et al., 1989; Axelrod et al., 1990; Armentano et al., 1990; Yao et al., 1991; Yao and Kurachi, 1992; Gerrard et al., 1993; Kay et al., 1993) (in many cases the vectors described by Miller and Rosman, 1989, were used) or adenoviral vectors (Smith et al., 1993) either under the control of the viral LTR or a heterologous viral promoter. A variety of (primary) cell types including fibroblasts, myoblasts, hepatocytes and keratinocytes were targeted, in order to eventually facilitate an *ex vivo* approach in the majority of cases.

In all cases, the various viral promoters (MMLV-LTR/SV40 early/CMV immediate early promoters) achieved high levels of factor IX expression *in vitro*. Their use *in vivo*, was however followed by inactivation of the promoters leading to only short term expression (St. Louis and Verma, 1988; Palmer et al., 1989). A systematic study

carried out by Palmer et al. (1991) revealed that hADA (human adenosine deaminase) expression in primary fibroblasts in an *in vivo* situation was down-regulated >1,500 fold in <1 month. No cell- or antibody-mediated immune response was detected and the survival of the transplanted cells was largely unaffected. When harvested and regrown *in vitro*, the cells failed to regain their transgene expression. Some evidence was presented in support of a methylation-mediated inactivation.

Intriguingly, although this phenomenon was encountered in fibroblasts (St. Louis and Verma, 1988; Palmer et al., 1989), primary myoblasts (Dai et al., 1992) and lymphoid cells (Osborne et al., 1990), similar (or even identical) viral vector/promoter constructs used in myeloid haematopoietic cells (Kaleko et al., 1990; Osborne et al., 1990), smooth muscle cells (Lynch et al., 1992) and hepatocytes (Kay et al., 1993) did not result in inactivation. Scharfmann et al. (1991) used a housekeeping gene (dihydrofolate reductase) promoter and achieved low but sustained levels of expression from mouse fibroblast implants. The same group then achieved long-term (>6 months) expression of factor IX from transplanted myoblasts after incorporating a 207 bp muscle creatine kinase (MCK) enhancer element into their retroviral-based, CMV promoter-driven factor IX cDNA in athymic mice (Dai et al., 1992). Similar success, using cellular regulatory sequences, was reported for the correction of the lysosomal storage defect in β -glucuronidase-deficient mice using genetically-modified fibroblasts *in vivo* (Mouller et al., 1993). Clearly the choice of cell-type is an important variable in determining the eventual success of a viral promoter *in vivo*.

The data for keratinocytes, although not substantial, would lead us to believe that viral promoters alone would not achieve long-term expression *in vivo*. Morgan et al. (1987) used a retroviral vector to transduce keratinocytes with a human growth hormone gene (hGH) and, after achieving efficient expression *in vitro*, grafted the cultured keratinocytes onto athymic mice. The cells reconstituted a normal epidermis, and hGH assayed in a biopsy of the graft showed very low levels, and no secreted hGH was detectable in the systemic circulation (using an assay with a sensitivity of 0.2 ng/ml). The same group later demonstrated *in vivo* secretion of hGH when an endogenous cellular promoter (mouse metallothionein) was used to drive expression of the transduced gene

(Teumer et al., 1990). Gerrard et al. (1993) using a retroviral vector with the factor IX cDNA under the control of the viral LTR, achieved very high levels of active factor IX expression in *in vitro* transduced primary keratinocytes. On transplanting these cells to a sub-dermal site in athymic mice, they could detect low levels of factor IX in the plasma, but expression ceased within 7 days. A report by Flowers et al. (1990) claimed to have achieved long-term expression of retrovirally transduced primary canine keratinocytes after grafting them *in vivo*. They used a recombinant retroviral vector containing the bacterial *neo* gene and ranging from 27-130 days after transplantation, these transduced cells were harvested and tested for formation of G418 resistant colonies *in vitro*. On average they isolated 0.6% (30/5119 in 10 separate biopsies) G418 resistant colonies (34% [10-76%] were resistant prior to transplantation). The findings from this sensitive assay system indicate a considerable reduction of expression *in vivo*, and indeed, this number of cells continuing to express the foreign gene probably also existed in the experiments performed by Morgan et al. (1987) and Gerrard et al. (1993). Although these studies demonstrate the ability of keratinocytes to secrete the active forms of proteins and the successful transfer across the dermo-epidermal basement membrane, viral vectors alone would probably not be sufficient to achieve sustained expression in transduced primary keratinocytes *in vivo*.

1.10.3 THE USE OF CYTOKERATIN GENE PROMOTERS

The cytokeratins (or keratins) are a family of at least 20 different intermediate filament (IF) cytoskeletal proteins expressed in keratinocytes. On the basis of biochemical, immunological and nucleic acid sequence data, the keratins have been classified into two main groups: the acidic (type I) and the more basic (type II) (Moll et al., 1982). This subdivision has a functional significance, since at least one representative of either type is required for the formation of the heterotypic subunit complexes of cytokeratin IF. Different "expression-pairs" of keratins are expressed depending on the state of differentiation of the keratinocytes. *In vivo*, in keratinocytes of the mammalian epidermis, cells of the basal layer express just one pair of cytokeratins, K5 and K14, whereas the differentiated suprabasal layers express K1 (variably K2) and K10 (and K11).

The latter pair of keratins have therefore been regarded as markers for "terminal differentiation" or "keratinisation". Epidermal keratinocytes in culture, however, express in addition to the basal keratins, a set of smaller keratins, K6 and K16 (and K17). These are referred to as the "hyperproliferative keratins".

Keratin gene expression seems to be regulated in a tissue-, differentiation- and development-specific fashion (Fuchs, 1988). Little is known so far regarding keratinocyte transcription factors, but the number of keratin genes cloned and characterised have served as a useful tool for comparative sequence analyses. Certain conclusions can be drawn on the basis of these sequence analyses, (a) the genes for type I and II keratins show striking similarities in their genomic organisation (especially regarding intron positions in the region coding for the central "α-helical rod" domains) (reviewed in Steinert and Roop, 1988) (b) various keratin genes co-expressed in the epidermis do not show extensive sequence homology in their coding/non-coding regions (except for certain shared motifs in their 5' UTRs) (Blessing et al., 1987; Steinert and Roop, 1988; Fuchs, 1990; Leask et al., 1991) and (c) the corresponding genes for human, murine and bovine keratins expressed in a similar differentiation-specific manner show extensive sequence homology (Blessing et al., 1987; Reiger et al., 1988; Fuchs, 1990).

In accordance with these observations the human K10 (HK10) gene and the bovine and murine counterparts, BKVIb and murine M_r 59,000 keratin genes (MK10) respectively, show extensive sequence homology that extends to both the coding and noncoding sequences (Reiger et al., 1988). Exons 2-6 which code for the α-helical rod domains are identical in size and exons 1, 7 and 8 are also very similar in size. They show 95% amino acid homology in their entire coding sequence which is not only confined to the usually conserved "α-helical rod" domain but also in the "head", "tail" and "spacer" regions. Exon 8, which codes for only the last amino acid (Tyr) and forms the 3'UTR in all three genes, also shows only minimal size variation among them. Remarkably it was also observed that the intron sizes are very similar between the species and surprisingly share extensive regions of sequence homology. It was noted that there were islands of >80% sequence identity scattered in regions with 40-60% identity. For example, the entire intron 2 in HK10 and BKVIb showed 74.5% sequence identity. None of this

intronic sequence identity was attributable to repeat elements. This extensive sequence homology of introns, in the absence of any obvious selective pressure may indicate that some of these sequences are likely to have a role in regulation of gene expression. This is known to be true for the first introns of many genes, for example: α1(I) collagen gene (Bornstein et al., 1987), troponin I gene (Konieczny and Emerson, 1987), human growth hormone gene (Moore et al., 1985), human smooth muscle α-actin gene (Nakano et al., 1991) and the muscle creatine kinase gene (Sternberg et al., 1988). Transcriptional enhancer-like elements have been reported to exist in introns other than the first and even in the 3' region of a gene, for example in the second intron and 3' to the structural gene for human β-globin (Behringer et al., 1987).

The 5' and 3' regions of the K10 genes in different species also shows extensive sequence homology. The HK10 and BKVIb genes in the region of the 3'UTR and 275 bp downstream of the polyadenylation signal, showed 90% sequence identity. In the region 1.15 kb upstream from the translation initiation site, HK10 and BKVIb show 73% sequence identity (83% similarity).

In common with all characterised cytokeratin genes the HK10, BKVIb and MK10 genes have a typical TATA-box situated 26 nucleotides upstream from the cap site. Also present 78 nucleotides upstream from the TATA-box, is the "AAPuCCAAA" consensus motif, seen upstream of the TATA-box in all epidermally-expressed keratin genes (Blessing et al., 1987) and in the involucrin gene (another epidermally-expressed gene; Eckert and Green, 1986). Significantly, this motif is not seen in non-epidermal keratin genes (Blessing et al., 1987). An AP2 binding site demonstrated in the 5' upstream region of the genes for K14, K5, K1 and K6b (Leask et al., 1991) does not exist in the HK10, BKVIb or MK10 genes. It is perhaps not unusual to have this lack of extensive homology in the 5'UTRs of genes co-expressed in the same tissue but at different stages of differentiation, as evidenced in the co-ordinately expressed genes coding for the several constituents of striated muscle and in the α-actin genes of chicken skeletal and cardiac muscle (co-expressed in skeletal muscle during embryonic development; Eldridge et al., 1985).

Since keratins can account for 30-85% (depending upon their stage of differentiation) of the total protein of an epithelial cell, their genes might be expected to have relatively strong promoter/enhancer elements (Fuchs, 1988). It has been regarded that the tissue- and differentiation-specific expression of keratin genes is mainly regulated at the transcriptional level (Blessing et al., 1989; Lersch et al., 1989; Stellmach et al., 1991). Practically, the 5' upstream sequences of various keratin genes, presumably containing the *cis*-acting elements required for efficient and tissue-specific expression, have been used to achieve keratinocyte-specific expression *in vitro* and *in vivo* (in transgenic mice). Blessing et al. (1989) used CAT (chloramphenicol acetyltransferase) transfection assays in primary epithelial cultures to demonstrate the tissue-specific promoter function of 5' upstream sequences of the BKIV (human keratin 6) gene. They also demonstrated the enhancer function of a 425 nucleotide sequence by upregulation of reporter expression by placing it downstream and in an opposite orientation to the reporter gene. Various groups have achieved keratinocyte-specific expression of either mutant/normal keratin genes (HK14 [Vassar et al., 1989] ; HK10 [Fuchs et al., 1992]; HK1 [Rosenthal et al., 1991]) or heterologous genes (BKVIb-mutant H-*ras* gene [Bailleul et al., 1990]; HK14-TGFα [Vassar et al., 1991]; BKIII/IV-BMP4 [Blessing et al., 1993]; HK1-TGFβ1 [Sellheyer et al., 1993]; BKVIb-TGFβ1 [Cui et al., 1994]) in transgenic mice. In every case, variable lengths (>1 kb) of 5' upstream regions of the named cytokeratin genes were used and keratinocyte- and differentiation-specific expression was demonstrated.

The gene promoter/enhancer and upstream *cis*-acting regulatory elements of the bovine counterpart of human keratin 10 i.e., BKVIb was acquired from Dr. Jose Jorcano (Madrid, Spain). Use of the same promoter (BKVIb) has been shown to successfully direct suprabasal keratinocyte-specific expression of a mutant H-*ras* gene (Bailleul et al., 1990) and of mutant TGFβ1 (Cui et al., 1994), in transgenic mice. In this project, the use of the BKVIb promoter to target expression of human clotting factor IX in a keratinocyte-specific fashion was investigated.

1.10.4 HaCaT CELLS AS THE KERATINOCYTE TARGET

As described previously, epidermal keratinocytes in culture switch off the differentiation-specific keratins and induce expression of a set of smaller keratins (K6 and K16) in addition to the basally-expressed keratins (K5 and K14). It is however possible to achieve the expression of differentiation-specific keratins, K1 and K10, by modifying the culture conditions. High calcium concentration (>0.10 mM; Yuspa et al., 1989), vitamin A depletion (Fuchs and Green, 1981; Kopan et al., 1987) and culturing at an air-liquid interface (Asselineau et al., 1986) or transfer to an *in vivo* site serve to stimulate cultured primary keratinocytes to synthesise K1 and K10.

The HaCaT cell line is a spontaneously transformed human epithelial cell line derived from normal adult skin (Boukamp et al., 1988). HaCaT cells are non-tumorigenic and maintain a full epidermal differentiation capacity. When transplanted onto nude mice, like normal keratinocytes, they reform an orderly structured and differentiated epidermal tissue.

HaCaT cells are a particularly attractive model to study the expression of differentiation-specific keratins because of their largely preserved normal epidermal characteristics (including a normal capacity to differentiate) and due to their unusual capacity to synthesise a broad spectrum of keratins including the differentiation-specific keratins under conventional culture conditions and low calcium levels (0.05 mM) (Ryle et al., 1989). At low cell densities they express K5, K6, K7, K8, K14, K16, K17, K18 and K19. With increasing cell densities they express more of K5, K6, K14, and K16 and begin to express K4, K13 and K15. At very high cell densities (at and after confluence) HaCaT cells begin to express the differentiation-specific keratins, K1 and K10. While K1 is detectable at confluence, there is a significant delay in the expression of K10. Moreover, K10 is expressed at considerably lower quantities than K1.

The expression of these keratins is strongly modulated by environmental conditions including cell density (Ryle et al., 1989). As witnessed by *en face* immunofluoresence, the increase of K1 and K10 expression is due to the increasing numbers of positively-stained cells and not due to a gradual increase of expression in all

cells. K1 and K10 expression in HaCaT cells is shown not to be related to stratification and is largely not affected by passage number.

Clearly, primary keratinocytes and other known keratinocyte cell lines would not be appropriate to assess the expression of the K10 promoter (in a tissue-specific manner) in conventional cultures. The HaCaT cell line was therefore used as the model to test the ability of the K10 promoter to drive expression of human factor IX *in vitro* in a keratinocyte-specific manner. Bearing in mind that the expression levels obtained would be suboptimal, the cell line should still allow the comparison of efficiency of different gene constructs in directing expression. The HaCaT cell line was kindly provided by Dr. N.E. Fusenig (Heidelberg).

1.11 SPECIFIC AIMS OF THIS PROJECT

The two overall objectives of this project are: (I) to carry out a comprehensive search for mutations in the factor VIII gene of patients with haemophilia A and (II) to investigate the possibility of keratinocyte-specific expression of factor IX as a potential method for somatic cell gene therapy for haemophilia B.

The specific aims of this project are as follows:

1.11.1 Characterisation of mutations in the factor VIII gene

(1) To detect mutations in the entire essential sequence of the factor VIII gene:

(i) Design PCR primers spanning the entire essential sequence of the factor VIII gene to enable the amplification, mutation screening and direct sequencing of alterations within these regions.

(ii) Detect and characterise gross alterations of the factor VIII mRNA/gene by RT-PCR/DNA-PCR followed by direct sequencing, and identify and accurately localise small sequence alterations by chemical mismatch cleavage (CCM) analysis followed by direct sequencing.

(iii) Characterise mutations in selected individual exons reported to harbour a disproportionately higher number of mutations for their size, using single strand conformation polymorphism (SSCP) analysis.

(2) To analyse the mutations detected:

(i) Correlate the genotype with the phenotype by analysis of residues affected by mutations. For example, to look for evolutionary conservation and involvement with known or predicted functional domains and analyse gross alterations of the factor VIII gene/mRNA for their predicted effects on the primary structure of factor VIII.

(ii) Analyse the sequence in the immediate vicinity of mutations to decipher probable mechanisms of mutagenesis.

1.11.2 Keratinocyte-specific expression of factor IX

(1) To create the basic cytokeratin 10 promoter-driven factor IX gene expression construct:

Directionally subclone the human factor IX coding sequence (FIX) into a gene expression construct containing the cytokeratin 10 (K10) promoter and SV40 viral small t-antigen intron and polyadenylation signal (SV40 polyA cassette).

(2) To introduce favourable sequence modifications into the basic K10-FIX construct:

(i) Engineer a favourable sequence context around the "AUG" initiation codon for efficient translation initiation (Kozak, 1991a,b) using PCR-mediated site-directed mutagenesis.

(ii) Repair the defective 5' untranslated sequence of the K10 promoter in the basic construct by insertion of synthetic oligonucleotides with the appropriate sequence (partial gene synthesis).

(iii) Replace the SV40 polyA cassette at the 3' end of the K10-FIX construct with the natural 3' end of the cytokeratin 10 gene.

(3) To verify the sequence of the various gene constructs before utilising them for gene expression studies:

(i) Confirm the successful incorporation of the various sequence manipulations by direct sequencing.

(ii) Detect mutations (if any) in inserts generated by PCR, using CCM analysis and direct sequencing.

(4) To transfect keratinocytes *in vitro*:

(i) Transfect a spontaneously transformed, non-tumorigenic human keratinocyte cell line (HaCaT cells) which is known to express a wide range of cytokeratins (including K10) *in vitro* (Boukamp et al., 1988; Ryle et al., 1989).

(ii) Linearise the gene constructs prior to transfection, to avoid disruption of the transcriptional unit during random integration of the constructs into the host cell genome.

(iii) Transfect using cationic-liposomes to achieve a high efficiency of transfection with minimal disruption of the physiology of the transfected cells.

(iv) Assess the efficiency of transfection using a control *lacZ* reporter construct.

(v) Co-transfect with a selectable marker to enrich for cells containing the stably integrated transgene and to isolate individual clones of stably transfected cells.

(vi) Confirm the successful integration of the transgene.

(5) To analyse the absolute and relative levels of factor IX expression achieved by the various gene constructs:

(i) Analyse expression at the transcriptional level by detection of factor IX mRNA using RT-PCR and Northern blotting.

(ii) Analyse expression at the translational level by detection of factor IX protein secreted into the tissue culture medium using the sensitive and quantitative enzyme linked immunosorbent assay (ELISA). Indirectly assess the ability of HaCaT cells to produce biologically active factor IX by using the "A7" antibody which detects only the γ-carboxylated form of factor IX (Smith et al., 1986).

(iv) Comparatively assess the effects of the site-directed sequence modifications introduced into the gene constructs using the results of Northern blotting and ELISA.

(6) To assess the ability of the K10-FIX construct to achieve keratinocyte-specific expression *in vivo* in transgenic mice: (in collaboration with Dr. Y. Alexander).

CHAPTER: 2

MATERIALS AND METHODS

2.0 MATERIALS AND METHODS

2.1 INTRODUCTION

All materials used throughout this project were sterilised to prevent contamination or degradation of nucleic acid samples. Solutions and plasticware used in DNA analysis were autoclaved before use. Gloves were worn at all times. Materials employed in procedures which involved the use of RNA were treated as follows: (a) glassware was washed thoroughly with distilled water and baked for a minimum of four hours at 180°C, (b) plasticware was submerged in distilled water containing freshly added 0.01% DEPC (diethyl pyrocarbonate; Sigma) for a minimum of four hours and sterilised by autoclaving and (c) all solutions, except Tris-based solutions and organic solvents, were treated with 0.01% DEPC before being autoclaved.

Contamination in PCR due to "carryover" was avoided by diligently separating the pre- and post-PCR stages. All procedures involving bacterial cultures and blood samples were carried out using aseptic techniques. Tissue culture was also carried out under aseptic conditions in a dedicated laboratory and flow hood.

2.2 PATIENT INFORMATION

A total of twenty-five unrelated haemophilia A patients were analysed in this project. A total of thirteen patients (group 1) were analysed for mutations in the entire essential region of the factor VIII gene and twelve (group 2) were analysed for mutations in certain selected exons. Fresh blood samples from the patients in group 1 were acquired from the Regional Haemophilia Centres at the Glasgow Royal Infirmary (West of Scotland), the Leeds General Infirmary and St. James' University Hospital (Yorkshire, England). Analysis of the patients in Group 2 was performed on DNA samples obtained from the West of Scotland Regional Genetics Service (Yorkhill Hospital, Glasgow, Scotland).

Of the thirteen patients in group 1, eight had severe haemophilia A and two of the ten from whom inhibitor data was available had developed inhibitors to exogenous factor VIII. Group 2 consisted of six severely and six mild to moderately affected haemophilia A patients.

2.3 EXTRACTION OF DNA AND RNA

2.3.1 EXTRACTION OF DNA

In this project, genomic DNA was extracted from whole blood samples, cultured cells and mouse tail-tips. The preparation of small and large scale plasmid DNA from bacterial cultures is described in sections 2.6.6 and 2.6.7.

Extraction of DNA from peripheral leukocytes

Genomic DNA was extracted from peripheral blood samples using a variation of the method described by Kunkel et al. (1977). Two volumes of ice cold lysis mix (0.32M Sucrose, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100) were added to the blood sample. The mixture was inverted a few times and centrifuged at 2,500 rpm for 10 minutes at 4°C in an IEC DPR-6000 refrigerated centrifuge. The supernatant was discarded and the pellet was dissolved in 3 ml nuclei lysis mix (10 mM Tris-HCl, pH 8, 0.4M NaCl, 2 mM EDTA), 200 µl 10% SDS, 100 µl proteinase K (Boehringer Mannheim; 10 mg/ml), and was incubated overnight at 37°C. 1 ml of 6M NaCl solution was added to the overnight digest, and the mixture was vigorously mixed and centrifuged at 2,500 rpm for 10 minutes. The supernatant was transferred to a fresh tube where two extractions were performed with equal quantities of phenol:chloroform and one with chloroform:isoamyl alcohol. DNA was precipitated from the aqueous phase with a tenth volume of 3M sodium acetate (pH 5.5) and two volumes of ethanol. The DNA was spooled out on a terminally-sealed pasteur pipette, washed in 70% ethanol, air-dried and dissolved in a suitable volume of T.E. buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA).

Extraction of DNA from tissue culture cells

Genomic DNA was extracted from cultured cells grown in 80 cm^2 flasks by a scaled-up version of the method described by Laird et al. (1991). The tissue culture medium was aspirated and the cells were washed twice in PBS (Gibco, BRL). 5 ml of lysis buffer (100 mM Tris-HCl, pH 8, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, Proteinase K 100 μg/ml) was added to the flask, which was followed by a 3 hour incubation at 37°C. An equal volume of isopropanol was added and the flask was placed on a shaker for 15-20 minutes until the precipitation was complete. The DNA was spooled out, washed in 70% ethanol, air-dried and dissolved in a suitable volume of T.E. buffer.

Extraction of DNA from tail-tip biopsies

Genomic DNA was extracted from tail biopsies as described by Laird et al. (1991), using the same solutions as described above. Tail tip biopsies (~1 cm) were transferred to a tube containing 0.5 ml lysis buffer, immediately upon cutting. They were then incubated overnight in a 55°C water bath with continuous shaking. Following complete lysis, the tubes were vortexed briefly and centrifuged on a bench top centrifuge at 12,000 rpm for 10 minutes to obtain a firm pellet. The supernatant was transferred to pre-labelled tubes containing 0.5 ml isopropanol. The DNA was spooled out, washed in 70% ethanol, air-dried and dissolved in a suitable volume of T.E. buffer.

2.3.2 EXTRACTION OF RNA

Total cellular RNA was extracted from peripheral blood lymphocytes and a cultured keratinocyte cell line. The acid guanidinium thiocyanate-phenol-chloroform extraction method described by Chomczynski and Sacchi (1987) was used. Peripheral blood lymphocytes were initially separated from whole blood by centrifugation through 'Histopaque 1077' (Sigma) followed by rinsing with cold PBS. RNA from cultured cells was extracted whilst still in the tissue culture flasks. In both cases, the cells were vigorously mixed with solution D (4M guanidinium thiocyanate, 25 mM sodium citrate, 0.5% sarcosyl and freshly added 0.1 mM β-mercaptoethanol), and extraction carried out by the addition of a tenth volume of 2M sodium acetate (pH 4.0), an equal volume of

water-saturated phenol and a fifth volume of chloroform:isoamyl alcohol (49:1). The contents were mixed by vortexing and incubated on ice for 15 minutes. The aqueous phase obtained after centrifuging the mixture at 10,000 g for 20 minutes at 4°C, was transferred to a fresh tube, and an equal volume of isopropanol added. The contents were mixed, and precipitation was carried out at -20°C for at least an hour. The mixture was centrifuged, as before, and the pellet redissolved in 600 µl solution D. The solution was reprecipitated with isopropanol and the RNA pellet was dissolved in a suitable volume of distilled water that had been treated with DEPC and autoclaved.

The integrity of the extracted RNA was assessed by electrophoresis on a 1.5% agarose gel containing 5.2% formaldehyde and 1xMOPS buffer (10x buffer: 200 mM MOPS, 50 mM sodium acetate pH 7, 10 mM EDTA). RNA samples were prepared for electrophoresis by adding an equal volume of 50% formamide, 5.2% formaldehyde in 1xMOPS buffer. The samples were heated at 70°C and quenched on ice. A sixth volume of loading dye (20% ficoll, 10 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol) was added and the samples were loaded on the gel. Electrophoresis was carried out (at 100 volts for 4-8 hours) in 1xMOPS buffer which was circulated to prevent the depletion of ions at the anode. An RNA ladder (Gibco RRL) was run alongside the samples for use as a marker.

2.4 PCR TECHNIQUES AND METHODOLOGY

2.4.1 OLIGONUCLEOTIDE DESIGN AND SYNTHESIS

In this project oligonucleotides were designed for use in (i) conventional PCRs, (ii) allele-specific PCRs, (iii) PCR cloning, (iv) site-directed mutagenesis, (v) recombinant PCR or splicing by overlap extension (SOE; Horton et al., 1989) and (vi) partial gene synthesis.

In every case, oligonucleotides were synthesised on an ABI 391 automated DNA synthesiser. Concentrated ammonia was used to elute the oligonucleotides from the solid support and deprotection was carried out at 55°C for 16 hours. A rapid method was

employed for the purification of deprotected oligonucleotides, using n-butanol (Sawadogo and van Dyke, 1991). Briefly, 100 µl of the crude deprotected oligonucleotide solution in ammonia was mixed with 1000 µl of n-butanol, vortexed and centrifuged for 5 minutes at 12,000 rpm. The pellet obtained was air-dried and resuspended in 100 µl distilled water and the n-butanol extraction was repeated once more. The pellet was finally dissolved in 100 µl distilled water, quantified and stored at -20°C until required.

Oligonucleotides synthesised for PCR were designed using the computer program, OLIGO™ version 3.4 (© Wojciech Rychlik; MedProbe). This program allowed the design of PCR primers that (i) were not self-complementary i.e., did not form internal duplex structures primer-dimers, (ii) would form specific and stable duplexes with the template DNA, (iii) as a pair, were compatible in PCR i.e., were not complementary at their 3' ends (to avoid primer-dimers) and had similar melting temperatures and duplex stabilities.

Oligonucleotides for other specific uses, either PCR related or otherwise, were designed to have the appropriate additional sequence features. For example, primers for cloning were designed with the sequence of the appropriate restriction enzyme recognition sequence incorporated at the 5' ends (e.g. primer F9A was designed to introduce a *Kpn*I site immediately 5' to the "AUG" initiation codon of the factor IX coding sequence; section 3.5.2; Scharf et al., 1986). Anchor sequences (5 - 6 nucleotides, GC-rich) were added 5' to the restriction enzyme recognition sequences in order to achieve efficient digestion at the termini of PCR products. Also, the appropriate base substitutions were incorporated into the PCR primer in order to facilitate site-directed mutagenesis (e.g. primer F9G4 was designed to produce the Q-45V mutation in the factor IX coding sequence; section 3.5.3). For SOE, primers were designed so as to generate PCR products with complementary sequence at their ends. When SOE was required for the joining of two unrelated sequences, then the primers were designed with additional sequences at their 5' ends so as to make the unrelated sequences overlap at one end (Horton et al., 1989; Higuchi, 1989). An example of this can be seen in figure 31 (section 3.5.5) where the factor IX coding sequence is spliced to a segment at the 3' end

of the bovine cytokeratin VIb gene (BKVIb). When SOE was required for parts of the same sequence, as in the case of the three segments of the factor IX coding sequence, no special sequence modification was required as long as overlapping PCR products were generated (figure 21, section 3.5.1). For partial gene synthesis, oligonucleotides were designed with the appropriate sequence that needed to be inserted, flanked by appropriate restriction enzyme and anchor sequences. An example of this can be seen in section 3.5.4, where a pair of complementary oligonucleotides (50 nucleotides long) were used to repair the region of the 5'UTR in the recombinant factor IX expression construct.

Primers designed during the course of this project are listed in tables 1, 2, 5 & 6.

2.4.2 DNA-PCR

In general, all PCRs for the amplification of targets in a DNA template were performed over 30 cycles, in a final volume of 50 µl. PCRs were performed either on a PHC I or III automated thermocycler. All reagents were obtained from Boehringer Mannheim. A typical PCR mix contained: 1xPCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin), each primer in a final concentration of 1 µM, each dNTP in a final concentration of 250 µM, 250 ng of DNA and 2 units *Taq* DNA polymerase.

PCR mixes were always assembled on ice using a dedicated set of pipettes. Buffer, primers and dNTPs were added into pre-labelled 0.5 ml microfuge tubes on ice, and the volume was made up to 48 µl with distilled water. The mixture was then UV-irradiated for 10 minutes (GRI-Amplirad). After addition of the DNA template (1 µl), the contents were briefly mixed by vortexing and centrifuged. *Taq* polymerase was added and after a brief vortex and centrifugation, light mineral oil was layered on top of the reaction mix. The tubes were immediately placed on a preheated (>90°C) thermocycler.

The DNA-PCR thermocycle profile was different for the two thermocyclers used. Since the PHC I thermal block is cooled by tap water and is much slower than the PHC III at temperature variation, longer times had to be set in the program. A typical cycle on a PHC I consisted of 94°C for 1 minute (denaturation), 56 - 64°C (depending on the primer pair) for 1 minute (primer annealing) and 72°C for 1-2 minutes (primer extension).

The denaturation and annealing steps were reduced to 30 seconds each, in the case of the PHC III thermocycler. A 10 minute final extension step (at 72°C) was performed after the 30 cycle PCR was complete.

PCR products were visualised by agarose gel electrophoresis (in 1xTBE), ethidium bromide staining and transillumination on a UV-lamp (Fotodyne, UV-transilluminator). Permanent records of these gels were made using either a polaroid camera and film or a gel documentation system (UVP Imagestore 5000) linked to a thermal printer.

2.4.3 HIGH FIDELITY PCRs USING *Taq* POLYMERASE

Due to the lack of 3' to 5' exonuclease activity in *Taq* polymerase, a high rate of random mutagenesis was experienced during the making of the basic construct (section 3.5.2), using the conventional PCR protocol described in the preceding section (2.4.2). Specific measures were taken (in subsequent subcloning experiments) in order to reduce the probability of extension at a misincorporated nucleotide, and also to reduce the DNA damage errors caused by high temperatures (Eckert and Kunkel, 1992). These included: (i) reducing the number of cycles to 25; (ii) reducing the final dNTP concentration five fold, to 50 μM and (iii) reducing the denaturation temperature and duration to 91°C for 30 seconds. As seen in section 3.5.3, this seemed to have a beneficial effect.

2.4.4 RT-PCR AND NESTED PCR

1 μg of total cellular RNA was mixed with 10 pM of a suitable reverse primer in a 9 μl volume in DEPC-treated water. This was heated to 90°C for 5 minutes followed by a 10 minute incubation at 65°C. To this mix was added, 4 μl 5X reverse transcriptase buffer (Gibco, BRL; 250 mM Tris-HCl, pH 8, 375 mM KCl, 15 mM $MgCl_2$), 2 μl 0.1M DTT, 250 μM dNTPs and 200 units of Moloney murine leukaemia virus (MMLV) reverse transcriptase (Gibco, BRL). The volume was made up to 20 μl and the mixture was incubated at 37°C for 45 minutes. Reverse transcriptase was inactivated by heating at 95°C for 5 minutes (Sellner et al., 1992). Half of this reverse-transcribed mixture was used as a template for PCR analysis.

When amplifying transcripts in very low abundance, as in the case of ectopic transcripts, it was necessary to perform a further PCR amplification using a nested set of primers. 2μl of the original PCR product was then used as the template for a subsequent nested PCR amplification. This effectively doubled the number of PCR cycles and tremendously increased the sensitivity of the assay (Roberts et al., 1990).

2.4.5 GENERATION OF RADIOACTIVELY-LABELLED PCR PRODUCTS

This was required for the generation of radioactive probe for chemical mismatch cleavage analysis (CCM) and prior to single strand conformation polymorphism analysis (SSCP). In the former case, primers were initially 5' end-labelled using T4 polynucleotide kinase (NBL) and [γ-^{32}P] ATP (Amersham). This was performed in a 10 μl volume (for each primer) in 1x polynuceotide kinase buffer (10x buffer: 500 mM Tris-HCl, pH 7.5, 0.1M DTT, 0.1M $MgCl_2$) by adding: 50 pM primer, 5-10 μCi [γ-^{32}P] ATP (3000 Ci/mmol; 10 Ci/ml) and 10 units of T4 polynucleotide kinase. The mixture was incubated at 37°C for 45 minutes. Half of this mixture was used as primer for a PCR reaction, which was otherwise performed as in section 2.4.2.

Generation of radiolabelled PCR products suitable for SSCP analysis was performed using internal incorporation labelling of [α-^{32}P] dCTP. This was performed by preparing the PCR pre-mix in exactly the same way as in section 2.4.2, except for a ten fold reduction in the final concentration of cold dCTP. To this was added 0.5 - 1.0 μCi [α-^{32}P] dCTP (3000 Ci/mmol; 10 Ci/ml), and PCR carried out as before.

2.4.6 ALLELE-SPECIFIC PCR AMPLIFICATION

In this method, the fact that the specificity and efficiency of primer extension is highly dependent on the 3' terminal 1-2 nucleotides, was exploited (Wu et al., 1989; Sarkar et al., 1990). An example of this is illustrated in figure 15, where the two reverse primers, 6N and 6M were specific for the normal (A) and mutant base (G) at position +3 in the 5' splice site of intron 6, respectively. A multiplex reaction was performed, using 6X and 6Y along with either 6N or 6M in order to have an internal positive control. 6N and 6M were designed so as to have similar melting temperatures to 6X and 6Y and so

that none of the four would form dimers with each other, thus making it possible to use them in a multiplex PCR assay.

PCRs were performed as in section 2.4.2, except that three PCR primers were added in each reaction (6XNY or 6XMY). At first a low annealing temperature was used. The annealing temperature was gradually increased by 1°C at a time, until it was possible to achieve the desired discrimination between the two alleles. Satisfactory allele specificity was achieved at 65°C, although the PCR products were rather faint at this temperature. Due to the variability in the temperature profiles generated from one thermocycler to another, the same PCR machine was used during the optimisation. The PHC III thermocycler was used, as unlike the PHC I, it was not dependent on tap water to cool the thermal block.

2.4.7 ASYMMETRIC PCR

Single stranded templates used for sequencing were generated by asymmetric PCR amplifications (Gyllensten and Ehrlich, 1988) as previously described (Mgone et al, 1992). Approximately 50 ng of gel-purified double-stranded PCR product (obtained by passive elution or adsorption to glass milk [geneclean II]; section 2.5.2) was used as a template for asymmetric PCR, which was performed by limiting one primer to between 50 to 100 times less concentration than the other (0.5 to 1.0 pM). As there was only a linear accumulation of PCR product generated by a single primer after the limiting primer was exhausted, the number of cycles was increased to 40. The single and double stranded products were visualised by ethidium bromide staining following agarose gel electrophoresis.

2.4.8 RECOMBINANT PCR

Two methods were employed in order to achieve recombination via PCR. These included: (i) splicing by overlap extension (SOE; Horton et al., 1989; Higuchi, 1989) and (ii) ligation followed by PCR. An example of this can be seen in figure 31 (section 3.5.5), where the factor IX coding sequence is spliced to a segment of the 3' end of the bovine

cytokeratin VIb gene (BKVIb). The principle of primer design in the two methods has been described in section 2.4.1.

In SOE, 50 ng of each purified PCR segment was combined and used as a template for PCR. A PCR performed using primers that amplified across the region of the sequence designed to overlap, enabled the joining of the two unrelated sequences. In the latter method, equimolar concentrations of digested PCR products (section 2.6.1) were initially ligated to each other using T4 DNA ligase (section 2.6.2). 50 to 100 ng of this ligation mixture was then used as a template for PCR using primers flanking the ligation site.

2.4.9 DIRECT PCR SCREENING FOR RECOMBINANT CLONES

Bacterial colonies that were formed after transfection with ligated recombinant plasmids were screened directly for the presence of the desired insert. Primers were designed to be specific to the sequence of the insert and to bridge at least one ligation site. This helped to prevent amplification from unligated insert present on the plate.

The PCR pre-mix was made in the same way as before. An autoclaved toothpick was used to scrape about a third to a sixth of an isolated bacterial colony. The toothpick was then dipped into the PCR pre-mix and the same toothpick was also used to grow a 10 ml culture (L-broth containing ampicillin; section 2.6.6). The mixture was vortexed and heated for 10 minutes at 95°C. After chilling on ice, *Taq* polymerase was added and PCR performed as in section 2.4.2. Plasmid DNA was extracted from the corresponding 10 ml cultures (section 2.6.6) that gave positive PCR results, indicating the presence of the desired recombinant plasmid. The preliminary screening result was then confirmed by restriction mapping (section 3.5.2), CCM (sections 3.5.2, 3.5.3) and direct sequencing (sections 3.5.3, 3.5.4).

2.4.10 PREPARATION OF DOUBLE-STRANDED OLIGONUCLEOTIDES

Double-stranded oligonucleotides were made to repair the sequence of the 5'UTR in the K10-factor IX expression constructs, as described in section 3.5.4. Two complementary oligonucleotides, 50 nucleotides long, were designed with the appropriate

sequence, flanking restriction enzyme sites and anchor sequences (K10A and K10B). After purifying them by n-butanol extraction, equimolar proportions were mixed together in a solution of 1x One-Phor-All *PLUS* buffer (Pharmacia; 10x buffer: 100 mM Tris-acetate, pH 7.5, 100 mM Magnesium acetate, 500 mM potassium acetate). The mixture was heated to 75°C for 5 minutes and then gradually cooled to room temperature over 30 minutes. The annealed double stranded oligonucleotide in restriction enzyme buffer was thus ready for digestion.

2.5 METHODS FOR MUTATION DETECTION

Three methods were used for the screening and eventual characterisation of mutations. Single strand conformation polymorphism (SSCP) analysis and chemical cleavage of mismatches (CCM) were used as mutation screening methods. Positive screening results were characterised by direct sequencing of single stranded template generated by asymmetric PCR amplifications (section 2.4.7).

2.5.1 SINGLE STRAND CONFORMATION POLYMORPHISM (SSCP) ANALYSIS

This method, for the detection of single base substitutions, was described by Orita et al. (1989). Incorporation of [α-^{32}P] dCTP during PCR amplification was used to obtain radiolabelled PCR products suitable for analysis (section 2.4.5). These were diluted 1 in 10 with T.E. buffer and a further 1 in 5 with formamide loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). The diluted products were denatured at 95°C for 10 minutes, and 10 µl of each sample was applied to a native 6% polyacrylamide gel (49:1, bisacrylamide:acrylamide). A 0.4 mm thick gel was polymerised in a 40 cm X 20 cm sequencing gel cast (Sequigen; Biorad). The gel was prepared and electrophoresed in 1xTBE buffer (89mM Tris borate, 1mM EDTA buffer, pH 8.0).

Two different electrophoretic conditions were used in order to increase the sensitivity of the SSCP analysis. In one method, the 6% native polyacrylamide gel was run at 4°C (in a cold room) and electrophoresis carried out for < 6 hours. In the other method, a similar gel containing 5% glycerol was electrophoresed at room temperature for over 16 hours.

Following the completion of electrophoresis, the gel was transferred to a 3MM Whatmann paper and dried under vacuum at 80°C (Biorad 583 gel dryer) for about 30 minutes. The gel was then exposed to X-ray film (X-Omat AR diagnostic imaging film; Kodak) at -70°C for 4 - 16 hours.

2.5.2 CHEMICAL CLEAVAGE OF MISMATCHES (CCM)

The CCM protocol involved five separate steps; (i) preparation of labelled probe and unlabelled test DNA, (ii) formation of heteroduplexes, (iii) chemical modification, (iv) chemical cleavage of mismatches and (v) electrophoresis and autoradiography.

Preparation of labelled probe and unlabelled test DNA

End-labelled wild-type PCR products that were used as probes in CCM were generated as described in section 2.4.5. Unlabelled test PCR products were prepared as described in section 2.4.2. The PCR products were purified (i.e. separated from unincorporated cold and radiolabelled nucleotides and primers) by adsorption to glassmilk, using a Geneclean II kit (BIO 101). Unlabelled PCR products were initially electrophoresed on a low melting point agarose gel (NuSieve; FMC) and the required products were excised from it. The gel slice, containing the DNA to be eluted, was melted in 3 volumes of 6M NaI solution (at 45-55°C). The DNA was then adsorbed to 10 µl of glassmilk. The glassmilk-DNA complex was washed three times with the commercially supplied "new wash buffer" to remove any agarose. DNA was eluted in 15 µl T.E. buffer at 65°C over 10 minutes. The radiolabelled wild-type PCR product was mixed directly with 3 volumes of the 6M NaI solution and the same method was followed. The elution of a typical PCR product yielded a concentration of approximately 100 ng/µl DNA.

Formation of heteroduplexes

15 ng of labelled probe (10^5 cpm) was mixed with 150-200 ng of each test DNA in a 20 µl volume of 1x annealing buffer (10x buffer: 1M Tris-HCl, pH 8, 3M NaCl). Heteroduplex formation was carried out in 0.5 ml microfuge tubes. 50 µl mineral oil was layered on to the mixture and it was heated to 99°C for 10 minutes and then placed in a 65°C water bath for two hours. The heteroduplexes were transferred to 1.5 ml siliconised microfuge tubes to be precipitated. To each tube was added, 3 µl of glycogen (Boehringer Mannheim; 20 mg/ml) and 750 µl of precipitation mix (63 mM sodium acetate, 20 µM EDTA, 80% ethanol) and the mixture was incubated at -70°C for 15 minutes. The tubes were centrifuged at 12,000 rpm for 10 minutes, the pellet was rinsed with 70% ethanol, air dried and resuspended in 14 µl T.E. buffer.

Chemical modification using hydroxylamine and osmium tetroxide

Each heteroduplex was split between two tubes (7µl each) i.e. one tube each for the hydroxylamine and osmium tetroxide reactions. For the hydroxylamine reaction, 1.39 G of hydroxylamine hydrochloride was mixed in 1.6 ml of pre-warmed (37°C) distilled water. The pH of the solution was adjusted to 6.0 by the addition of diethylamine (about 1.5 ml) and the solution was stored at 4°C for up to 7-10 days. 20 µl of this solution was added to the heteroduplex and incubation carried out at 37°C for 2 hours.

For the osmium tetroxide (OsO_4) reaction, a commercially available 4% OsO_4 solution (Aldrich; which was stored at 4°C for up to 3 months) was used. A 10x reaction mix was prepared by mixing 1.5 µl osmium tetroxide, 6.75 µl pyridine and 154 µl T.E. buffer on ice. 18 µl of this mix was then added to 7 µl of the heteroduplex. The mixture was incubated at 37°C for 1 - 2 hours.

Both the hydroxylamine and osmium tetroxide reactions were stopped by transferring the tubes onto ice which was followed by the addition of 750 µl of precipitation mix (63 mM sodium acetate, 20 µM EDTA, 80% ethanol). The DNA was precipitated at -70°C for 15 minutes and centrifuged at 12,000 rpm to obtain a pellet, which was rinsed and briefly air dried.

Piperidine cleavage of the chemically modified mismatch

50 µl of a 10% piperidine solution was added to each pellet and the tubes were vortexed for up to one minute. They were then incubated at 90°C for 30 minutes. The tubes were chilled on ice and the DNA was precipitated as before and the pellet was resuspended in 15 µl formamide loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol).

Gel electrophoresis and autoradiography

The resuspended pellet was heated to 95°C for 10 minutes and applied to an 8% denaturing polyacrylamide gel containing 7M urea. The gel mixture was prepared and polymerised in a sequencing gel apparatus (Sequigen; Biorad) as previously described. A labelled 1 kb DNA ladder (Gibco, BRL) was also denatured and electrophoresed on the gel in order to permit accurate size estimation of the cleavage products. Electrophoresis was carried out at 35 watts constant power till the bromophenol blue reached the end of the gel. The gel was transferred onto a pre-cut piece of 3MM Whatmann paper and autoradiographed using Kodak X-Omat AR film for a period of 12-16 hours.

2.5.3 SEQUENCING OF PCR PRODUCTS

All sequencing reactions performed in this project involved a prior PCR amplification. The PCR product was used as a template for asymmetric PCR amplification to generate single stranded DNA suitable for direct sequencing, as described in section 2.4.7. Sequencing was carried out by the dideoxy chain termination method (Sanger et al., 1977) using the Sequenase version 2.0 sequencing kit (USB). Products of the asymmetric PCR were precipitated using a final concentration of 2M ammonium acetate and an equal volume of isopropanol, in order to preferentially precipitate the PCR products and remove the unincorporated dNTPs and primers. The DNA pellet was washed in 70% ethanol and dissolved in 14 µl distilled water. 7 µl was used for each sequencing reaction.

The samples were treated exactly according to the protocol provided by the manufacturers of the Sequenase kit. 1 pM of the appropriate sequencing primer and 2 µl of 5x Sequenase reaction buffer (to give a final concentration of 40 mM Tris-HCl pH 7.5,

20 mM $MgCl_2$, 50 mM NaCl) were added to the DNA template (in a final volume of 10 µl). The sample was incubated at 65°C for 2 minutes and then allowed to cool slowly to < 35°C over 20 to 30 minutes. 1 µl of 0.1M DTT, 2 µl of diluted labelling mix (diluted 1:5 in distilled water), 0.5 µl of [α-^{35}S] dATP (1000 Ci/mmol; 10 µCi/µl) and 2 µl (3.25 units) of diluted Sequenase enzyme (T7 DNA polymerase; 13 units/µl; diluted 1:8 in enzyme dilution buffer [10mM Tris-HCl, pH 7.5, 5mM DTT, 0.5 mg/ml BSA]) were added to the mixture and this was kept at room temperature for 2-5 minutes. Meanwhile the four termination mixes (containing 80 µM of each dNTP and 8 µM of each ddNTP) were prewarmed at 37°C, to which 3.5 µl of the labelling reaction was added. The reaction was allowed to continue for 5 minutes before being terminated by the addition of formamide dye stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol).

The samples were heated at 75°C for 2 minutes before applying them to an 8% denaturing polyacrylamide gel containing 7M urea and which had been preheated to 50° C. The gel was electrophoresed at constant power in order to maintain the gel temperature at 50°C. The sequencing gel was fixed in a solution of 10% methanol:10% acetic acid for 10 minutes before drying in a vacuum gel dryer (Biorad) at 80°C for 30 minutes. The dried gel was exposed for 16-18 hours to Kodak X-Omat AR film in a cassette using intensifying screens. The sequence was analysed manually.

2.6 RECOMBINANT DNA METHODS

2.6.1 RESTRICTION ENDONUCLEASE DIGESTS OF dsDNA

All restriction enzymes were obtained from Pharmacia, because of the compatibility of almost all their enzymes with one reaction buffer, One-Phor-All *PLUS* (10x buffer: 100 mM Tris-acetate, pH 7.5, 100 mM Magnesium acetate, 500 mM potassium acetate). All digestions were performed according to the manufacturer's recommendations. Digestion at internal sites in plasmids or PCR products were

performed using 1-2 units/μg of DNA for 1 hour. Digestion at termini of PCR products was performed using at least 3 units/μg of DNA for a minimum of 3-5 hours.

2.6.2 LIGATION OF DNA FRAGMENTS

As all ligations were performed at unique cohesive ends on either end of the insert, no end-filling or treatment with phosphatase was necessary. DNA fragments to be ligated were simply precipitated prior to the ligation reaction. Restriction-digested PCR products were purified by agarose gel electrophoresis followed by elution with adsorption to glassmilk (Geneclean II; as in section 2.5.2), in order to remove the digested termini and unincorporated primers and dNTPs.

Ligations were performed with a 3-5:1 molar excess of insert to vector. The ligation reaction consisted of 50-100 ng of linearised vector mixed with approximately one tenth of the purified digested PCR product (insert), 1 unit of T4 DNA ligase (1 unit/μl; Boehringer Mannheim) in 10 μl of 1x ligase buffer (10x buffer: Tris-HCl 660 mM, pH 7.5, 50 mM DTT, 10 mM ATP). Ligation was performed for 16 hours at room temperature (22°C), as recommended by the supplier.

2.6.3 PREPARATION OF COMPETENT *E. Coli* CELLS

A glycerol stock of *E.Coli* LE392 was used to inoculate 10 ml L-broth (1% bactotryptone, 0.5% yeast extract, 1% NaCl pH 7.0, antibiotic free) which was incubated at 37°C overnight with orbital shaking (at 225 rpm). 1 ml of this overnight culture was used to inoculate 50 ml of L-broth and incubation was carried out at 37°C for approximately 2 hours. When the optical density at 600nm was approximately 0.3, the culture was transferred to a 50 ml Falcon centrifuge tube and the cells were pelleted by centrifugation at 2,500 rpm for 10 minutes at 4°C in an IEC DPR-6000 centrifuge. The cell pellet was resuspended in 10 ml 10 mM $MgSO_4$ and incubated on ice for 20 minutes. The cells were again pelleted as before and resuspended in 5 ml chilled 50 mM $CaCl_2$, incubated on ice for at least 30 minutes, and were then ready for use in the transformation protocol.

2.6.4 TRANSFORMATION OF COMPETENT CELLS

Plasmids were transformed into *E.Coli* LE392. Half the ligation mixture was added to 100 µl of ice-cold competent LE392 cells in a 1.5 ml microfuge tube. The cells and DNA mixture were incubated on ice for 30 minutes after which, they were given a heat-shock of 1 minute at 42°C. The cells were allowed to recover for 2 minutes at 20°C before adding 800 µl of L-broth (prewarmed to room temperature) and incubating at 37° C for 45 minutes. The cells were briefly pelleted for 20 seconds, and all but ~100 µl of medium was poured off. The cell pellet was resuspended in this residual 100 µl of medium and spread on L-agar plates (L-broth, 1.5% agar) containing 100 µg/ml ampicillin. The culture was allowed to dry on the agar and the plates were inverted and incubated overnight at 37°C. Colonies were visible the following morning.

2.6.5 SCREENING FOR PUTATIVE RECOMBINANTS

L-agar plates were poured containing ampicillin (100 µg/ml), which only allowed the selection of those clones containing closed circular plasmid DNA with an intact ampicillin resistance gene. Analysis of recombinant colonies was initially performed by direct PCR screening as described in section 2.4.9. Plasmid DNA extracted from cultures (containing ampicillin) of the various positive colonies (by PCR) was used to perform detailed restriction mapping.

As all the inserts were generated by PCR using the proof-reading deficient *Taq* polymerase, it was necessary to rule out the presence of random mutagenesis in the selected clones (see section 3.5.2). A CCM analysis was therefore performed on at least 5-6 positive clones. Furthermore, for the confirmation of the site-directed mutagenesis (Q-45V; section 3.5.3) and the repair of the K10 5'UTR sequence (section 3.5.4) it was necessary to confirm these small sequence alterations by direct sequencing. For this purpose, flanking PCR primers were designed and direct sequencing of asymmetric PCR products was performed as described in section 2.5.3.

2.6.6 SMALL SCALE PREPARATION OF PLASMID DNA

Small scale plasmid DNA was prepared by the alkali-lysis method (Sambrook et al., 1989). An isolated bacterial clone was picked with a sterile toothpick and used to inoculate a 10 ml L-broth culture (containing 100 μg/ml ampicillin), which was incubated at 37°C overnight with orbital shaking. 1.5 ml of this overnight culture was briefly pelleted in a 1.5 ml microfuge tube. The supernatant was removed, and the pellet resuspended in 100 μl of solution A (50 mM glucose, 25 mM Tris-HCl, pH 8, 10 mM EDTA), to which 2 mg/ml lysozyme was freshly added. 200 μl of solution B (0.2 M NaOH, 1% SDS) was added and the tube mixed by rocking. A clear but viscous solution was obtained after incubating for 5 minutes on ice. 150 μl of solution C (5 M potassium acetate, 11.5% glacial acetic acid to get pH 4.8) was then added, followed by inversion and a quick vortex to mix the contents. The tube was held on ice for 5 minutes to allow the DNA and protein to precipitate, then centrifuged at 12,000 rpm for a minute. The supernatant was transferred to a fresh 1.5 ml tube and phenol:chloroform extraction was performed. The solution was then treated with RNase at a final concentration of 10 μg/ml at 37°C for 30 minutes. The plasmid DNA was then precipitated with ammonium acetate and ethanol and the pellet was dissolved in 50 μl T.E. buffer.

2.6.7 LARGE SCALE PREPARATION OF PLASMID DNA

Large scale plasmid DNA was prepared using Qiagen™ columns. DNA produced by this method was suitable for eukaryotic cell transfection without further purification. A single transformed bacterial colony was used to inoculate a 10 ml L-broth culture containing 100 μg/ml ampicillin. After overnight incubation, 1 ml was used to inoculate a 500 ml culture which was incubated at 37°C overnight by shaking (225 rpm). Bacterial cells were harvested by centrifugation at 4,000 rpm for 10 minutes at 4°C in a Sorvall rotor. The supernatant was gently resuspended in solution I (50 mM glucose, 25 mM Tris-HCl, pH 8, 10 mM EDTA; 2 ml / 50 ml culture) to which 2 mg/ml lysozyme had freshly been added. The cell suspension was incubated on ice for 5 minutes. Freshly-made solution II (0.4 M NaOH, 1% SDS) was added (4 ml / 50 ml culture) and the mixture was gently inverted until it turned clear. This mixture was incubated on ice for a further 5

minutes. Solution III (5 M potassium acetate, 11.5% glacial acetic acid to get pH 4.8; 6 ml / 50 ml culture) was added and the solution mixed gently until a heavy white protein precipitate was formed. The solution was incubated on ice for 30 minutes and then centrifuged at 5,000 rpm for 10 minutes at 4°C to pellet the white precipitate (bacterial chromosomal DNA). The supernatant was added to a Qiagen™ column, equilibrated with 10 ml QBT (750 mM NaCl, 50 mM MOPS, 15% ethanol, pH 7, 0.015% Triton-X), and was allowed to flow by gravity. 30 ml buffer QC (1 M NaCl, 50 mM MOPS, 15% ethanol, pH 7) was added and allowed to flow through by gravity. The plasmid DNA was eluted with buffer QF (1.25 M NaCl, 50 mM MOPS, 15% ethanol, pH 7) and the eluant mixed with 0.7 volumes of isopropanol to precipitate the plasmid DNA. The mixture was centrifuged at 5,000 rpm for 30 minutes followed by one wash in 70% ethanol. The plasmid DNA pellet was resuspended in 500 µl of T.E. buffer.

2.7 GENES AND VECTORS

The factor IX coding sequence used in the construction of the various K10-factor IX expression vectors was obtained from the retroviral vector pLXSN5'G3'cVI (pLIXSNL-2 in Gerrard et al., 1993). This was constructed by Dr. Ann Gerrard by inserting the factor IX cDNA from p5'G3'cVI (Anson et al., 1985) into the polylinker site (X) of the retroviral vector pLXSN (Miller and Rossman, 1989), and kindly provided by Professor G.G. Brownlee (Oxford University). In this project, primers were designed to amplify the entire factor IX coding sequence from the "AUG" initiation codon to 9 nucleotides downstream of the "UAA" translation termination codon (1.4 kb).

A cloned DNA segment containing 5 kb of upstream sequence of the bovine cytokeratin VI gene (BKVIb; from the genomic clone λKBVIb; Lehnert et al., 1984; Reiger et al., 1985), just short (eight nucleotides) of the natural BKVIb cap site (Blessing et al., 1989; called the K10 promoter because it is the bovine counterpart of the human cytokeratin 10 gene [K10]; Bailleul et al., 1990) was obtained from Jose Jorcano (Madrid, Spain). A 0.85 kb, SV40 polyadenylation cassette (SV40 polyA) containing the SV40 small-t antigen intron (66 nucleotides) and polyadenylation signal sequence was obtained from pRSVβglobin (Mulligan et al., 1979). The K10 promoter and SV40 polyA

sequence were directionally subcloned into the multiple cloning site (MCS) of pIC20H leaving a few restriction enzyme sites of the MCS for insertion of cDNAs. This expression vector (K10-polyA cassette) was constructed by Dr. D. Fowlis in our department. This was used as the backbone for the various K10-factor IX expression constructs made in this project. The genomic BKVIb clone (λKBVIb; Lehnert et al., 1984; Reiger et al., 1985) was used as a template for PCR, in order to obtain sequences at the 3' end of the gene (see section 3.5.5).

The plasmid pIRV-Neo-Act-lac Z, which contains the *E.Coli lacZ* gene under the control of the rat β-actin promoter (Beddington et al., 1989), was used to optimise liposome-mediated transfection of linearised DNA fragments into a transformed human keratinocyte (HaCaT) cell line (Boukamp et al., 1988). This plasmid was linearised (9.6 kb) at a *Not*I site in the vector and used for transfection (see section 3.6.1).

The various K10-factor IX constructs were co-transfected with pSV2Neo (Southern and Berg, 1982) to confer G418 (a neomycin analogue) resistance as a selective marker. pSV2Neo contains the coding sequence of the neomycin resistance gene driven by a SV40 viral promoter. Prior to its use in co-transfection, pSV2Neo was linearised (using *Eco*RI) and used in a 20 fold reduced concentration, in order to increase the probability of successful co-selection.

2.8 RNA ANALYSIS

The extraction of total cellular RNA and confirmation of its structural integrity by denaturing agarose gel electrophoresis have been described in section 2.3.2.

2.8.1 TRANSFER OF RNA TO NYLON MEMBRANES

For Northern analysis, RNA samples (10 μg) were prepared and electrophoresed as described in section 2.3.2. When the gel had run a sufficient distance (as judged from the migration of the loading dye) it was washed in 2-3 changes of distilled water for 30 minutes each, in order to remove the formaldehyde. The marker lanes were then cut from

the rest of the gel. The gel strip containing the marker lane was then stained in ethidium bromide (5 μg/ml) for 1 hour, followed by destaining in distilled water over 16 hours, viewed on a UV-transilluminator and photographed alongside a scale.

The main gel was placed on two sheets of 3MM Whatmann paper on a platform, which were pre-wetted in 2xSSC and made to function as a wick by dipping into a reservoir containing 10xSSC (1.5 mM NaCl, 150 mM sodium citrate). The edges of the gel were shrouded by plastic film in order to prevent the 10xSSC solution bypassing the gel into the blotting materials. A piece of the nylon membrane (Hybond-N; Amersham) and two sheets of 3MM Whatmann paper were cut to the size of the gel. The blotting membrane was carefully placed over the gel avoiding the trapping of any air bubbles between them. The two sheets of filter paper were soaked in 2xSSC and similarly placed over the nylon membrane. A pack of dry paper towels was laid on top of the filter paper. Finally, a weight (~1 Kg) was placed over the towels. Transfer was carried out overnight at room temperature. The transferred RNA was cross-linked onto the nylon membrane by placing it on a UV-transilluminator for 4 minutes.

2.8.2 RANDOM PRIMED LABELLING OF DNA

DNA segments to be used as a probe were labelled with [α-^{32}P] dCTP (Amersham) using a random primer labelling kit (Boehringer Mannheim). 100 ng of the double stranded DNA segment was denatured by heating, and to it were added 2 μl of (0.5mM solution) each dNTP (except dCTP), 4 μl of the reaction mix (containing reaction buffer and random hexanucleotides), 2 units of Klenow enzyme and 50 μCi of [α -^{32}P] dCTP (3000 Ci/mmol; 10 Ci/ml). The mixture was incubated at 37°C for 30 minutes. Unincorporated nucleotides were removed by passing the sample through a Nick column (Pharmacia). The specific activity of the probe was measured and it was used in a concentration of 2 X 10^6 counts/ml after heating at 100°C for 10 minutes.

2.8.3 NORTHERN HYBRIDISATION

The nylon membrane and a suitably sized mesh were pre-wet in 2xSSC. The two were rolled together and placed inside a Hybaid hybridisation bottle. To the bottle, 10-15

ml (5-10 ml for a small bottle) prehybridisation solution (5xSSC, 1xDenhardt's solution, 100 μg/ml salmon sperm DNA, 0.5% SDS, 50% formamide) was added, and incubation carried out in a Hybaid rotating oven at 42°C for at least 1 hour. The prehybridisation solution was then replaced by a similar quantity of hybridisation solution (5xSSC, 1xDenhardt's solution, 100 μg/ml salmon sperm DNA, 0.5% SDS, 50% formamide, 10% dextran sulphate, 2 X 10^6 counts/ml denatured probe) and incubation carried out overnight at 42°C in a rotating oven.

2.8.4 POST-HYBRIDISATION WASHES

The hybridisation solution was replaced by 2xSSC, 0.1% SDS and the first wash was carried out in the hybaid bottle in the 42°C oven. The washing procedure after this was dependent on the counts read on the membrane. In most cases a 15 minute wash with 0.2xSSC, 0.1% SDS at 65°C in a shaking water bath, followed by one or two such washes at room temperature on a shaker, were adequate to get the counts down to < 10 cpm. The hybridised membrane was then exposed to Kodak X-Omat autoradiographic film.

2.9 TISSUE CULTURE

2.9.1 INTRODUCTION

Cells were maintained at 37°C in a humidified atmosphere containing 5% CO_2, in 80cm^2 tissue culture flasks. Cells were grown in "Special Liquid Medium" (formulated for keratinocyte cultures by Dr. Ian Freshney; supplied by Gibco, BRL), supplemented with 10% foetal calf serum (FCS; Gibco, BRL; Cat. No. 10099-075), 10 mM L-glutamine, 100 units/ml penicillin and 100 μg/ml streptomycin. Cultures were usually split every 3-4 days by washing cells once each in 10 ml PBS (Gibco, BRL; Cat. No. 14040-091) and versene (1:5000; Gibco, BRL; Cat. No. 15040-033), followed by incubation in 2.5 ml trypsin solution (0.25%; Gibco, BRL; Cat. No. 25050-022) for 5-10 minutes at 37 °C. Cells were observed under low magnification to determine if trypsinisation was

complete before the addition of 7.5 ml fresh tissue culture medium containing 10% FCS to inactivate trypsin. Cell numbers were estimated using an improved "Neubauer" haemocytometer (BDH) and fresh tissue culture flasks were seeded with 5 X 10^5 cells per 80 cm^2.

2.9.2 CRYOPRESERVATION OF EUKARYOTIC CELLS

Cells were grown to 80% confluency and trypsinised as described in 2.9.1. After inactivating the trypsin with fresh serum-containing medium, cells were centrifuged at 1,000 rpm for 5 minutes. The supernatant was discarded and cells were resuspended in 1 ml of 10% DMSO in tissue culture medium. The cells were slowly cooled by storage in Nunc cryostat tubes, wrapped in bubblepak at -70°C for two days and then transferred to liquid nitrogen stores. To regrow cells frozen by this method, the tubes were thawed quickly at 37°C, added to fresh medium, centrifuged at 1,000 rpm for 5 minutes and replated. The following day the media was aspirated off, cells were washed in PBS and fresh media added.

2.9.3 LIPOSOME MEDIATED TRANSFECTION OF HaCaT CELLS

For this purpose a commercially available liposome transfection reagent, DOTAP (Boehringer Mannheim) was used exactly according to the supplier's recommendations. HaCaT cells were transfected at 60-80% confluence in 80cm^2 tissue culture flasks. For each flask of cells, 70 µl of DOTAP was made up to 250 µl with the Hepes-buffered saline (Sigma; HBS; 20 mmM Hepes, 150 mM NaCl) in a polystyrene reaction vial. To avoid contact with air the DOTAP solution was aspirated through the rubber seal using a sterile cannula. The DNA to be transfected was diluted up to a final volume of 250 µl with HBS in another polystyrene reaction vial. For optimisation of transfection of linearised plasmids into HaCaT cells, a *lac Z* reporter construct (pIRV-Neo-Act-lac Z; Beddington et al., 1989) was used in concentrations of 1 µg, 5 µg and 10 µg and transfections were carried out in duplicate. As seen in section 3.6.1, 10 µg DNA showed the maximum transfection efficiency and therefore 10 µg was used for the transfection of all K10-factor IX constructs. Both the solutions were mixed and incubated at room

temperature for 10 minutes. Meanwhile, the cells were washed twice in PBS and 12 ml tissue culture medium was added to the flask containing cells to be transfected. The liposome-DNA solution was then added to the flask. Cells were incubated at 37°C for 24 hours in 5% CO_2. The medium was then replaced with fresh medium and incubated for a further 48 hours.

2.9.4 DETERMINATION OF THE EFFICIENCY OF TRANSFECTION

The plasmid pIRV-Neo-Act-lac Z (Beddington et al., 1989), containing the *E.Coli lac Z* gene driven by the rat β-actin promoter was used to determine the efficiency of transfection. Transfection was carried out in duplicate as described in 2.9.3. Transfected cells were rinsed twice in PBS and fixed in 0.2% glutaraldehyde, 2 mM $MgCl_2$, 5 mM EGTA in PBS for 5 minutes at 4°C. The fix was aspirated from the flasks and cells washed twice in 2 mM $MgCl_2$, 0.1% sodium desoxycholate, 0.02% NP40, 0.05% BSA in PBS. The reaction mix (1 mg/ml X-Gal, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 2 mM $MgCl_2$ in PBS) was then added to the flask and incubation carried out at 37°C for 24 hours. The number of blue-stained cells were counted.

2.9.5 SELECTION OF STABLE CLONES

In order to have a selectable marker for identifying transfected cells, they were co-transfected with pSV2Neo (Southern and Berg, 1982) to confer G418 (a neomycin analogue) resistance. pSV2Neo contains the neomycin resistance gene driven by a SV40 viral promoter. Prior to its use in transfection, pSV2Neo was also linearised (using *Eco*RI). As the selectable marker was not linked to the expression vectors, in order to increase the probability of co-selection with the factor IX constructs, the latter were used in a 20 fold excess concentration over that of pSV2Neo.

A protocol identical to that used for the transfection of the linearised *lacZ* reporter construct (as described in 2.9.4) was used for the factor IX expression constructs. Transfections were carried out using 10 μg of the appropriate factor IX

construct along with 0.5 μg of pSV2Neo. Control transfections were also performed using only pSV2Neo, K10-FIX II and liposomes without DNA.

72 hours after transfection the confluent flasks were split 1 in 5. Two parts, were stored in liquid nitrogen (in 10% DMSO) and two plated on to fresh 80 cm^2 tissue culture flasks. The fifth part was plated onto 10 tissue culture petridishes (100 x 20 mm). Cells were grown in medium containing 400 μg/ml G418 sulphate (Geneticin®; Gibco, BRL). Flasks and petridishes were thoroughly washed (to remove dead cells) and fresh, G418-containing tissue culture medium was replaced every 3-4 days. After about 20 days (5-6 washes), individual clones of cells were observed in those that had been transfected with pSV2Neo. Cells that were transfected with only K10-FIX II or liposome without DNA had all died by this stage.

Cells in the petridishes were grown till the average clone contained about 50 cells when they were ready for picking. Clones in 80 cm^2 flasks were grown in 400 ng/ml G418-containing medium until they overlapped and reached subconfluency. They were then trypsinised and split into fresh 80 cm^2 flasks, and maintained in 200 ng/ml G418 for a further 2-3 passages. At this stage they were used for analysis of factor IX gene expression and the G418 selection was withdrawn. These cells were used for analysis as "pools" of transfected clones.

Individual stably-transfected clones were isolated for each K10-factor IX construct. The method involved the use of cloning rings and the protocol described by Malcolm et al. (1987) was followed. Clones in petridishes that had approximately 50 cells each and were adequately separated from other clones, were selected for the cloning procedure. They were first identified under low power magnification on an inverted microscope and were encircled (~0.5 cm diameter) on the underside of the dish using a marker pen, by inserting the pen tip between the underside of the dish and the objective. A total of 10-15 suitable clones were selected for each transfected factor IX expression construct (not more than 3-5 per dish).

The wide ends of plastic pipette tips (blue) were excised and used as cloning rings. After suitable clones had been selected, cells in the petridishes were washed once each with PBS and versene. The flat (uncut) open surface of the ring was coated with

autoclaved silicone grease. Cloning rings were placed on each clone using the marked circles to identify exact positions and tapped firmly in place. About 3-5 clones were selected per dish and all were treated promptly and simultaneously to avoid cells from drying. A drop of trypsin (~100 µl) was added to each ring (using plugged pipette tips) and incubation carried out for about 5 minutes at 37°C. 200 µl of fresh serum containing medium was added to each ring and the total solution was flushed up and down a few times to free cells. These were transferred to individual wells in a 24-multiwell dish. Cells were incubated at 37°C in 5% CO_2 in a combination of conditioned and fresh medium (1 part:2 parts), containing G418 at a final concentration of 200 ng/ml. Cells were not disturbed for a period of one week. After that, the cells were washed and fresh medium containing 200 ng/ml G418 was replaced. Clones in wells that reached subconfluency, were trypsinised and transferred to individual petridishes (35 x 10 mm). Culture volumes were gradually scaled up from these petridishes to 25 cm^2 flasks and finally to 80 cm^2 flasks. Throughout the scaling up process, cells were grown in medium containing 200 ng/ml G418. This selection was finally withdrawn at the 80 cm^2 flask stage, when the cells were ready for gene expression analysis.

2.9.6 COLLECTION OF CONDITIONED MEDIA

Transfected HaCaT cells (pools or individual clones) were grown in media containing 500 ng/ml vitamin K (Sigma). For the collection of conditioned media, transfected HaCaT cells were washed four times in PBS and incubated in serum-free "Special Liquid Medium" for 24 hours at 37°C. The conditioned medium was cleared by centrifugation at 1,000 rpm for 10 minutes and stored on ice until being loaded on to microtitre plates for ELISA (usually within an hour).

2.10 ENZYME LINKED IMMUNOSORBENT ASSAY OF FACTOR IX

Enzyme Linked Immunosorbent Assay (ELISA) was used to detect secreted factor IX protein in conditioned media. Two monoclonal antibodies were used in a

"sandwich" ELISA for the detection of factor IX antigen. The two antibodies used were A5 (anti-heavy chain of factor IXa) and A7 (anti-light chain of factor IXa) (Smith and Ono, 1984; Smith, 1988) which were generously supplied to us by Dr. K. Smith (University of New Mexico, USA). The A7 antibody detects the calcium-dependent epitope in the light chain (*gla* region) of factor IX and distinguishes between biologically-active factor IX that contains γ-carboxyglutamic acids and non-activated factor IX, which does not (Smith et al., 1986; Smith, 1988). The protocol used in this study was essentially as described by Smith (1988). The A7 antibody was biotinylated and detected using avidin-peroxidase. As this assay detects the active form of factor IX it was used as an indirect measure of biological activity of the factor IX produced by the expression constructs (Gerrard et al., 1993).

Polystyrene flat bottom microtitre plates (Nunc) were used in the assay. All additions were made in 100 µl volumes using an 8-channel multipipette. The plates were coated with the A5 antibody at a concentration of 0.01 mg/ml in 0.05 M $NaHCO_3$ buffer at pH 8.6 (made by adding a small amount of dilute NaOH). The plates were incubated for 4 hours at 37°C or overnight at 4°C. The plates were washed three times in wash buffer I (0.1 M NaCl, 0.05 M Tris-HCl, pH 7.5, 0.1% Tween 20). The concentration of Tween 20 was ten fold more than that recommended by Smith (1988), in order to reduce the background due to non-specific binding (Smith, personal communication). Washing was performed using two different methods. One included the immersion of the plate in a container that was placed on a shaker, followed by tapping the plates to empty the contents of the wells. The other involved the use of a plastic squeezable bottle, with the water jet washing individual wells. The plates were blocked with 5% BSA (Fatty Acid free; Boehringer Mannheim) in 0.1 M NaCl, 0.05 M Tris-HCl, pH 7.5 with 0.05% azide for 2 hours at room temperature or overnight at 4°C. Plates were washed three times in wash buffer I.

Serial dilutions of conditioned media collected from the various transfected cells (and an untransfected negative control) were added in triplicate. To make a standard curve, serial dilutions of commercially available pure factor IX (Dakopatts) to yield concentrations between 0-50 ng/ml were also loaded in triplicate. Dilutions were made in

a buffer containing 0.1 M NaCl, 0.05 M Tris-HCl, pH 7.5, 1 mg/ml BSA. Incubation was carried out for 3 hours at 37°C or overnight at 4°C. Plates were then washed six times in wash buffer II (0.1 M NaCl, 0.05 M Tris-HCl, pH 7.5, 0.01 M $CaCl_2$, 0.1% Tween 20).

Antibody A7 was diluted to 1 nM (1:6667 for a 1 mg/ml solution) in 0.1 M NaCl, 0.05 M Tris-HCl, pH 7.5, 0.005 M $CaCl_2$, 1 mg/ml BSA and 100 µl was added to each well. Incubation was carried out for 2-5 hours at 37°C or overnight at 4°C. Plates were then washed six times with wash buffer II. To each well was added, avidin-peroxidase (Boehringer Mannheim) diluted 1:10,000 in buffer containing 0.1 M NaCl, 0.05 M Tris-HCl, pH 7.5, 0.005 M $CaCl_2$, 1 mg/ml BSA, 0.05% Tween 20. Incubation was carried out for 45 minutes at room temperature. Plates were washed six times in wash buffer II. A solution of 0.1 M sodium citrate (pH 4.5) with 1 mg/ml OPD (1,2-Phenylenediamine; Dakopatts) and 0.033% hydrogen peroxide was added to each well. The solution and plates were shielded from light with an aluminium foil cover. Usually in about 5 minutes the colour development was adequate and the reaction was stopped by the addition of 2.5 M H_2SO_4. Optical densities (OD) were recorded on a microtitre plate reader at 490 nm.

CHAPTER: 3

RESULTS

3.0 RESULTS

3.1. AMPLIFICATION OF THE SEQUENCE OF THE FACTOR VIII GENE

The sequence of the factor VIII gene was amplified using both, genomic DNA and total cellular RNA as templates obtained from peripheral blood lymphocytes. All 26 exons except the large, central exon 14, were amplified by RT-PCR followed by nested PCR of the ectopic factor VIII mRNA transcripts. Selected individual exons (including exon 14) were amplified using flanking intronic primers and genomic DNA as the template. The putative promoter and polyadenylation signal sequence of the factor VIII gene were also amplified using appropriate flanking primers.

3.1.1. AMPLIFICATION OF THE ENTIRE ESSENTIAL SEQUENCE

The entire coding sequence, putative promoter and polyadenylation/cleavage sequence, representing the entire essential sequence of the factor VIII gene (figure 3), was amplified in 13 unrelated and randomly selected haemophilia A patients. Twenty-five of the twenty-six exons were amplified in four overlapping PCR fragments. These were obtained by RT-PCR followed by nested PCR amplification of ectopic factor VIII transcripts. The four RT-PCR segments (as seen in figure 3) are named R1 to R4 ('R' indicating the use of RNA as template) from 5' to 3' of the factor VIII cDNA and spanned the following factor VIII nucleotide numbers, R1: -50 to 1223; R2: 1114 to 2282; R3: 5000 to 6171 and R4: 6071 to 7143 (numbering according to Wood et al., 1984). Primer sequences for the amplification of all segments except R3 were as described by Naylor et al. (1991) and obtained from Professor F. Giannelli (personal communication; appendix I). Primers for the amplification of segment R3 were designed using OLIGO™ version 3.4 (section 2.4.1; table 1).

The entire exon 14 was initially amplified from genomic DNA in a single PCR product of 3.2 kb, using flanking intronic primers. This large PCR fragment was then amplified in two segments using nested primers. The two segments were named D1 and

Table 1: Primer sequences spanning the entire factor VIII coding sequence

Name	Orientation	Sequence (5' - 3')	mRNA Position[a]
R1-E	Forward	ATC ACC TTT TCA ACA TCG CT	224 - 244
R1-G	Forward	CCA GTC AAA GGG AGA AAG AA	410 - 430
R1-F	Reverse	CCA ACT TTT CCC TTC ATC AA	662 - 682
R1-H	Reverse	CTG TCC AAG GTC CAT CAA GA	953 - 973
R2-E	Forward	CAG CGG ATT GGT AGG AAG TA	1306 - 1326
R2-G	Forward	TTC AAG GAG ATT ACC AAA AG	1518 - 1538
R2-F	Reverse	TTG GGG AGA AAG CGT TGT AT	1828 - 1848
R2-H	Reverse	AGA ATG GGA ATA GGG TGA GT	2058 - 2078
R3-A	Forward	ACC CAC CAG TCT TGA AAC GC	4973 - 4993
R3-B	Reverse	TGT CCT GAA GCT GTA ATC TG	6163 - 6183
R3-C	Forward	GGG AAA TAA CTC GTA CTA CT	5000 - 5020
R3-D	Reverse	GTA ATC TGA AAA TCT CTA AT	6151 - 6171
R3-E	Forward	GGC TCT GGG ATT ATG GGA TG	5171 - 5191
R3-G	Forward	GCC TCT CGT CCC TAT TCC TT	5392 - 5412
R3-F	Reverse	GAT GGT GAA AAA CAG AGC AA	5684 - 5704
R3-H	Reverse	CAA AAA CAC CTG GAT AGA GA	5988 - 6008
R4-E	Forward	CCA GCC TCT ACA TCT CTC AG	6338 - 6358
R4-G	Reverse	CTG AGA GAT GTA GAG GCT GG	6338 - 6358
R4-F	Forward	CAG GTG AAT AAT CCA AAA GA	6721 - 6741
22Fd	Forward	TCC ACT GGA ACC TTA ATG GT	6412 - 6432
22Rev	Reverse	ACC ATT AAG GTT CCA GTG GA	6412 - 6432
23Fd	Forward	ATA GCA TTC GCA GCA CTC TT	6524 - 6544
23Rev	Reverse	AAG AGT GCT GCG AAT GCT AT	6524 - 6544
D2-C	Forward	GAA TGT AGT TTT GCC TCA GA	3678 - 3698
D2-D	Reverse	TTT CAC AAG CGT TCA GGG AC	4854 - 4874
D2-E	Forward	CGC AAC GTA GTA AGA GAG CT	3980 - 4000
D2-F	Reverse	ACC CAT TGC TAG TTT CCG TA	4578 - 4598

R1 - R4 represent the four segments amplified by RT-PCR and D1 & D2 the two halves of exon 14 amplified by DNA-PCR. R1, 2 and 4 were amplified using primer sequences described by Naylor et al. (1991, 1993a). R3-A, B, C & D were designed for the initial RT-PCR amplification of that segment. All other primer sequences listed are for the nested PCR amplifications of the named segments. All forward and reverse primers within each segment have been specifically designed to be compatible for PCR when used in any combination. Primers 22/23-Fd/Rev were used for the detection of the characteristic intron 22 mRNA abnormality (Naylor et al., 1992, 1993a) along with R4-E, F & G (figure 19). [a]Numbering is according to Wood et al. (1984).

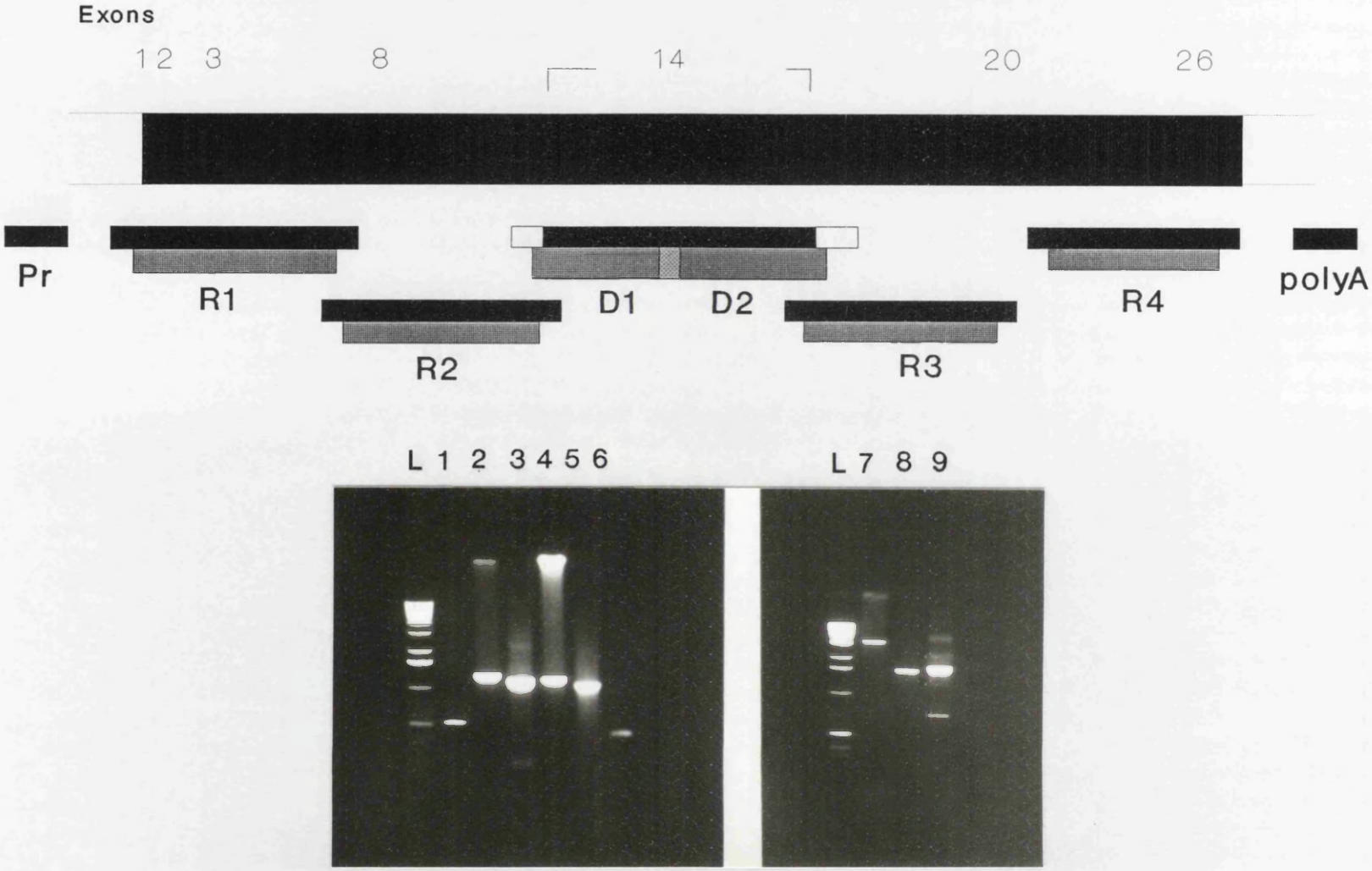
Figure 3: Amplification of the essential sequences of the factor VIII gene

A diagrammatic representation of the factor VIII cDNA amplified in 8 overlapping segments is shown. The positions of selected exons are indicated. The segments amplified by RT-PCR (R1-R4) followed by nested PCR (indicated by hatched boxes) of the ectopic factor VIII transcripts are shown. Exon 14 is amplified in two overlapping nested segments (D1 and D2; indicated by hatched boxes) from the full-length exon 14 PCR product. Two additional segments amplified by DNA-PCR, one around the putative promoter region (Pr) and the other around the polyadenylation signal sequence (polyA) are indicated.

The products of amplification resolved by agarose gel electrophoresis are shown. Lanes 2-5 show the RT-PCR segments R1-R4 respectively. Lanes 1 and 6 show the segments Pr and polyA, respectively. Lane 7 shows the 3.2 kb full-length exon 14 PCR product and lanes 8 and 9 show segments D1 and D2, respectively.

Factor VIII cDNA

Amplification of the entire essential sequence



D2 ('D' indicating the use of DNA as template). Primer sequences for the amplification of exon 14 were as described by Higuchi et al. (1991a) (GC-clamps were excluded) and those for segments D1 and D2 were as described by Naylor et al. (1991) and also obtained from Professor F. Giannelli (personal communication; appendix I). Segments D1 and D2 spanned the following factor VIII nucleotide positions, D1: 2114 to 3692 and D2: 3537 to 5219 (numbering according to Wood et al., 1984). Two additional DNA-PCRs were performed in order to amplify the putative promoter and polyadenylation sequence using primer sequences described by Higuchi et al. (1991b). The nucleotide positions spanned by these primers were as follows: -505 to +16 for the promoter and 8665 to 9056 for the polyadenylation sequence (these numbers are according to Gitschier et al., 1984). The amplification of the entire exon 14, segments D1 and D2, the additional segments for the promoter and polyadenylation sequence, are also shown in figure 3.

All the primer sequences obtained from published data or by personal communication have been listed in appendix I.

Additional oligonucleotides were designed (section 2.4.1) for further nested PCR amplifications of segments R1-4 and D2. These were so designed, as to be compatible for PCR in any combination of forward and reverse primers, within any one segment. The primer sequences, their orientation and nucleotide positions are listed in table 1. They have been named according to the segment they map to, followed by an 'E', 'F', 'G' or 'H' within R1-4 and 'C', 'D', 'E' or 'F' within D2. These oligonucleotides were also used as primers for direct sequencing. Primers 22 Fd/Rev and 23 Fd/Rev (table 1) were specifically designed to amplify across the junction of exons 22 and 23, for the diagnosis of the common intron 22 mRNA abnormality (Naylor et al., 1992; 1993a) seen in about half of all severe haemophilia A cases.

3.1.2. AMPLIFICATION OF SELECTED INDIVIDUAL EXONS

Exons 4, 7, 8, 11, 12 and 16 of the factor VIII gene were amplified in 12 unrelated haemophilia A patients selected at random. These exons ranged in size from 151 bp (exon 12) to 262 bp (exon 8) and were all amplified as PCR products of <400 bp in length, thus being amenable for PCR-SSCP analysis (Orita et al., 1989). Other

individual exons were also amplified as required for the sequencing and confirmation of mutations detected in the ectopic factor VIII transcripts. Flanking intronic primers described by Higuchi et al. (1991a), but excluding the GC-clamps, were used for this purpose (appendix I).

3.2. CHARACTERISATION OF MUTATIONS IN THE FACTOR VIII GENE

Two distinct strategies were employed for the identification of mutations in the factor VIII gene. The choice depended on whether an RNA sample was available from the patient. Twelve unrelated haemophilia A patients were screened for mutations by PCR-SSCP of six exons, four of which were selected for their reported non-random concentration of mutations (Tuddenham et al. 1991). Thirteen other unrelated patients were screened by chemical mismatch cleavage (CCM) analysis of the entire essential sequence of the factor VIII gene. This comprehensive search for mutations was possible due to the availability of fresh blood samples and thus RNA samples from these individuals. All positive results obtained from the initial mutation screening method (SSCP/CCM), were eventually characterised by direct sequencing of asymmetrically amplified PCR products. The mutations characterised, are described in separate sub-sections, depending on the initial screening method used.

3.2.1. PCR-SSCP OF SELECTED EXONS

Exons 4, 7, 8, 11, 12 and 16 of the factor VIII gene were screened by PCR-SSCP. Twelve unrelated haemophilia A patients were screened by this method. Two different electrophoretic conditions were used (as described in section 2.5.1), in order to maximise the sensitivity of SSCP.

Mobility shifts on SSCP analysis

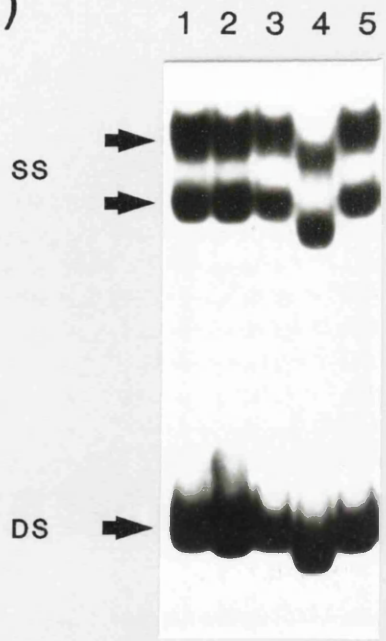
In the twelve unrelated individuals screened, two positive mobility shifts were detected. Both of these involved exon 4. Lane 4 in figure 4A shows altered electrophoretic mobilities at the level of both, single and double stranded DNA. The

Figure 4: SSCP analysis of exon 4

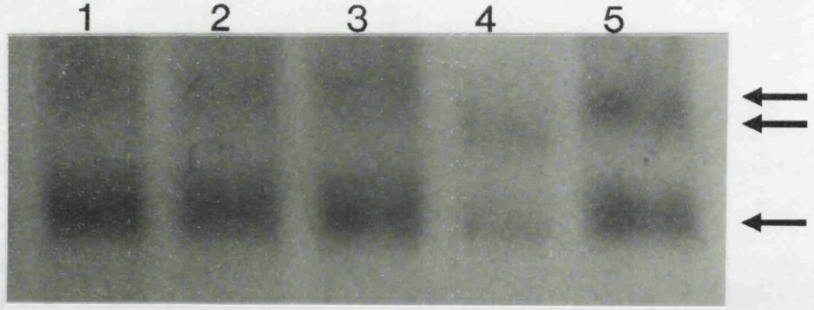
4A. SSCP analysis of exon 4 showing altered mobility of both the single (SS) and double (DS) strands in lane 4 (GLA14).

4B. SSCP analysis of exon 4 showing altered mobility of the single strands in lane 4, in another patient (GLA15). The altered mobility is indicated by arrows.

(A)



(B)



presence of a faster moving double stranded band suggested the presence of a small deletion. Lane 4 in figure 4B reveals a mobility shift in the single strands of the exon 4 PCR product of another patient. Samples in figure 4A were electrophoresed on a non-denaturing polyacrylamide gel at 4°C (in a cold room) whereas samples in figure 4B were electrophoresed at room temperature for >16 hours on a similar gel, containing 5% glycerol.

Characterisation of a 5 bp deletion in exon 4

The cause of the mobility shift seen in figure 4A was characterised by the direct sequencing of exon 4 in that individual (GLA14). Figure 5 reveals a 5 bp deletion of "TACCT" involving nucleotide positions 519 to 523 in exon 4 of the factor VIII gene. This involved amino acid positions 154, 155 and 156.

Segregation of the deletion for genetic diagnosis

The 5 bp deletion was resolved on a 10% denaturing polyacrylamide gel. This enabled the detection of individuals with the deletion and thus offered an unequivocal means of carrier diagnosis in female relatives. Figure 6 demonstrates the segregation of the deletion for this purpose.

3.2.2. CCM ANALYSIS OF THE ENTIRE ESSENTIAL SEQUENCE

The entire essential sequence was amplified in thirteen unrelated haemophilia A patients. Those individuals who had normal sized PCR products for each of the eight segments were assumed to have point mutations or other smaller gene alterations and subjected to chemical mismatch cleavage (CCM) analysis (Cotton et al., 1988).

The products of Chemical Mismatch Cleavage analysis

Nine of the thirteen patients analysed by this strategy revealed normal sized PCR products spanning the entire essential sequence (all except GLA3, 8, 11 and 12). All eight segments (R1-4, D1-2, promoter and polyadenylation sequences) from each of the nine patients and seven (normal) segments from those four who had one segment grossly abnormal, were subjected to CCM analysis. A total of thirteen cleavage products were detected. All the normal sized segments detected in patients that had a segment grossly altered, showed no cleavage products. At least one cleavage product was detected in

Figure 5: Partial sequence analysis of exon 4 showing del519TACCT

Partial sequence analysis of exon 4 identifying the cause of the abnormal SSCP result in lane 4 in figure 4A, to be a 5 bp deletion (TACCT) involving amino acid residues 154, 155 and 156. The position of the deleted bases is indicated alongside a normal control sequence (C) and by an arrow against the patient's sequence (P). The alteration in the translational reading frame resulting from the 5 base deletion is also shown.

Figure 6: Segregation analysis of del519TACCT

Resolution of the 5 base pair deletion on a 10% denaturing polyacrylamide gel for direct detection of carriers of this mutation. The results of the resolution on a denaturing polyacrylamide gel are shown for the various family members. All the female relatives possess the patients deleted allele (Del) in addition to the normal allele (N) indicating their haemophilia A carrier status.

Figure 5

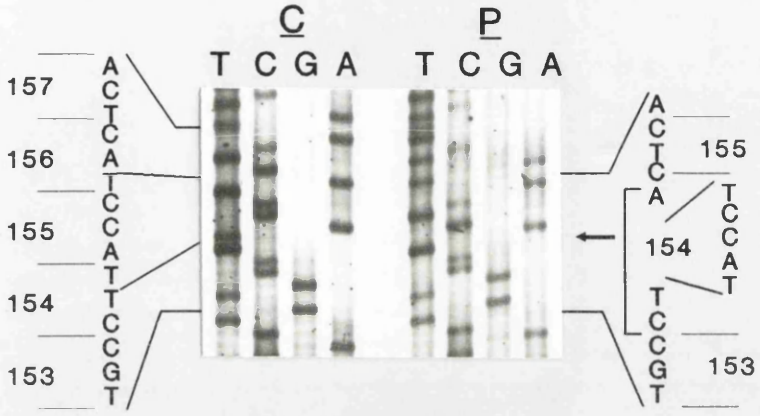


Figure 6

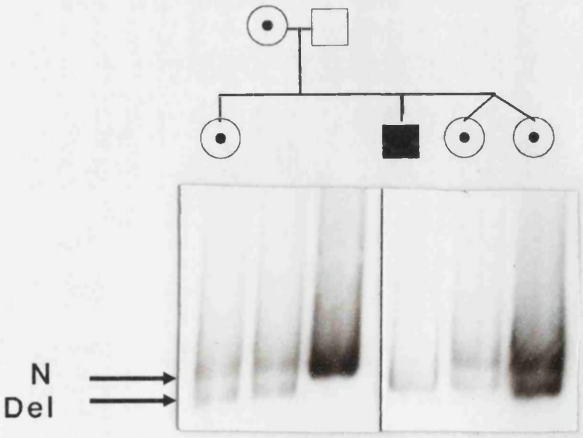


Figure 7 (A to G): CCM analysis of the factor VIII cDNA

Results of the chemical cleavage analysis of the various indicated segments of the factor VIII coding sequence. The modifying chemical used to obtain the cleavage is mentioned. The position and mutation responsible for each cleavage product are shown by arrows. In those cases where the cleavage product has not been completely characterised, the exonic location is mentioned within parentheses. Note that in lane 1 in 7B, two cleavage products in segment R3 are seen in the same individual (GLA1), only one of which has been characterised by sequencing (A1834T). Lane 3 in 7G, shows two cleavage products in patient GLA6, who also showed a cleavage product seen in lane 3 of 7B (5697delC). The cleavage products mapping to exon 14 and 21 seen in lane 4 of 7C and lane 5 of 7G were both identified in patient GLA10.

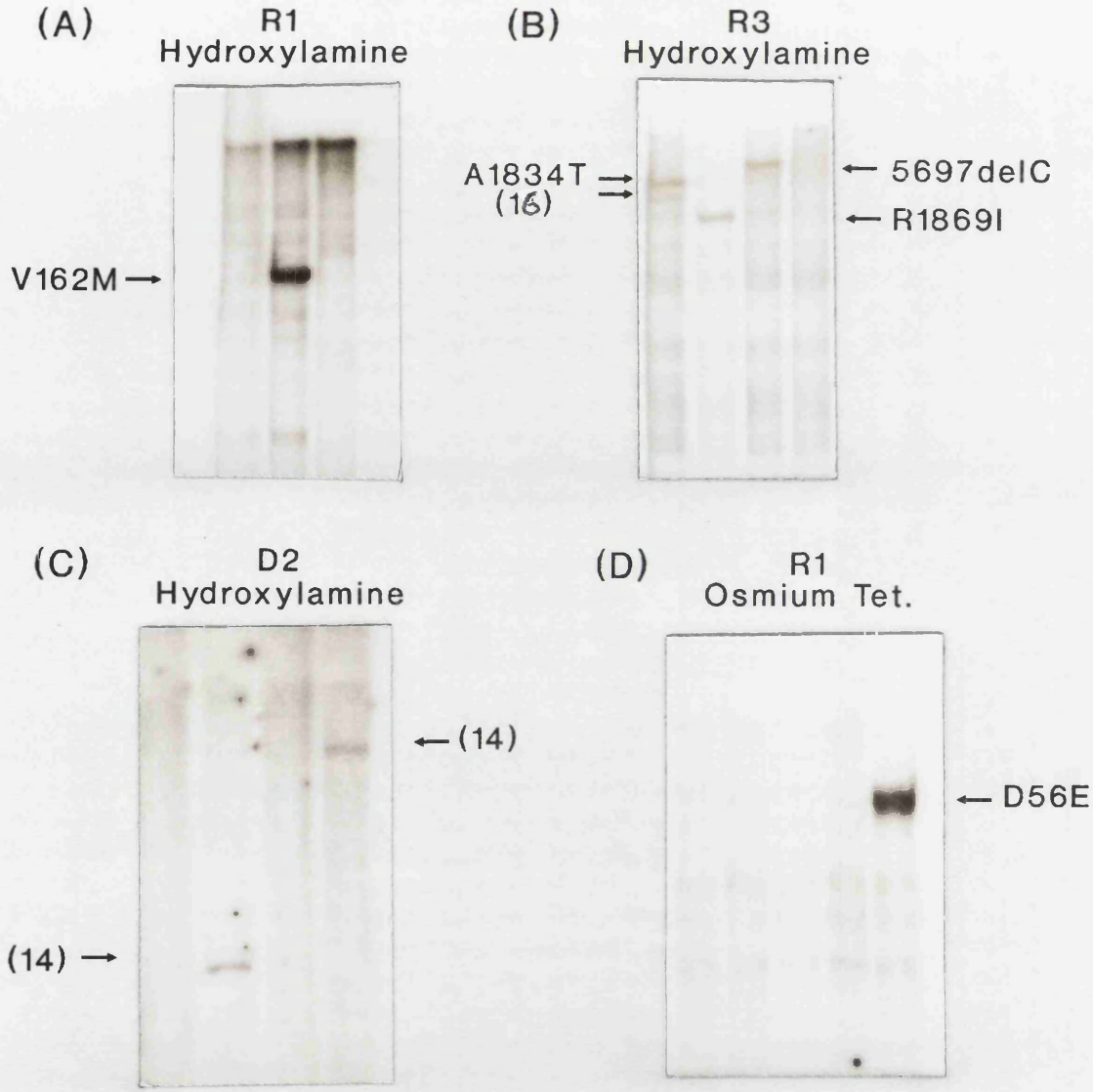
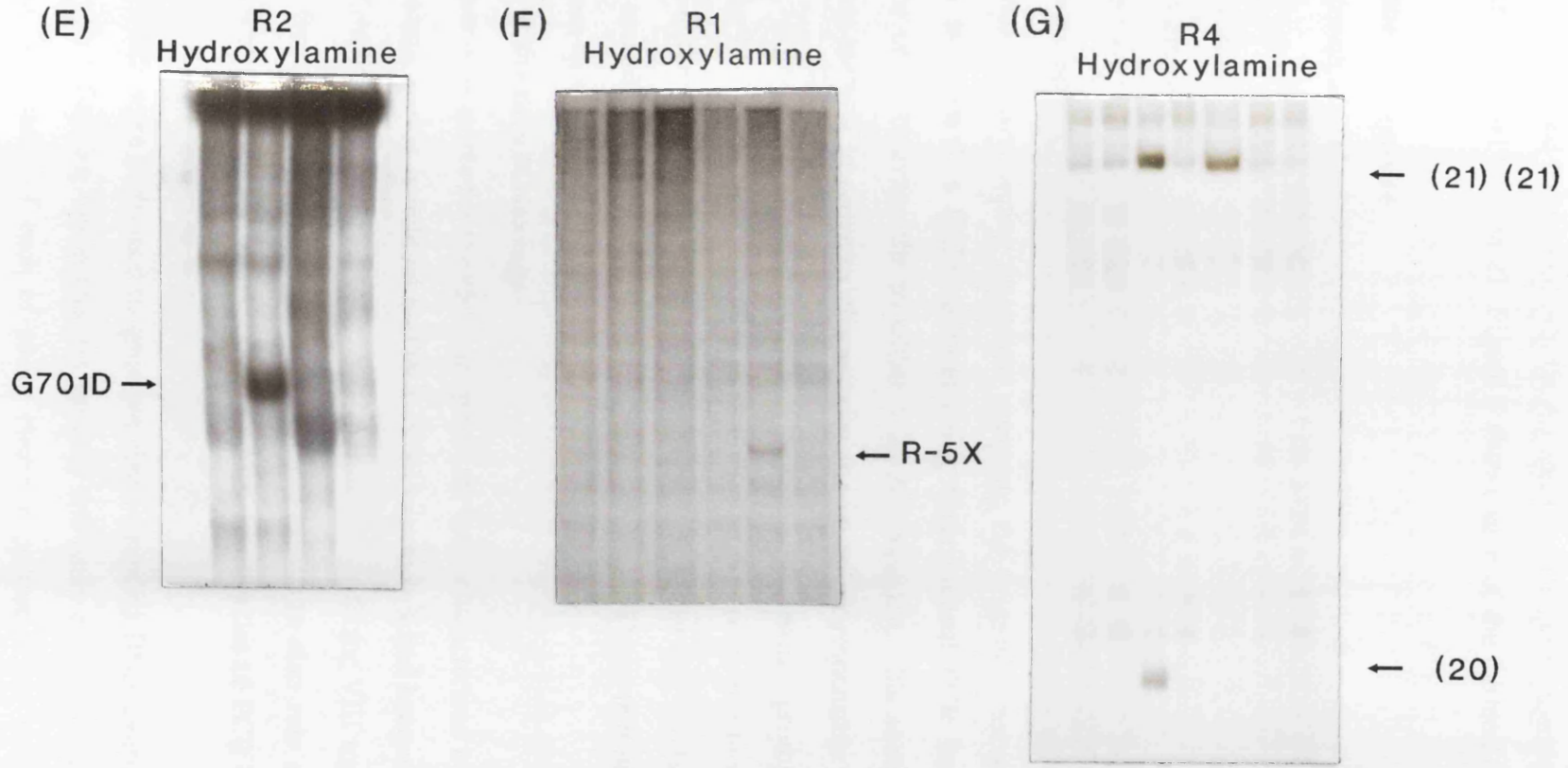

(A)
R1
Hydroxylamine
V162M
(B)
R3
Hydroxylamine
A1834T
(16)
5697delC
R1869I
(C)
D2
Hydroxylamine
(14)
(14)
(D)
R1
Osmium Tet.
D56E

Figure 7 (E, F, G): CCM analysis of the factor VIII cDNA (continued)


(E)
R2
Hydroxylamine
G701D
(F)
R1
Hydroxylamine
R-5X
(G)
R4
Hydroxylamine
(21) (21)
(20)

each of the other nine patients. Three patients had more than one cleavage product, with two each in GLA1 and 10 and three in GLA6. GLA6 and GLA10 had an identical sized cleavage product that mapped to exon 21. Figure 7 shows each of the thirteen cleavage products.

Fine mapping of the cleavage site

The probe for CCM analysis was generated by radiolabelling both the 5' ends with [γ-^{32}P] ATP and T4 polynucleotide kinase. This meant that the cleavage product could represent the radiolabel on either end of the probe. The mutation therefore mapped to any of two possible locations, assuming either end to be labelled. In order to avoid the possible need to perform two sequencing reactions to characterise the mutation, a second CCM analysis was performed in each case. For this, nested PCR primers were used to generate two internal and overlapping fragments spanning the originally cleaved PCR fragment. As seen in figure 8, a CCM reaction using these internal PCR fragments mapped the location of the cleavage site to within a 20-30 bp region. This necessitated only one sequencing reaction in order to characterise the mutation responsible for the cleavage product. This exercise of fine-mapping of the initial cleavage product was performed for all the 13 results obtained. In this way all thirteen cleavage products were mapped to specific regions in exons of the factor VIII gene. The cleavage products are denoted according to their exonic locations in figure 7 and table 4. Those that have not yet been characterised by sequencing are marked within parentheses.

Direct sequencing of the sites of cleavage

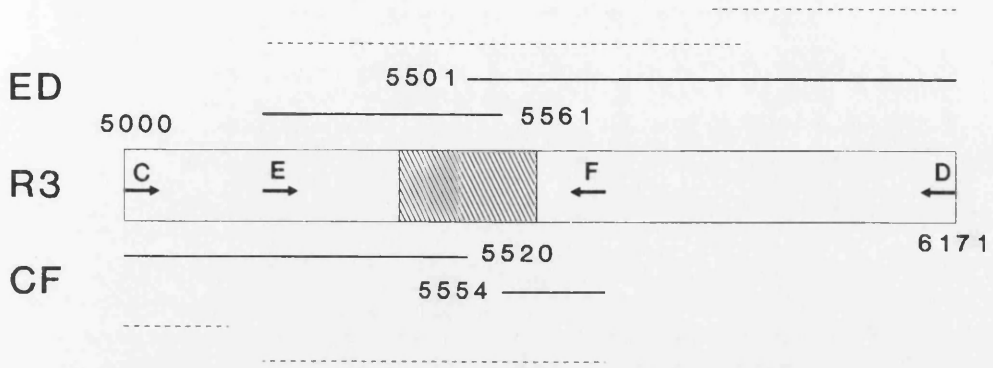
The mutations responsible for the cleavage products were characterised by direct sequencing of the exonic region (amplified by DNA-PCR) to which it had been mapped. This served to confirm that the mutation that existed in the ectopic factor VIII transcript also existed in the factor VIII gene itself. Furthermore, it served to also rule out the likelihood of a misincorporation error introduced during the early cycles of PCR by *Taq* DNA polymerase.

Asymmetric PCRs were performed to generate single stranded DNA template for direct sequencing, using flanking intron-specific primers for each exon shown to be carrying a mutation. The cause of each of seven cleavage products was completely

Figure 8: Fine mapping of cleavage sites

This figure shows a diagrammatic representation and a CCM analysis using nested primers within segment R3 (position 5000-6171) to precisely localise the two cleavage products seen in lane 1 of figure 7B. The position and orientation of the nested primers R3-E (E) and R3-F (F) are indicated. The hatched box represents exon 16 (position 5374-5586).

The cleavage products obtained for R3-CF and R3-ED and their sizes are indicated by arrows. L=radiolabelled 1 kb ladder. Localisation of the cleavage site depends on the size of the cleavage product and can be at any of two sites assuming either end of the product to be labelled. The uninterrupted and dotted lines seen above and below segment R3, represent the two possible positions calculated for each of the two nested segments, R3-ED and R3-CF. The uninterrupted lines precisely localise the cleavage product to two sites in exon 16. One site is in the vicinity of positions 5501-5520 and the other is between positions 5554-5561. Only one of these mutations has been characterised which is a missense mutation (A1834T) at position 5557.


ED
5501
5000
5561
R3
C
E
F
D
6171
CF
5520
5554

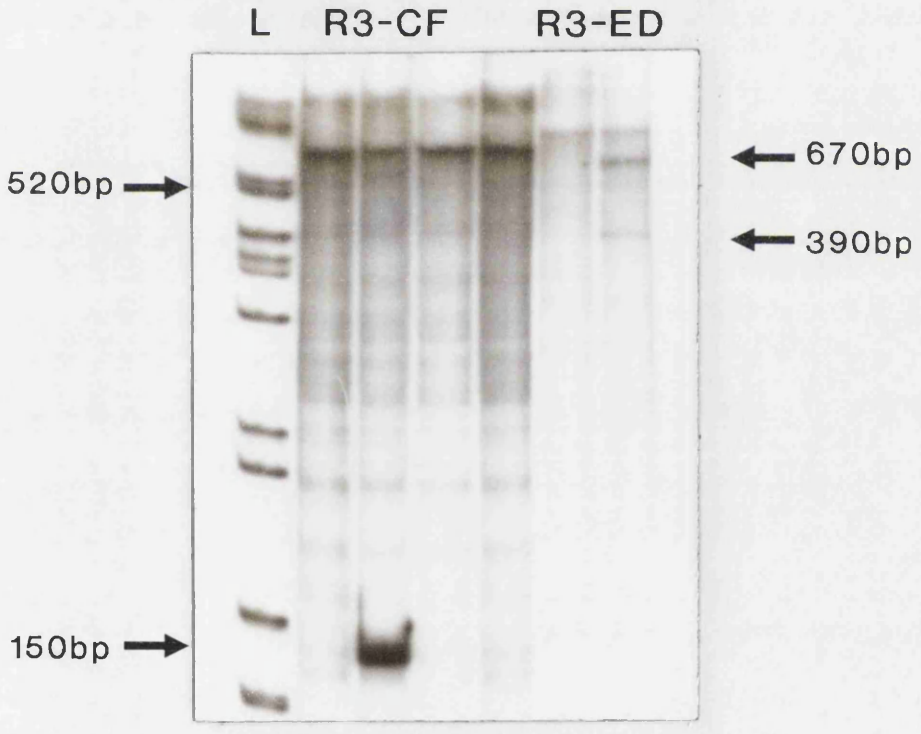

L
R3-CF
R3-ED
520bp
150bp
670bp
390bp

Figure 9 (A to G): Characterisation of cleavage products by direct sequencing

The results of direct sequencing of the various mutations detected as cleavage products on CCM analysis in Figure 7 are shown. 'P' and 'C' represent patient and control sequences, respectively. The mutation and the order of loading (T, C, G, A) are indicated along the top of each sequence. The mutation is indicated by the change (in the codon) shown at the appropriate position along the patient's sequence.

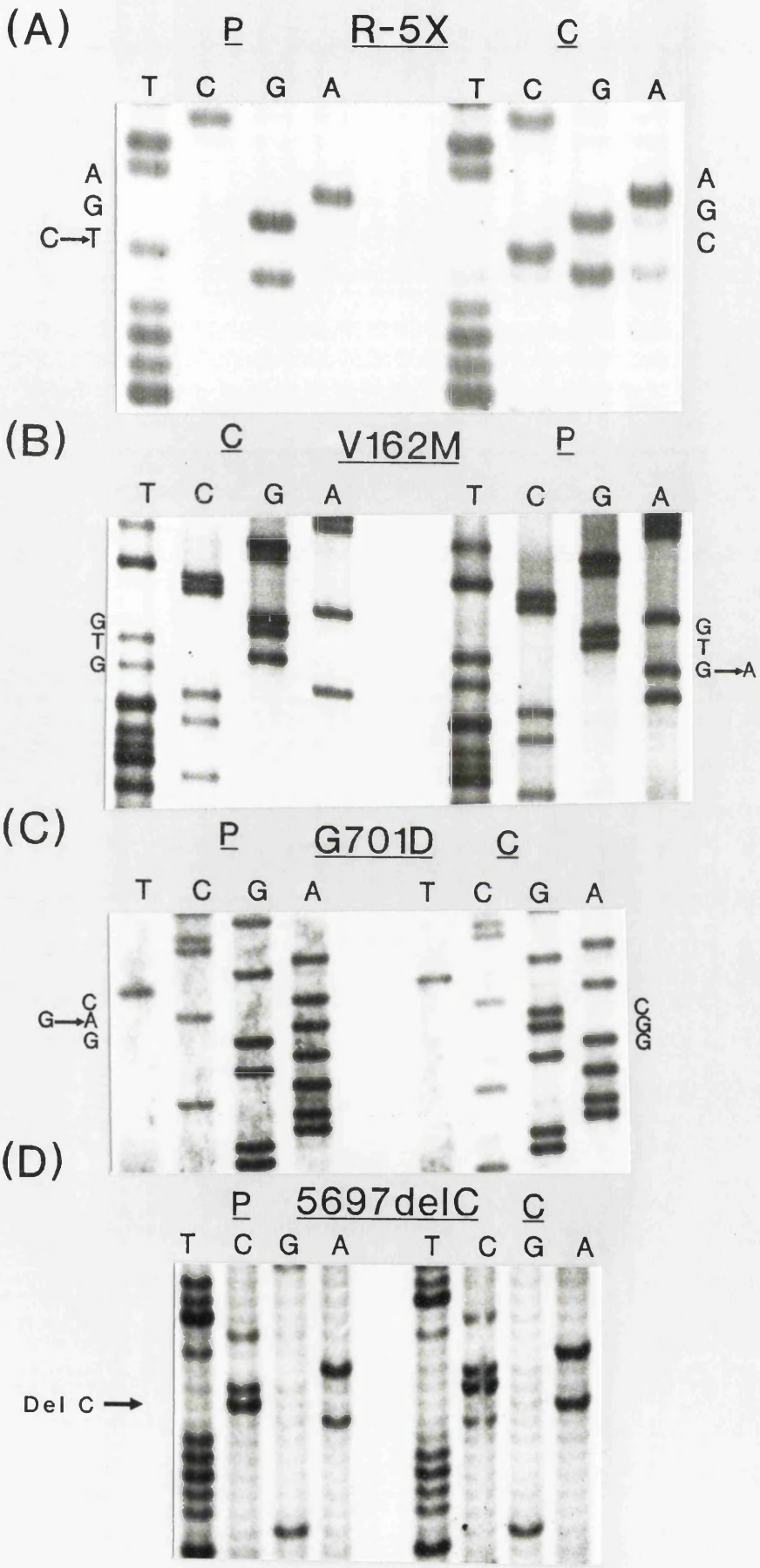
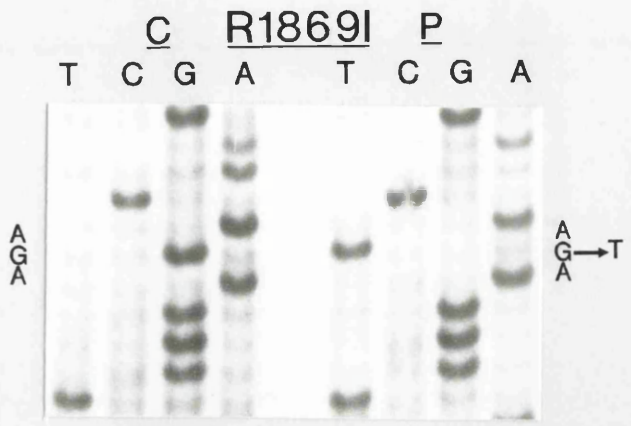

(A)
P
R-5X
C
T C G A T C G A
A
G
C→T
A
G
C
(B)
C
V162M
P
T C G A T C G A
G
T
G
G
T
G→A
(C)
P
G701D
C
T C G A T C G A
G→C
A
G
C
G
G
(D)
P
5697delC
C
T C G A T C G A
Del C →

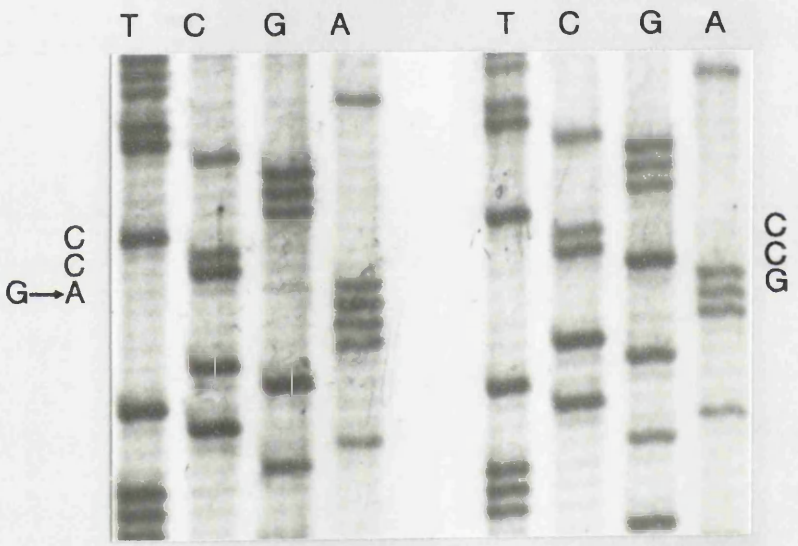
Figure 9 (E, F, G): Characterisation of cleavage products by direct sequencing (continued)

The T→A mutation (D56E) is not very clear, but due to (i) its position exactly at the site where the cleavage product mapped, (ii) the modifying chemical being osmium tetroxide (see figure 7D), which detects T/A→N substitutions and (iii) the presence of no other change in the vicinity, it is beleived that it is the change responsible for the cleavage product in segment R1 of patient GLA13.

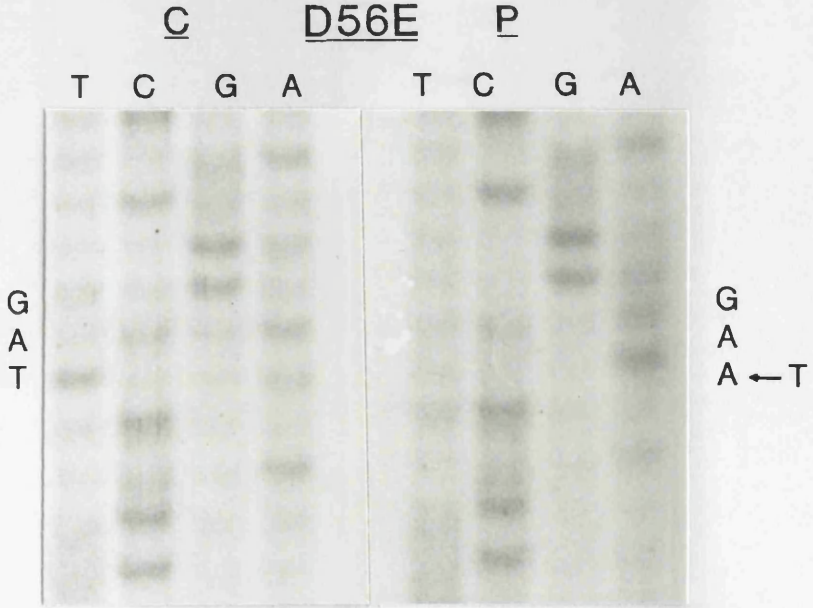
(E)


C R1869I P
T C G A T C G A
A G A
A G → T
A

(F)


P A1834T C
T C G A T C G A
C C G → A
C C G

(G)


C D56E P
T C G A T C G A
G A T
G A A ← T

characterised, in seven different haemophilia A patients. The mutations thus characterised include, six single base substitutions and a single base pair deletion. The single base substitutions detected were, five missense mutations: D56E, V162M, G701D, A1834T and R1869I and one nonsense mutation: R-5X. The deletion was that of a single "C" nucleotide at position 5697 (5697delC), in exon 17 of the factor VIII gene. The sequences of all the point mutations characterised, along with some surrounding normal sequence, are illustrated in figure 9. The details regarding the base substitutions, predicted amino acid changes, nucleotide and codon numbers involved, are listed in table 4.

3.2.3. GROSS ABNORMALITIES OF THE FACTOR VIII mRNA

Following the initial PCR amplification spanning the entire factor VIII cDNA, three patients (GLA3, 8 & 11) consistently showed products of abnormal size in one RT-PCR segment each. In addition, one repeatedly resisted the amplification of segment R4 (GLA12). These were subsequently investigated and the underlying causes for the mRNA abnormalities determined.

An identical error of splicing resulting in shortening of the factor VIII cDNA in the region comprising segment 'R1' in two patients

Of the three patients who showed products of abnormal size on amplification, two patients (GLA3 and 8) seemed to have an identical shortening of segment R1. As seen in figure 10, RT-PCR of segment R1 showed an approximately 200 bp shortening in both patients compared with the normal. The altered sized transcripts in both patients were identical and reproducible (on three separate amplifications). No normal or other transcripts of aberrant size were co-amplified.

Since GLA3 had an affected brother and two affected maternal uncles, his mother was an obligate carrier. GLA8 on the other hand, was a sporadic case and so a sample of RNA from his mother (MP1) was investigated similarly. As seen in figure 10, she also showed the altered transcript in addition to the normal one from her normal X-chromosome, indicating that she was a carrier.

Figure 10: RT-PCR of segment R1 showing loss of 200 bp

The figure shows the products of RT-PCR of the factor VIII mRNA between positions -50 and 1223, resolved by agarose gel electrophoresis. The normal control (N) shows the expected size of 1.3 kb, whereas both patients P1 (GLA3) and P2 (GLA8) show a shorter 1.1 kb product. The mother of patient P1, MP1 shows the normal and altered transcripts identifying her as a carrier.

Figure 11: Sequence analysis of segment R1 showing skipping of exons 5 and 6

Partial sequence of the altered size transcript reveals the exclusion of sequence between mRNA positions 602 and 787 and the consequent splicing together of exons 4 and 7. The arrow indicates the abnormal splice junction.

Figure 11

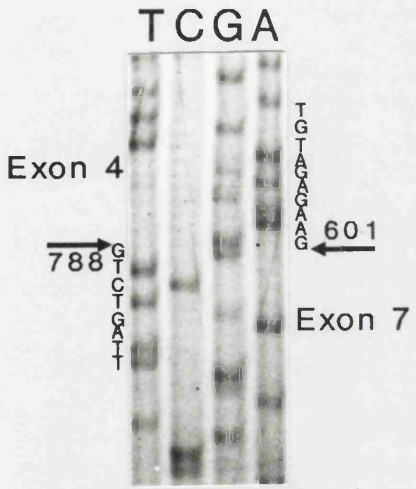


Figure 10

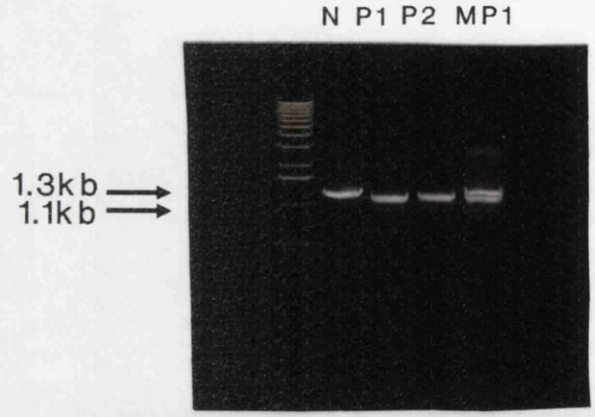


Table 2: Primers used in the characterisation of the intron 6 splice donor mutation

Primer	Orientation	Sequence (5' to 3')
6X	Forward	CAT GAG ACA CCA TGC TTA GCT GAC TCT A
6Y	Reverse	AAC TCT GGT GCT GAA TTT GGA AGA CCC T
6N	Reverse	AGA TTG TTG AGC AGG TGT GTA CAT
6M	Reverse	AGA TTG TTG AGC AGG TGT GTA CAC
5Hyb	Forward	GCC AAG GAA AAG ACA CAG AC
6Hyb	Forward	GGC ACT CAG AAA CAA AGA AC

6X and 6Y are intron-based primers for the amplification of exon 6. 6N and 6M are allele-specific primers for the mutant and normal alleles of 787+3A to G. 5Hyb and 6Hyb are primers used for the identification of some normally spliced factor VIII transcripts in GLA3 and GLA8. The nucleotide positions of 5Hyb and 6Hyb are 610 (exon 5-specific) and 680 (exon 6-specific), respectively.

The altered transcript was sequenced following an asymmetric PCR amplification. As seen in figure 11, this revealed a missing 186 bp between nucleotide positions 602 and 787 of the mature factor VIII mRNA. This corresponded to an exon skipping event involving exons 5 and 6, with the consequent perfect splicing together of exons 4 and 7.

Exons 5 and 6 amplified normally from the DNA of both patients, GLA3 and 8 and corresponded in size to those from a normal control, indicating that they were not deleted from the factor VIII gene (figure 12).

The splice sites flanking exons 5 and 6 were sequenced following asymmetric PCR amplifications of these exons. This revealed an A→G substitution at position +3 in the splice donor site of intron 6. As seen in figure 13, both patients were hemizygous for the base substitution and the mother of GLA8 (MP1) was heterozygous for the mutation, confirming her carrier status.

Additional nested PCRs were performed using primers complementary to exons 5 and 6 (5Hyb and 6Hyb, table 2). The RT-PCR product of segment R1 (which is itself a product of a nested PCR) was used as the template for this reaction. As seen in figure 14, this enabled the detection of transcripts containing sequences from both of these exons, i.e. a small fraction of full-length transcripts. These were not seen after the initial PCR amplification, even when the number of cycles were increased from 30 to 40.

As no restriction site was either lost or created by this splice mutation, allele specific primers (Wu et al., 1989; Sarkar et al., 1990) were designed that were specific for either the normal or mutant allele, in order to screen normal males for this mutation. For this purpose, a base substitution was made at the 3' end of the primers so that they would then be specific for either the normal or mutant allele (6N & 6M, respectively, table 2). As depicted in figure 15, both these primers were used in a reverse orientation along with a common forward primer in intron 5, just flanking exon 6 (6X). Multiplex PCR amplifications were carried out using three primers in each reaction. Either 6N or 6M was used along with 6X and 6Y (6X & 6Y are intronic primers flanking exon 6, table 2). 6X and 6Y produced a PCR product in every individual and was used as an internal positive control. 6N or 6M along with 6X produced an amplification in only those who were hemizygous for the normal (A) or mutant (G) base at +3 in the intron 6 splice donor

Figure 12: Genomic amplification of exons 5 and 6

This figure shows the successful amplification of exons 4, 5 and 6 by DNA-PCR in patients P1 (GLA3) and P2 (GLA8) using flanking intronic primers.

Figure 13: Sequence analysis of the intron 6 splice donor site

This figure shows the sequence of the splice donor site of intron 6. A hemizygous A to G substitution at position +3 is seen in both patients, P1 (GLA3) and P2 (GLA8), and the mother of patient P1 (MP1) shows a heterozygous A to G change at the same position, confirming her carrier status. The boundary between exon 6 and intron 6 is indicated along the sequence.

Figure 13

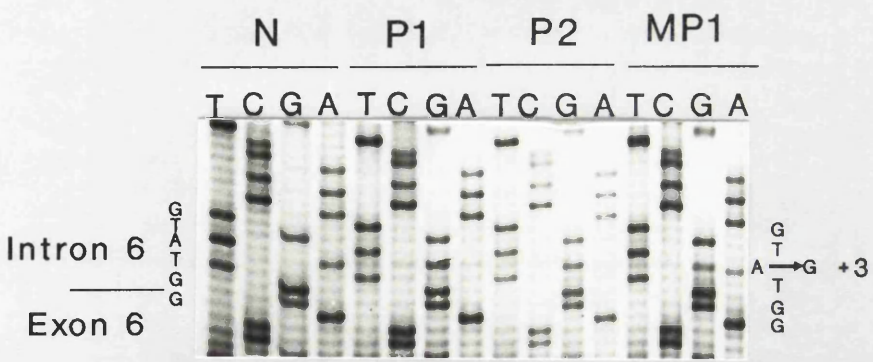


Figure 12

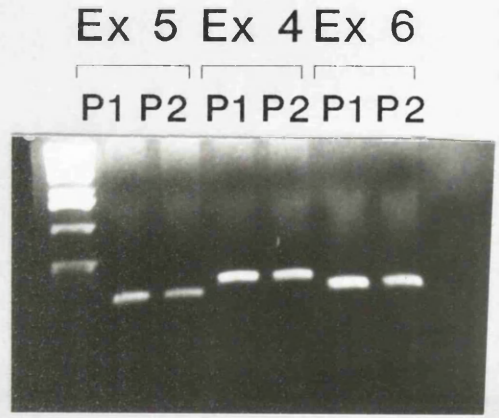


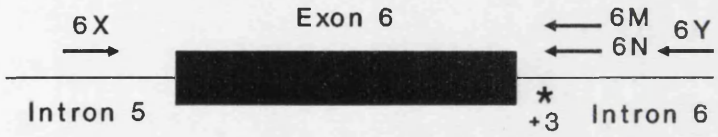
Figure 14: Demonstration of a small fraction of full-length factor VIII mRNA by nested PCR

The diagrammatic representation of the RT-PCR segment R1 is shown, with the hatched boxes representing exons 5 and 6. 5Hyb and 6Hyb are two forward primers specific to exons 5 and 6, respectively, and are used in a nested PCR along with the reverse primer R1-D. Both patients, P1 (GLA3) and P2 (GLA8) show the expected normal sized product for the two PCRs, indicating the presence of some full-length factor VIII mRNA.

Figure 15: Screening of 787+3A to G by allele-specific PCR amplification

A diagrammatic representation of the allele specific PCR amplification for the 787+3A to G mutation, and the results obtained on agarose gel electrophoresis are shown. 6N and 6M represent allele-specific primers for amplification of the normal (N) and mutant (M) alleles, respectively. 6X and 6Y are flanking intronic primers for the amplification of exon 6. The mutant allele-specific reaction (6XMY) reveals the presence of the mutant allele in both patients P1 (GLA3) and P2 (GLA8). The normal allele-specific reaction (6XNY) shows the presence of two bands in the normal lane (N) and the expected single exon 6 band in the two patients P1 and P2.

Figure 15



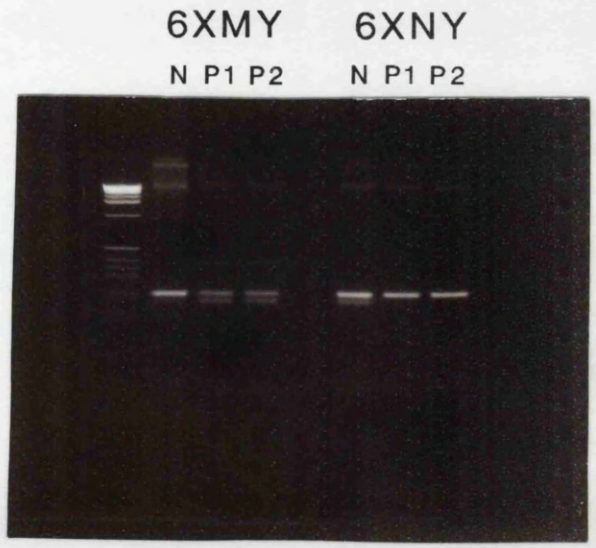
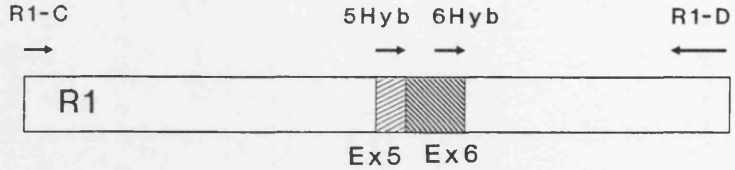
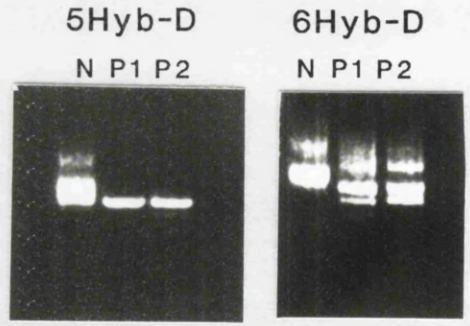


Figure 14





site, respectively. After optimising the multiplex allele-specific PCR (figure 15), a total of 24 other haemophilia A patients and 60 unrelated, normal male controls were screened for the presence of the mutation. All showed the absence of this mutation.

Gross deletion of an exon resulting in shortening of the factor VIII cDNA in the region comprising segment 'R3'

Patient GLA11 consistently showed a shortening of segment R3 (three separate amplifications). As seen in figure 16, electrophoresis on a 1% agarose gel revealed an approximately 200 bp shortening of segment R3. The altered transcript was not accompanied by either normal or other aberrantly sized products.

Detailed restriction mapping of segment R3 was carried out. This revealed that the sequence originating from exon 16 was deleted from the RT-PCR product. A partial map depicting this is shown in figure 17.

In order to determine the cause of this exclusion of exon 16 from the mature factor VIII mRNA, an attempt was made to amplify exon 16 from the factor VIII gene, using flanking intronic primers and genomic DNA as the template. As seen in figure 18, it was not possible to amplify exon 16 in patient GLA11. It was, however, possible to amplify exons 14, 15, 17, 18, 19 and 20, the other exons contributing to the RT-PCR segment R3. These findings were consistent with a gross deletion of exon 16 from the factor VIII gene, and indicates that the proximal and distal breakpoints are likely to be in introns 15 and 16, respectively. This deletion is therefore predicted to be at least 0.35 kb up to a maximum of 1.8 kb.

The characteristic intron 22 mRNA abnormality

One of the severe cases, GLA12, repeatedly resisted the RT-PCR amplification of segment R4. Since this segment spans the junction of exons 22 and 23, it was therefore suggestive of the presence of the intron 22-mediated inversion, that is known to affect half of all severe haemophilia A cases. Using the same method described by Naylor et al. (1992, 1993a) (but using different primer sequences, table 1), successful RT-PCR amplifications on either side of the exon 22-23 junction and a failure of amplification across it, was demonstrated (figure 19). This was consistent with the diagnosis of the commonly occurring DNA inversion that disrupts the factor VIII mRNA in this region.

Figure 16: RT-PCR of segment R3 showing loss of 200 bp

The diagrammatic representation of the RT-PCR segment R3 and its exonic composition is shown. The position of primers R3-C (C), R3-E (E), R3-F (F), and R3-D (D) and their orientation, is shown with respect to exon 16 (5374-5586). The PCRs performed using the nested primers show the loss of approximately 200 bp (Del; indicated by arrows) that maps between primers R3-E and R3-F (positions 5171-5704).

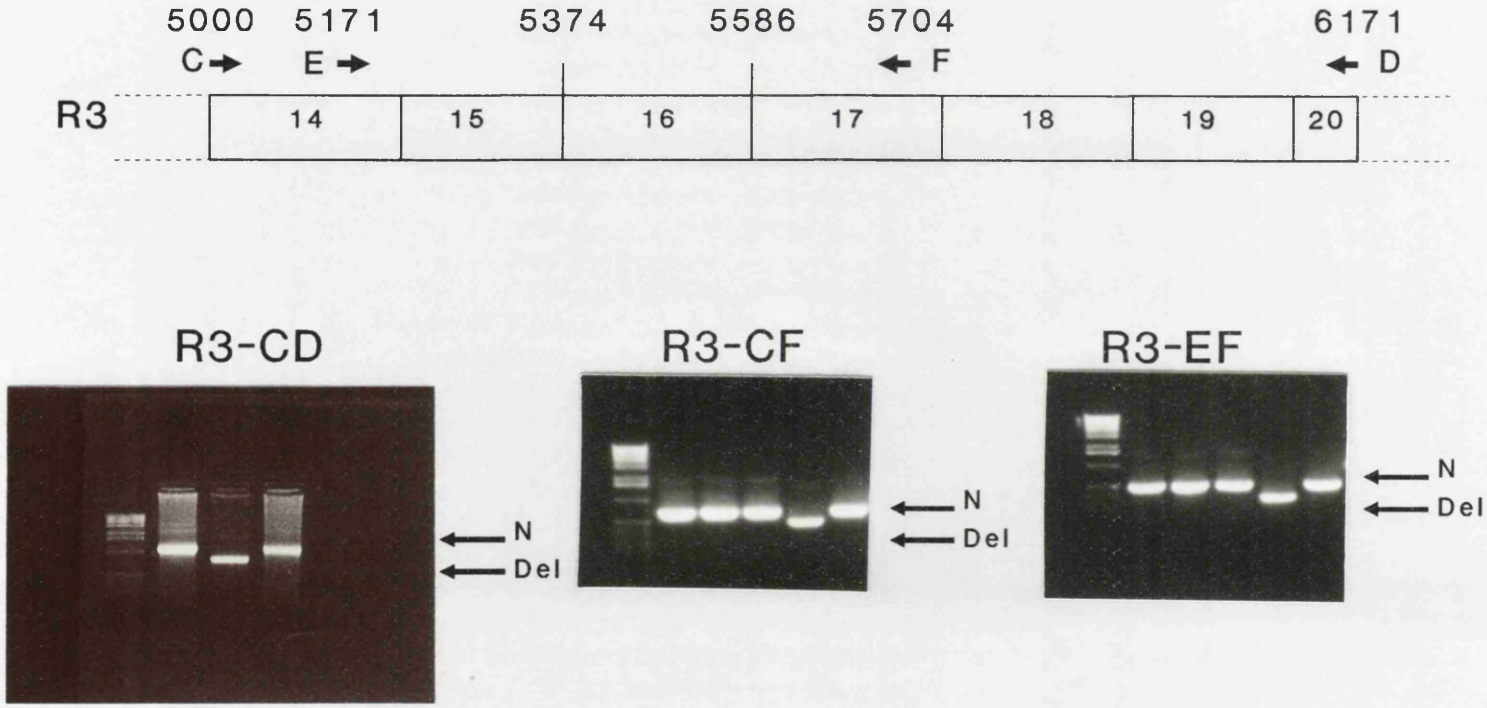

5000
5171
5374
5586
5704
6171
C
E
F
D
R3
14
15
16
17
18
19
20
R3-CD
R3-CF
R3-EF
N
Del
N
Del
N
Del

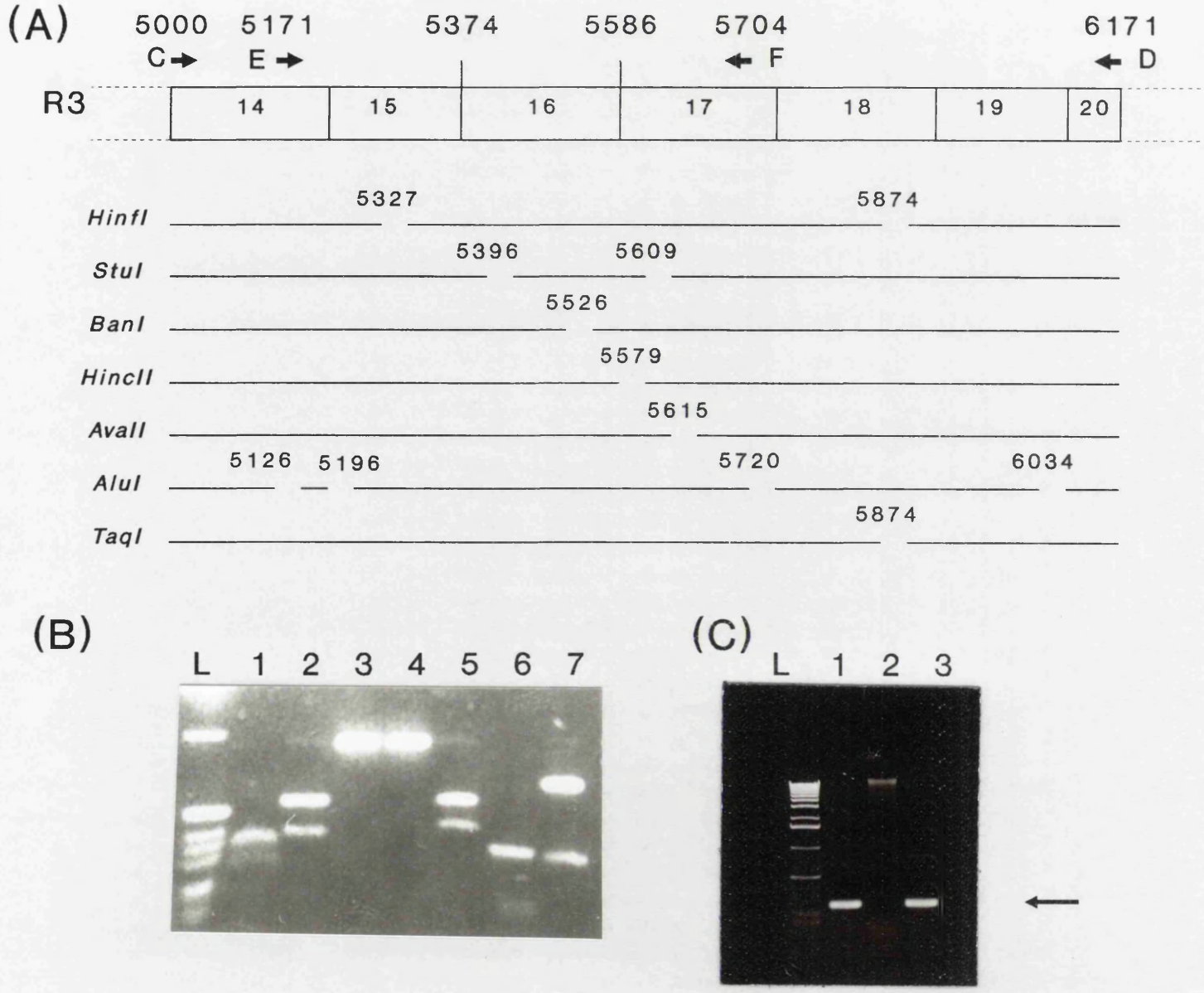
Figure 17: Restriction mapping of segment R3 showing loss of exon 16

17A. The diagrammatic representation of the RT-PCR segment R3 and its exonic composition is shown. The position of primers R3-C (C), R3-E (E), R3-F (F), and R3-D (D) and their orientation, is shown with respect to exon 16 (5374-5586). The expected cleavage sites for the seven restriction enzymes used for digestion are shown.

17B. The figure shows the results of the restriction mapping of the shortened segment R3 in patient GLA11, using enzymes *Hinf* I, *Stu* I, *Ban* I, *Hinc* II, *Ava* II, *Alu* I and *Taq* I in lanes 1 to 7, respectively. L=1 kb ladder. *Ban* I and *Hinc* II are expected to cut only in exon 16 (*Hinc* II is expected to cut near the 3' end of exon 16 at position 5579), and have not resulted in cleavage of this segment (lanes 3 and 4). *Stu* I is expected to cut near the 5' end of exon 16 (position 5396), and digestion with this enzyme (lane 2) indicates the abolition of this site. The digest in lane 1 indicates that the *Hinf* I site (position 5327) near the 3' end of exon 15 is intact. The digest in lane 5 indicates that the *Ava* II restriction site (position 5615) near the 5' end of exon 17 is also intact. The results of this restriction mapping along with the results in figure 16 (R3-EF) indicate that the shortening of segment R3 is likely to be due to a missing exon 16 (213 bp).

Figure 18: Deletion of exon 16.

The amplification of exon 16 by DNA-PCR using flanking intronic primers is shown. Lanes 1 and 3 represent the amplification products from normal controls and lane 2 shows non-amplification of exon 16 in GLA11.


(A)
5000
5171
5374
5586
5704
6171
C
E
F
D
R3
14
15
16
17
18
19
20
HinfI
5327
5874
StuI
5396
5609
BanI
5526
HincII
5579
AvaII
5615
AluI
5126
5196
5720
6034
TaqI
5874
(B)
L 1 2 3 4 5 6 7
(C)
L 1 2 3

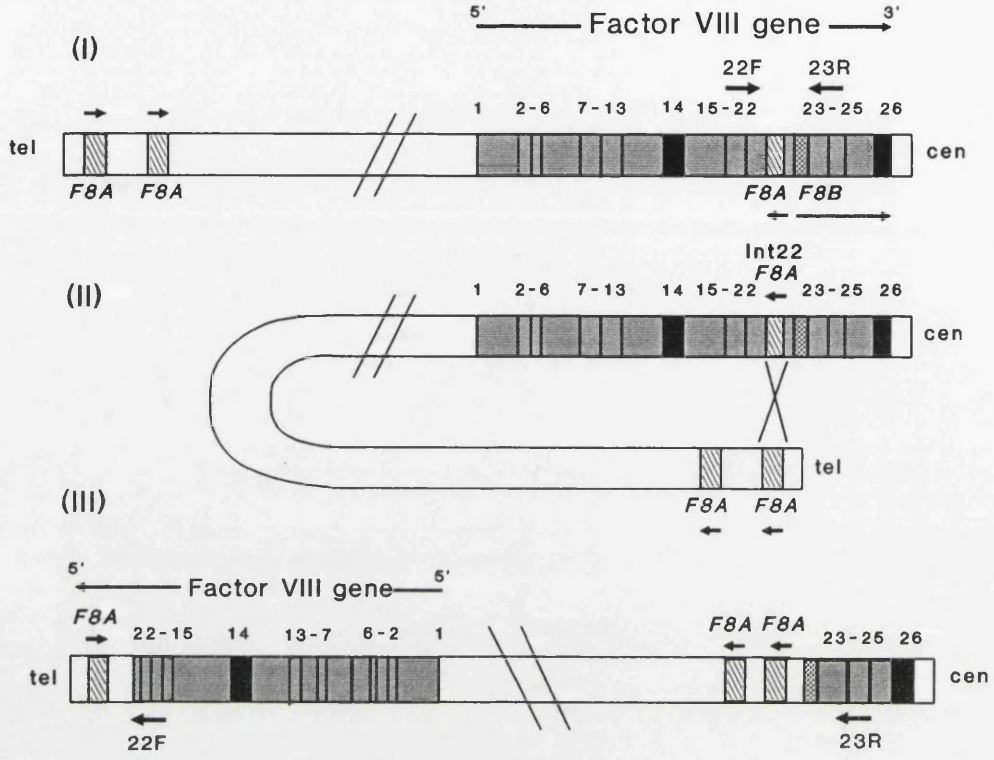
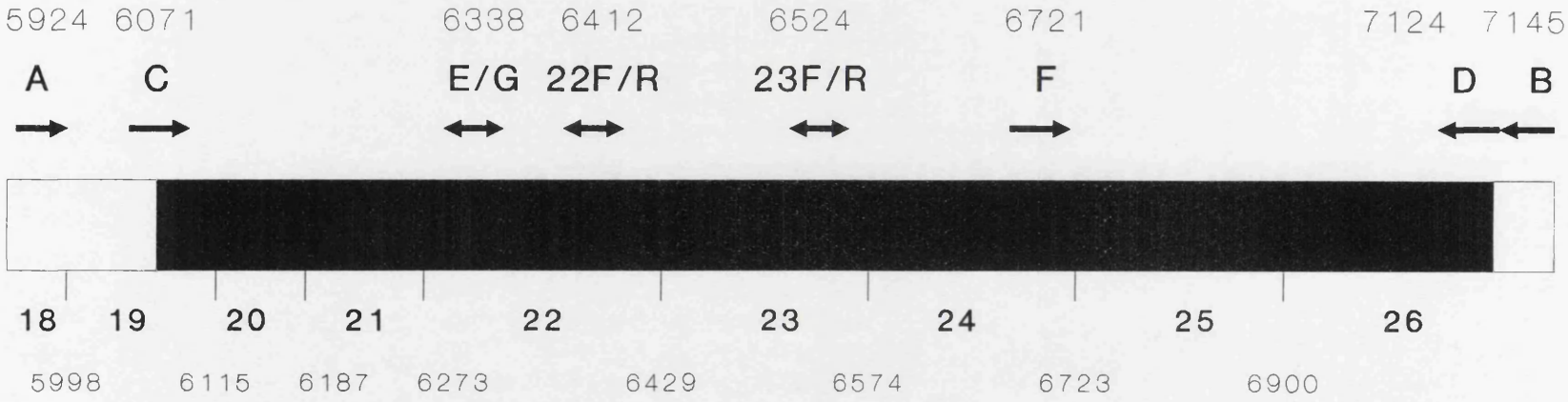
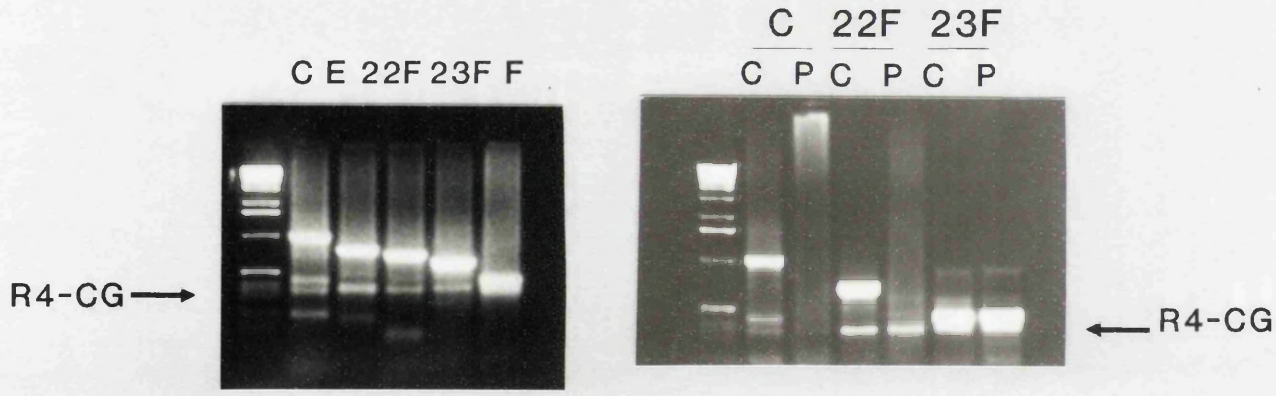


Figure 19A: Diagrammatic representation of the intron 22-mediated inversion (adapted from Lakich et al., 1993)
tel & cen = telomere and centromere of the X-chromosome; Int22=intron 22; F8A (hatched box) and F8B (crossed box) are the two genes embedded in intron 22 (orientations are shown); Nested primers 22F and 23R (along with their orientations) are indicated. **(I)** The genomic organisation of the factor VIII gene. The two F8A copies situated 500 kb 5' to the factor VIII gene are shown. **(II)** The intrachromosomal recombination mediated by the homologous sequences inside and outside of the factor VIII gene is shown (the commoner recombination, mediated by the distal copy of F8A is shown in this figure). **(III)** The product of the inversion after the homologous recombination. The separation of the previously juxtaposed primers, 22F and 23R, is depicted.

Figure 19B: Detection of the characteristic intron 22-mediated mRNA abnormality
A diagrammatic representation of segment R4 with its exonic composition is shown. The positions and orientations of the various nested primers are indicated. The agarose gel to the left shows the optimisation of the PCRs using various forward nested primers (R4-C, R4-E, 22F, 23F, R4-F) with R4-D, multiplexed with R4-C and R4-G, which acts as the internal positive control. The agarose gel to the right shows the various PCRs used to diagnose the intron 22-mediated mRNA abnormality. The control (C) and patient (P) lanes, and the various forward primers used, are indicated along the top of the gel. In all reactions the reverse primer used was R4-D. The use of forward primers R4-C and 22F, 5' to the junction of exons 22 and 23, demonstrate non-amplification in the patient, whereas the use of 23F shows amplification in both patient and control. This demonstrates the inability to amplify across the junction of exons 22 and 23. The constant amplification using primers R4-C with R4-G (indicated by the arrow) and 23F with R4-D, demonstrate the ability to amplify on either side of the junction of exons 22 and 23.

R4: Intron 22 mRNA abnormality





At least six different studies have shown that this inversion is responsible for about 45% of all severe haemophilia A cases (Lakich et al, 1993; Naylor et al, 1993a; Goodeve et al, 1994; Collins et al, 1994; Ljung, 1994; Tizzano et al, 1994). In the present study, this abnormality was detected in only one of eight severe cases. This discrepancy is most likely to have arisen due to the small number of patients analysed and the consequent misrepresentation of its frequency. It has not yet been determined if the inversion in this patient involves the proximal or distal homologous sequence (Lakich et al, 1993).

3.3. ANALYSIS OF THE MUTATIONS DETECTED IN THE FACTOR VIII GENE

The five missense mutations detected in this study were analysed for their potential role in the pathogenesis of haemophilia A, especially in the context of the observed clinical and biochemical severity. Two attributes were tested, *viz.* evolutionary sequence conservation and the involvement with consensus sequence patterns catalogued in the PROSITE dictionary of protein sites and patterns (Release 10.1; Bairoch, 1993). Two computer programs, namely, "PileUp" and "MOTIFS" (Smith et al., 1990) were used for this purpose. Both are part of the GCG (Genetics Computers Group; Devereux et al., 1984) package of sequence analysis programs, and were used on a VAX/VMS computer.

Furthermore, an attempt was made to objectively analyse the relevant splice donor sites, by assigning "consensus values" and to study in general their base composition in order to understand the mechanism of the double exon skipping event resulting from the intron 6 splice donor mutation. The results obtained from the sequence analyses are illustrated in this section and a detailed discussion is presented in sections 4.1.1 and 4.1.2.

Finally, in order to discover clues to the possible mechanisms for the origin of the two small deletions and six point mutations (excluding the splice site mutation) detected, the sequence at the site of the mutation and in the immediate vicinity was analysed. The observations and their implications are discussed in section 4.1.2.

3.3.1. MISSENSE MUTATIONS: ANALYSIS FOR EVOLUTIONARY SEQUENCE CONSERVATION

Computer-aided analysis of the factor VIII amino acid sequence has revealed the presence of a triplicated homologous segment about 330 residues in length, called the "A" domains. Similar triplicated "A" domains have been demonstrated in factor V and ceruloplasmin. There is between 30-40% sequence identity amongst the various "A" domains of these three proteins (Vehar et al., 1984; Kochinsky et al., 1986; Kane and Davie, 1988). Since all five missense mutations affected only the three "A" domains of factor VIII, these were aligned with the other homologous "A" domains of human factor VIII, clotting factor V and ceruloplasmin. Murine factor VIII, the only other factor VIII sequence in the GenBank/EMBL database was also included in the analysis.

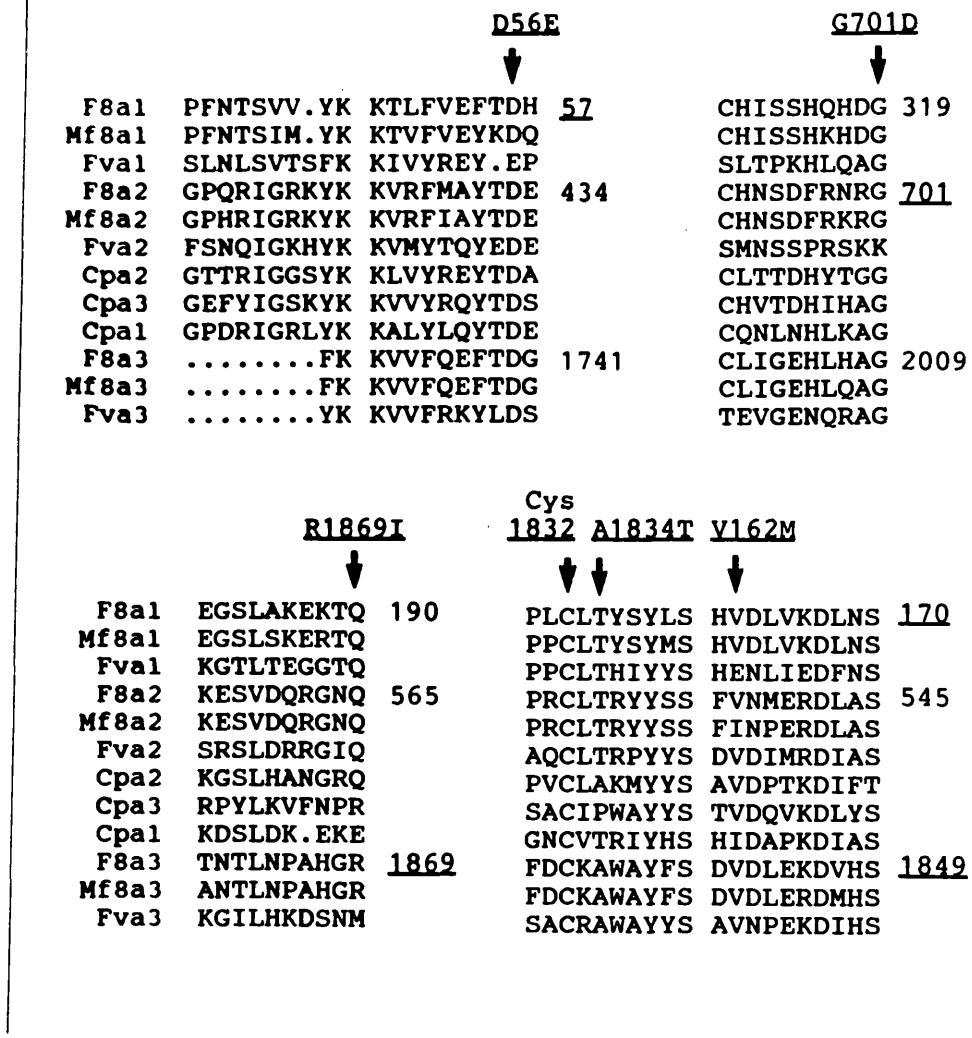
The program "PileUp" from the GCG package was used to align the three "A" domains of the four proteins i.e. twelve in all. Figure 20 shows that three of the missense mutations (D56E, V162M, G701D) are very highly conserved, occurring in at least 9 of the 12 "A" domains analysed. R1869 on the other hand occurred in 3 of the 12 domains, being conserved only in the "A3" domains of human factor VIII, ceruloplasmin and murine factor VIII. Interestingly, the A1834T mutation occurred as an Alanine in 4/12 and Threonine in 7/12 of the "A" domains (see discussion).

3.3.2. MISSENSE MUTATIONS: ANALYSIS FOR THEIR PRESENCE IN CONSENSUS SEQUENCES

The five missense mutations were also analysed for their involvement of common functional domains, mapped on the human factor VIII sequence, using the "MOTIFS" sequence analysis program (Smith et al., 1990). This program scans through a protein sequence for patterns catalogued and defined in the PROSITE dictionary of protein sites and patterns (Release 10.1; Bairoch, 1993). This analysis revealed that the V162M mutation involved a casein kinase II phosphorylation site "SHVD" spanning amino acid positions 160 through to 163. The G701D mutation was seen to affect a multicopper oxidase I signature sequence "GLWILGCHNSDFRNRGMTALL" spanning amino acid positions 686 through to 706 (see discussion).

Figure 20: Analysis of missense mutations for evolutionary sequence conservation

This figure shows the partial depiction of a "PileUp" analysis of the homologous "A" domains (three each) of human (F8) and murine (Mf8) factor VIII, factor V (Fv) and ceruloplasmin (Cp) (12 "A" domains in total). The coding sequence of these proteins were obtained from GenBank/EMBL using the accession numbers X01179, L05573, M14335 and M13699, respectively. Each missense mutation is denoted by an arrow pointing at the appropriate position above the aligned sequences. The position of the highly conserved cysteine residue at 1832 is also shown (see discussion). The amino acid position in the three "A" domains of human factor VIII (F8a1, a2 & a3) are denoted at the end of the respective aligned sequence of the domain. The domain in which the indicated mutations lie is marked by underlining the corresponding amino acid number at the end of the aligned sequence. Numbering is according to Wood et al (1984).


D56E
G701D
F8a1 PFNTSVV.YK KTLFVEFTDH 57 CHISSHQHDG 319
Mf8a1 PFNTSIM.YK KTVFVEYKDQ CHISSHKHDG
Fva1 SLNLSVTSFK KIVYREY.EP SLTPKHLQAG
F8a2 GPQRIGRKYK KVRFMAYTDE 434 CHNSDFRNRG 701
Mf8a2 GPHRIGRKYK KVRFIAYTDE CHNSDFRKRG
Fva2 FSNQIGKHYK KVMYTQYEDE SMNSSPRSKK
Cpa2 GTTRIGGSYK KLVYREYTDA CLTTDHYTGG
Cpa3 GEFYIGSKYK KVVYRQYTDS CHVTDHIHAG
Cpa1 GPDRIGRLYK KALYLQYTDE CQNLNHLKAG
F8a3 .....FK KVVFQEFTDG 1741 CLIGEHLHAG 2009
Mf8a3 .....FK KVVFQEFTDG CLIGEHLQAG
Fva3 .....YK KVVFRKYLDS TEVGENQRAG
R1869I
Cys
1832
A1834T
V162M
F8a1 EGSLAKEKTQ 190 PLCLTYSYLS HVDLVKDLNS 170
Mf8a1 EGSLSKERTQ PPCLTYSYMS HVDLVKDLNS
Fva1 KGTLTEGGTQ PPCLTHIYYS HENLIEDFNS
F8a2 KESVDQRGNQ 565 PRCLTRYYSS FVNMERDLAS 545
Mf8a2 KESVDQRGNQ PRCLTRYYSS FINPERDLAS
Fva2 SRSLDRRGIQ AQCLTRPYYS DVDIMRDIAS
Cpa2 KGSLHANGRQ PVCLAKMYYS AVDPTKDIFT
Cpa3 RPYLKVFNPR SACIPWAYYS TVDQVKDLYS
Cpa1 KDSLDK.EKE GNCVTRIYHS HIDAPKDIAS
F8a3 TNTLNPAHGR 1869 FDCKAWAYFS DVDLEKDVHS 1849
Mf8a3 ANTLNPAHGR FDCKAWAYFS DVDLERDMHS
Fva3 KGILHKDSNM SACRAWAYYS AVNPEKDIHS

3.3.3 ANALYSIS OF THE INTRON 6 SPLICE DONOR MUTATION

Consensus values were calculated for the splice donor sites of introns 4, 5, 6 and mutant 6, according to the method described by Shapiro and Senapathy (1987). This was based on the conservation of bases at the 8 positions most likely to contribute 5' splice recognition determinants i.e. position -2 through to +6. These values ranging from 0 to 100%, were used as relative indicators of the efficacy of splice sites. Briefly, the formula used that gives a score between 0 and 100 was:

$$CV = 100\ (t - \min t) / (\max t - \min t)$$

where "min t" and "max t" are the sums of the lowest and highest possible total of percentages in each of the eight positions (-2 to +6) at the splice donor site. "t" is the total of percentages for the same eight positions observed in the splice donor sequence in question. Instead of using the "min t" and "max t" values for the various GenBank files used by Shapiro and Senapathy (1987), these were calculated for the factor VIII gene from the published splice donor sequences of its 25 introns (Gitschier et al., 1984). The values used in the above calculation were therefore, min t = 12 and max t = 624. As seen in table 3, the consensus value of the mutant intron 6 splice donor site showed a 9% reduction from that of the normal sequence. Perhaps also of significance was the observation of a consensus value of 100 for the intron 4 splice donor compared to 90.84 for the intron 5 counterpart. Evaluation of the intron 5 splice donor sequence also revealed a "T" at position +3 which is only observed in 2% of all primate 5' splice sites. In addition to the reduction in efficiency of the intron 6 splice donor site caused by the splice mutation, these findings also suggest that the intron 5 splice donor site may be less effective than that of intron 4 (see discussion).

3.4 SUMMARY OF MUTATIONS DETECTED IN THE FACTOR VIII GENE

Table 4 summarises all the factor VIII mutations that were detected in this study. A total of 4 gross mRNA abnormalities, 13 positive CCM and 2 positive SSCP results were detected while screening for mutations within the factor VIII gene during the course of this project. A total of 11 different and potentially pathogenic mutations have been

Table 3. Evaluation of the splice donor sites of introns 4, 5, 6 and mutant 6

5' SS	Percentages at the 5' SS positions	t	CV
	-2 -1 +1 +2 +3 +4 +5 +6		
Intron 4	A G G T A A G T		
	52 88 100 100 68 84 76 56	624	100
Intron 5	A G G T T A G T		
	52 88 100 100 12 84 76 56	568	90.84
Intron 6 normal	A G G T A T G T		
	52 88 100 100 68 16 76 56	556	88.88
Intron 6 mutant	A G G T G T G T		
	52 88 100 100 20 16 76 56	508	81.04

5' SS = 5' splice site or splice donor site; t = the sum total of the percentages observed for the various bases occuring at the eight positions in the splice donor site, most likely to contribute 5' splice recognition determinants; CV = consensus value of Shapiro and Senapathy (1987).

fully characterised. These include 7 single base substitutions, 3 deletions and 1 gross mRNA abnormality due to the intron 22-mediated inversion (Naylor et al., 1992, 1993a; Lakich et al., 1993). The 7 single base substitutions include 5 missense (D56E, V162M, G701D, A1834T and R1869I), 1 nonsense (R-5X) and 1 splice donor site mutation in intron 6 (787+3A→G). The 3 deletions include a single base deletion in exon 17 (5697delC), a five base pair deletion in exon 4 (519delTACCT) and a gross deletion of exon 16. A further seven mutations have been mapped to specific exons of the factor VIII gene using CCM/SSCP analysis, but not yet characterised by sequencing (shown within parentheses).

Eight of these are novel changes affecting the factor VIII gene. Of the recurrent mutations, both V162M (Tuddenham et al., 1991; Diamond et al., 1992) and R-5X (Pattinson et al., 1990a; Reiner and Thompson, 1992) have each been reported twice before. The same intron 6 splice donor mutation (787+3A→G) was seen in two patients in this series (GLA3 and GLA8) and as discussed in section 4.1.2, it is unlikely that this represents a true recurrence. Of the eight severe haemophilia A cases in whom a complete ascertainment of mutations was carried out, only one showed the characteristic intron 22 mRNA abnormality. Only one of the seven different single base substitutions detected, occurred in a hypermutable CpG dinucleotide. The findings of only 1 in 8 severe case with the intron 22 mutation and 1 of 7 single base substitutions involving a CpG dinucleotide, go in favour of this small group of patients not being representative of the population. The distribution of mutations among the various exons was not totally random. If the tentatively mapped mutations are also included, exons 4, 16, 17 and 21 were seen to be affected by two mutations each. Three mutations were mapped to exon 14. Since GLA6 had a single base deletion in exon 17, it is very likely that the cleavage product in exon 21 is silent/conservative. Interestingly, both GLA6 and GLA10 have an identical cleavage product that maps to exon 21 (GLA10 also has another CCM result that maps to exon 14) where there are no known factor VIII polymorphisms (Tuddenham et al., 1991; Lin et al., 1993).

Table 4: Summary of factor VIII mutations detected in this study

Case	Screening method	Exon/ Intron[a]	Nucleotide No.[b]	Nucleotide change	Amino acid change[b]	FVIII domain[a]
GLA1	CCM	16 (16)	5557	GCC→ACC	Ala1834Thr [c]	A3 (A3)
GLA2	CCM	17	5663	AGA→ATA	Arg1869Ile [c]	A3
GLA3	RT-PCR	Intron 6	787 + 3	GTA→GTG [c,d]	Skipping of ex 5 & 6 (62 aa)	A1
GLA4	CCM	14	2159	GGC→GAC	Gly701Asp [c]	A2
GLA5	CCM	4	541	GTG→ATG	Val162Met [e]	A1
GLA6	CCM	17 (20, 21)	5697	5697delC [c]		A3 (C1)
GLA7	CCM	(14)			N/A	(B)
GLA8	RT-PCR	Intron 6	787 + 3	GTA→GTG [c,d]	Skipping of ex 5 & 6 (62 aa)	A1
GLA9	CCM	1	43	CGA→TGA [g]	Arg-5Stop [f]	signal peptide
GLA10	CCM	(14, 21)			N/A	(B, C1)
GLA11	RT-PCR	16	5374-5586	del exon 16 [c]	71 aa deleted	A3
GLA12	RT-PCR	N/A		Intron 22 mRNA abnormality		
GLA13	CCM	2	225	GAT→GAA	Asp56Glu [c]	A1
GLA14	SSCP	4	519-523	519delTACCT [c]		A1
GLA15	SSCP	(4)			N/A	(A1)

ex=exon; aa=amino acid; [a]Exons and FVIII domains in parentheses represent those chemical cleavage/SSCP results that have been mapped but not yet fully characterised. [b]Numbering according to Wood et al. (1984). [c]Novel mutations. [d]Identical splice mutation in two patients. [e]Mutation reported twice before (Tuddenham et al., 1991; Diamond et al., 1992). [f]Mutation reported twice before (Pattinson et al., 1990a; Reiner and Thompson et al., 1992). [g]Involves a hypermutable CpG dinucleotide.

The eleven potentially pathogenic mutations that were characterised in this study have been listed in table 10, along with their effect on the phenotype and the predicted deleterious effects on the factor VIII polypeptide.

3.5. FACTOR IX EXPRESSION CONSTRUCTS

The second objective of the project was to investigate the use of keratinocytes as a delivery system for ectopically expressed therapeutic gene products. An attempt was made to target the expression of clotting factor IX in a keratinocyte-specific fashion with the use of a cytokeratin gene promoter. With this aim, a variety of bovine cytokeratin VI promoter (BKVIb; Blessing et al., 1989; Bailleul et al., 1990) driven, human factor IX cDNA expression constructs were made. It was known that the *cis*-acting elements located in a 5 kb region upstream of the translation initiation site would be required to achieve expression in a tissue and differentiation-specific manner, in suprabasal keratinocytes (Bailleul et al., 1990), and it was hoped that levels high enough to be of therapeutic value might be obtained. BKVIb is the bovine counterpart of human cytokeratin 10 (K10) gene (Reiger et al., 1988). The two genes share a high level of sequence homology and have an identical expression pattern (in suprabasal keratinocytes). The expression constructs created in the present study, have been named with the prefix "K10" since the human nomenclature is the most widely used and accepted (Moll et al., 1982).

A "basic construct" was first created which was followed by various sequence manipulations that were facilitated by PCR and synthetic oligonucleotides.

3.5.1. SEQUENCE VERIFICATION OF THE HUMAN FACTOR IX cDNA CLONE

A possibility existed that the available human factor IX cDNA clone, pLXSN5'G3'cVI (Gift from Professor Brownlee at Oxford University; Anson et al., 1985), might contain point mutations introduced during the propagation of the bacterial

host. This would be critical for recombinant factor IX protein expression studies. Thus, a CCM analysis was performed in order to compare the sequence of the human factor IX cDNA clone with that of the wild-type sequence. Ectopic factor IX transcripts amplified from total human peripheral lymphocytic RNA was used as the wild-type sequence. The factor IX coding sequence was amplified using RT-PCR followed by nested PCR. As seen in figure 21, the entire factor IX coding sequence was initially amplified in three segments. The PCR primers used are listed in table 5 (A/B are primers used in the external reaction and C/D for the nested PCR). This initial amplification was followed by "splicing by overlap extension" (SOE; Horton et al., 1989), in order to combine the three individual segments and achieve the amplification of the full length factor IX coding sequence (figure 21B).

This full-length, wild-type factor IX coding sequence was used as the probe in a CCM analysis against the human factor IX cDNA insert in plasmid pLXSN5'G3'cVI. As seen in figure 22, no cleavage product was obtained, confirming that the sequence of the factor IX cDNA clone was identical to the wild type sequence. Furthermore, no cleavage product was observed that would correspond to the junctions of the SOE. These might be predicted to arise from the non template-directed addition of a deoxyadenosine (A) residue at the 3' ends of duplex PCR products by *Taq* DNA polymerase (Clarke, 1988). The high annealing temperatures and low concentration of dNTPs used during the SOE may have prevented the priming with those strands that had the non-complementary base at their 3' ends (as in allele-specific amplifications; Wu et al., 1989; Sarkar et al., 1990). Also analysis of the predicted mismatches resulting from the added adenosine residues at the 3' end of the primers in SOE revealed that they were either A-G or A-A, both of which are inefficiently extended by *Taq* polymerase (Kwok et al., 1990).

3.5.2. THE BASIC K10-FACTOR IX cDNA CONSTRUCT

A cloned DNA segment containing 5 kb of upstream sequence of the bovine cytokeratin VI gene (BKVIb), just short (eight nucleotides) of the natural BKVIb cap site (K10 promoter; Blessing et al., 1989; Bailleul et al., 1990) was obtained from Jose Jorcano (Madrid, Spain). A 0.85 kb, SV40 polyadenylation cassette (SV40 polyA)

Table 5: Primer sequences for the amplification of ectopic factor IX transcripts

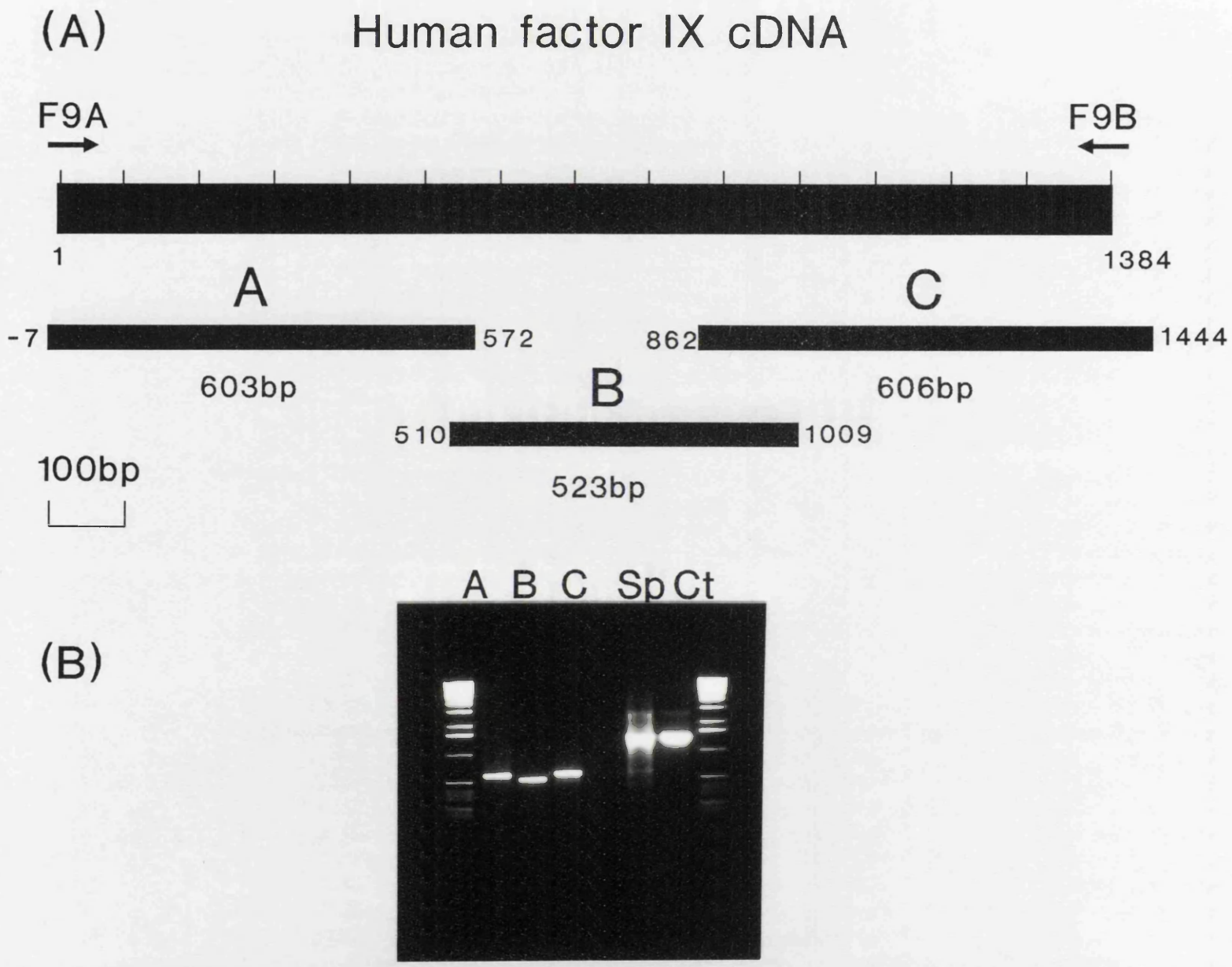
Name	Orientation	Sequence (5' - 3')	Position[a]
F9IA	Forward	GCT AGC AAA GGT TAT GCA GCG CGT	-13 - 11
F9IB	Reverse	TGA TTG GGT GCT TTG AGT GAT GTT	637 - 660
F9IC	Forward	AAA GGT TAT GCA GCG CGT GAA CAT	-7 - 17
F9ID	Reverse	CAT CAG GAA AAA CAG TCT CAG CAC	572 - 595
F9IIA	Forward	AAA ATA GTG CTG ATA ACA AGG TGG	437 - 460
F9IIB	Reverse	TTC CCC AGC CAC TTA CAT AGC CAG	1049 - 1072
F9IIC	Forward	TGA ACC AGC AGT GCC ATT TCC ATG	510 - 533
F9IID	Reverse	GAT GTT CGT GTA TTC CTT GTC AGC	1009 - 1032
F9IIIA	Forward	GGA GAC AGA ACA TAC AGA GCA AAA	855 - 878
F9IIIB	Reverse	CTA ACA AAA GAT GGG AAA GTG ATT	1450 - 1473
F9IIIC	Forward	GAA CAT ACA GAG CAA AAG CGA AAT	862 - 885
F9IIID	Reverse	AAA GAT GGG AAA GTG ATT AGT TAG	1444 - 1467

Primers named I, II and III are for the amplification of segments A, B and C, in figure 21, respectively. Primers that end with B were used for reverse transcription. Primers ending with A and B were used for the initial PCR reaction and those ending with C and D for the nested PCR. [a]Numbering is according to GenBank/EMBL accession number J00136.

Figure 21: Amplification of the full-length ectopic human factor IX transcript by SOE-PCR

21A. This figure shows the diagrammatic representation of the factor IX cDNA and the positions of the primers used to amplify the factor IX ectopic transcript in three overlapping segments (A, B and C). The position and orientation of the primers F9A and F9B, spanning the entire factor IX coding sequence, are also indicated.

21B. The three overlapping RT-PCR segments (A, B and C) spanning the entire factor IX coding sequence are resolved by agarose gel electrophoresis. "Ct", represents the control PCR performed using primers F9A and F9B to amplify the cloned factor IX cDNA. "Sp", indicates the product of SOE-PCR using primers F9A and F9B, and segments A, B and C as templates. Lane "Sp" indicates a product similar in size to that of the control PCR (Ct). Faint bands at the level of segments "B" and "C" are also noted in the lane marked "Sp".


(A)
Human factor IX cDNA
F9A
F9B
1
1384
A
C
-7
572
862
1444
603bp
606bp
B
510
1009
523bp
100bp
(B)
A B C Sp Ct

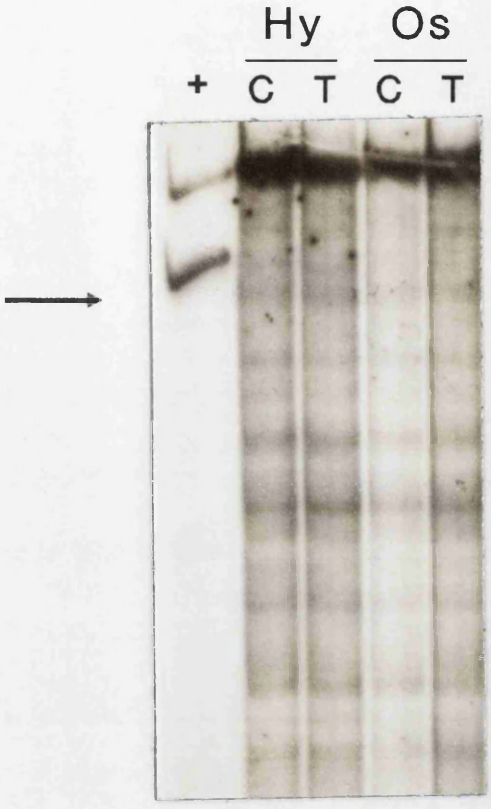


Figure 22: CCM analysis of the cloned factor IX cDNA

The figure shows the CCM analysis of the human factor IX coding sequence in pLXSN5'G3'cVI (pLIXSNL-2 in Gerrard et al., 1993). The normal human factor IX coding sequence was obtained by RT-PCR of ectopic factor IX transcripts (figure 21). The modifying chemicals are indicated as Hy=hydroxylamine and Os=osmium tetroxide. The lane marked "+" is a positive control (cleavage product indicated by an arrow). The control (C) and test (T) lanes are indicated along the top margin.

Table 6: Oligonucleotides used in cloning / sequence modifications:

Name	Orientation	Sequence (5' - 3')
F9A	Forward	CTC GGG GTA CCA TGC AGC GCG TGA ACA TGA T
F9B	Reverse	AAC CTT GGA AAT CCA TCT TTC ATT AAG TGA G
F9G4	Forward	CTC GGG GTA CCA TGG TCC GCG TGA ACA TGA T
3'K10	Forward	GCA AAA AGA AAA CCA AGC AAG TAA G
5'FIX	Reverse	AAT TTT GTT GGC GTT TTC ATG ATC A
K10SCR	Forward	ACT CCC TGA GCT AAA CAG CAT CAC
K10A	Forward	CAC GGG GGA TCC CAC TCC CTG AGC TAA ACA GCA TCA CCG GTA CCG GGC AC
K10B	Reverse	GTG CCC CCT CTT GTG AGG GAC TCG ATT TGT CGT AGT GGC CAT GGC CCG TG
3'MAN-A	Forward	CAC GGG GGA TCC CAC TCC CTG AGC TAA ACA G
3'MAN-B	Reverse	CTA GTT TCT GCT GAC CTT GGT TTA AGT GAG GAA ACA AAA AAG GAA
3'MAN-C	Forward	CTT TTT TGT TTC CTC ACT TAA ACC AAG GTC AGC AGA AAC TAG CTG
3'MAN-D	Reverse	GAA GCC TCG CGA GAA TTC GAC TTA ATT CAA AGC CAG AGA A
K10X	Forward	GCT GGG ATC GAT ACC AAG GTC AGC AGA AAC TAG CTG

containing an SV40 polyadenylation signal sequence and small-t intron (66 nucleotides) was obtained from pRSVβglobin (Mulligan et al., 1979). The K10 promoter and SV40 polyA sequence were subcloned into the multiple cloning site (MCS) of pIC20H in the correct orientation with a few restriction enzyme sites of the MCS left in between them (figure 23A). This expression vector (K10-polyA cassette) was constructed by Dr. D. Fowlis in our department. The chance presence of a consensus cap sequence 24 nucleotides downstream of the BKVIb TATA box (as part of the MCS; instead of 26 nucleotides as in the normal BKVIb gene) and its functional capacity were demonstrated following the use of the same K10 promoter/enhancer to drive the expression of mutant H-ras and TGFβ1 in transgenic mice (Bailleul et al., 1990; Cui et al., 1994). *Cis*-acting elements required for suprabasal keratinocyte-specific expression are present in a 3.4 kb segment of this promoter (with the same 3' end as the promoter used here; Bailleul et al., 1990) but absent in a 1.2 kb segment of immediate upstream sequence in the human K10 gene (Jiang et al., 1991). The small t-intron in the SV40 polyA cassette was included to provide the non-specific effect of efficient mRNA processing and transport attributable to the presence of heterologous introns in expression vectors (Hamer and Lader, 1979; Gruss and Khoury, 1980; Mulligan and Berg, 1980; Villarreal and White, 1983) and for transcriptional enhancement in transgenic mice (Brinster et al., 1988).

Two oligonucleotide primers were designed to facilitate the amplification and cloning of the entire human factor IX coding sequence (figure 23B). *Kpn*I and *Cla*I recognition sequences were incorporated into the forward (F9A) and reverse (F9B) primers respectively (table 6), because these were unique sites available for cloning between the K10 promoter and SV40 polyA sequence. The *Kpn*I recognition sequence was incorporated into the upstream primer, immediately flanking the 'ATG' translation initiation codon. A single base substitution (C→G) was introduced into the downstream primer, nine nucleotides downstream of the 'UAA' translation termination codon, so as to produce a *Cla*I recognition sequence. GC-rich anchor sequences were also incorporated into the primer design to allow efficient digestion at the termini. The plasmid pLXSN5'G3'cVI was used as a template for PCR. A 1.4 kb PCR product was obtained (figure 23C), and the *Kpn*I / *Cla*I restriction-digested PCR product was gel purified and

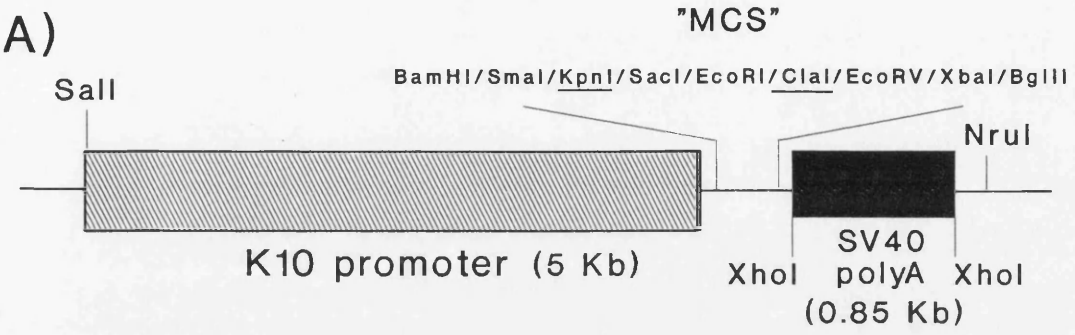
Figure 23: Construction design of the basic K10-FIX gene construct

23A. The figure shows a diagrammatic representation of the K10 polyA cassette. The relative positions and sizes of the K10 promoter and SV40 polyA cassette (SV40 polyA) are indicated. The restriction enzyme recognition sites as part of the multiple cloning site (MCS) of the pIC20H plasmid are also shown.

23B. This figure demonstrates the design of the primers (F9A and F9B) use to facilitate the cloning of the human factor IX coding sequence into the K10 polyA cassette. F9A, which primes at the 5' end, has a *Kpn*I recognition site immediately flanking the "ATG" initiation codon. F9B, which primes at the 3' end, has a single base substitution (C→G) situated 9 nucleotides 3' to the translation termination codon (TAA) to create a *Cla*I recognition site. Both primers have sufficient sequence flanking the restriction enzyme sites to facilitate efficient digestion at the termini of the resultant PCR product.

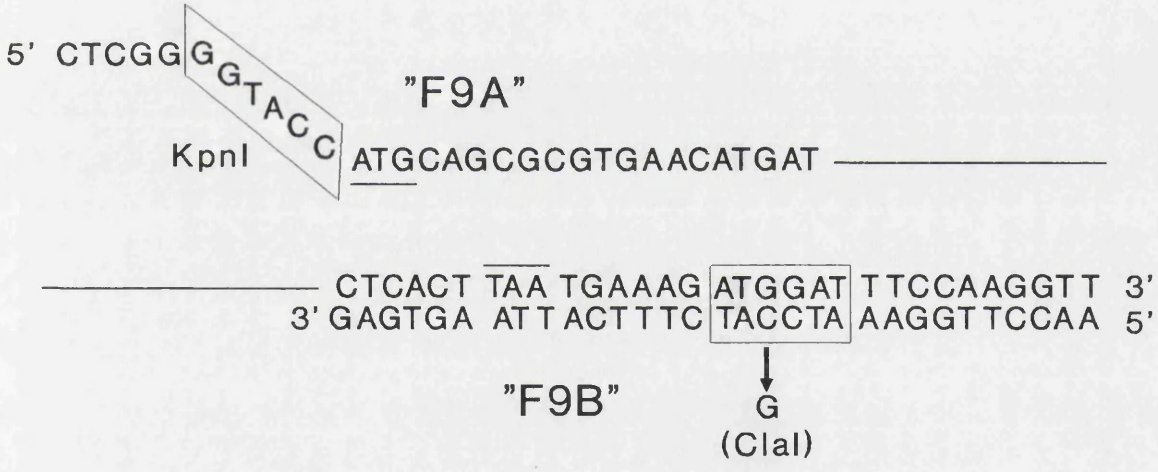
23C. This figure shows the amplification of the complete factor IX coding sequence (1.4 kb) using primers F9A and F9B.

(A)



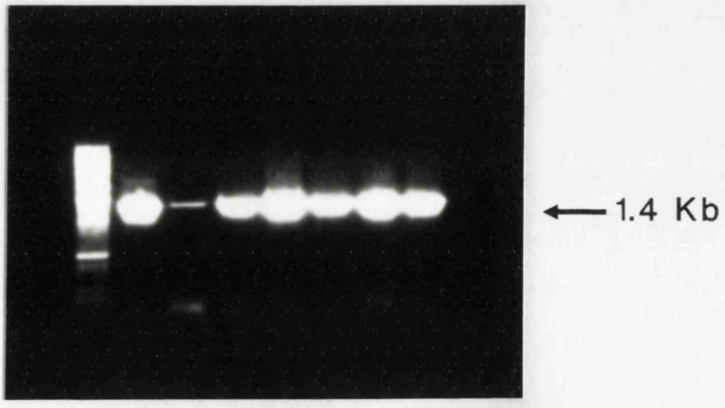
K10 polyA cassette
(in pIC20H)

(B)



(C)

F9A/B PCR



ligated into the MCS of the K10-polyA cassette. After transformation into *E. Coli* LE392, six out of twenty-five colonies picked, were found to contain the desired insert by direct PCR screening, which was then confirmed by restriction mapping (figure 24). These subclones were named K10-FIX-I to VI.

As *Taq* polymerase was used for PCR and no specific measures were taken to minimise the infidelity of the polymerisation, it was imperative to verify the sequence of the individual clones. A CCM analysis was performed, using the factor IX cDNA clone in pLXSN5'G3'cVI as the wild-type probe. Figure 25 shows cleavage products in two of the six clones analysed, one each by osmium tetroxide and hydroxylamine modification respectively. K10-FIX-I and VI therefore represent examples of PCR-induced mutagenesis. It was important to analyse the mutation in K10-FIX-I, since this construct had already been used for generation of transgenic mice (in collaboration with Dr. Y. Alexander). Following a precise localisation of the mutation in K10-FIX-I by an overlapping nested CCM analysis (as in section 3.2.2), it was characterised by direct sequencing of the products of an asymmetric PCR. As seen in figure 26, clone I had a single base substitution, T→A (nucleotide 575; numbering according to Anson et al., 1984), which changes a serine to threonine at position 136 (amino acid numbering is according to Anson et al., 1984) in exon VI of the factor IX gene, mapping within the activation domain of the factor IX polypeptide. Over 70% of mutations resulting from the lack of 3'→5' exonuclease activity are T→C substitutions resulting from efficient extension at T-dGTP mispairs (Tindall and Kunkel, 1988; Bebenek et al., 1990; Eckert and Kunkel, 1992). Another cause for mutagenesis in PCR includes heat damage causing cytosine deamination that result in C→T / G→A mutations (Eckert and Kunkel, 1992). None of these mechanisms explain the error observed in K10-FIX-I. For generation of further gene constructs K10-FIX-II was selected for use as the basic construct. since this was shown not to contain mutations.

The following observations can be made about the basic construct from figure 28: (a) The "AUG" initiation codon of the factor IX coding sequence is the first in the leader sequence, but is not in an appropriate "context" (Kozak, 1991a,b); (b) There is no natural cap site, either from the K10 or the factor IX gene, though a consensus cap site is

Figure 24: Restriction analysis of the basic K10-FIX gene construct

24A. This figure shows a diagrammatic representation of the K10 polyA cassette and the restriction enzyme recognition sites available in the MCS. K10=K10 promoter; polyA=SV40 polyA cassette.

24B. This figure shows a diagrammatic representation of the basic K10-FIX gene construct and the restriction enzyme sites used to confirm the subcloning of the factor IX coding sequence into the K10 polyA cassette.

24C. Partial restriction mapping of the basic K10-FIX gene construct to confirm the successful subcloning of the factor IX coding sequence. Lanes 1, 3, 5 and 7 are digests of the K10 polyA cassette. Lanes 2, 4, 6 and 8 are digestions of the K10-FIX-II gene construct. The restriction enzymes used include, *Kpn*I/*Nru*I (lanes 1, 2); *Eco*RV (lanes 3, 4); *Sac*I (lanes 5, 6) and *Eco*RI (lanes 7, 8). The *Kpn*I/*Nru*I double digest demonstrates the expansion due to a 1.4 kb insert. The *Eco*RV digest demonstrates that the expansion is due to the subcloning of the factor IX coding sequence and serves to confirm the orientation of the insert. Lanes 5-8 demonstrate the abolition of the recognition sites in the MCS as a result of the insertion between the *Kpn*I and *Cla*I site.

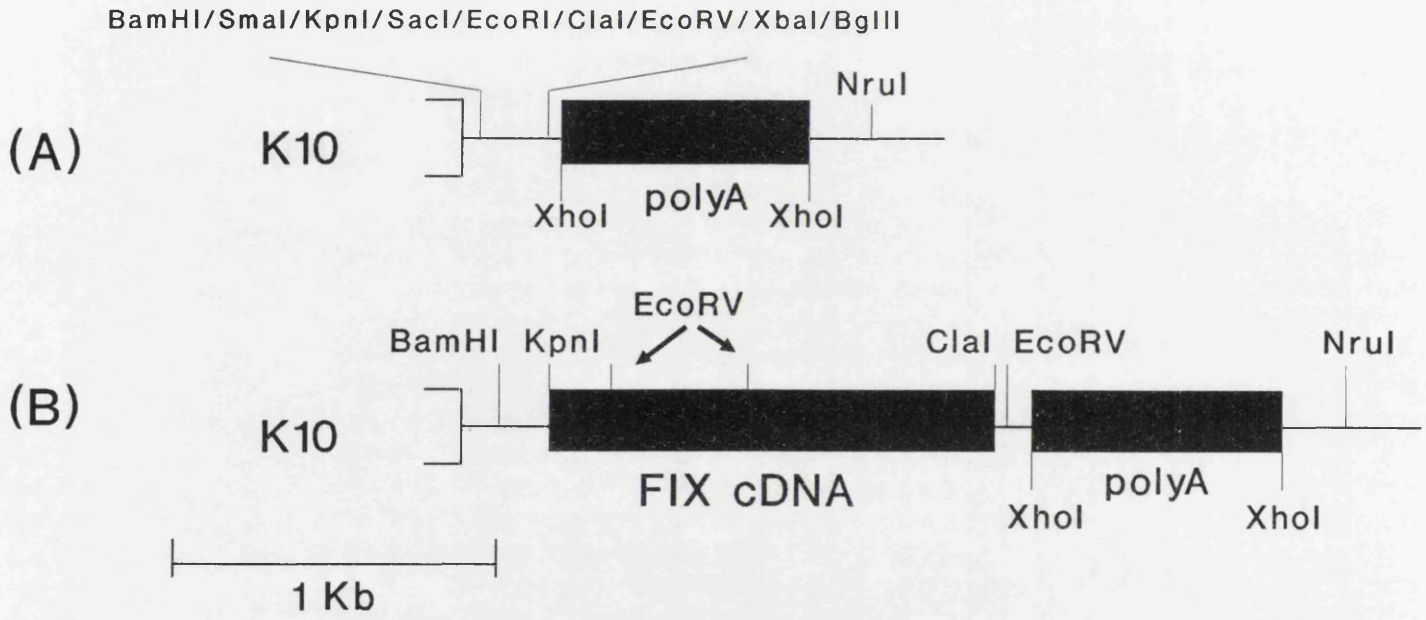

BamHI/SmaI/KpnI/SacI/EcoRI/ClaI/EcoRV/XbaI/BglII
(A)
K10
NruI
XhoI
polyA
XhoI
EcoRV
BamHI
KpnI
ClaI
EcoRV
NruI
(B)
K10
FIX cDNA
XhoI
polyA
XhoI
1 Kb
(C)
1 2 3 4
5 6 7 8

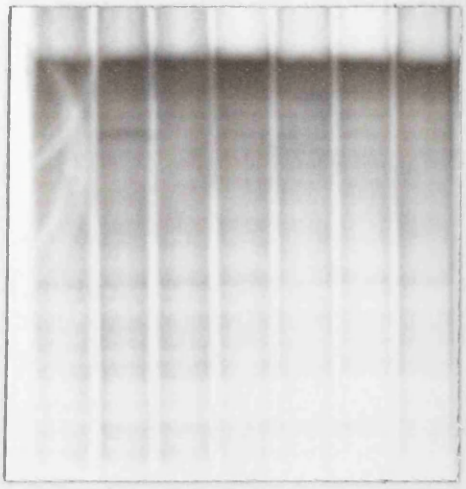
Figure 25: CCM analysis of the basic K10-FIX gene constructs

This figure shows the result of the CCM analysis of the six positive constructs (I-VI) isolated after the subcloning of the human factor IX coding sequence into the K10 polyA cassette (K10-FIX-I to VI). The modifying chemical used in each reaction is indicated. C=Negative control; L=radiolabelled 1 kb ladder. Cleavage products (indicated by arrows) are seen in K10-FIX-I (A) and K10-FIX-VI (B), detected by osmium tetroxide and hydroxylamine modifications, respectively.

(A)

Osmium tetroxide

C I II III IV V VI



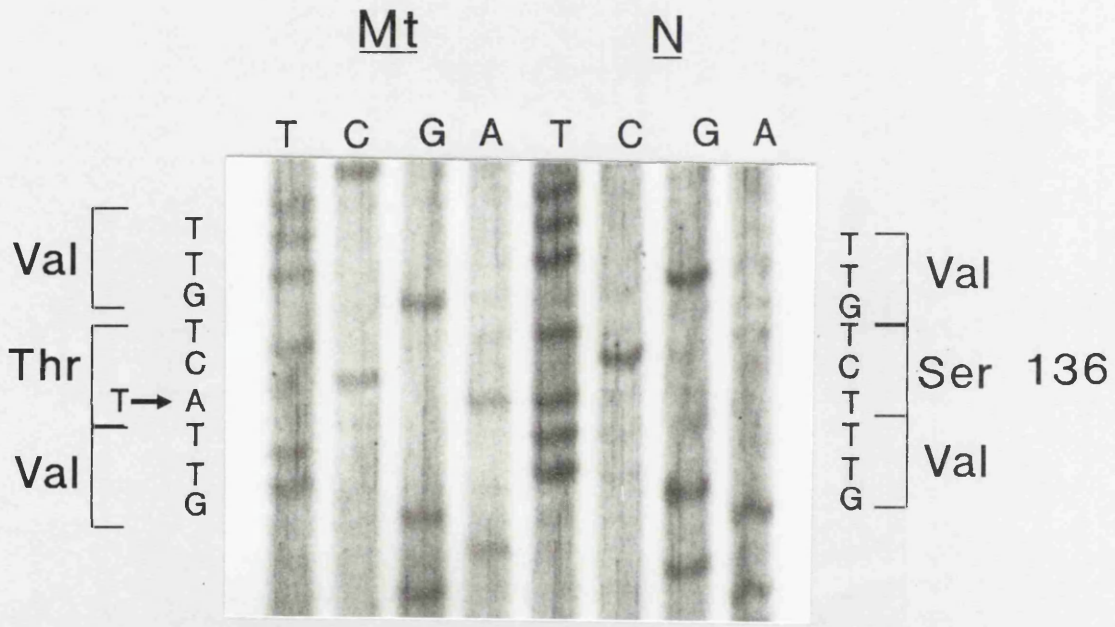
(B)

Hydroxylamine

L C I II III IV V VI

Figure 26: Sequence analysis of the K10-FIX-I construct showing the S136T mutation

This figure shows partial sequence analysis surrounding the cleavage site (detected in figure 25A) in the factor IX coding sequence of K10-FIX-I. Mt=mutant sequence; N=normal sequence. Analysis of the sequence shows a T→A transversion at mRNA position 575, which results in a serine to threonine missense mutation at amino acid 136 in exon VI of the factor IX gene (numbering according to Anson et al., 1984).



T→A 575 / Ser 136 Thr / Exon VI

provided within the MCS; (c) The 5'UTR (leader) which is formed by the pIC20H polylinker sequence, is only 14 nucleotides long and has a GC content of ~71% (10/14). All of these features would contribute in considerably reducing the efficiency of translational initiation by interfering with scanning of the leader sequence (by the 40S ribosomal subunit/initiation complex) and the efficient recognition of the "AUG" initiator (Kozak, 1991a,b). Sequence modifications were therefore designed to minimise, as much as possible, the apparently negative effects of these unfavourable sequence contexts.

3.5.3. ENGINEERING OF A FAVOURABLE CONSENSUS SEQUENCE FOR TRANSLATION INITIATION

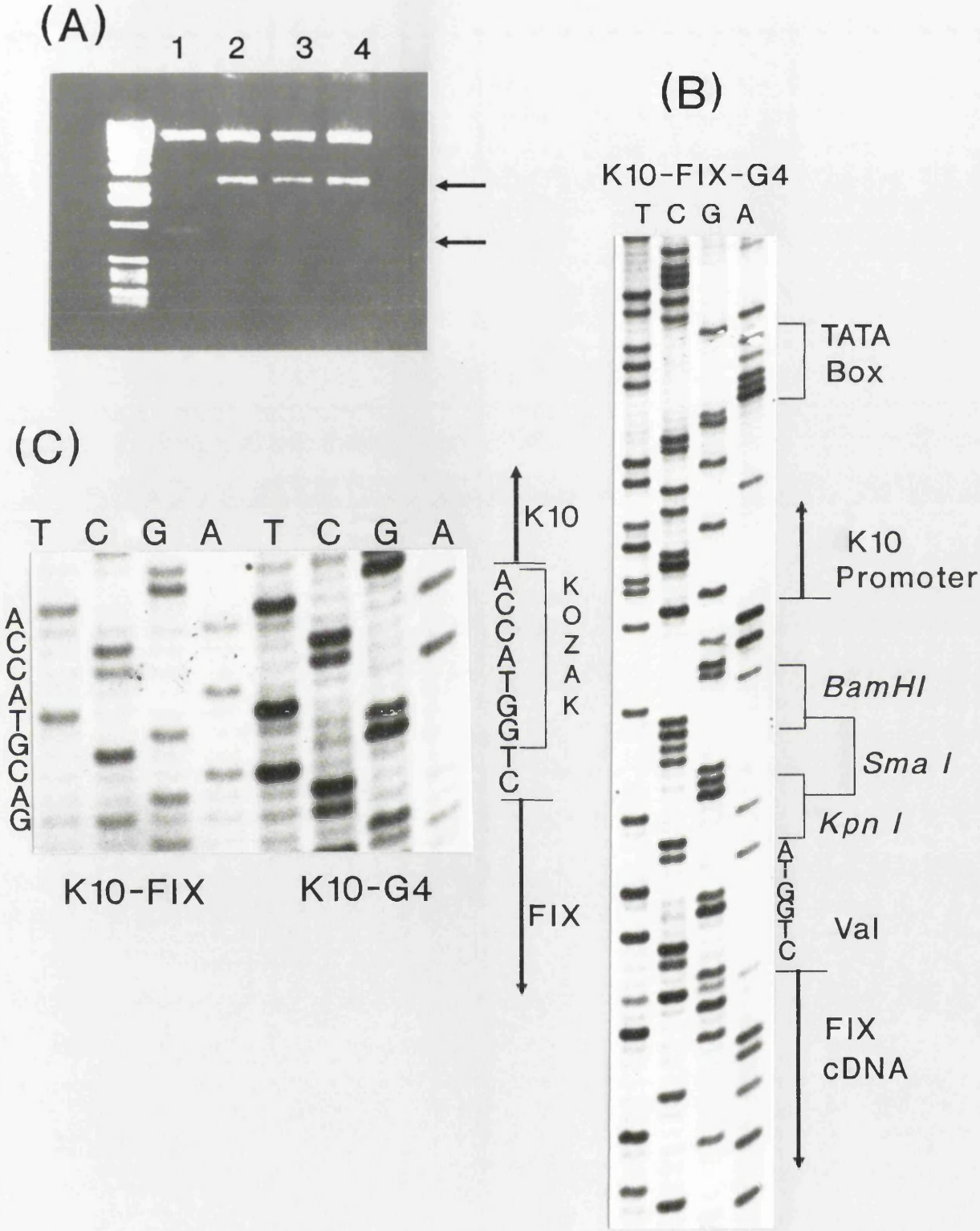
Kozak (1986a, 1987, 1989a,c) demonstrated that for the most efficient initiation of translation, the consensus sequence GCC$^{A}/_{G}$CCAUGG, must exist around the initiation codon in eukaryotic mRNAs. Two positions were found to be most critical, namely: a purine (usually an 'A') at -3 and a 'G' at +4 (where the 'A' of AUG is +1) and, in fact, most vertebrate RNAs conform to the consensus sequence $^{A}/_{G}$NNAUGG. As seen in figure 27, the sequence surrounding the initiation codon in the basic construct (ACCAUGC) does not have a 'G' at position +4. The choice of *Kpn*I as the enzyme recognition sequence immediately upstream from the AUG codon, provided a purine (A) at position -3.

In order to theoretically increase the translational efficiency of the factor IX basic construct, a 'C' nucleotide at position +4 was replaced by a 'G'. To avoid altering the amino acid charge at position 2 (Gln to Glu) it was decided to cause a more conservative change by altering the entire sequence of the second codon (position -45 according to Anson et al., 1984), so that it would read as GUC (Val) instead of the normal CAG (Gln) (Q-45V). A forward primer that was otherwise identical to F9A (used for the basic construct), except for the 'CAG→GUC' change at positions +4, +5 and +6 in the coding sequence (positions 15-17 in the primer), was designed for this purpose (F9G4, table 6). This altered factor IX coding sequence was amplified and cloned into the K10-polyA cassette in the same way as for the basic construct. Seven out of twenty-five colonies

Figure 27: Site-directed mutagenesis to introduce a favourable sequence context for translational initiation

27A. This figure shows a *Kpn*I/*Nru*I double restriction digest to demonstrate the expansion of the K10 polyA cassette due to a 1.4 kb insert (as in figure 24). Lane 1 contains DNA from the K10 polyA cassette and lanes 2 and 3 contain DNA from K10-FIX-I and II, respectively. Lane 4 contains DNA from the K10-FIX-G4 construct which shows a similar pattern to those of the two latter constructs.

27B & C. These figures show the sequence surrounding the junction of the K10-promoter and the factor IX coding sequence and confirm the incorporation of the site-directed mutagenesis involving the second codon of the factor IX cDNA in the K10-FIX-G4 construct. The order of loading of the sequence is indicated along the top margin. The restriction enzyme recognition sequences in the MCS are shown. The "TATA" box in the K10 promoter and the translation initiation site of the factor IX cDNA are indicated. Figure 27C shows the change in the sequence of the second codon from CAG (Gln; in K10-FIX-II) to GTC (Val; in K10-FIX-G4) (Q-45V) and the attainment of a favourable consensus sequence (described by Kozak, 1991a,b) for translation initiation ($^{A}/_{G}$CCATGG).


(A)
1 2 3 4
(B)
K10-FIX-G4
T C G A
TATA
Box
K10
Promoter
BamHI
Sma I
Kpn I
Val
FIX
cDNA
(C)
T C G A T C G A
ACCATGCAG
K10-FIX
K10-G4
K10
ACCATGGTC
KOZAK
FIX

screened by PCR were positive and on restriction mapping revealed the presence of the correct insert.

In order to confirm that PCR mediated, site-directed mutagenesis had been successful, the region surrounding the AUG initiation codon was sequenced. A forward primer (3'K10, table 6) near the 3' end of the truncated K10 promoter (59 bp upstream from the BKVIb TATA box) and a reverse primer at position 96 in the factor IX cDNA (5'FIX) were synthesised. Asymmetric PCR was followed by direct sequencing. Figure 27 demonstrates the successful modification of the sequence at the second codon of the factor IX cDNA and the consequent attainment of a 'G' at +4 and therefore a favourable translation initiation consensus sequence (ANNAUGG). To signify the mutagenesis of 'C →G' at position +4, this construct was named K10-FIX-G4.

Due to the high mutability rate encountered during the making of the basic construct, conditions for PCR were modified so as to permit a higher fidelity of polymerisation (section 2.4.3). The number of cycles were curtailed to 25, a five-fold reduced concentration of dNTPs was used (50 μM) and the duration and temperature of the denaturation step at the beginning of each cycle were reduced. A subsequent chemical mismatch cleavage analysis revealed no mutations in the coding sequence of all seven positive clones. The polymerase error rate was calculated as described by Eckert and Kunkel (1992). In the making of the basic construct, the error frequency (f) after 30 cycles (n) was 1/4194 nucleotides synthesised (2 mutations in 6 X 1398 nucleotides screened). Using the formula f = np/2 the error rate was estimated as 1/63,000 per nucleotide synthesised per PCR cycle. In the case of K10-FIX-G4, it was estimated to be <1/147,000 (error rate <1/9786 nucleotides), which represents a 2.3 fold reduction in polymerase infidelity.

3.5.4. ENGINEERING THE 5' UNTRANSLATED SEQUENCE

A moderately long and unstructured leader sequence is ideal for efficient translational initiation (Kozak, 1991a,b,c,d, 1992, 1994). Recognition of the first "AUG" is impaired if it is positioned too close to the cap site (≤12 nucleotides) and a proportional imrovement takes place when it is elongated (with sequences of low GC

content) over a range of 17-80 nucleotides. The leader sequence of K10-FIX-II and K10-FIX-G4 is 14 nucleotides long with a high GC content of ~71% (10/14) and is theoretically not an optimal leader sequence. The normal BKVIb/HK10 leader sequence is 26 nucleotides long and has a GC content of ~54% (14/26) as would be expected for a highly expressed gene (Kozak, 1991a,b). Furthermore, the leader sequence of BKVIb and HK10 are very highly conserved and are virtually identical (25/26, except for an A→G transition at -17). The high level of sequence conservation may also indicate the presence of primary sequences that could enhance gene expression (Reiger et al., 1988).

In order to theoretically improve the translational efficiency of the K10-FIX-II and K10-FIX-G4 constructs, by altering the ribosomal binding site to that of BKVIb, the natural cap site and 5'UTR sequence of the BKVIb gene (K10CAP and K105'UTR) was inserted into the MCS, between the *Bam*HI and *Kpn*I sites (figure 28 A & B). Two complementary oligonucleotides (K10A and K10B, table 6) were synthesised with the appropriate sequence (figure 28B) i.e. the K10 CAP site and 5'UTR sequence flanked by the recognition sequence for *Bam*HI and *Kpn*I (table 6). Suitable anchor sequences were also included to enable efficient restriction digestion.

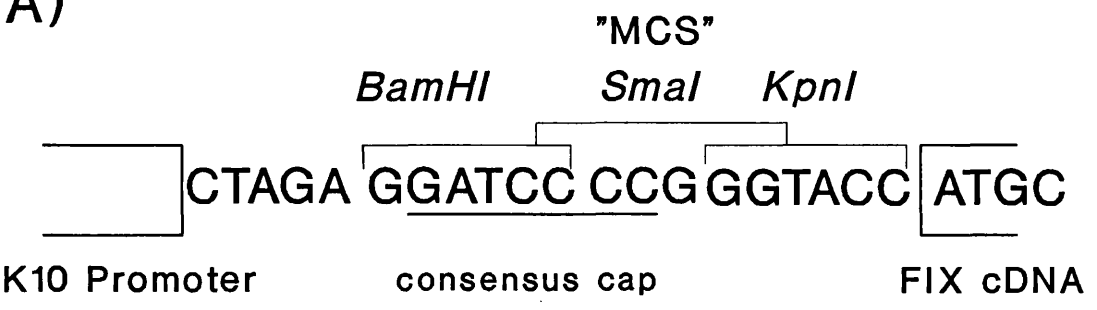
Equimolar concentrations of the two oligonucleotides were annealed together, digested and ligated into suitably-linearised K10-FIX-II and K10-FIX-G4 constructs. Individual colonies were screened directly by PCR. A forward primer specific to the inserted sequence (K10SCR, table 6) was synthesised for this purpose (underlined in figure 28B). In order to confirm the oligonucleotide-mediated partial gene synthesis, PCR primers flanking the region (3'K10 and 5'FIX, table 6) were used to sequence clones that were positive for the insert. Figures 29 and 30 show the successful incorporation of the required sequence between the *Bam*HI and *Kpn*I sites. Figure 29 also shows the sequence of one of the (seven positive) clones (RB15) that had the anchor sequence incorporated into the construct, immediately upstream of the translation start site. This was due to incomplete restriction digestion and illustrates the importance of using excess enzyme and prolonged duration for digestion at termini of DNA segments (Clackson et al., 1992). As the clones K10-FIX-II and K10-FIX-G4 were previously shown not to have any mutations in the factor IX coding sequence, no further analysis was necessary.

Figure 28: Construction design for repair of the 5' untranslated region (5'UTR)

28A. This figure shows a diagrammatic representation of the junction between the K10 promoter and the factor IX coding sequence in the basic K10-FIX gene construct. The enzyme recognition sequences in the MCS are shown. A consensus cap site sequence is underlined in the MCS.

28B. This figure shows the two complementary synthetic oligonucleotides designed to repair the 5'UTR of the basic K10-FIX and K10-FIX-G4 gene constructs. The sequence of the K10 cap site and K10 5'UTR are shown. The sequence was flanked by recognition sites for *BamH*I and *Kpn*I to facilitate the directional cloning of this synthetic DNA segment into the MCS of K10-FIX-II and K10-FIX-G4. GC-rich anchor sequences are incorporated in the oligonuleotide design to enable efficient digestion at the termini. The underlined sequence represents a forward PCR primer (K10SCR) used for screening bacterial colonies that contained plasmids with the synthetic insert.

(A)



5'UTR of K10-FIX-I/II

(B)

Anchor	*BamHI*	K10 Cap	K10 5'UTR	*KpnI*	Anchor
CACGGG	GGATCC	CACTCCC	TGAGCTAAACAGCATCACC	GGTACC	GGGCAC
GTGCCC	CCTAGG	GTGAGGGACTCGATTTGTCGTAGTGG		CCATGG	CCCGTG

K10 A / B

Figure 29: Sequence analysis confirming the repair of the 5'UTR in K10-FIX-II

29A. This figure shows the product of PCR using primers 3K10 and 5FIX, which span the junction between the (3' end of the) K10 promoter and (5' end of the) factor IX cDNA. The product in lane 1 is from K10-FIX-II and in lane 2 is from K10RB-FIX. The arrows demonstrate the resolution of the expansion seen due to the insertion of the 26 bp synthetic sequence.

29B. This figure demonstrates the sequence across the junctions of the K10 promoter and the factor IX cDNA. The order of loading in each track is indicated along the top margin. K10RB=K10RB-FIX gene construct; K10-G4=K10-FIX-G4 gene construct. The TATA box in the K10 promoter and the translational initiation site of the factor IX cDNA are indicated. Analysis of the sequence demonstrates the successful incorporation of the synthetic DNA segment (representing the natural K10 cap site and 5'UTR) between the *Bam*HI and *Kpn*I recognition sites (and are indicated alongside the sequence).

29C. This figure shows the sequence of the same region of one of seven positive K10RB-FIX clones (RB15). Analysis of the sequence demonstrates the incorporation of the anchor sequence (just prior to the start of the factor IX coding sequence) into the construct, due to poor digestion at the terminus (indicated alongside the sequence).

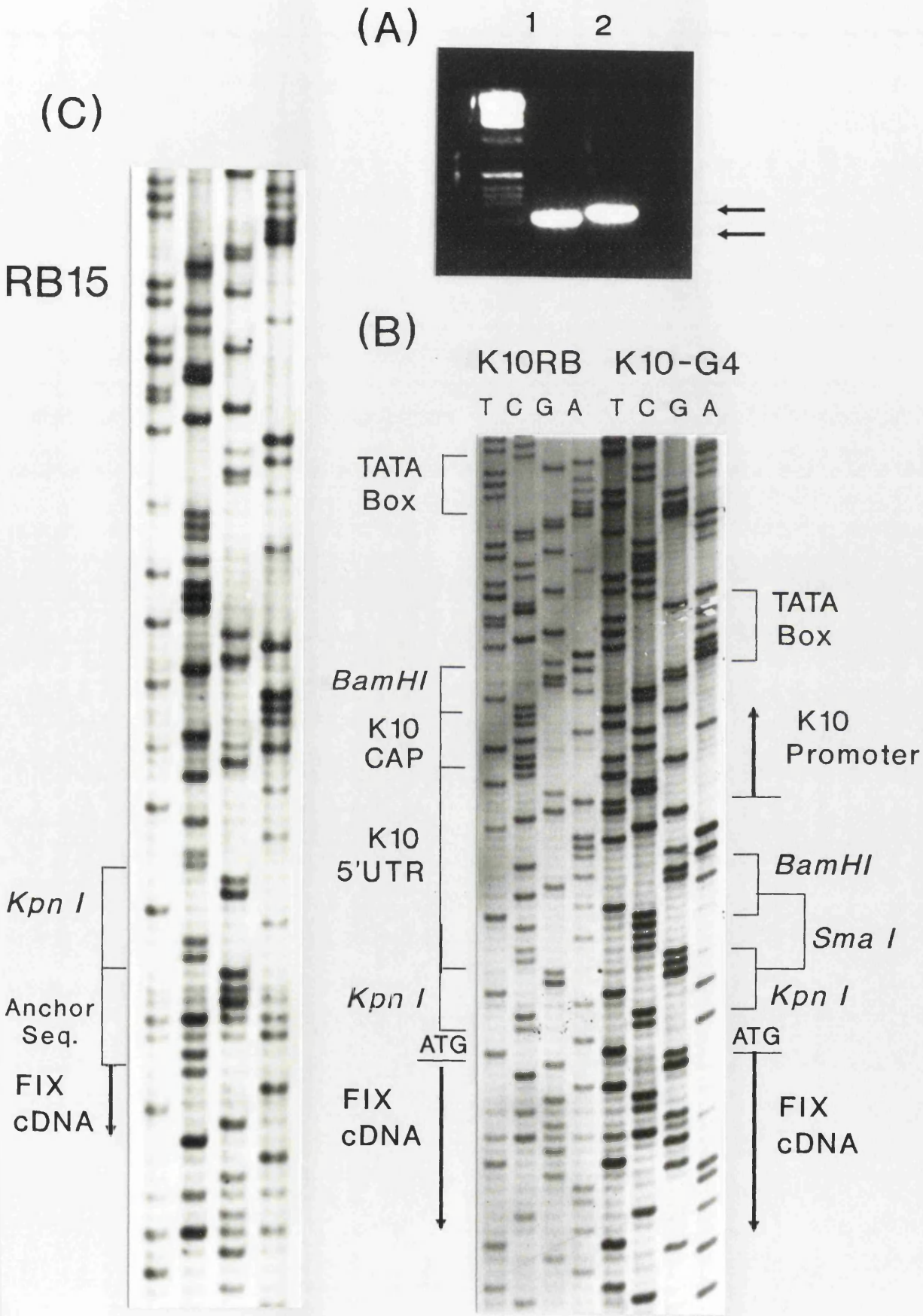
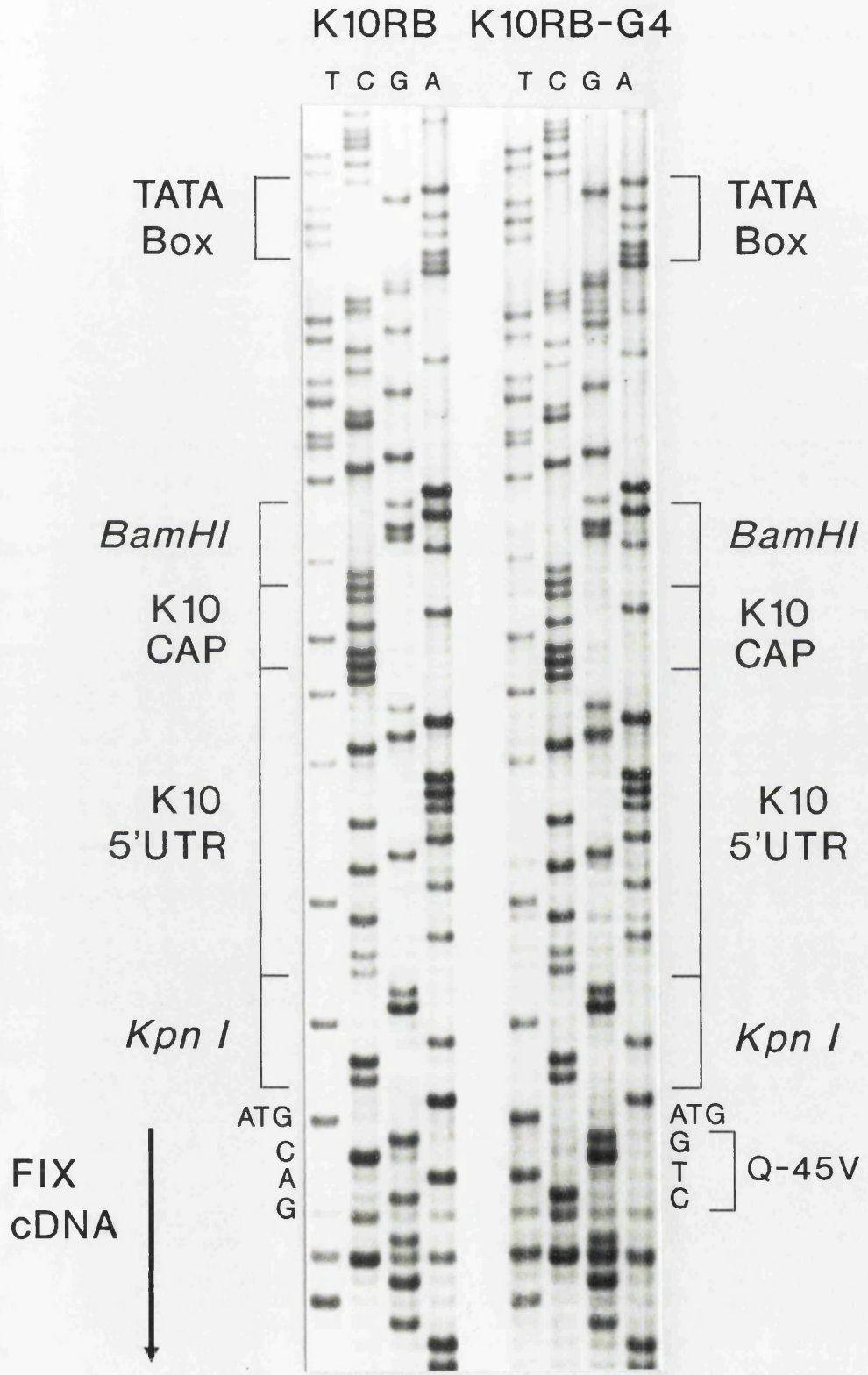

(A)
1
2
(B)
K10RB
K10-G4
T C G A T C G A
TATA Box
BamHI
K10 CAP
K10 5'UTR
Kpn I
ATG
FIX cDNA
TATA Box
K10 Promoter
BamHI
Sma I
Kpn I
ATG
FIX cDNA
(C)
RB15
Kpn I
Anchor Seq.
FIX cDNA

Figure 30: Sequence analysis confirming the repair of the 5'UTR in K10-FIX-G4

This figure demonstrates the sequence across the junctions of the K10 promoter and the factor IX cDNA. The order of loading in each track is indicated along the top margin. K10RB=K10RB-FIX gene construct; K10RB-G4=K10RB-FIX-G4 gene construct. The TATA box in the K10 promoter and the translational initiation site of the factor IX cDNA are indicated. Analysis of the sequence demonstrates the successful incorporation of the synthetic DNA segment (representing the natural K10 cap site and 5'UTR) between the *Bam*HI and *Kpn*I recognition sites (and are indicated alongside the sequence). The site-directed mutagenesis to achieve a favourable translation initiation sequence context surrounding the "AUG" initiation codon is also seen in the K10RB-FIX-G4 construct (Q-45V).


K10RB
K10RB-G4
T C G A
T C G A
TATA Box
TATA Box
BamHI
BamHI
K10 CAP
K10 CAP
K10 5'UTR
K10 5'UTR
Kpn I
Kpn I
ATG
C
A
G
ATG
G
T
C
Q-45V
FIX cDNA

To signify the intended enhanced ribosomal binding capacity of the K10-FIX-II and K10-FIX-G4 constructs, these constructs were named K10RB-FIX and K10RB-FIX-G4 respectively. These engineered constructs would now express mRNAs with a natural K10 cap site and a leader sequence 32 nucleotides long with a GC content of ~56% (18/32).

3.5.5. ENGINEERING THE 3' UNTRANSLATED SEQUENCE

The C-terminal codon of the natural BKVIb gene encodes a tyrosine. In the BKVIb gene, this codon (Tyr) is separated from the preceding codons by intron 7 (also seen in the corresponding human and murine genes; Reiger et al., 1988). The polyadenylation sequence is present 356 bases downstream from the translation termination codon and the cleavage site is a further 18 bases downstream. It has been suggested that the 3' end of some keratin genes are essential for efficient transcription (D. Roop, personal communication). In order to mimic the natural 3' end of the BKVIb gene, the SV40 polyA cassette was replaced with a segment of genomic DNA including intron 7 (along with its splicing determinants) and 285 bp of sequence distal to the BKVIb polyadenylation signal. In some gene constructs, use of the SV40 polyA cassette has resulted in abberant splicing (Huang and Gorman, 1990) and its replacement would obviate this variable. In this construct therefore, intron 7 of the BKVIb gene would serve to enhance mRNA processing and transport (Hamer and Lader, 1979; Gruss and Khoury, 1980; Mulligan and Berg, 1980; Villarreal and White, 1983; Brinster et al., 1988). The required sequences were amplified using the cloned BKVIb gene (in λKBVIb; Lehnert et al., 1984) as the template for PCR and termed the K10int7 cassette.

Two recombinant PCR strategies were used to facilitate the joining of the factor IX coding sequence to the K10int7 cassette. One method involved an initial ligation step followed by a second PCR in order to generate one PCR fragment. The other method involved PCR amplification using primers that resulted in SOE (Horton et al., 1989). Both methods involved the initial PCR amplification of the two segments to be joined *viz.* the complete factor IX coding sequence (with the previously engineered K10 CAP and 5'UTR sequences) and the K10int7 cassette.

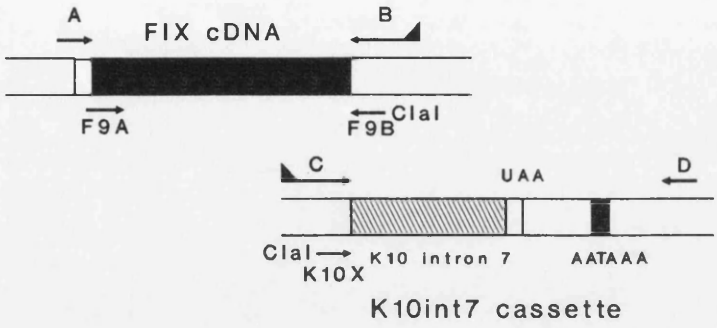
Figure 31: Construction design and manipulation of the 3'UTR

31A. This figure shows a diagrammatic representation of the factor IX cDNA and the 3' end of the K10 gene. The regions of the K10 gene shown in this figure include intron 7, the translation termination signal in exon 8 (which has only 4 coding nucleotides) and the polyadenylation signal. Together these comprise the 1 kb segment named the K10int7 cassette. The primers used to amplify the factor IX cDNA and the K10int7 cassette and to achieve the sequence manipulation are indicated. A, B, C and D are the 3'MAN primers listed in table 6. The *Cla*I recognition sites incorporated in the primers F9B and K10X are indicated. The solid triangles on primers B and C represent the complementarity of sequence to facilitate SOE.

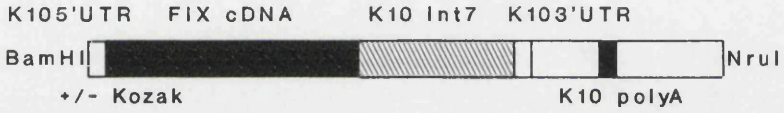
31B. This figure shows a diagrammatic representation of the joining of the factor IX coding sequence with the K10int7 cassette, using primers 3'MAN A and D. This was achieved in two ways, (i) ligation of the two PCR fragments at the *Cla*I sites in F9B and K10X and (ii) SOE using the complementary sequences in primers 3'MAN B and C (described in section 3.5.5). The recognition sequence for *Bam*HI (in A) and *Nru*I (in D) and the presence or absence of the Kozak consensus sequence (in K10RB-FIX and K10RB-FIX-G4, respectively) are indicated.

31C. This figure shows the PCR-based manipulations performed to join the two segments illustrated in 31A. L=1 kb ladder. Lanes 1 and 2 are PCR amplifications of the K10RB-FIX and K10RB-FIX-G4 constructs respectively (1.4 kb) . Lane 3 demonstrates the PCR amplification of the K10int7 cassette (~1 kb). Lanes 4 and 5 show products of PCR amplification (~2.4 kb) using primers 3'MAN A and D. The template used in lane 4 was obtained by ligation at the *Cla*I sites in F9B and K10X. The template used in lane 5 was obtained by SOE using the primers 3'MAN A-B and C-D followed by a second amplification using 3'MAN A-D.

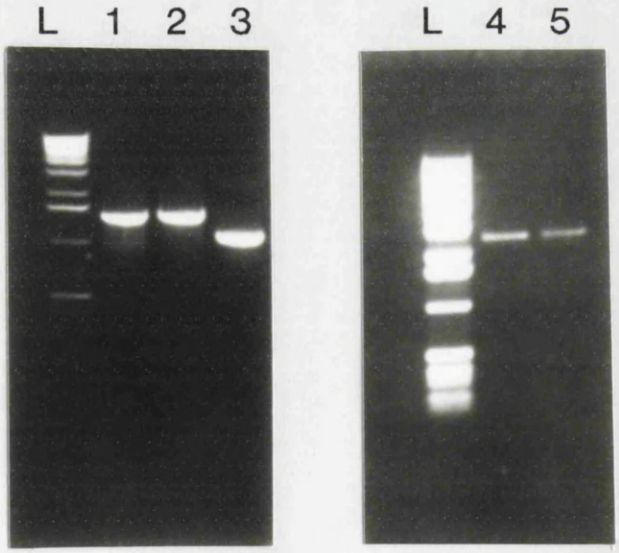
(A)


A
FIX cDNA
B
F9A
F9B
ClaI
C
UAA
D
ClaI
K10X
K10 intron 7
AATAAA
K10int7 cassette

(B)


K10 5'UTR
FIX cDNA
K10 Int7
K10 3'UTR
BamHI
NruI
+/- Kozak
K10 polyA

(C)


L 1 2 3
L 4 5

Five new primers were designed. Four were called 3'MAN (for 3' manipulation) primers 'A' to 'D' (labelled as A, B, C & D in figure 31) and the fifth was named K10X. The primer F9B, designed for cloning of the basic construct, was also used. The sequences of these primers are listed in table 6. 3'MAN A and D, which were common to both the methods and as seen in figure 31, were used for the final amplification of the combined PCR product. They were designed with restriction enzyme sequences incorporated in their sequence (*Bam*HI in 'A' and *Eco*RI and *Nru*I in 'D') in order to facilitate the subcloning of the final product.

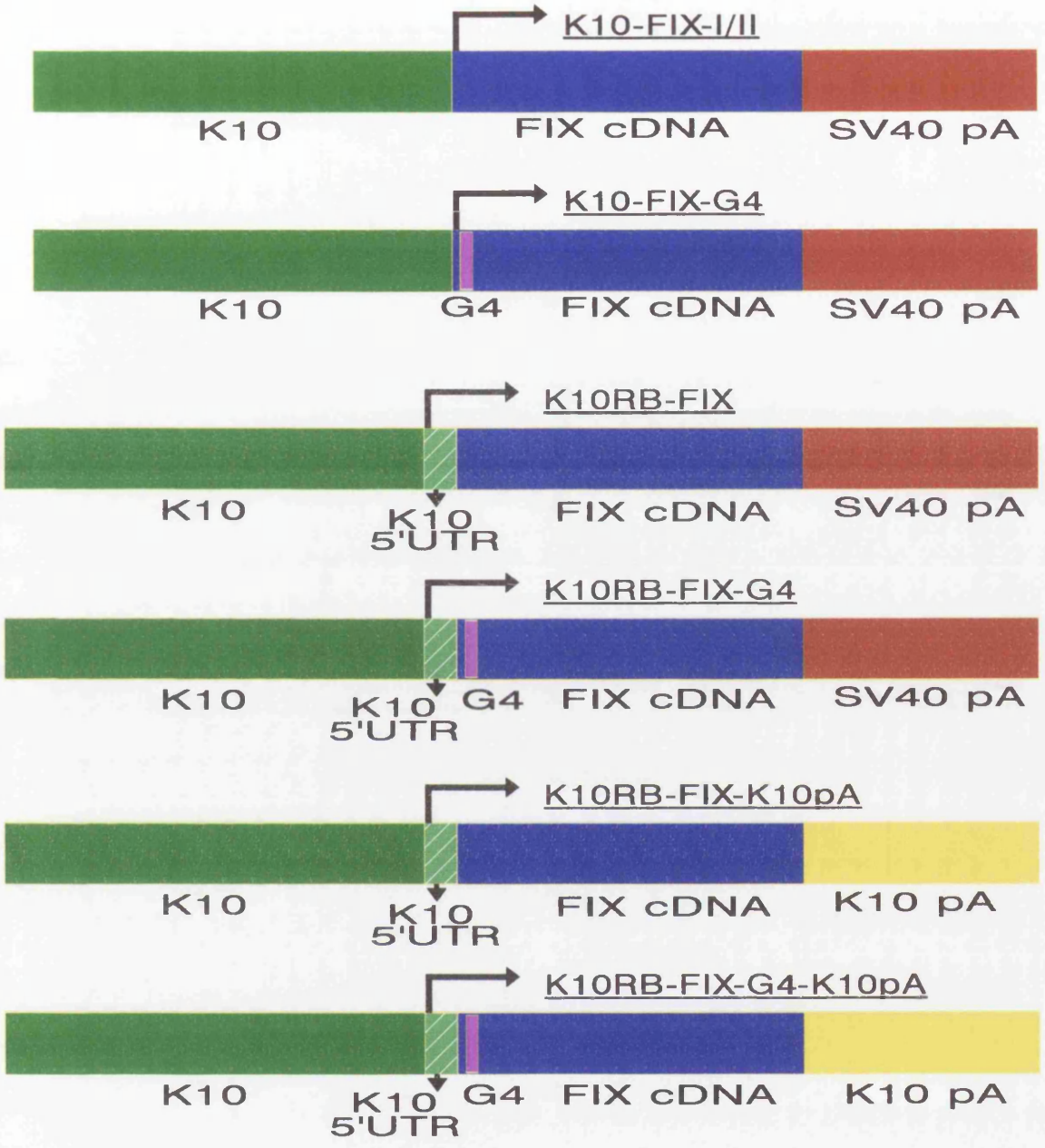
In the first method (figure 31), two PCRs were performed; (i) primers 3'MAN A and F9B were used to amplify the factor IX coding sequence (1.4 kb) using either K10RB-FIX or K10RB-FIX-G4 as templates and (ii) primers K10X and 3'MAN D were used to amplify the K10int7 cassette (1.0 kb) using the cloned BKVIb gene (λKBVIb; Lehnert et al., 1984; Reiger et al., 1985) as the template. Primers F9B and K10X had *Cla*I recognition sequences incorporated in their sequence so that both PCR products would have a *Cla*I recognition sequence at one end (3' end of the former and 5' end of the latter). The PCR products were digested, gel purified and ligated at their *Cla*I ends. This ligation mixture was then used as the template for a second PCR, using primers 3'MAN A and D to achieve the joining into one DNA fragment (2.4 kb).

In the second method (figure 31), the same two initial PCRs were performed on identical plasmid templates, except for the replacement of primers F9B and K10X with 3'MAN B and C, respectively. Primers 3'MAN C & D were designed in such a way that they would achieve complementarity of the two PCR products at their ends, thus enabling SOE. Both PCR products were purified following a gel electrophoresis and mixed in equal quantities. The mixture was denatured and allowed to reanneal slowly. Due to the sequence of 3'MAN B and C, the otherwise unrelated PCR products would overlap at their ends. This was used as a template for PCR using the same external primers, 3'MAN A and D, which resulted in the desired product by SOE.

At the conclusion of this work, the K10int7 cassette had been successfully engineered 3' to the factor IX coding sequence, but not yet cloned into a plasmid vector.

Figure 32: Summary of the various K10-FIX gene expression constructs

This figure shows a diagrammatic representation of the various K10 promoter-based factor IX gene expression constructs. The K10 promoter, factor IX cDNA, SV40 polyA cassette and K10int7 cassette are depicted as green, purple, red and yellow filled boxes, respectively. The repaired 5'UTR is represented by a green hatched box and the manipulation to achieve a favourable translation initiation sequence by a pink box. The transcription start sites are indicated.


K10-FIX-I/II
K10
FIX cDNA
SV40 pA
K10-FIX-G4
K10
G4
FIX cDNA
SV40 pA
K10RB-FIX
K10
K10
5'UTR
FIX cDNA
SV40 pA
K10RB-FIX-G4
K10
K10
5'UTR
G4
FIX cDNA
SV40 pA
K10RB-FIX-K10pA
K10
K10
5'UTR
FIX cDNA
K10 pA
K10RB-FIX-G4-K10pA
K10
K10
5'UTR
G4
FIX cDNA
K10 pA

3.5.6 SUMMARY OF THE VARIOUS K10-FACTOR IX GENE CONSTRUCTS

Figure 32 shows all the constructs containing the factor IX coding sequence driven by the K10 promoter/enhancer element.

3.6. CATIONIC LIPOSOME-MEDIATED TRANSFECTION OF HaCaT CELLS

Cationic liposomes have been used for the transfection of a wide variety of established cell lines and primary cells either transiently or permanently (Loeffler and Behr, 1993). These newer cationic-lipid vectors have overcome most of the drawbacks associated with the earlier liposomal preparations (Nabel and Felgner, 1993). There are several advantages to this system of gene transfer: (i) it is highly efficient; (ii) no upper limit exists for the size of DNA; (iii) it is non-toxic; (iv) the cells do not have to be actively dividing; (v) it is applicable to a wide variety of cell types and (v) it is an effective method of gene transfer *in vivo* (approved for clinical trials; Nabel and Felgner, 1993) and a particularly attractive method to directly target keratinocytes.

3.6.1. OPTIMISATION OF THE TRANSFECTION PROTOCOL WITH A LINEARISED *lacZ* REPORTER CONSTRUCT

Even though better transfection efficiencies are obtained using supercoiled plasmids, to minimise the possibility of interruption of the transcriptional unit of the recombinant expression constructs during random integration into the genomes of the transfected cells, all the constructs were initially linearised. To ascertain the optimum quantity of linearised plasmid DNA and the expected efficiency of transfection obtained for the HaCaT cell line, a *lacZ* reporter construct pIRV-Neo-Act-lac Z (Beddington et al., 1989) was used. In this construct, the expression of *lacZ* was driven by the β-actin promoter. Digestion with *Not*I released a 9.6 kb linearised DNA fragment which was extensively dialysed against T.E. buffer prior to its use in transfection.

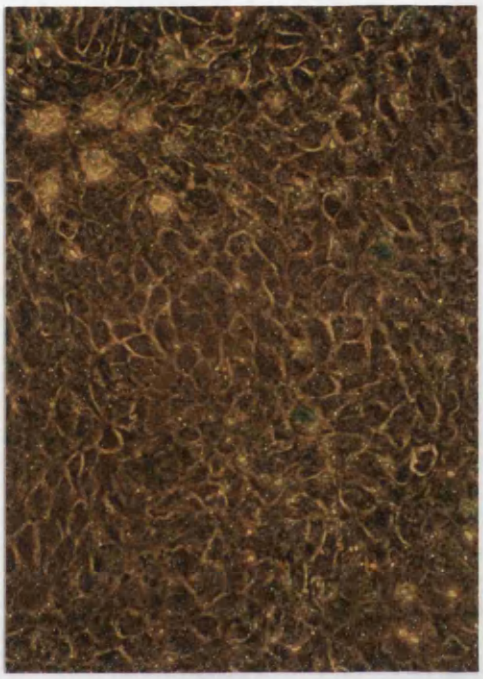
1 μg, 5 μg and 10 μg of linearised plasmid DNA were transfected into subconfluent HaCaT cells in individual 80 cm^2 tissue culture flasks (in duplicate). The

Figure 33: Optimisation of liposome-mediated transfection of HaCaT cells using a *LacZ* reporter construct

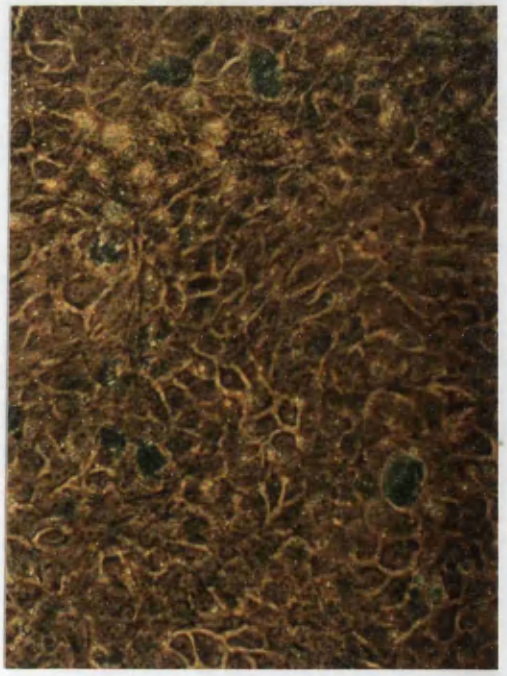
33A-D. These figures show the results of the staining for β-galactosidase activity following transfection with a reporter *LacZ* construct (pIRV-Neo-Act-LacZ) to estimate the transfection efficiency achieved using a cationic liposome preparation (DOTAP). 33A was the negative control and was exposed to liposomes without DNA. 1 μg, 5 μg and 10 μg of DNA was used for transfections in 33B, C and D, respectively. The transfection efficiency was estimated at ~1% in 33B, ~3% in 33C and 5-6% in 33D.

(A)

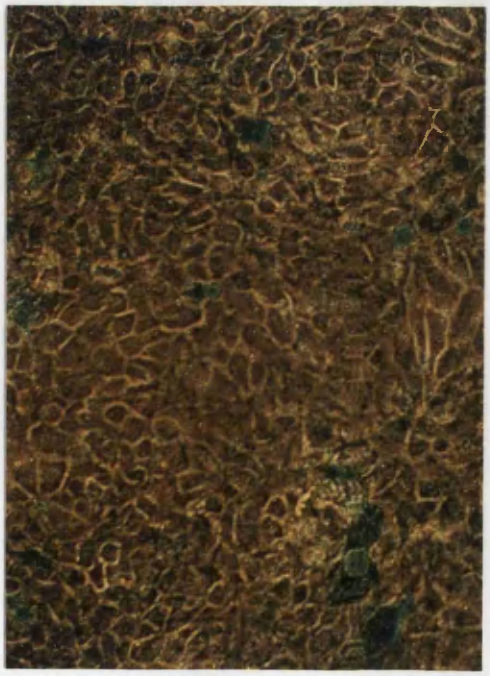
(B)



(C)



(D)



transfected cells were stained for β-galactosidase activity, 72 hours after the transfection (48 hours after the withdrawal of the DNA-liposome mixture). The results of a negative control (which was exposed to liposomes without DNA) and the three different concentrations of DNA transfected, are shown in figure 33. Transfection efficiencies were estimated for the various transfections. A linear increase in the percentage of blue cells with increasing concentrations of DNA was observed. Transfection of 10 μg of linearised plasmid DNA gave a maximum transfection efficiency of 5-6%, whereas 1 μg gave an efficiency of 1-2%. Loeffler and Behr reported an average transfection efficiency of 15% using cationic liposomes for a wide variety of cells. The lower transfection efficiency observed in this study is likely to be due to the use of linearised plasmids (Kriegler, 1990a).

3.6.2. TRANSFECTION OF HaCaT CELLS WITH THE VARIOUS FACTOR IX EXPRESSION CONSTRUCTS

The various K10 promoter-factor IX cDNA constructs included: K10-FIX-I (which had a missense S136T mutation in the factor IX coding sequence), K10-FIX-II (the basic construct), K10-FIX-G4, K10RB-FIX and K10RB-FIX-G4. All five constructs were linearised at the *Sal*I site at the 5' end of the K10 promoter and extensively dialysed against T.E. buffer prior to use in the transfection protocol (section 2.9.3). In order to have a selectable marker for identifying transfected cells, they were co-transfected with pSV2Neo (also linearised; Southern and Berg, 1982) to confer G418 resistance (section 2.9.5). Transfections were carried out in duplicate using 10 μg of the appropriate factor IX construct (section 3.6.1) along with 0.5 μg of pSV2Neo. Immediately after the transfection protocol, selection was applied using G418 (see section 2.9.5). Two separate pooled samples of clones were thus obtained for each transfected factor IX expression construct.

3.6.3. CLONING OF INDIVIDUAL STABLE TRANSFECTANTS

Clones in petridishes that had approximately 50 cells each and were adequately separated from other clones, were used for the cloning procedure. The method using

"cloning rings" is described in detail in section 2.9.5 (Malcolm et al., 1987). Selected clones that survived, were gradually scaled up from small multiwelled plates to 80 cm^2 flasks maintaining the G418 selective pressure throughout. Table 7 summarises the number of clones picked for each factor IX construct and the numbers that survived the procedure. On average, the success rate of the cloning procedure was ~26%. Both individual transfected clones and pools were used for further analysis.

3.6.4. ASSESSMENT OF STABLE INTEGRATION OF THE TRANSFECTED DNA

Genomic DNA was extracted from the various transfected clones and pools of HaCaT cells (table 7) and from control (untransfected) HaCaT cells. Primers F9A and F9B, which amplify the entire factor IX coding sequence were used to test for the successful integration of the transfected constructs (figure 34). These primers would fail to amplify the endogenous human factor IX sequences as the corresponding genomic fragment (>30 kb), would be far beyond the size range for conventional PCR amplifications. As cells used for the DNA extraction had been passaged at least five generations for pools and three for individual stable clones, this was a reliable indication of stable integration. The presence of an amplified 1.4 kb PCR product indicated that at least some intact copies of the complete factor IX coding sequence had been integrated per pool of transfected clones and at least one copy per cell for the individual stable clones. All the pools and a total of fifteen stably transfected (G418 resistant) clones were found to be positive (table 7). As seen in figure 34, DNA from untransfected HaCaT cells failed to show a PCR product.

3.7. ASSESSMENT OF FACTOR IX EXPRESSION: RNA ANALYSIS

Total cellular RNA was extracted from fully confluent flasks of the transfected and control HaCaT cells using the acid guanidinium thiocyanate-phenol-chloroform

Table 7: Summary of transfected HaCaT cells

EXPRESSION CONSTRUCT	POOLS	CLONES[a]	CLONES[b]	CLONES[c]
K10-FIX-I	2	12	3	2
K10-FIX-II	2	15	4	3
K10-FIX-G4	2	11	3	2
K10RB-FIX	2	10	2	0
K10RB-FIX-G4	2	12	3	2

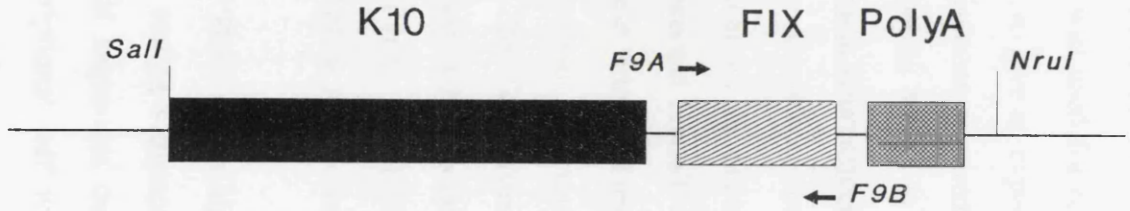
[a]Number of clones picked for analysis; [b]Number of clones that survived the cloning procedure. All tested positive with the F9A/B PCR (figure 34); [c]Number of clones analysed by RT-PCR (all positive; figure 35) and ELISA.

Figure 34: PCR analysis of transfected cells showing integration of the transgene

34A. This shows a diagrammatic representation of the primers used for the screening PCR to assess for integration of the transgene. The expected product of amplification is 1.4 kb.

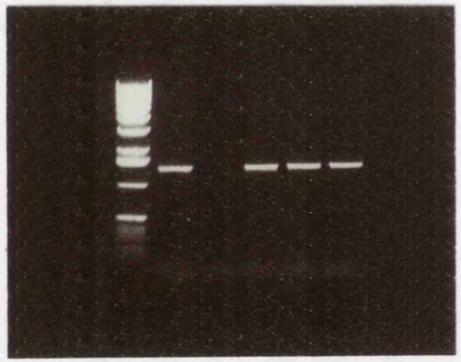
34B & C. These figures show the results of the screening of the various transfected HaCaT cell lines (pools and individual clones). L=1 kb ladder; +=positive control DNA (cloned factor IX cDNA); H=DNA from untransfected HaCaT cells. In 34B, I=K10-FIX-I (pool); II=K10-FIX-II (pool); G4=K10-FIX-G4 (pool). In 34C, lane 1=K10RB-FIX (pool); lane 2=K10RB-FIX-G4 (pool); Lanes 3-5=K10-FIX-I clones 1-3; Lanes 6-9=K10-FIX-II clones 1-4; Lanes 10-12=K10-FIX-G4 clones 1-3; Lanes 13&14=K10RB-FIX clones 1 and 2; Lanes 15-17=K10RB-FIX-G4 clones 1-3.

(A)

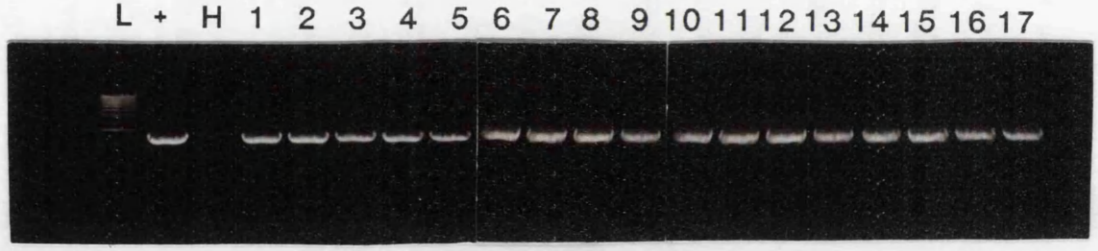

K10
FIX
PolyA
SalI
F9A
NruI
F9B
1Kb

(B)

F9A/B


L + H I II G4

(C)


L + H 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

method (Chomczynski and Sacchi, 1987). Two methods were used to assess the expression of factor IX mRNA in the transfected cells.

3.7.1. FACTOR IX mRNA DETECTION BY RT-PCR ANALYSIS

Total cellular RNA (0.5 μg) extracted from the transfected HaCaT cells were used as a template for RT-PCR analysis. Primers previously designed for the amplification of the human factor IX cDNA were used for this purpose (table 5). Primer F9IIB was used for reverse transcription followed thereafter by PCR using F9IIC and F9IID, to give an expected product of 523 bp (figure 35). As opposed to the nested PCR amplifications required for the detection of ectopic factor IX transcripts in peripheral lymphocytic RNA, in this analysis no second (nested) round of amplification was performed. Untransfected HaCaT cells and HaCaT cells transfected with pSV2Neo were used as negative controls. Primers specific for the actin cDNA were used for RT-PCR amplification, in order to confirm the integrity of the RNA extracted from the untransfected HaCaT cells. Factor IX mRNA was detected by RT-PCR, in all tested pools and individual transfected clones containing a factor IX construct (table 7), but was not detected in untransfected (or transfected with pSV2Neo) HaCaT cellular RNA (figure 35). To confirm that the expected PCR product arose from the factor IX mRNA expressed in these cells and not from contaminating DNA in the RNA preparation, PCRs using F9IIC and F9IID were simultaneously carried out without a prior reverse transcription stage. None of the RNA samples tested were positive for this PCR (figure 35).

These results also confirm the stable integration of the transgene and the integrity of the coding sequence. In addition, the RT-PCR results indicate the integrity of the essential sequences outside the coding sequence and it can be concluded that the "transcriptional unit" was intact in each case. These results do not quantify the factor IX expression from the transfected cell lines for which Northern blot analysis and ELISAs were performed.

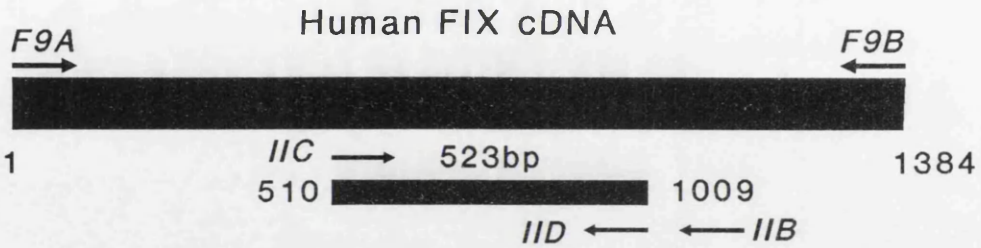
Figure 35: Analysis of factor IX mRNA expression by RT-PCR

35A. This shows a diagrammatic representation of the primers used for the RT-PCR to detect factor IX mRNA. F9IIB (IIB) was used for reverse transcription and F9IIC (IIC) and F9IID (IID) for the PCR. The expected product of amplification is 523 bp.

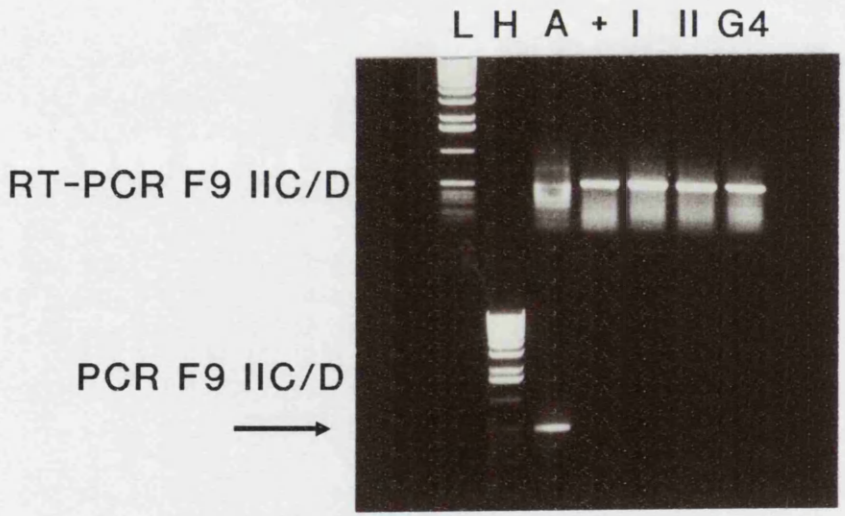
35B. This figure shows the result of the screening of three pools of transfected HaCaT clones. L=1 kb ladder; H=RNA from untransfected HaCaT cells; A=RT-PCR to detect actin mRNA in the untransfected HaCaT RNA sample used in lane H; +=plasmid DNA from the cloned factor IX cDNA. In the upper panel labelled "RT-PCR F9IICD", the RNA samples were subjected to reverse transcription prior to PCR. In the lower panel (indicated by an arrow; marked "PCR F9IICD") the same samples were subjected to PCR without prior reverse transcription. I=K10-FIX-I (pool); II=K10-FIX-II (pool); G4 (K10-FIX-G4 (pool). The first lane in the lower panel (after the size marker; marked by the arrow) is a positive control (cloned factor IX cDNA). The pooled samples in the lower panel show no products of amplification, indicating the absence of contaminating DNA.

35C. This figure shows the results of the screening of the various transfected HaCaT cell lines (pools and individual clones) by RT-PCR. L=1 kb ladder; H=RNA from untransfected HaCaT cells; N=RNA from HaCaT cells transfected with pSV2Neo; lane 1=K10RB-FIX (pool); lane 2=K10RB-FIX-G4 (pool); Lanes 3&4=K10-FIX-I clones 1 and 2; Lanes 5-7=K10-FIX-II clones 2-4; Lanes 8&9=K10-FIX-G4 clones 1 and 3; Lanes 10&11=K10RB-FIX-G4 clones 1 and 3.

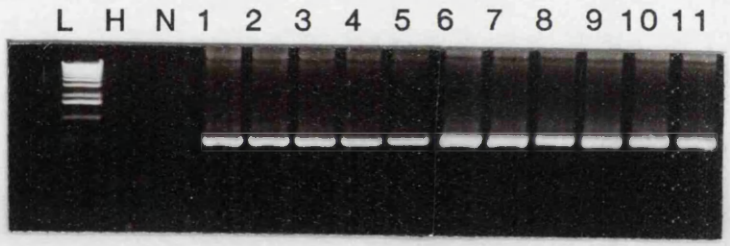
(A)


Human FIX cDNA
F9A
F9B
1
IIC
523bp
1384
510
1009
IID
IIB

(B)


L H A + I II G4
RT-PCR F9 IIC/D
PCR F9 IIC/D

(C)


L H N 1 2 3 4 5 6 7 8 9 10 11

3.7.2. FACTOR IX mRNA DETECTION BY NORTHERN BLOT ANALYSIS

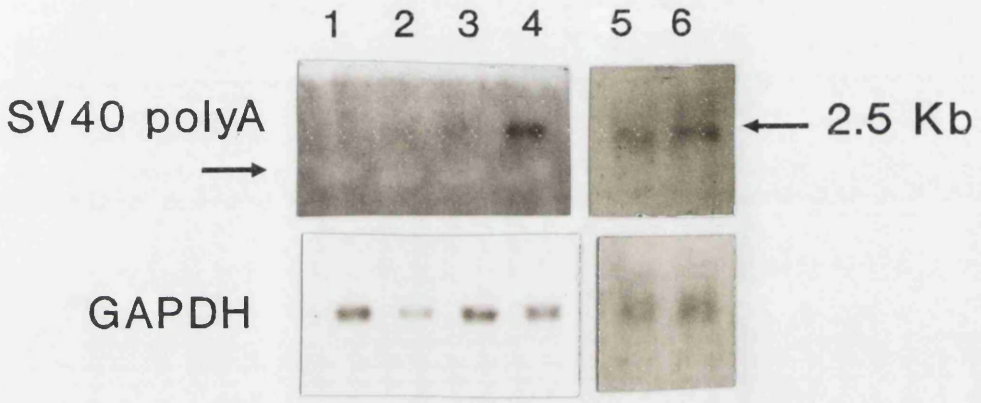
Total cellular RNA (10 μg) from the five pools of transfected factor IX constructs was used for Northern blot analysis. An RNA sample from untransfected HaCaT cells was used as a negative control. In order to specifically detect the mRNA expressed by the transfected constructs, the SV40 polyA cassette was used to probe the Northern blot. Transcripts from a ubiquitiously transcribed gene, GAPDH, were analysed in order to assess the relative quantities loaded and the structural integrity of the RNA samples used in the Northern analysis.

Figure 36 demonstrates the detection of a ~2.5 kb recombinant human factor IX transcript in total cellular RNA extracted from pools of clones transfected with the five constructs. The GAPDH probe revealed that RNA in the lane not showing factor IX transcripts was intact and that similar quantities of RNA were loaded in all lanes. K10-FIX-I and K10-FIX-II gave very faint, though detectable, factor IX bands on Northern analysis, compared with K10-FIX-G4, K10RB-FIX and K10RB-FIX-G4, which in turn were similar to each other. All constructs produced a transcript of the same size (~2.5 kb) and no alternative transcripts were seen. The results with the SV40 polyA probe were obtained after a five day exposure compared with the GAPDH probe which was obtained after 24 hours, indicating low levels of factor IX expression.

RNA extracted from the epidermis of transgenic mice created using the K10-FIX-I construct, in collaboration with Dr. Y. Alexander, also gave the same result. The presence of only one band indicates that there were no alternate transcripts and that within the limited resolution of the agarose gel, the factor IX expression from the various constructs utilised one site each for transcriptional initiation and termination. All the constructs used for transfection had almost identical sequences in all regions except in the 5'UTR. The 5'UTR of the K10-FIX-I, K10-FIX-II and K10-FIX-G4 constructs is predicted to be 14 nucleotides and that of K10RB-FIX and K10RB-FIX-G4 constructs to be 32 nucleotides long. The factor IX coding sequence in the various constructs is 1398 nucleotides (of which, 9 nucleotides are from the factor IX 3'UTR) and the 3'UTR sequence contributed by the SV40 polyA cassette is predicted to be 761 nucleotides (assuming the 66 nucleotide small t-intron to be normally spliced out). The total length of

Figure 36: Analysis of factor IX mRNA expression by Northern blotting

This figure shows the detection of factor IX mRNA by Northern blot analysis of the pools of transfected HaCaT cells. Lane 1=RNA from untransfected HaCaT cells; Lane 2=K10-FIX-I; Lane 3=K10-FIX-II; Lane 4=K10-FIX-G4; Lane 5=K10RB-FIX; Lane 6=K10RB-FIX-G4. The individual lanes can be distinguished by the loading dye shown by the arrow. Probing the Northern transfer with the SV40 polyA cassette reveals a ~2.5 kb transcript (indicated by an arrow) in all transfected samples (faint signals in lanes 2 and 3). The same filters were re-probed with a GAPDH probe to determine the integrity and relative loading in the lanes.


1 2 3 4 5 6
SV40 polyA
2.5 Kb
GAPDH

the factor IX mRNA transcript from the two types of constructs is therefore predicted to be 2174 and 2191 nucleotides in addition to the poly A tail, respectively. As the Northern blot revealed a single band of ~2.5 kb, it indicates that the poly A tail is approximately 300 nucleotides long. It is noteworthy that due to the low resolution of the agarose gel it is not possible to detect the predicted 17 nucleotide difference among the two classes of transcript, nor is it possible to determine if the small t-intron (66 nucleotides long) in the SV40 polyA cassette was successfully spliced out.

3.8. ASSESSMENT OF FACTOR IX EXPRESSION: TRANSLATIONAL ANALYSIS.

Enzyme Linked Immunosorbent Assay (ELISA) was used to detect secreted factor IX protein secreted into conditioned media by HaCaT cells transfected with the various gene expression constructs. The same assay was also used to estimate the levels of factor IX in the plasma of transgenic mice (created with the K10-FIX-I construct).

3.8.1. ENZYME LINKED IMMUNOSORBENT ASSAY OF SECRETED FACTOR IX

Two monoclonal antibodies, namely 'A5' and 'A7' were used in a "sandwich" ELISA for the detection of factor IX antigen. The antibodies and protocol for the ELISA were provided by Dr. Kenneth Smith (University of New Mexico, USA) and are described in section 2.10. For factor IX to be biologically active, the γ-carboxylation of 12 glutamyl residues near the N-terminus is essential (reviewed by Vermeer et al., 1990). Since Vitamin K is required for this post-translational modification, it was supplemented in the tissue culture medium. The antibody (A7) used in the ELISA detected a calcium-dependent epitope in the light chain and was used to specifically estimate γ-carboxylated factor IX (Smith and Ono, 1984; Smith, 1988) and thus determine the capacity of keratinocytes to secrete the biologically active protein. Transfected cells were made to condition serum-free medium (for 24 hours) for the assay. Even though this may have

had some depressing effect on the metabolism of the cells (probably slightly affecting the factor IX synthesis and/or secretion), this was done to avoid any cross-reaction with bovine factor IX present in the supplementing serum. Furthermore, Smith et al. (1993) observed a 40% suppression of the ELISA values due to the presence of (some factors in) mouse plasma. This effect would also be avoided by use of serum-free medium for sample collection.

Cells in fully confluent 80 cm^2 flasks were incubated for 24 hours in serum-free, vitamin K containing tissue culture medium. Serial dilutions of conditioned media (in triplicate) were used in the ELISA. Serial dilutions of commercially available pure factor IX antigen (Dakopatts) was used to plot a standard curve between 0 to 50 ng/ml. The biotinylated 'A7' antibody bound to factor IX was detected by a chromogenic assay, and the optical densities (at 490 nm) obtained for the conditioned media were plotted on the standard curve obtained with serial dilutions of pure factor IX.

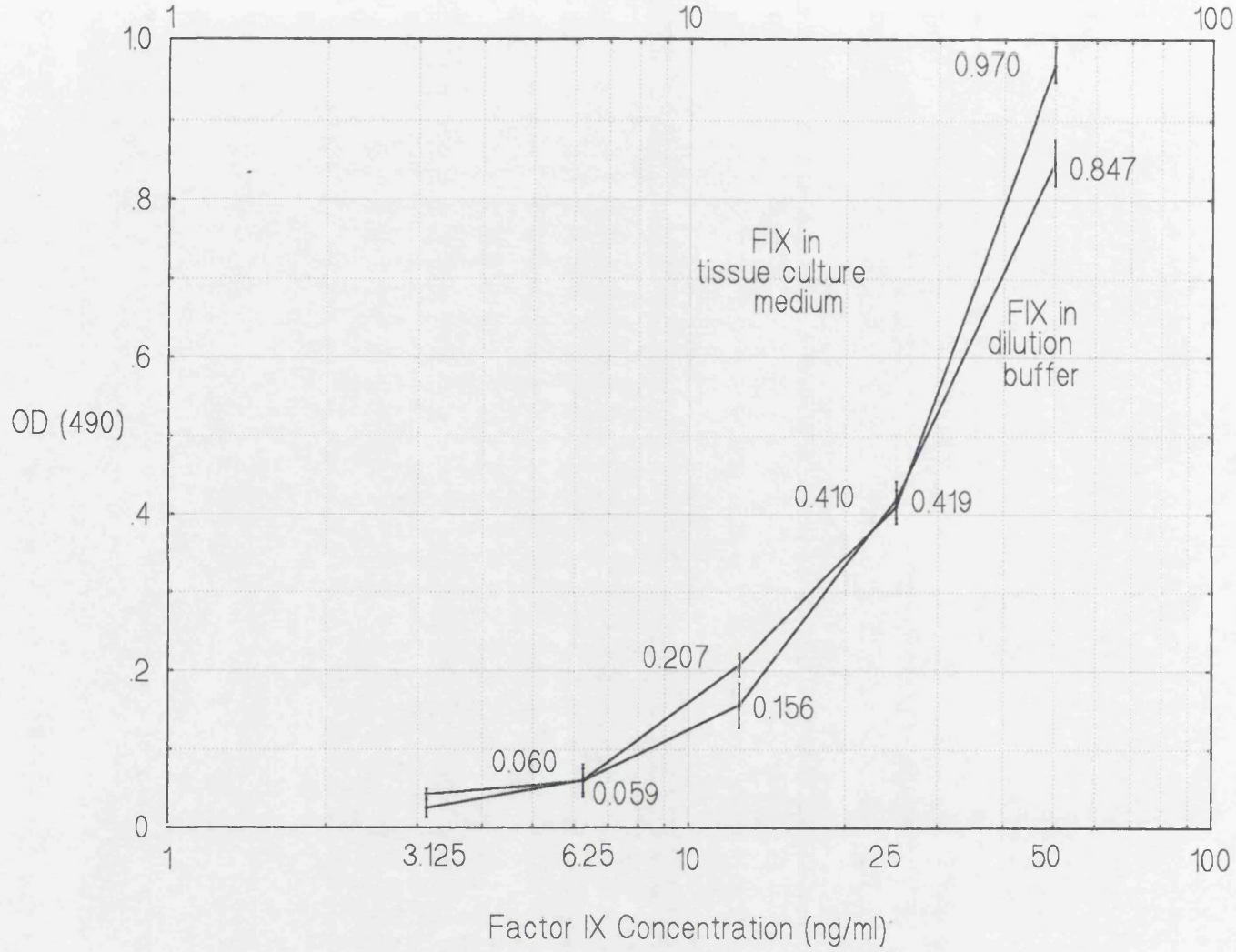
To plot a standard curve, serial dilutions of pure factor IX were made in the recommended dilution buffer (section 2.10). A control experiment was performed to rule out any inhibition/enhancement of the optical densities (OD) due to the high proportion of tissue culture medium in the undiluted, 1 in 2 and 1 in 4 diluted conditioned media samples. Figure 37 shows the standard curves plotted for the serial dilutions of pure factor IX in tissue culture medium (serum free) and in the recommended dilution buffer. No significant difference in the curve profile was apparent.

Figures 38-41 show the plotting of OD (490) readings obtained from the secreted factor IX in conditioned media from the various transfected cell lines (pools and clones; tables 7 and 8). In addition, figures 38 and 39 also show the curves obtained for dilutions of conditioned media obtained from the pooled samples of K10-FIX-G4, K10-FIX-II and K10-FIX-I. These curves indicate that serial dilutions of the conditioned media showed the expected reductions in factor IX concentrations (figure 38) and although dilutions of up to 1 in 4 plotted in the linear portion of the standard curve in the case of K10-FIX-G4, this was only possible up to a 1 in 2 dilution for K10-FIX-I and K10-FIX-II (figure 38, 39). The OD readings used for the estimation of factor IX secreted by all transfected pooled clones plotted in the linear portion of the standard curve (figures 38-41). While

Figure 37: Effect of tissue culture medium on ELISA

X-axis (logarithmic scale), factor IX concentration (ng/ml); Y-axis, optical density at 490 nm (OD 490). Serial dilutions of pure factor IX (50 ng/ml to 3.125 ng/ml) were made in tissue culture medium and sample dilution buffer (Smith, 1988), and the two curves obtained on plotting the optical densitometric readings are seen. The error bars represent variation of ±2 standard deviations.

Figure 37


1
10
100
1.0
.8
.6
.4
.2
0
OD (490)
0.970
0.847
FIX in
tissue culture
medium
FIX in
dilution
buffer
0.410
0.419
0.207
0.156
0.060
0.059
1
3.125
6.25
10
25
50
100
Factor IX Concentration (ng/ml)

each of the ELISAs seen in figures 38-41 had different sensitivities, in general it was always possible to detect factor IX levels above 1.5-3 ng/ml. Conditioned media collected from untransfected HaCaT cells consistently gave OD readings that were below the background.

Figure 40 demonstrates the effect of confluency and storage of conditioned media, on the factor IX levels measured. G4-80 represents the same K10-FIX-G4/Pool in this figure (including the generation), except that the flask was 80% confluent at the time of sampling. Also in figure 40, G4-5D represents the same K10-FIX-G4/Pool sample assayed in figure 38 that had been stored at 4°C for 5 days. Bearing in mind the relative inaccuracy of the readings in this region of the standard curve, they both nevertheless represent a significant drop from the K10-FIX-G4/Pool readings in figures 38 and 40.

The actual OD (490) readings, factor IX concentrations in the various conditioned media samples and the factor IX secretion calculated as ng/million cells/24 hours are listed in table 8. The measure of variance is indicated by the ±2 standard deviations listed within parentheses.

In an attempt to assess any difference that may have arisen due to the sequence differences in the constructs, a comparative analysis of factor IX expression from the various transfected cells was performed (section 4.2.1). In order to avoid errors due to possible variations amongst the individual clones, only pooled clones were used for this analysis. Figure 42 shows the relative expression levels obtained, along with error bars demonstrating the variation of ±2 standard deviations. A student *t* test was performed to estimate the significance of the differences observed in the average expression levels (appendix II). Listed in table 9 are, the mean expression levels, ±2 standard deviations and the result of the student *t* test. The rest of the factor IX constructs were compared with the expression achieved by the basic (K10-FIX-II) and K10-FIX-G4 constructs. The results demonstrate the following: (a) no significant difference existed between the basic construct (K10-FIX-II) and that with the S136T mutation (K10-FIX-I); (b) All the three manipulated constructs i.e. K10-FIX-G4, K10RB-FIX and K10RB-FIX-G4, showed significantly higher expression than the basic construct and (c) No significant difference was observed amongst the three manipulated constructs.

Figure 38: ELISA I

X1-axis (logarithmic scale), factor IX concentration (ng/ml), used for plotting the standard curve of serial dilutions of pure factor IX; X2-axis, dilution factor (1 = undiluted, 1/2 = 1:1 dilution, etc.) of conditioned tissue culture medium from transfected HaCaT cells; Y-axis, optical density at 490 nm (OD 490). The X2-axis was used to plot the densitometric readings of serial dilutions of conditioned medium from cells (pooled clones) transfected with K10-FIX-G4 and K10-FIX-II. The OD readings (summarised in table 8) were used to calculate the factor IX concentration (ng/ml) using the standard curve, and the readings are tabulated in the figure (inset). Abbreviations used in the table are, Diln=dilution factor; Obs.=observed reading; Exp.=expected reading (from the dilution factor). The readings obtained (Obs.) for the conditioned medium from cells transfected with K10-FIX-G4 were above the background up to 1 in 8 dilution and correlated with the expected readings (Exp.). Similar correlation was seen for K10-FIX-II up to 1 in 4 dilution. It can be seen from the curves, that in the case of K10-FIX-G4, serial dilutions up to 1 in 4 plotted in the linear portion of the standard curve, whereas for K10-FIX-II this was only possible up to a 1 in 2 dilution.

Figure 38

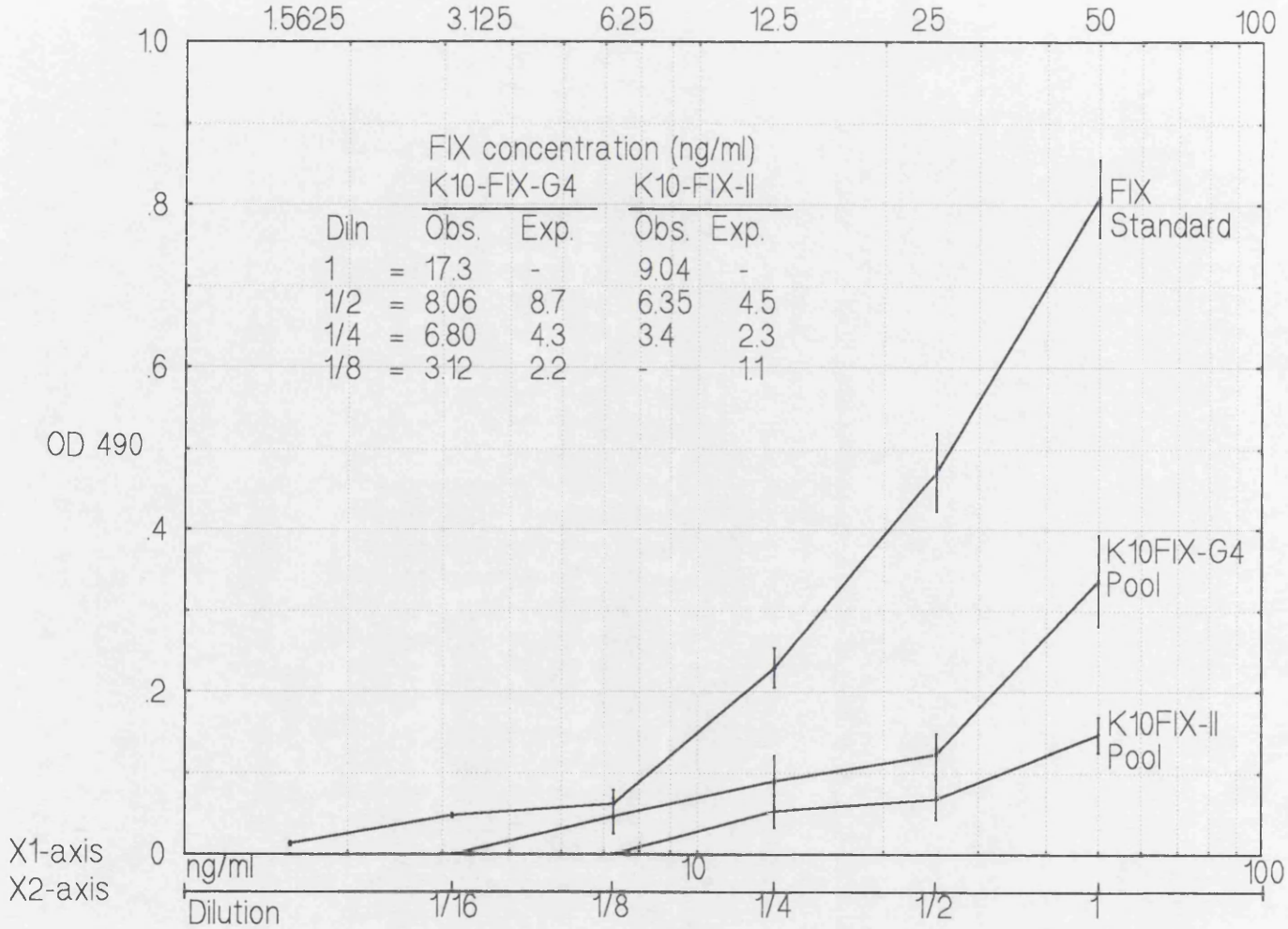

1.5625
3.125
6.25
12.5
25
50
100
1.0
.8
.6
.4
.2
0
OD 490
FIX concentration (ng/ml)
K10-FIX-G4
K10-FIX-II
Diln
Obs.
Exp.
Obs.
Exp.
1 = 17.3 - 9.04 -
1/2 = 8.06 8.7 6.35 4.5
1/4 = 6.80 4.3 3.4 2.3
1/8 = 3.12 2.2 - 1.1
FIX Standard
K10FIX-G4 Pool
K10FIX-II Pool
X1-axis
ng/ml
10
100
X2-axis
Dilution
1/16
1/8
1/4
1/2
1

Figure 39: ELISA II

X1, X2 and Y axes are identical to those in figure 38. The factor IX standard curve and curve for serial dilutions of conditioned medium from cells transfected with K10-FIX-I, were also plotted as in figure 38. In the case of K10-FIX-I, even though dilutions up to 1 in 4 gave readings above background, only the undiluted and 1 in 2 diluted samples plotted in the linear portion of the standard curve. OD (490) readings obtained for the various individual clones of transfected cells (K10-FIX-I/II) are plotted on the standard curve and the factor IX concentrations (ng/ml) calculated from the X1-axis, are tabulated (inset). I and II indicate clones of K10-FIX-I and K10-FIX-II, respectively. The individual clones are numbered (for example, II/4 is the 4th stably transfected clone of K10-FIX-II). All observed OD (490) readings, calculated factor IX concentrations (ng/ml) and factor IX expression levels (ng/million cells/24 hours) are summarised in table 8.

Figure 39

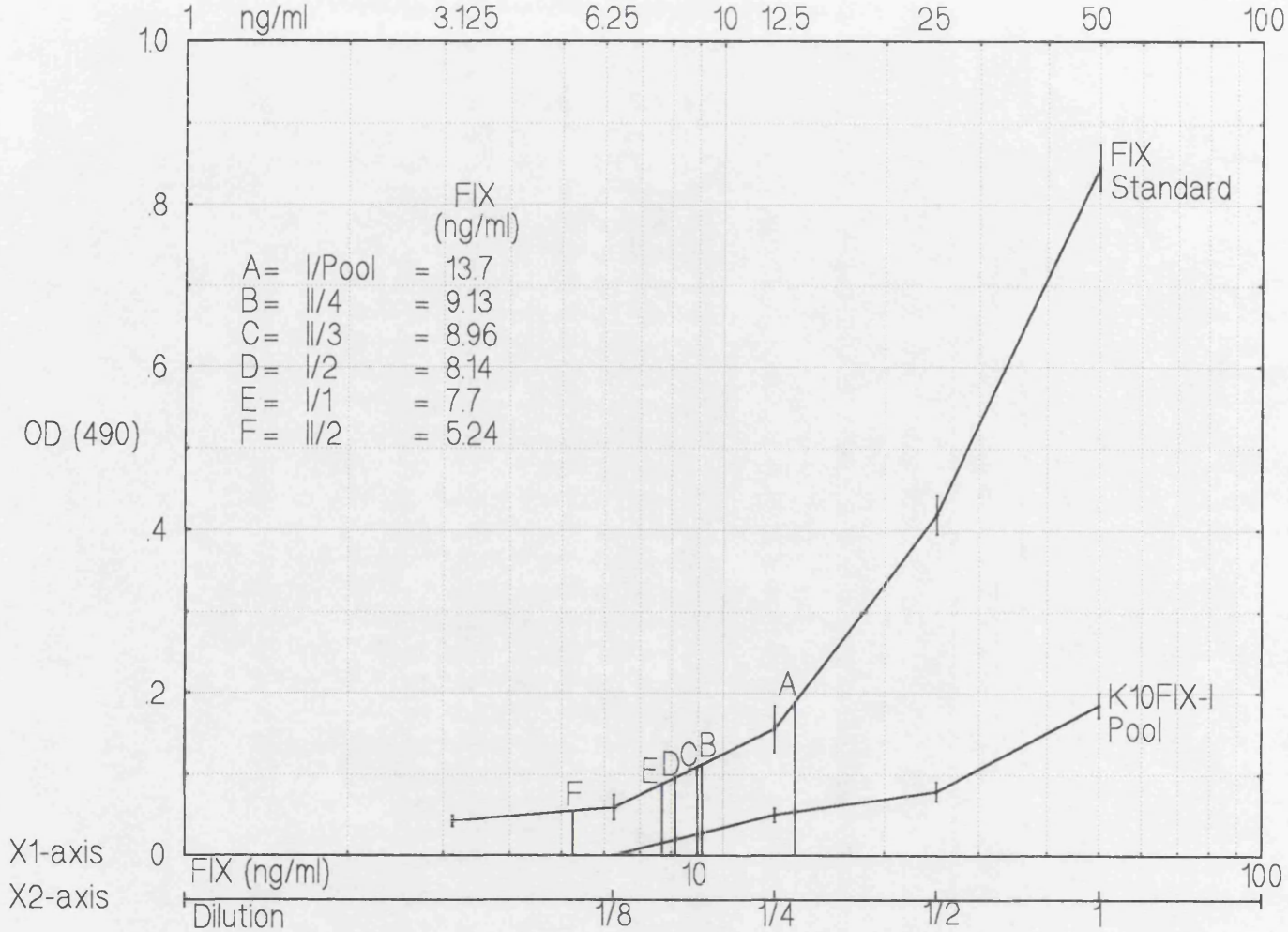

1 ng/ml
3.125
6.25
10
12.5
25
50
100
1.0
.8
.6
.4
.2
0
OD (490)
FIX
(ng/ml)
A= I/Pool = 13.7
B= II/4 = 9.13
C= II/3 = 8.96
D= I/2 = 8.14
E= I/1 = 7.7
F= II/2 = 5.24
FIX
Standard
K10FIX-I
Pool
A
B
C
D
E
F
X1-axis
FIX (ng/ml)
10
100
X2-axis
Dilution
1/8
1/4
1/2
1

Figure 40: ELISA III

The X and Y axes are similar to those in figure 37. OD (490) readings obtained for the conditioned media (summarised in table 8) from cells transfected with K10-FIX-G4 (pools and individual clones) are plotted on the factor IX standard curve. The estimated concentration of factor IX (ng/ml) in the various media are tabulated in the figure (inset). G4=K10-FIX-G4; G4-5D represents conditioned medium from K10-FIX-G4 transfected pool (the same sample used in the ELISA in figure 38) stored for 5 days at 4°C; G4-80 represents the same K10-FIX-G4 pool of transfected cells used in this ELISA, but the conditioned medium was collected from 80% confluent flasks.

Figure 40

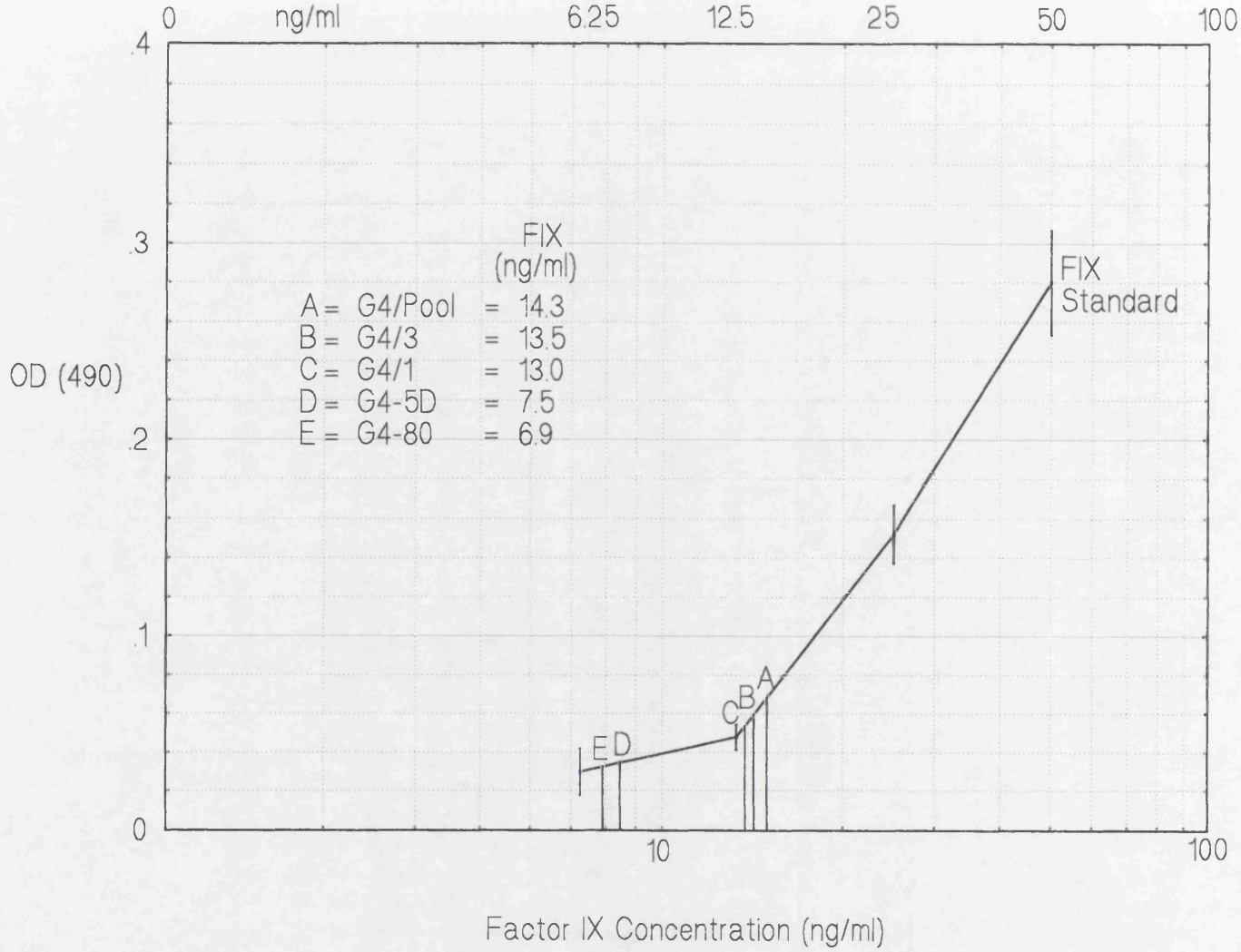

0
ng/ml
6.25
12.5
25
50
100
.4
.3
.2
.1
0
OD (490)
FIX
(ng/ml)
A = G4/Pool = 14.3
B = G4/3 = 13.5
C = G4/1 = 13.0
D = G4-5D = 7.5
E = G4-80 = 6.9
FIX
Standard
A
B
C
D
E
10
100
Factor IX Concentration (ng/ml)

Figure 41: ELISA IV

The X and Y axes are similar to those in figure 37. OD (490) readings obtained for the conditioned media (summarised in table 8) from cells transfected with K10RB-FIX-G4 (pool and individual clones) and K10RB-FIX (pool) are plotted on the factor IX standard curve. The reading obtained from a pooled plasma sample of Y-line transgenic mice is also plotted in this figure. The estimated concentration of factor IX (ng/ml) in the various samples are tabulated in the figure (inset).

Figure 41

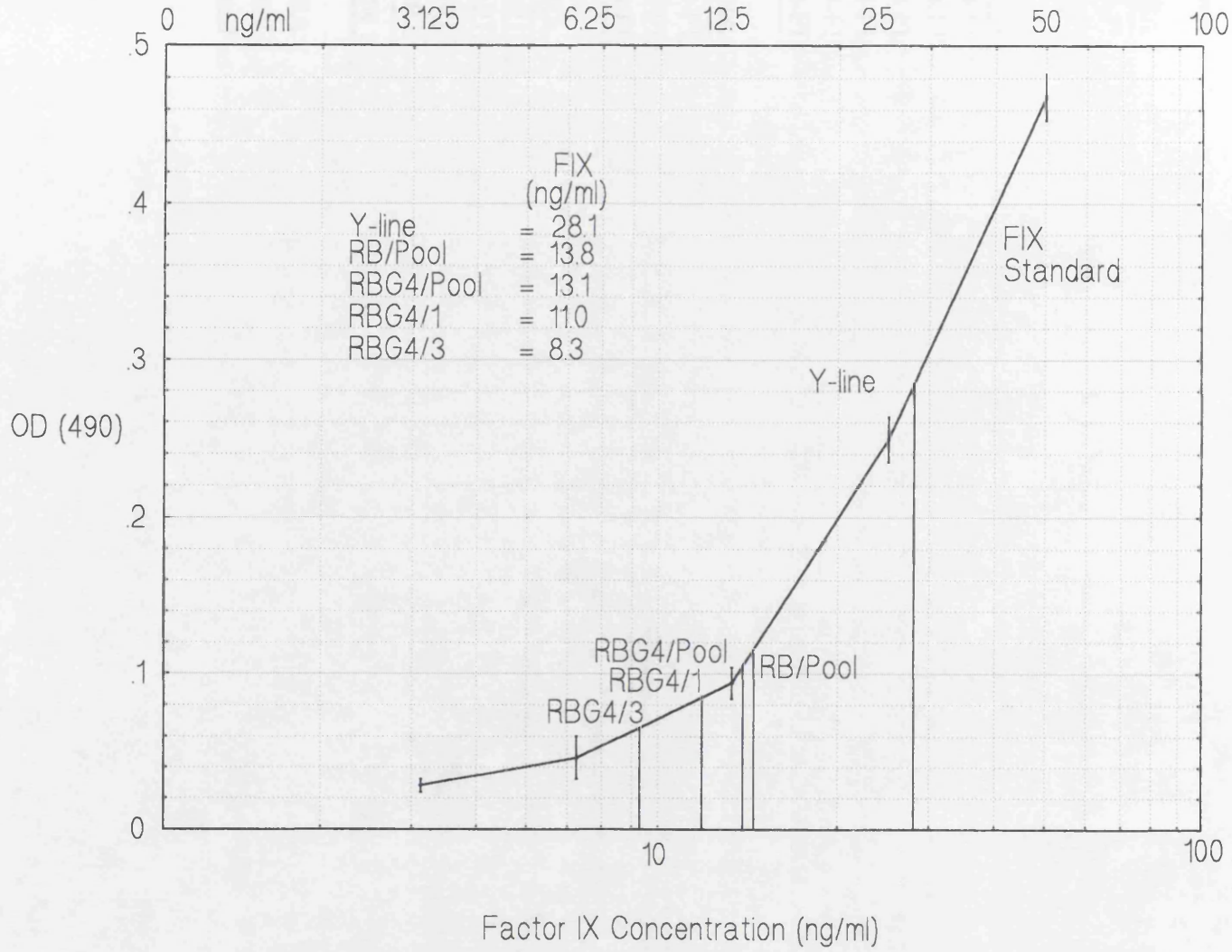

0 ng/ml 3.125 6.25 12.5 25 50 100
.5
.4
.3
.2
.1
0
OD (490)
FIX
(ng/ml)
Y-line = 28.1
RB/Pool = 13.8
RBG4/Pool = 13.1
RBG4/1 = 11.0
RBG4/3 = 8.3
FIX
Standard
Y-line
RBG4/Pool
RBG4/1
RB/Pool
RBG4/3
10
100
Factor IX Concentration (ng/ml)

Table 8: Summary of factor IX ELISA results

Factor IX construct (Clone / Pool)	OD(490) (±2SD)	Factor IX concentration ng/ml (±2SD)	Factor IX secretion ng/10^6 cells/24 hours (±2SD)
ELISA I:			
K10-FIX-II/Pool	0.137 (0.102-0.172)	9.04 (7.66-10.42)	18.1 (15.3-20.8)
K10-FIX-G4/Pool	0.337 (0.241-0.433)	17.30 (12.46-22.14)	34.6 (24.9-44.3)
ELISA II:			
K10-FIX-I/Pool	0.185 (0.159-0.211)	13.70 (10.18-17.22)	22.8 (16.9-28.7)
K10-FIX-I/clone 1	0.090 (0.058-0.122)	7.70 (5.60-9.80)	12.8 (9.3-16.4)
K10-FIX-I/clone 2	0.096 (0.062-0.130)	8.14 (6.16-10.12)	13.6 (10.3-16.9)
K10-FIX-II/clone 2	0.050 (0.026-0.074)	5.24 (<3.13-7.10)	8.7 (<5.2-11.8)
K10-FIX-II/clone 3	0.102 (0.062-0.142)	8.96 (6.48-11.44)	14.9 (10.8-19.1)
K10-FIX-II/clone 4	0.111 (0.081-0.141)	9.13 (7.29-10.97)	15.2 (12.2-18.3)
ELISA III:			
K10-FIX-G4/Pool	0.065 (0.045-0.085)	14.30 (12.34-16.26)	30.6 (26.4-34.9)
K10-FIX-G4/clone 1	0.050 (0.045-0.055)	13.00 (12.18-13.82)	27.9 (26.1-29.6)
K10-FIX-G4/clone 3	0.058 (0.045-0.071)	13.50 (12.20-14.80)	28.9 (26.1-31.7)
ELISA IV:			
K10RB-FIX/Pool	0.110 (0.092-0.128)	13.80 (12.76-14.84)	29.6 (27.3-31.8)
K10RB-FIX-G4/Pool	0.100 (0.073-0.128)	13.10 (12.12-14.08)	28.1 (25.9-30.2)
K10RB-FIX-G4/clone 1	0.081 (0.060-0.102)	11.00 (7.60-14.40)	23.6 (16.3-30.9)
K10RB-FIX-G4/clone 3	0.060 (0.051-0.069)	8.30 (6.60-10.00)	17.8 (14.1-21.5)
Y-line transgenic (plasma)	0.286 (0.254-0.318)	28.1 (22.14-34.06)	N/A

ELISA I to IV represent the readings in figures 38 to 41, respectively. Numbers within parentheses represent variations by ±2 standard deviations. Y-line transgenic mice were created using the K10-FIX-I gene construct.

Figure 42: Comparative analysis of factor IX expression by the various constructs

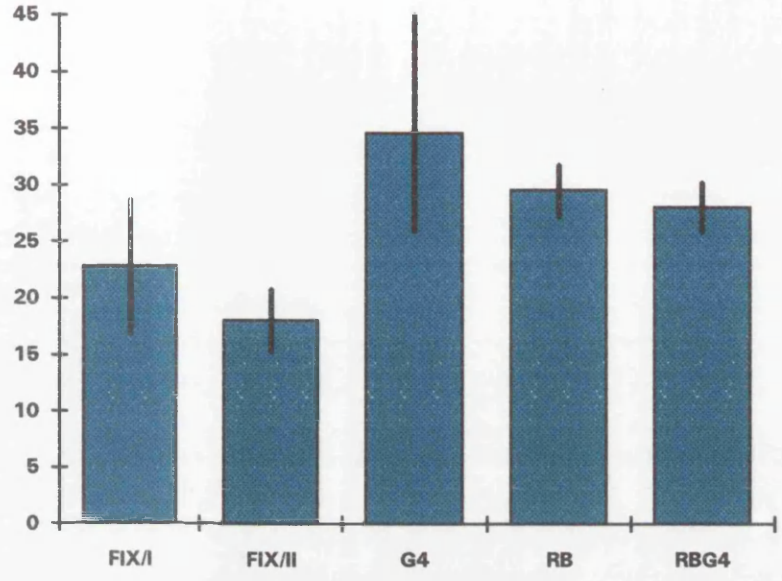
The relative expression of factor IX resulting from the pools of HaCaT cells transfected with the various gene constructs is shown. FIX/I=K10-FIX-I; FIX/II=K10-FIX-II; G4=K10-FIX-G4; RB=K10RB-FIX; RBG4=K10RB-FIX-G4. The error bars depict variation of ±2 standard deviations.

Figure 43: PCR screening of tail-tip biopsies from potential transgenic mice

This figure shows the 1.4 kb PCR product (indicated by the arrow) obtained on screening DNA samples extracted from tail-tip biopsies of potential transgenic mice. L=1 kb ladder; Lane 1=positive control; Lanes 3, 9, 10 and 11 are DNA samples from mice with the transgene.

Factor IX Expression

ng/10^6 cells/24 hours



Transfected HaCaT Pools

F9A/B PCR

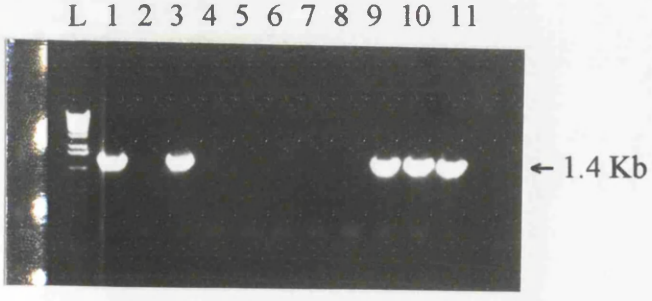


Table 9: Comparative analysis of factor IX expression

Factor IX construct	Factor IX secretion ng/10^6 cells/24 hours (±2SD)	Comparison with K10-FIX-II/Pool (*t* test)	Comparison with K10-FIX-G4/Pool (*t* test)
K10-FIX-I/Pool	22.8 (16.9-28.7)	$p < 0.1$	$p < 0.02$
K10-FIX-II/Pool	18.1 (15.3-20.8)	N/A	$p < 0.01$
K10-FIX-G4/Pool	34.6 (26.0-45.0)	$p < 0.01$	N/A
K10RB-FIX/Pool	29.6 (27.3-31.8)	$p < 0.01$	$p < 0.2$
K10RB-FIX-G4/Pool	28.1 (25.9-30.2)	$p < 0.01$	$p < 0.1$

The factor IX expression achieved by the transfected HaCaT pools of the various gene constructs are listed. The results of the *t* test of significance are shown in the third and fourth columns. The mean expression levels obtained for the basic construct (K10-FIX-II) and the highest expressing construct (K10-FIX-G4) were compared with those of the other constructs. N/A=not applicable.

3.8.2 ASSESSMENT OF SECRETED FACTOR IX IN SERUM FROM TRANSGENIC MICE:

Analysis of expression of the K10-factor IX constructs *in vivo* was performed in collaboration with Dr. Y. Alexander in our department. The construct K10-FIX-I was used for this purpose. The transcriptional unit (7.3 kb *Sal*I / *Nru*I fragment) was separated from plasmid sequences and gel-purified for microinjection (as described in section 2.5.2). Three lines of transgenic mice (X, Y and Z lines respectively) were created by Dr. Y. Alexander. Tail DNA preparations were used to assess the integration of the transgene. For this, PCR screening was performed using primers F9A and F9B (Figure 43). Dr. Alexander performed Southern blot analysis to confirm integration of the transgene. A ~2.5 kb recombinant factor IX transcript was also obtained from total cellular RNA extracted from transgenic mouse epidermis (and not from normal controls) and in situ hybridisation confirmed expression in suprabasal epidermal keratinocytes.

Pooled plasma samples from the Y-line transgenic mice were analysed for secreted factor IX using ELISA. Samples were obtained by cardiac puncture and supplied by Dr. Alexander. Figure 41 shows that the mean concentration of factor IX assayed from the plasma of Y-line mice was 28.1 ng/ml (22.1-34.1 ng/ml). These results demonstrate the efficacy of the basic construct to express γ-carboxylated factor IX in a keratinocyte-specific fashion and perhaps also the ability of keratinocytes to deliver it to the systemic circulation.

CHAPTER: 4

DISCUSSION

4.0 DISCUSSION

4.1 ANALYSIS OF MUTATIONS IN HAEMOPHILIA A

Haemophilia A is caused by a heterogeneous group of deleterious mutations in the factor VIII gene. Besides the mutational hotspot responsible for disease in ~20% of haemophilia A patients, the vast majority of mutations include single base substitutions. The identification of naturally occurring mutations in the factor VIII gene offers an insight into the biological properties of this large and complex protein. In a directed search for factor VIII mutations, eleven different and potentially pathogenic mutations were completely characterised in this study. These have been summarised in table 10. Over 150 mutations have now been reported in the factor VIII gene and the spectrum and distribution of these is relatively well characterised (Tuddenham et al., 1991). However, a relative deficit exists in the knowledge regarding structure-function correlates in factor VIII, and in this regard naturally occurring mutations provide a potential source of useful information. Analysis of the mutations characterised in this study and the sequence in their immediate vicinity, revealed interesting findings regarding phenotypic correlation and the mechanism of mutagenesis.

4.1.1 GENOTYPE-PHENOTYPE CORRELATION

For practical purposes the mutations detected in this study can be classified as being (i) disruptive i.e., predicted to cause either truncation or internal exclusion of amino acids from the mature factor VIII polypeptide and (ii) non-disruptive, which include all the missense mutations. In this study, 5 missense and 6 disruptive mutations (of which the intron 22-mediated inversion was one) were identified. While it is clear how the disruptive mutations and missense mutations affecting sites of established functional significance result in a haemophilic phenotype, it is more difficult to explain the role of the vast majority of missense mutations in disease causation. The main reason for this inability to attribute disease to missense mutations stems from the fact that not much is

Table 10: Summary of factor VIII mutations characterised in this study

Case	FVIII:C (IU/dl)	In	Exon/ Intron	Nucleotide No. [a]	Nucleotide change	Amino acid change [a]	FVIII domain	Effect on FVIII
GLA1	<1	NK	16	5557	GCC→ACC	Ala1834Thr [b]	A3	
GLA2	<1	NK	17	5663	AGA→ATA	Arg1869Ile [b]	A3	
GLA3	3-4	-	Intron 6	787 + 3	GTA→GTG [b,c]	Skipping of exons 5 and 6		62 a.a. deleted from the A1 domain
GLA4	2	+	14	2159	GGC→GAC	Gly701Asp [b]	A2	Multicopper oxidase I site mutation
GLA5	9	-	4	541	GTG→ATG	Val162Met [d]	A1	Casein kinase phosphorylation site mutation
GLA6	<1	+	17	5697	5697delC [b]		A3	Truncation (partial A3, C1 and C2)
GLA9	<1	-	1	43	CGA→TGA [f]	Arg-5Stop [e]	signal pep	Severe truncation (A1,A2,B,A3,C1,C2)
GLA11	1	-	16	5374-5586	Gross deletion of exon 16 [b]		A3	71 a.a. deleted from A3
GLA12	<1	NK	N/A		Intron 22 mRNA abnormality			Truncation (partial C1 and C2)
GLA13	3	-	2	225	GAT→GAA	Asp56Glu [b]	A1	
GLA14	<1	NK	4	519-523	519delTACCT [b]		A1	Severe truncation (A1,A2,B,A3,C1,C2)

In=Inhibitors; NK=not known; a.a.=amino acid; [a]Numbering according to Wood et al. (1984). [b]Novel mutations. [c]Identical splice mutation in two patients. [d]Mutation reported twice before (Tuddenham et al., 1991; Diamond et al., 1992). [e]Mutation reported twice before (Pattinson et al., 1990a; Reiner and Thompson et al., 1992). [f]Involves a hypermutable CpG dinucleotide.

known about structure-function correlates in factor VIII. These have been reviewed (Pittman and Kaufman, 1989) and the salient features are described in section 1.3.2.

In this section, a genotype-phenotype correlation is attempted for the two classes of mutations detected. By a prediction of the products of translation, the exact extent of the loss of amino acid residues or truncation of the factor VIII molecule is estimated. An attempt is also made to elucidate the pathogenicity of the missense mutations detected.

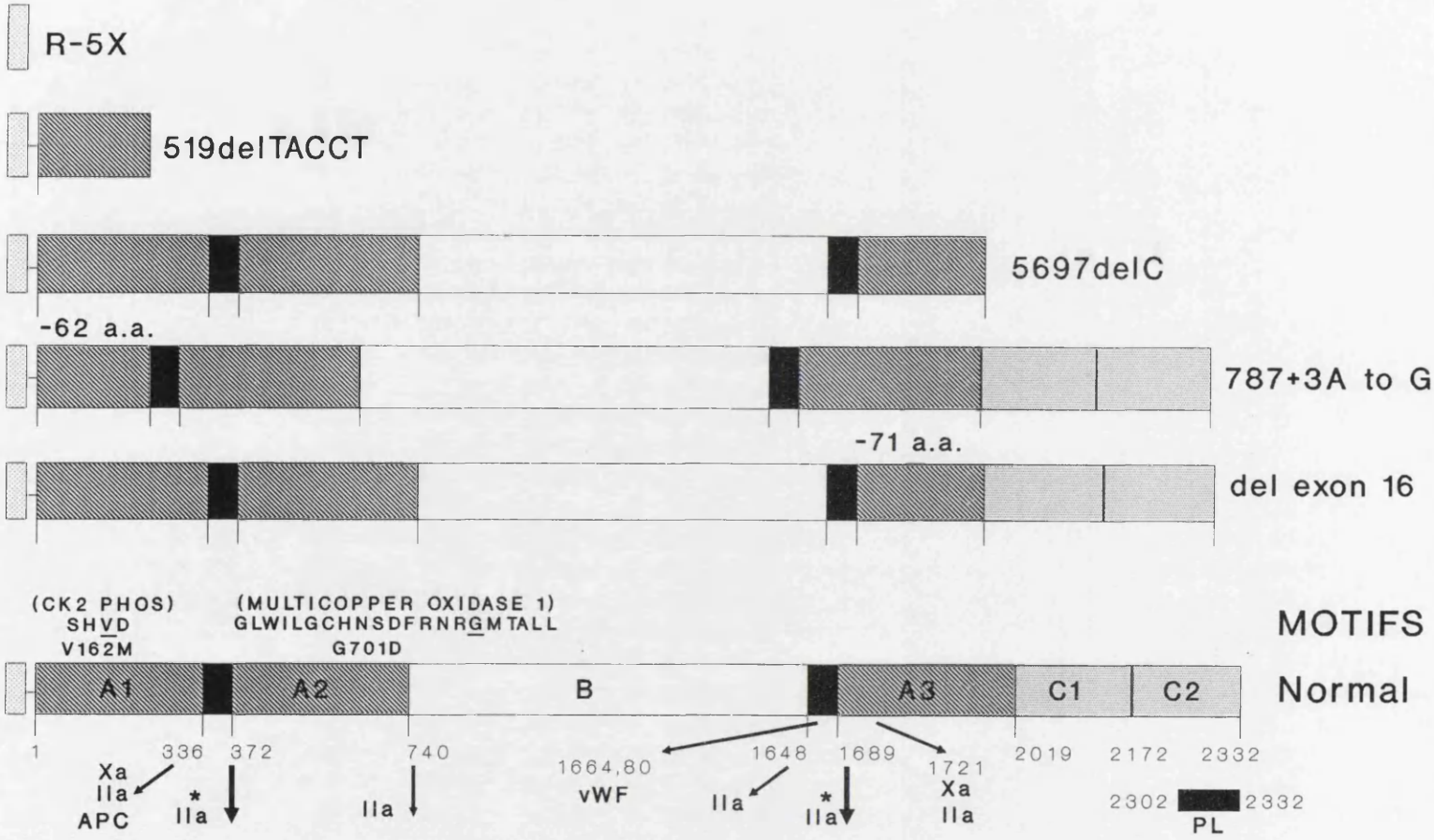
Prediction of the products of translation

In six of the mutations detected, there is predicted to be varying degrees of truncation or exclusion of amino acids from the product of translation of the factor VIII mRNA. These mutations include two single base substitutions (nonsense mutation: R-5X; intron 6 splice donor site mutation: 787+3A→G), three deletions (5697delC, 519delTACCT, deletion of exon 16) and the characteristic intron 22-mediated inversion. Figure 44 shows a diagrammatic representation of the normal and various abnormal products of translation for the disruptive mutations detected. The prediction of the products of translation was made assuming that the severely truncated products would be stable. It is however likely that the prematurely terminated products may be rapidly degraded in the cell (Goldberg and St. John, 1976) thus making some of these predictions purely theoretical. The normal factor VIII product of translation is shown with its characteristic domain structure, the sites of cleavage, vWF binding and phospholipid binding (see section 1.3.2).

The nonsense mutation at codon 15 (Arg) in the signal peptide (R-5X) is predicted to result in severe truncation and virtually no circulating normal factor VIII due to the termination of translation before the start of the mature factor VIII polypeptide. This mutation was detected in a 15 year old severe haemophilia A patient (GLA9). Of the seven point mutations, this was the only one that occurred in a hypermutable CpG dinucleotide and is likely to be due to a methylation mediated mutagenic event (Cooper and Youssoufian, 1988; Cooper and Krawczak, 1990). This is the third independent report of the same R-5X mutation (Pattinson et al., 1990a; Reiner and Thompson, 1992). This patient has not developed inhibitors to therapeutically administered factor VIII.

Figure 44: Diagrammatic representation of truncated products of translation due to factor VIII mutations

This figure shows a diagrammatic representation of the (internally) shortened or truncated products of translation caused by five mutations detected in this study. The causative mutation is indicated alongside each anomalous product. The normal domain structure of factor VIII and the various cleavage and binding sites are shown (numbering is according to Wood et al., 1984). The positive results of analysis by "MOTIFS" are also indicated at the appropriate location on the normal structure of factor VIII. The amino acid sequence of the putative functional domain detected by the computer program "MOTIFS", are indicated with the affected residue being underlined. APC=Activated protein C; vWF=von Willebrand factor; PL=phospholipid. The thrombin (IIa) cleavage sites marked by the asterisks are those that are essential for cofactor activation.


R-5X
519delTACCT
5697delC
-62 a.a.
787+3A to G
-71 a.a.
del exon 16
(CK2 PHOS)
SHVD
V162M
(MULTICOPPER OXIDASE 1)
GLWILGCHNSDFRNRGMTALL
G701D
MOTIFS
Normal
A1
A2
B
A3
C1
C2
1
336
372
740
1664,80
1648
1689
1721
2019
2172
2332
Xa
IIa
APC
*
IIa
IIa
vWF
IIa
*
IIa
Xa
IIa
2302
2332
PL

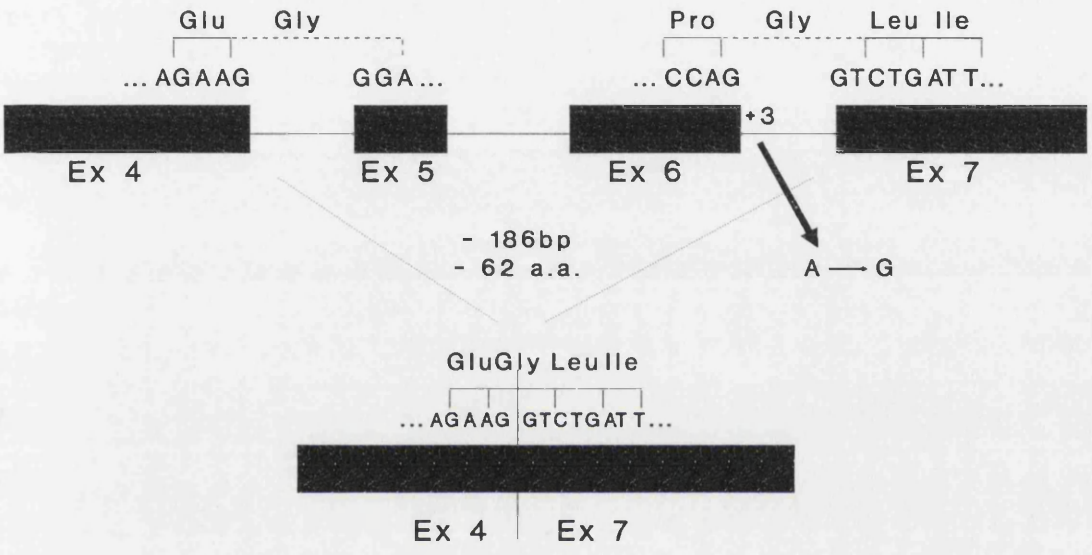
A single base substitution at position +3 in the splice donor site of intron 6 (787+3A→G), that caused a skipping of exons 5 and 6 in the mature factor VIII mRNA, was detected in two apparently unrelated haemophilia A patients (GLA3 and GLA8) in this series (section 3.2.3). Even though nested PCRs using primers specific for exons 5 and 6 could detect some percentage of full length transcript, it can be assumed from the RT-PCR result of segment R1 that >95% of the factor VIII mRNA produced is devoid of exon 5 and 6 sequence. This exon skipping event has left the translational frame intact and the resultant in-frame "deletion" of 186 bp in the mature mRNA can be predicted to cause a shortening of the mature factor VIII polypeptide by 62 amino acid residues (figure 45). This shortening will affect the conserved "A1" domain of the factor VIII polypeptide (Vehar et al., 1984).

Clinically, both patients with this mutation have disease of moderate severity and residual factor VIII activity of 3 IU/dl. It is interesting to note that a patient lacking exons 5 and 6 in the mature factor VIII mRNA due to a gross DNA deletion has previously been reported to have severe haemophilia A with FVIII:C <1 IU/dl (Naylor et al., 1993a). The relatively higher coagulant activity observed in both patients in the present study is probably due to the small component of correctly processed mRNA detected by nested PCRs with primers specific for exons 5 and 6. The above predicted alteration in the factor VIII protein does not result in a deletion of any known functional site (e.g. thrombin cleavage site), or conserved cysteine residues which are known to be important in protein folding and maintenance of the three-dimensional structure. This "in-frame" shortening of the conserved "A1" domain has therefore resulted in disease of moderate severity, in spite of a marked deficiency of normal sized transcripts from the mutant gene.

SSCP analysis and direct sequencing of exon 4 revealed a 5 base pair deletion of "TACCT" involving nucleotide positions 519 to 523 and amino acid positions 154, 155 and 156. The resultant shift in the reading frame is predicted to result in the translation of 25 altered amino acids followed by a premature stop codon at position 179. This is predicted to result in a severely truncated factor VIII polypeptide, 178 amino acids long, terminating approximately midway through the conserved "A1" domain (figure 44, Vehar

Figure 45: Diagrammatic representation of 787+3A to G resulting in skipping of exons 5 and 6

This figure shows a diagrammatic representation of the splice donor site mutation in intron 6 (787+3A to G) resulting in a skipping of exons 5 and 6 from the mature factor VIII mRNA. The codons at the junctions of exons 4, 5, 6 and 7 are shown. This demonstrates the "in-frame" exclusion of 62 codons (186 nucleotides) from the mature factor VIII mRNA. a.a.=amino acid; Ex=exon.


Glu Gly
... AGAAG
GGA ...
Pro Gly Leu Ile
... CCAG
GTCTGATT...
+3
Ex 4
Ex 5
Ex 6
Ex 7
- 186bp
- 62 a.a.
A → G
GluGly LeuIle
... AGAAG GTCTGATT...
Ex 4
Ex 7

et al., 1984). The patient is a 10 year old boy of Pakistani descent, with severe haemophilia A (FVIII:C = 1 IU/dl) and with no family history of the same.

The frameshift produced by the single base deletion in exon 17 (5697delC) is predicted to result in a severe truncation of the factor VIII polypeptide. The deletion of C-5697, which is the third base in codon 1880 (Phe), would result in the translation of 44 altered amino acids before terminating prematurely at a stop codon in the altered reading frame at position 1925. This truncated factor VIII of 1924 residues (instead of 2332) if stable, will be devoid of the carboxy half of "A3" and the two "C" domains which are thought to be essential for phospholipid binding (Arai et al, 1989b; Foster et al, 1990; figure 44). This mutation has resulted in a severe phenotype with FVIII:C <1 IU/dl and the patient has developed inhibitors to exogenous factor VIII (GLA6, table 10).

One patient (GLA11) showed a gross deletion that resulted in the exclusion of exon 16 from the mature factor VIII mRNA. This does not alter the translational reading frame and is predicted to result in a factor VIII polypeptide shorter by 71 amino acids (1773-1843) within the "A3" domain (figure 44). This mutation has resulted in a severe phenotype. The reason for this could be the exclusion of the highly conserved cysteine residue at position 1832 (see "pileUp" analysis in figure 20). The characteristic distribution of the cysteine residues at similar positions within related domains of factor VIII (Vehar et al, 1984), indicates that Cys-1832 plays a crucial role in maintaining the conformation of the post-translationally processed factor VIII by disulphide bridging.

The common inversion, mediated by the sequence of *F8A* in intron 22, causes a major disruption of the factor VIII gene (see figure 19) and results in the characteristic mRNA abnormality at the junction of exons 22 and 23 (Naylor et al, 1992; 1993a). This mutation always results in a severe phenotype and indeed was seen in one of the eight severe cases analysed in this study.

Pathogenicity of missense mutations

Missense mutations are readily associated with the disease phenotype when they alter an important functional residue, as is seen in the case of recurrent missense mutations affecting Arg-372 and Arg-1689, both of which are thrombin activation sites and thus critical for factor VIII function. However, such structure-function correlations

are somewhat limited at present and more information is usually necessary in the case of the vast majority of missense mutations. In general, the following information is helpful in determining the pathogenicity of missense mutations that do not clearly affect residues with well established function: (i) absence of any other change, (ii) a high level of evolutionary conservation, (iii) previous independent occurrence of the same mutation in an unrelated patient, (iv) failure to observe the same mutation in a sufficiently large sample of normal controls, (v) the novel appearance and subsequent cosegregation of the mutation with the disease phenotype in the family, and finally, more definitive evidence can be had from (vi) functional studies i.e., when recombinant techniques are used to show the production of a mutant protein from a mutant clone and correction of the defect is achieved by transfer of the normal clone in *in vitro* experiments.

A total of five missense mutations were characterised in this study. These include D56E, V162M, G701D, A1834T and R1869I. Four of these mutations were the only sequence alterations detected after screening the entire essential sequence of the factor VIII gene and are therefore likely to be the cause of haemophilia A in these patients. Two of these have resulted in severe disease (G701D, R1869I) and one each in a moderate (D56E) and mild (V162M) phenotype.

Three main attributes were used to assess the significance of the observed missense mutations in this study. These included (i) the degree of evolutionary sequence conservation of the mutated amino acid residue in both, the genes known to be homologous to factor VIII and conservation in factor VIII sequences from other mammalian species, (ii) involvement of protein sequence patterns (or motifs) that are linked with certain functions or that determine sites of post-translational modification, as catalogued in the PROSITE dictionary (Bairoch, 1993), and (iii) the assessment of the effect of a particular amino acid substitution, as judged from their chemical properties.

An in-depth study of the phenotypic effect of missense mutations in the factor IX gene was reported by Bottema et al. (1991). They noted that mutations at "generic" residues (i.e. those conserved in related homologous serine proteases and in the factor IX genes of other mammalian species) invariably resulted in disease and mutations in those residues that were only homologous to the other factor IX genes were about six-fold less

likely to cause disease. Mutations at non-conserved residues on the other hand were some 33-fold less likely to result in disease. They estimated that about 40% of all missense mutations would cause disease and the remaining 60% would only affect "spacer" regions. In the case of haemophilia B, patients are ascertained only if there is a significant loss of factor IX procoagulant activity (Sommer et al., 1992). Haemophilia A and B are clinically and genetically alike and, a similar degree of inactivation of the wild-type protein is required for the clinical ascertainment of the disease. If it were extrapolated from the case of the factor IX gene, determination of the evolutionary sequence conservation is likely to be a useful exercise by helping to rule out the affection of relatively unimportant "spacer" regions (Bottema et al., 1991). This analysis is especially important for missense mutations detected by a screening strategy that has a near total ascertainment of mutations (as evidenced from the three patients with two or more cleavage products in table 4).

All five missense mutations that were detected in this study occurred in the three conserved "A" domains of factor VIII. The "PileUp" program (in the GCG package of sequence analysis programs) was therefore used to align the three homologous "A" domains of human factor VIII (Vehar et al., 1984) with those of factor V (Kane and Davie, 1988), ceruloplasmin (Kochinsky et al., 1986) and murine factor VIII (Elder et al., 1993) (figure 20). This was done in order to examine the degree of evolutionary sequence conservation of the affected residues in the 12 repeated "A" domains of the above named proteins. The rationale being that, the more conserved the amino acid, the more likely it is to be important for maintaining the structural/functional integrity of factor VIII (Diamond et al., 1992).

Since 1988, a compilation of sites and patterns (or motifs) in protein sequences, that determine specific and common functions or sites of post-translational modification has been developed, called the PROSITE dictionary of sites and patterns in proteins (Bairoch, 1993). Release 10.1 (April, 1993) contained 803 different patterns and these are periodically reviewed and updated (Bairoch, 1993). The program "MOTIFS" (Smith et al., 1990) searches for sequence motifs by scanning through a given amino acid sequence for patterns catalogued and defined in the PROSITE dictionary. The five

missense mutations were thus analysed for their involvement with common functional domains, as mapped on the human factor VIII sequence.

The effect of (missense) mutations on protein stability and function has been reviewed (Pakula and Sauer, 1989; Alber, 1989) and it is seen that the phenotypic consequences of a given mutation depends not only on the nature of the amino acid substitution, but also on its location within the protein. Most mutations appear to affect activity of the protein indirectly via effects on protein structure or stability. These include for example, those that are buried in the protein structure (core residues), those involved in disulphide, hydrogen or electrostatic bonding and those amino acids with special structural properties such as glycine and proline. This can be appreciated from the observation of a wide variety of missense mutations resulting in haemophilia A, where only a fraction of these affect known sites of function (e.g. Arg-372 and Arg-1689). Due to the lack of knowledge regarding the structure of the factor VIII molecule, the position of the mutation in the polypeptide itself is not sufficient to predict the effect of the mutation. In the course of this discussion, even though comments are made regarding the alteration in chemical attributes (as a result of a missense mutation), the true significance of these cannot really be appreciated. Also as many residues could potentially tolerate changes better than others, a comprehensive cataloguing and analysis of natural mutations in the factor VIII gene (as in Tuddenham et al., 1991) will enable the delineation of such "spacer" regions (as was estimated in factor IX, Bottema et al., 1991).

The G701D mutation was the only sequence alteration detected in a patient with severe haemophilia A (GLA4). It was shown to affect a highly conserved residue, conserved in all but one of the 12 homologous "A" domains (F5A2, figure 20) analysed by "PileUp". Analysis using "MOTIFS" revealed that this mutation affected a multicopper oxidase I consensus pattern (Messerschmidt and Huber, 1990; Ouzounis and Sander, 1991; figure 44), spanning factor VIII residues 686 through to 706, with the Gly at 701 being an invariable determinant of this sequence pattern. There were three multicopper oxidase I signatures in the factor VIII amino acid sequence, detected by this method, with one occurring in each of the three "A" domains as follows, A1: 304-324, A2: 686-706 and A3: 1994-2014 (Vehar et al. [1984] used homology to plastocyanin, and

demonstrated putative ligands for type 1 copper at the same locations, but only in the A1 and A3 domains). These patterns are observed in a variety of proteins that are known or thought to have a copper binding function (laccase, ascorbate oxidase, ceruloplasmin and copper resistance protein A) (Messerschmidt and Huber, 1990) and may represent the domains responsible for a putative metal binding function of factor VIII (Vehar et al., 1984; Pittman and Kaufman, 1989). This missense mutation represents a change from a neutral to acidic amino acid. The amino acid glycine, which lacks a β-carbon can assume many backbone dihedral angles, not possible by other amino acids. Due to this, glycine is used in certain types of reverse turns where other amino acids are not suitable. Replacing glycine in such turns with any other amino acid would be expected to have a destabilising effect (Pakula and Sauer, 1989). Due to the lack of structural information for the factor VIII protein, it is however not possible to determine if the Gly-701 residue is used in this fashion. Missense mutations are not generally associated with the development of inhibitors to exogenously administered factor VIII. However, two examples of this have been reported, R2209Q and W2229C, and are believed to alter epitopes required for the state of immune tolerance (Tuddenham et al., 1991). G701D in patient GLA4 has led to the development of inhibitors (9.4 Bethesda units) to therapeutic administration of factor VIII.

The V162M mutation has been reported twice before and was noted to result in mild and moderate haemophilia A in these two cases (Tuddenham et al., 1991; Diamond et al., 1992). The patient in this study has a mild phenotype with FVIII:C = 9 IU/dl (GLA5, table 10). Analysis using "MOTIFS" revealed the involvement of a casein kinase II phosphorylation site (amino acid positions 160-163; figure 44). It is known that acidic and basic residues increase and decrease the substrate specificity for this serine/threonine kinase, respectively (Pinna, 1990). The significance of the mutation in this respect is therefore doubtful, because valine and methionine are both neutral amino acids. This mutation was however shown to involve a conserved residue (9 of 12 domains, figure 20) and was the only mutation found in this patient.

The R1869I mutation was at a much less conserved site, being seen in only the "A3" domains of human factor VIII, ceruloplasmin and murine factor VIII (figure 20).

R1869I represents a change in the class of amino acid from basic to neutral (and hydrophobic) and in view of this being the only change detected in this severe haemophiliac (GLA2), signifies the biological importance of this residue. Its conservation in the three above mentioned "A3" domains may indicate that it plays a specific role in that part of the molecule.

Lin et al. (1993) reported a polymorphic conservative change, D56V (GAT→ GTT at nucleotide 224), in the Chinese population. This change was also noted between the amino acid sequences reported by the first two groups who cloned the factor VIII cDNA (Wood et al., 1984; Toole et al., 1984). One of the missense mutations in the present study affected a different nucleotide position in the same codon (GAT→GAA at nucleotide 225) and resulted in the replacement of the normally occurring aspartic acid residue with glutamic acid (D56E). This mutation affects a very highly conserved residue, occurring in 11 of the 12 domains (except F5A1, which in fact has an "E" at this position, figure 20). A change from aspartic acid to glutamic acid would normally be considered as a conservative change. This mutation was, however, the only sequence change encountered in this patient (GLA13) with moderate haemophilia A (FVIII:C = 3 IU/dl).

The fifth missense mutation, A1834T, was one of two chemical cleavage results seen in the same patient (GLA1). The second mutation, as yet uncharacterised, has been mapped to exon 16 by a further CCM of two overlapping and nested PCRs. Analysis of the A1834T mutation by PileUp, interestingly revealed that only 4 of the 12 "A" domains normally had an alanine, whereas 7 of the other 8 actually had a threonine at this position (figure 20). This can be interpreted as not being a significant mutation, since the modified residue is in fact the residue of choice in the sister domains. In this case therefore, the uncharacterised mutation mapped to exon 16 may be the cause of the severe haemophilia A. An alternate view could be that an alanine at position 1834 in the "A3" domain of factor VIII may play a unique role, being crucial for its procoagulant function. There are known examples of conservative missense mutations that can result in a disease phenotype, if they occupy sites that are key determinants of stability or function (Pakula and Sauer, 1989). Furthermore, even though both are neutral amino acids, alanine is polar and threonine is hydrophobic.

Table 11: Analysis of the missense mutations

Mutation	Phenotype*	In	PileUp[a,b]	MOTIFS [a,c] (PROSITE)	Alteration in ionic charge
D56E [d]	moderate	-	11/12		acidic→acidic
V162M [d]	mild	-	9/12	CK-2 [h] phosphorylation site	neutral, hydrophobic→neutral, hydrophobic
G701D [d]	severe	+	11/12	multicopper oxidase I site	neutral, polar→acidic
A1834T [e]	severe	NK	4/12 [f]		neutral, hydrophobic→neutral, polar
R1869I [d]	severe	NK	3/12 [g]		basic→neutral, hydrophobic

*See table 10 for FVIII:C levels; In=inhibitors to therapeutically administered factor VIII; NK=not known; [a]"PileUp" and "MOTIFS" are sequence analysis programs in the GCG package; [b]"PileUp" analysis was performed in order to assess sequence conservation among the homologous "A" domains of human and murine factor VIII, factor V and ceruloplasmin (12 domains in total); [c]"MOTIFS" was used to scan the factor VIII amino acid sequence for consensus sequences catalogued in the PROSITE dictionary of protein sites and patterns (Release 10.1; Bairoch, 1993); [d]These mutations were the only sequence alterations found in these four patients (GLA13, 5, 4 and 2); [e]This mutation was one of two chemical cleavage products characterised in patient GLA1; [f]The other 7/12 domains had a Threonine (T) residue at this position. [g]Only the "A3" domains of human factor VIII, murine factor VIII and ceruloplasmin had an arginine (R) residue at this position. [h]Casein kinase II.

The results of the analysis of phenotypic effects of missense mutations detected in this study are summarised in table 11.

4.1.2 MECHANISM OF MUTAGENESIS

The frequency and distribution of mutations in eukaryotic genomes is non-random. The reason for this is largely due to the strong influence of the local sequence environment in the pathogenesis of mutational lesions (Cooper and Krawczak, 1993). Single base substitutions are known to result either from chemical (deamination of 5-methylcytosine, Coulondre et al., 1978 and depurination, Loeb and Preston, 1986), physical (e.g. DNA slippage, Kunkel and Soni, 1988) or enzymatic (e.g. post-replicative mismatch repair or exonucleolytic proof-reading, Modrich, 1987; Loeb and Kunkel, 1982) mechanisms, all of which are known to be sequence-dependent (Cooper and Krawczak, 1993). Human gene deletions are also distributed non-randomly, and this is evident from their dissimilarity in frequency and distribution in X-linked recessive conditions of similar incidence. For example, 2.5% of haemophilia A patients have deletions of the factor VIII gene (Millar et al., 1990), whereas 84% of patients with steroid sulphatase deficiency have deletions (Ballabio et al., 1989). Moreover, hotspots have been demonstrated for deletion breakpoints in several genes, example dystrophin (Forrest et al., 1987, 1988; Dunnen et al., 1987) and low density lipoprotein receptor (Langlois et al., 1988). Although, Krawczak et al. (1992) demonstrated that single base substitutions affecting splice sites did not represent a non-random occurrence, the distribution of mutations within the splice sites was shown to have a definite non-random distribution. Perhaps the most obvious examples of non-randomness of mutational events in the human genome are the recently discovered hotspots namely, the *F8A*-mediated inversion in about half the severe cases of haemophilia A (Naylor et al., 1993a, b; Lakich et al., 1993), the CMT1A-REP-mediated 1.5 Mb duplication in 17p11.2 responsible for over 90% of Charcot-Marie-Tooth disease type 1A (Lupski et al., 1991; Pentao et al., 1992; Wise et al., 1993) and the trinucleotide repeat expansions responsible for most cases of fragile X syndrome (Fu et al., 1991; Verkerk et al., 1991; Oberle et al., 1991; Yu et al., 1991) and myotonic dystrophy (Harley et al., 1992; Buxton et al., 1992).

In a recent report, it was shown that the common intron 22-mediated inversion almost exclusively occurred during male meiosis (Rossiter et al., 1994). The patient in this study is the only affected male in his family, but his mother's twin sister (monozygotic) showed an abnormal coagulation profile, indicating her carrier status (C. Shiach, personal communication). This finding is consistent with those of Rossiter et al. (1994), who found that the mothers of all but one of 50 sporadic inversion cases in their study were carriers.

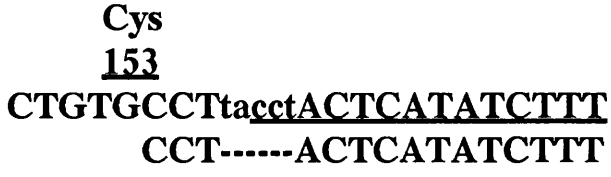
The small deletions, splice donor mutation and other single base substitutions detected in this study were evaluated for possible mechanistic explanations for the mutagenesis. The nucleotide sequence at and in the vicinity of the mutations was analysed for possible clues, in order to answer the question(s), "why and how did a particular mutation arise in that particular site?" Satisfactory answers were not available for all the mutations, but some interesting findings were observed in many.

The small deletions, 519delTACCT AND 5697delC

Analysis of the sequence in the vicinity of the 5 bp deletion in exon 4 (519delTACCT) revealed the presence of direct repeats of a 3 bp motif "CCT", one copy immediately preceding the deleted bases and the other as part of the deleted sequence (figure 46). The modified slipped-mispairing model for the generation of deletions at the DNA replication fork, as proposed by Krawczak and Cooper (1991), relies on there being homology between the sequence flanking a deletion and the neighbouring DNA sequence. As seen in figure 46, the underlined sequence would mispair with its displaced homologue (shown in the bottom line) if the deleted bases (seen as a hyphenated line under the lower case letters) were permitted to loop out during DNA replication and thereby get excised by DNA repair enzymes. This deletion is therefore concordant with the modified slipped-mispairing model proposed by Krawczak and Cooper (1991).

The sequence surrounding the deleted C-5697 (in exon 17) was similarly analysed and showed no evidence of direct or inverted repeats or the presence of quasi-palindromic and symmetric sequence elements that could be implicated in the mechanism of this deletional mutagenesis (Krawczak and Cooper, 1991).

Figure 46: "Slipped mispairing at DNA replication" in the pathogenesis of 519delTACCT



Analysis of the sequence in the vicinity of the 5 base pair deletion in exon 4. The normal DNA sequence flanking the deletion is shown in the top line with the deleted bases being indicated by lower case letters. The underlined sequence may mispair with its displaced homologue, shown in the lower line, if the deleted bases indicated by the hyphenated line (under the lower case letters) were to loop out at DNA replication. The position in the factor VIII coding sequence is denoted by the cysteine residue at 153 (TGC), which is the last normal amino acid residue encoded by the deleted allele.

The intron 6 splice donor mutation

The phenotypic effect of mutations affecting splice junctions include either exon skipping or activation of cryptic splice sites in the vicinity, with the former being more common. Mutations affecting splice donor sites almost invariably cause skipping of the preceding exon. In a recent review of 101 splice mutations (Krawczak et al., 1992) it was observed that 60% of 5' splice site mutations involve the invariant GT dinucleotide at positions +1 and +2. Many examples of 5' splice mutations have been reported that also affect positions -3, -1, +3, +5, +6 and +7 (Krawczak et al., 1992; Cooper and Krawczak, 1993). Of these, it was observed that positions -1 and +5 are most commonly affected which is perhaps a reflection of these being more highly conserved positions and their strong base-pairing with complementary region of U1 snRNA (Krawczak et al., 1992).

A number of mutations putatively affecting factor VIII mRNA splicing have been reported (Tuddenham et al., 1991; Higuchi et al., 1991a; Krawczak et al., 1992), but confirmation of their phenotypic effects have so far been difficult due to its hepatocyte-specific expression. A splice acceptor mutation producing exon skipping was detected using RT-PCR of the ectopically transcribed factor VIII mRNA from peripheral lymphocytes (Naylor et al., 1991). This method is now being increasingly used, especially for large genes transcribed in relatively inaccessible tissues and will result in many more splice mutations being reported with the consequent better understanding of splicing errors and their role in disease.

In this study, a single base substitution (A→G) was discovered at position +3 in the 5' splice site of intron 6, that caused a skipping of exons 5 and 6 in the mature factor VIII mRNA, in two apparently unrelated individuals. The following three questions were asked: (i) Is this splice donor mutation (at +3) the cause for the observed error of splicing? (ii) Why did exon 5 (with normal splice determinants) get skipped from the mature factor VIII transcript, along with exon 6 (with the splice mutation)? and (iii) Does the occurrence of the same mutation in two apparently unrelated individuals represent a true case of recurrence?

In the 101 splice mutations reviewed by Krawczak et al. (1992), only three examples of mutations affecting position +3 of the splice donor site were detected. This is

not surprising as the corresponding residue in U1 snRNA is a pseudouridine rather than a cytosine. The same reason accounts for no mutations being found at position +4 although the first case of this has recently been reported in the gene for neurofibromatosis type 1 (Hutter et al., 1994). In order to rule out the possibility of this mutation being a neutral polymorphism, a total of 24 other haemophilia A patients and 60 normal controls were screened for its presence. As no restriction enzyme site was abolished or created by this base substitution, allele-specific PCRs were designed for the normal and mutant alleles (Wu et al., 1989; Sarkar et al., 1990). None of those screened showed this mutation. Furthermore, consensus values (CV) were calculated for the splice donor sites of introns 4, 5, 6 and mutant 6, according to the method described by Shapiro and Senapathy (1987). These values were estimated on the basis of the conservation of bases at the 8 positions most likely to contribute 5' splice recognition determinants (position -2 through to +6) and were used as relative indicators of the efficacy of splice sites. As seen in table 3, the single base substitution at +3 results in a 9% reduction of the consensus value of the mutant intron 6 splice donor site compared to that of the normal sequence. These observations go in favour of this sequence alteration being responsible for the observed splicing error.

Direct sequencing of the major transcript showed that exon 5 was consistently skipped along with exon 6 in the mature factor VIII mRNA. Why this was so is not entirely clear, but may be due to one or a combination of the following explanations: (a) Previous examples of splice donor mutations have caused exon skipping of only the preceding exon, which is consistent with the exon definition model of splice site recognition (Robberson et al., 1990). Examples of multiple combinations of skipped adjacent exons have also been seen (albeit as part of a spectrum of associated minor transcripts) which have been attributed to secondary effects on mRNA processing (Naylor et al., 1991; Fisher et al., 1993). A consistently skipped exon 5 in the mature factor VIII transcript could have been due to these secondary changes even in the presence of normal splice recognition signals. (b, c) The average size of vertebrate internal exons is 137 bp (Hawkins, 1988) and experiments involving the shortening of internal constitutive exons (i.e. exons not normally alternatively spliced) to <51 bp have

resulted in skipping of the exon (Dominski and Kole, 1991). These observations suggest that the extreme proximity of the 3' and 5' splice sites impairs their recognition. Furthermore, it was observed by Sterner and Berget (1993) that splicing of exon 3 of the chicken fast skeletal troponin I gene (sTNI), only 7 bp in length, was dependent on the presence of the naturally occurring upstream exon 2. In the absence of exon 2 upstream from it, only an internal expansion of this "mini" exon 3 permitted its efficient splicing. Exon 5 is only 69 bp and is the smallest exon in the factor VIII gene (Gitschier et al., 1984). Its splicing may be similarly interlinked with that of its neighbouring exons (exon 6). (d) Estimation of the relative efficiency of the 5' splice sites (of introns 4,5,6 and mutant 6) by the method of Shapiro and Senapathy (1987) also gave some clues. Table 3 shows that the 5' splice site of intron 4 has a score of 100, whereas that of intron 5 was 90.84. This, along with the unusual finding of a "T" at +3 in the splice donor sequence of intron 5 (seen in only 2% of all primates; Shapiro and Senapathy, 1987) may have also contributed to the relative preference of the 5' splice site of intron 4 over that of intron 5, resulting in the additional skipping of exon 5 (figure 45).

So far, only three examples of established recurrence of a splice site mutation have been reported (Dunn et al., 1989; Matsuhita et al., 1989; Weil et al., 1990; Vasan et al., 1991; Chen et al., 1991; Krawczak et al., 1992). Two other cases of recurrence were inferred on the basis of RFLP haplotyping (Atweh et al., 1987; Wong et al., 1989). In the present study, two patients who were apparently unrelated on the basis of a careful four generation family history, were found to have identical mutations. Evidence was sought to prove that at least one of the two patients was a true sporadic case, thus indicating an independent origin for the two mutations. One of the patients (GLA3) was the only affected person in his family (GLA8 has an affected brother and two affected maternal uncles and so his mother is an obligate carrier). In order to determine if GLA3 was indeed a sporadic case, his mother (MP1) was similarly analysed. She was also shown to be a carrier for the same mutation. On performing a haplotype analysis using three intragenic factor VIII RFLPs and the closely linked, highly polymorphic St14 site (DXS52) (figure 47), it was not possible to rule out relatedness. Also this base substitution does not affect a hypermutable CpG dinucleotide as in one of the previously reported cases of a

Figure 47: Haplotype analysis of the two patients with the identical 787+3A to G mutation

This figure shows the result obtained for the PCR analysis of the highly polymorphic VNTR, St14 (DXS52), in the two patients with the same mutation (P1=GLA3; P2=GLA8). Both P1 and P2 have the same allele. L=1 kb ladder; C=control.

Table 12: Analysis of the sequence in the vicinity of the single base substitutions detected in this study

The table lists the results of the systematic analysis of the sequence in the immediate vicinity of the single base substitutions detected in this study (except 787+3A to G). The three mutations marked with an asterisk are those that fit the definition of a mutational hotspot (section 4.1.2). The normal base at the mutation site is indicated by bold typeface and the mutant base is indicated above it. The trinucleotide "CTT" topoisomerase I cleavage site was observed in the vicinity of each of the six mutations and are underlined and indicated by vertical arrows. Direct repeat elements or other sequence motifs (for example, "TGGA" which is the putative arrest site for DNA polymerase α, was noted in immediately flanking the V162M mutation site) are underlined. The site of the D56V polymorphism (which is immediately adjacent to the D56E mutation detected in this study) is also indicated.

Figure 47

St14 (DXS52)

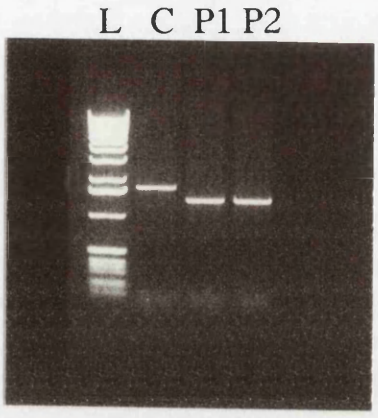


Table 12: Analysis of the sequence in the vicinity of single base substitutions

MUTATION	SURROUNDING SEQUENCE	Sequence features
R-5X *	↓ T ↓ 5' TGTGCCTTTTGCGATTCTGCTTTAGT 3' 3' ACACGGAAAACGCTAAGACGAAATCA 5'	"CpG" site, "CTT" site++
D56E *	TA ↓ 5' AGAATTCACGGATCACCTTTTCAAC 3' 3' TCTTAAGTGCCTAGTGGAAAAGTTG 5'	"TCAC" repeat "CTT" site+
V162M *	↓ A 5' TATCTTTCTCATGTGGACCTGGTAAAA 3' 3' ATAGAAAGAGTACACCTGGACCATTTT 5'	"TGGA" motif "CTT" site+
G701D	A ↓ 5' TTCGGAACAGAGGCATGACCGCCTTA 3' 3' AAGCCTTGTCTCCGTACTGGCGGAAT 5'	"CTT" site+
A1834T	↓A ↓ 5' TTGACTGCAAAGCCTGGGCTTATTTC 3' 3' AACTGACGTTTCGGACCCGAATAAAG 5'	"CTT" site++
R1869I	T ↓ 5' CTGCTCATGGGAGACAAGTGACAGTA 3' 3' GACGAGTACCCTCTGTTCACTGTCAT 5'	"GA" repeat "CTT" site+

recurrence (Dunn et al., 1989). The possibility of this mutation, in two patients in this study, representing a true recurrence, seems unlikely.

The single base pair substitutions

Single base-pair substitutions by far constitute the most common form of mutagenesis in the factor VIII gene. The most common and perhaps the best understood mechanism for the creation of single base substitutions is due to deamination of the hypermutable 5-methylcytosine (5mC) at CpG dinucleotides. Consequently the most common spontaneous single base substitutions observed include C→T and G→A at CpG dinucleotides. 38% of the point mutations reported in the database of mutations in haemophilia A represent these changes (Tuddenham et al., 1991). Cooper and Krawczak (1993) observed a frequency of methylation-mediated substitutions at CpG dinucleotides in 32.8% of a sample of 880 single base substitutions. This figure is similar to the study of factor VIII mutations by Higuchi et al. (1991a,b), who observed a frequency of 32%. The higher figure noted in the database represents (at least) in part the result of the deliberate search for mutations at CpG sites (Gitschier et al., 1985c, 1988; Pattinson et al., 1990a). Due to their hypermutability, these represent the most common cause for independent origins of the same point mutation in unrelated individuals. At least 16 of the CpG sites in the factor VIII gene have been observed in recurrent mutations (Tuddenham et al., 1991). In the present study, a nonsense mutation was detected at a CpG dinucleotide at codon -5 in the signal peptide (not one of the above 16 sites), coding for an arginine residue (CGA →TGA; R-5X). This mutation has been reported twice before (Pattinson et al., 1990a; Reiner and Thompson, 1992). In the absence of a haplotype analysis, it will be difficult to assess if this is truly a recurrence at this site.

Cooper and Krawczak (1990, 1993) have analysed an unbiased sample of 880 point mutations (including all such mutations reported up until May 1992 and excluding mutations reported either in the factor IX gene, as recurrences, at splice sites, in oncogenes or in the mitochondrial genome) and noted that 62% occurred in dinucleotides other than "CG". It was observed that a definite non-randomness existed in their distribution with respect to DNA sequence and several possible mechanisms were reviewed (Cooper and Krawczak, 1993). In the present study, 5 non-CpG site

substitutions were detected (excluding the splice site mutation). The sequence surrounding these mutations was similarly analysed. Table 12 lists all these mutations (including the R-5X mutation at a CpG site) along with a limited sequence flanking the site of the substitution.

In their analysis of single base substitutions, Cooper and Krawczak (1993) have defined as hotspots, those positions or codons that have been mutated more than once. Three of the mutations detected in the present study (R-5X, D56E, V162M) fit this definition (table 12). The R-5X mutation represents a recurrence at a CpG site. The V162M mutation has previously been reported to have recurred in two other unrelated individuals (Tuddenham et al., 1991; Diamond et al., 1992). The D56E mutation affects an adjacent position (nucleotide 224) in the same codon (Asp-56) that has a polymorphic site (at position 225, D56V; Lin et al., 1993).

Depurination is another important chemical mechanism for the causation of single base substitutions. Loeb and Preston (1986) showed that dG is the most commonly depurinated base. If depurination were really an important cause of point mutations then it would be expected that there should be an excess of G→A transitions (after excluding those at CpG sites). In fact of 591 non-CpG site mutations, 111 were G→A substitutions (Cooper and Krawczak, 1993). As seen in table 12, three of the five non-CpG mutations detected in this study were G→A transitions.

On investigating non-CpG point mutational hotspots (as defined previously), Cooper and Krawczak (1990, 1993) observed that certain types of sequences were seen overlapping or in the immediate vicinity of the substituted base. These included direct repeats, inverted repeats, a trinucleotide motif "CTT" which is a topoisomerase I cleavage site (Bullock et al., 1985) and two tetranucleotide elements "TCGA" and "TGGA". The latter motif also fits perfectly with the deletion hotspot consensus sequence previously described for human genes (Krawczak and Cooper, 1991) which is also the putative arrest site for DNA polymerase α, one of the major polymerases required for the replication and repair of vertebrate chromosomal DNA. These observations indicate that sites involved in the breakage and rejoining of DNA sequences are potential sites for single base substitutions.

As seen in table 12, analysis of the sequence in the immediate vicinity of the single base substitutions showed evidence of some of the possible mechanisms for the causation of these mutations as described by Cooper and Krawczak (1993). This study is the third independent report of the V162M missense mutation and it therefore fits the definition of a hotspot for point mutations. Sequence analysis surrounding this non-CpG site mutation revealed a tetranucleotide motif "TGGA" immediately flanking the G→A at position 541, which is known to be a putative arrest site (during DNA replication) for DNA polymerase α. In addition, the trinucleotide topoisomerase I cleavage site consensus sequence "CTT", was also noted in the immediate vicinity of the V162M mutation site. A tetranucleotide direct repeat "TCAC" was seen immediately flanking the base change responsible for D56E (and overlapping the polymorphic base change causing D56V) that may have mediated a four-base slippage producing the misincorporation error(s). A similar example has been noted for two missense mutations affecting codon 1922 of the factor VIII gene, which involve a direct repeat motif "CAAT" (Traystman et al., 1990; Higuchi et al., 1991b). The trinucleotide topoisomerase I cleavage site sequence "CTT", was also noted in the immediate vicinity of the D56E mutation. The R-5X and A1834 mutation sites were both flanked on either side by a "CTT" motif. Another such sequence was observed on the opposite strand as the R1869I mutation, which also occurred within a direct dinucleotide "GA" repeat sequence.

These observations do not provide enough evidence to explain the possible mechanisms of mutagenesis at these nucleotides. However, judging from the systematic analyses performed by Cooper and Krawczak (1993), it seems likely that these sequence elements observed at the sites of the mutations in this study, play at least some (if not a major) role in determining the choice of the mutated nucleotide. Perhaps, these sequence motifs could even be the reason for the two non-CpG hotspots, V162M and D56E.

4.1.3 EVALUATION OF THE STRATEGY FOR MUTATION DETECTION

The majority of cases of haemophilia A result from a variety of single base substitutions, distributed across the large coding sequence of the factor VIII gene. At the time of the design of this project, the common intron 22-mediated inversion had not yet

been discovered, but it was known that up to half the severe haemophilia A cases did not possess mutations in the regions screened (or by the method used) by Higuchi et al. (1991b). Any mutation screening strategy designed for the effective screening of the factor VIII gene, therefore had to screen the entire essential sequence of the gene, and be sensitive enough to enable the efficient detection of single base substitutions. The method used in this study was that described by Naylor et al. (1991) which is a rapid, simple and more cost-effective method of amplifying the entire essential sequence of the factor VIII gene and exploits the presence of ectopic transcripts in peripheral lymphocytes.

The choice of the mutation detection method was determined by the size of the PCR fragments to be screened. CCM analysis has been successfully used for segments up to 1.7 kb in length (Grompe, 1993) and was used for the screening of the entire coding sequence, which was amplified in segments >1 kb in length. Individual exons that were screened for mutations ranged in size from 151 bp (for exon 12) to 262 bp (for exon 8), and were all amplified as PCR products of less than 400 bp. SSCP analysis was used to screen these exons for mutations.

The results summarised in table 4, show that the strategy of RT-PCR / DNA-PCR followed by CCM analysis, was successful in detecting at least one mutation in each of the 13 patients thus screened, and that three of them showed more than one mutation each. Potentially pathogenic mutations were fully characterised in 11 of these 13 patients. It was also possible to detect an error of splicing, a gross deletion of an exon and the characteristic intron 22 mRNA abnormality (Naylor et al., 1992, 1993a). This strategy for mutation screening was therefore found to be very successful and it is assumed that a near total ascertainment of mutations was possible.

The estimation of the efficiency of mutation detection for the strategy involving the PCR-SSCP of individual exons was not as straightforward. Since each of the six exons were successfully amplified in each of the 12 patients, this ruled out the presence of any gross abnormalities involving these exons. Altogether, the six exons comprised 1276 bp or 17.9% of the factor VIII coding sequence. If it is assumed that (i) the distribution of small alterations in the factor VIII gene is random, (ii) gross rearrangements occur at a rate of just under 6% (Antonarakis and Kazazian, 1988) and (iii) the *F8A*-mediated DNA

inversion accounts for 20% of all haemophilia A patients (Lakich et al., 1993; Naylor et al., 1993a,b), then identification of mutations would be expected in 24% of the cases analysed. Using this method, a reproducible SSCP mobility shift was detected in two of twelve patients. This represents a detection rate of 17% (2/12) or a 70% mutation screening efficiency using SSCP in this sample. This observed efficiency is likely to be an overestimate due to (i) the observed non-random distribution of mutations involving these exons, (ii) a higher detection rate for small deletions/insertions (one of the mutations thus detected was a 5 bp deletion in exon 4) and (iii) the small number of cases in this study. It was also demonstrated that in order to improve the efficiency of SSCP analysis more than one gel electrophoretic condition was necessary.

In summary, even though CCM analysis is a lengthy and potentially hazardous procedure, its high sensitivity (>95%, and so its capacity to exclude the presence of mutations in regions found negative; Grompe, 1993), accuracy of localisation and ability to screen large segments, make it an ideal screening method for the factor VIII gene. SSCP on the other hand is a rapid, easy and reasonably sensitive screening method, but with limited utility in screening a gene like that of factor VIII.

If a mutation screening protocol were to be designed now, it would be appropriate to initially screen every severe haemophilia A patient for the presence of the common inversion. This can be performed by a Southern blot analysis (Goodeve et al., 1994) or by RT-PCR (Naylor et al., 1992, 1993a; see results). The former method requires only a DNA sample and is more definitive. It is also more informative as it is possible to directly diagnose female carriers and to tell whether the *F8A* gene in intron 22 recombines with the proximal or distal homologous sequence 5' to the factor VIII gene.

Analysis of the sequence in the vicinity of mutations characterised in this study, revealed the presence of several motifs that could explain the choice of the mutation site and perhaps also play a role in determining recurrences at some of these sites. These findings are in accordance with the observations made by Cooper and Krawczak (1993) regarding the role of the local DNA sequence in the pathogenesis of mutations.

In cases like that of the factor VIII gene, it is beneficial to catalogue all the naturally occuring mutations reported. This is necessary because of the difficulties

associated with detection of mutations in this large gene, and the consequent reporting of only few mutations by individual groups. A comprehensive analysis of these mutations will serve to ultimately characterise all sites that are functionally/structurally important and those sites that can tolerate changes better than others. In countries like the UK, where the health service is offered and controlled by one (or a few) central governing body(ies), mutations in the majority of known haemophilia A patients can be characterised and held in a database to enable a rapid and sensitive diagnostic genetic testing of relatives.

4.2 KERATINOCYTE-SPECIFIC EXPRESSION OF FACTOR IX FOR GENE THERAPY FOR HAEMOPHILIA B

The second objective of this project was to investigate the role of keratinocytes as a delivery system for ectopically synthesised proteins. A cytokeratin gene promoter was used to achieve keratinocyte-specific expression, to result in tissue-specific synthesis of a heterologous protein (section 1.10.3). In this project, the role of keratinocyte-specific expression of clotting factor IX was investigated as a potential method to deliver somatic gene therapy for haemophilia B. The reasons for gene therapy for haemophilia and the advantages of using keratinocytes as the target cell type have been discussed (sections 1.9.1 and 1.10.1). The reason for selecting a cellular promoter was to eventually achieve long-term expression *in vivo* (section 1.10.2). The HaCaT cell line was selected as the *in vitro* model for gene expression studies because of its ability to express a wide variety of keratins including the differentiation-specific K10 (section 1.10.4). The assessment of factor IX gene expression resulting from the various constructs and the overall evaluation of the strategy in relation to the recent progress made towards gene therapy for haemophilia B are discussed here.

4.2.1 KERATINOCYTE-SPECIFIC EXPRESSION OF FACTOR IX

Keratinocytes (HaCaT cells) secrete γ-carboxylated factor IX

Factor IX is primarily synthesised in the liver and subjected to at least four post-translational modifications before being secreted into the bloodstream (section 1.3.4). These modifications include glycosylation, proteolytic cleavage of the pre and propeptides, vitamin K-dependent γ-carboxylation of 12 N-terminal glutamic acid residues and β-hydroxylation of an aspartic acid residue at position 64. Due to the complex nature of post-translational modifications, factor IX production using recombinant DNA techniques requires the use of mammalian cells. Although many mammalian cell types including fibroblasts (St. Louis and Verma, 1988; Palmer et al., 1989; Axelrod et al., 1990; Scharfmann et al., 1991), endothelial cells (Yao et al., 1991), myoblasts (Dai et al., 1992; Yao and Kurachi, 1992) and hepatocytes (Armentano et al., 1990; Kay et al., 1993) have been successfully used for the production of biologically-active factor IX, many cells which only produce partially γ-carboxylated factor IX (with lower biological activity) have been reported (Anson et al., 1985; Busby et al., 1985; De La Salle et al., 1985; Kaufmann et al., 1986; Balland et al., 1988; Derian et al., 1989).

One of the aims of this project was to determine if keratinocytes are capable of carrying out γ-carboxylation of factor IX. All pools and individual clones of transfected cells were analysed for their ability to secrete γ-carboxylated factor IX into the serum-free tissue culture medium in the presence of vitamin K (as a catalyst for the vitamin K-dependent γ-glutamyl carboxylase; Wu et al., 1991). All transfected cells produced readings above the background (negative control) following a 24 hour incubation. The OD readings were converted to ng/million cells/24 hours in order to determine the relative efficiencies of factor IX secretion and the values so obtained ranged from 8.7 to 34.6 ng/million cells/24 hours (tables 8 and 9). The A7 antibody detects the calcium-dependent epitope in the light chain of factor IX and therefore distinguishes between biologically-active factor IX that contains γ-carboxyglutamic acids and factor IX which does not (Smith et al., 1986; Smith, 1988). This indicates that keratinocytes (HaCaT cells) have γ-glutamyl carboxylase activity and can secrete biologically active factor IX.

These results however do not determine the proportion of secreted factor IX that is γ-carboxylated and therefore cannot elucidate the efficiency of γ-carboxylation in keratinocytes. Recently, Gerrard et al. (1993) demonstrated the capacity of primary human keratinocytes to secrete γ-carboxylated factor IX. They used a retroviral vector to transduce primary keratinocytes and showed using the same A5/A7 sandwich ELISA that 88% and 22% of secreted factor IX was active in the presence or absence of vitamin K, respectively.

ELISAs were performed on platelet-poor plasma samples from the Y-line transgenic mice (with K10-FIX-I) which also revealed the presence of γ-carboxylated factor IX (table 8). From the results obtained by Bailleul et al. (1990), who used the same (BKVIb/K10) promoter, it would be expected that the foregut would also express factor IX. However, the detection of factor IX mRNA in the epidermis of these mice (by Northern blots) and the previously demonstrated ability of keratinocytes to secrete factor IX into the systemic circulation (Gerrard et al., 1993), suggests that the keratinocyte-specific expression was probably the major source of factor IX in the systemic circulation. These findings together prove that keratinocytes, which are an attractive cell type for *ex vivo* and *in vivo* somatic gene therapy, are suitable for the ectopic production of factor IX for the treatment of haemophilia B.

Comparative analysis of the various expression constructs

Northern blot analysis

While all transfected pools were positive using RT-PCR, it was not possible to determine the relative levels of factor IX mRNA expression. Northern analysis revealed that the pools of transfected K10-FIX-I and II (i.e. basic) constructs, expressed much less factor IX mRNA than the manipulated constructs (K10-FIX-G4, K10RB-FIX, K10RB-FIX-G4). The following questions needed to be answered: (i) is this result true i.e., are the number of cells expressing factor IX mRNA in each transfected pool the same (similar) and is the relative proportion of high, low and non expressing clones similar in each case? If yes, then (ii) does this difference in level of factor IX transcript represent lower transcription levels or a reduced mRNA stability for the basic construct?

First of all, as the pools analysed consisted of G418-resistant clones, most of the cells will contain the transfected expression vector. Furthermore, the initial transfection efficiencies are likely to be similar because the DNA preparations, purification and quantities used were identical and the transfections were performed simultaneously, using the same protocol and efficiencies estimated in duplicate. This means that the total number of cells transfected would be similar and therefore statistically also the distribution of high, low and non expressing clones. To increase the statistical significance of the results obtained from the analysis of the pooled samples, expression at the levels of RNA and protein were determined at relatively early passages after transfection and selection. From the number of cells transfected, the observed transfection efficiency and the number of passages, it is estimated that each pool of transfected cells used for gene expression analyses represented between 125-750 individual clones. It is also very unlikely that purely by chance both, K10-FIX-I and K10-FIX-II would have a higher proportion of low and non expressers and the other three manipulated constructs, more efficient expressers. It therefore seems very likely that there are similar numbers and distribution of cells expressing factor IX mRNA in the case of each construct i.e., that the result is a true one.

Is this reduced mRNA level due to a reduced transcription rate? The only difference between K10-FIX-II (basic construct) and K10-FIX-G4 is the presence of the induced mutation Q-45V (in order to achieve a 'G' at +4; section 3.5.3), the same difference that exists between K10RB-FIX and K10RB-FIX-G4. Northern analysis however revealed a large difference in the quantities of recombinant factor IX mRNA, only between the former two constructs and not between the latter. It is therefore unlikely that Q-45V could have resulted in this difference in factor IX mRNA abundance by an increase in the transcriptional level. However, nuclear run-on experiments will have to be performed to be absolutely certain.

The main difference that exists between the two groups of constructs with high and low mRNA levels is the sequence manipulation introduced to enhance their translational efficiency. It would be expected that the repaired K10 cap site and 5'UTR and the introduction of the 'G' at position +4 will improve the translatability of the

otherwise similarly transcribed mRNAs. In view of these observations it is proposed that the levels of mRNA are regulated at a post-transcriptional level, that is, due to a modulation of their stabilities, which in turn could be related to the process or product of translation. The modulation of stability of mRNAs in relation to translation is not unknown. Two well characterised examples include the regulation of mRNA for β-tubulin and histones. In the case of the β-tubulin mRNA, a set of elegant studies have shown that mRNA stability is determined by translation of a short length (~13) of nucleotides near the 5' end of the message (Yen et al., 1988). In the case of the histone mRNA the sequence responsible for mRNA stability was localised to a ~30 nucleotide segment in the 3'UTR. It was shown that premature termination of translation, ending before this segment led to abolition of the stability-modulating function (Graves et al., 1987; and reviewed in Marzluff and Pandey, 1988). It is significant to note in this respect that the 3'UTRs of the various constructs tested are identical. It is also unlikely that the Q-45V mutation (the only sequence difference between the basic construct and K10-FIX-G4) is directly responsible for this effect because this difference was not reproduced between K10RB-FIX and K10RB-FIX-G4.

It is widely accepted that efficiently translated mRNAs are more stable due to a protective effect of the polysome. In models describing mRNA degradation, the initial event is thought to be due to endonucleolytic cleavages proceeding in a 5' to 3' direction, following the last ribosome (Lewin, 1994). The released fragments are then degraded by exonucleolytic digestion in a 3' to 5' direction. Lim et al. (1992) demonstrated the occurrence of degraded β-globin mRNAs (with loss of varying lengths at the 5' end) in response to various nonsense mutations, in transgenic mice. Recently, Lynch et al. (1993) localised a 1.2 kb segment of the factor VIII coding sequence with the ability to act as a dominant inhibitor of RNA accumulation. Using nuclear run-on experiments it was previously shown that the factor VIII mRNA is transcribed at normal rates but does not accumulate accordingly, and the defect is therefore at a post-transcriptional level (Kaufmann et al., 1989). The evidence in favour of the process of translation in influencing the susceptibility of translated mRNAs to nucleolytic activity has been reviewed by Brawerman (1989) and Ross (1988).

In order to determine the integrity and relative quantities of RNA loaded in each lane, the same Northern filters were also hybridised with a ubiquitiously transcribed GAPDH cDNA probe. All lanes indicated that the RNA was largely intact and the difference in the quantities loaded was only minimal. The result of the GAPDH transcript was obtained after a 24 hour exposure and that of factor IX required a 5 day autoradiographic exposure. This indicates that the levels of factor IX mRNA (not necessarily the rate of transcription) is much lower than that of GAPDH.

Assay of secreted factor IX by ELISA

It was demonstrated that a structurally intact transgene had stably integrated in all pools and individual clones of transfected cells. Pools of transfected cells contain a mixture of high, low and non-expressing clones. To eliminate the effect produced by the varying proportion of expressers and to isolate at least some high expressers, individual stable clones were generated. It is also known that transgenes in individual stable clones are generally more stable (Mather, 1990). However, the low success rate of the cloning procedure allowed the isolation of only a few stable transfectants (table 7). ELISAs were performed on all pools and clones and this revealed no individual stably transfected high expressing clone i.e. expressing significantly more than the corresponding pooled sample (table 8). This is likely to be due to the low number of clones tested. It was assumed that the proportion of high, low and non-expressers in each transfected pool would not be significantly different in the case of each expression construct. Also, as mentioned before, expression analyses were performed at relatively early passages to maximise the statistical significance. It was estimated that the number of clones per individual pooled sample represented between 125-750 individual expressing clones. This would then mean that the expression levels obtained for a pool of transfected clones with a particular expression construct would reflect the expressivity of that particular construct, thus allowing a comparative analysis. For this purpose, a comparative study of the factor IX levels estimated (by ELISA) for each transfected pool of clones i.e. K10-FIX-I, K10-FIX-II, K10-FIX-G4, K10RB-FIX and K10RB-FIX-G4, was performed.

An analysis was performed to assess the following factors: (i) effect of the sequence manipulations introduced into the basic construct; (ii) identification of the most

efficient expression construct and (iii) effect of the S136T and Q-45V missense mutations.

The mean of the triplicate results of factor IX expression for each of the five pools and the variation by ±2SD are listed in table 9. In order to assess the effect of the sequence manipulations introduced into the basic construct (K10-FIX-II), a student *t* test of significance was performed. It is seen in table 9 that the basic construct (K10-FIX-II) expresses the least factor IX (18.1 ng/10^6 cells/24 hours) and that all the manipulated constructs (K10-FIX-G4, K10RB-FIX and K10RB-FIX-G4) express significantly higher levels ($p<0.01$). This indicates that both, the creation of a favourable sequence context around the "AUG" initiator codon and the introduction of the K10 cap site and 5'UTR into the basic construct were favourable modifications.

Even though it seems that the factor IX expression is higher in the K10-FIX-I (basic construct + S136T) construct (22 ng/10^6 cells/24 hours), the difference was not statistically significant ($p<0.1$). This indicates that the S136T missense mutation is likely to be conservative, as predicted from the similar properties of the normal and mutated amino acids (although a functional assay will have to be performed to be sure). Furthermore, the Y-line transgenic mice (created with this gene construct) were shown to have recombinant human factor IX at a level of 28 ng/ml in their plasma.

The highest expression of biologically active factor IX was obtained with the K10-FIX-G4 construct, which secreted 35 ng/10^6 cells/24 hours. The student *t* test revealed that this was significantly higher than the level obtained for the basic construct ($p<0.01$). Introduction of the 'G' at position +4 to achieve a favourable sequence around the "AUG" initiator codon had resulted in a missense mutation at the second codon of the factor IX mRNA (Q-45V). When mutating the 'C' at +4 to 'G', the alteration of the sequence at the second codon (-45; Anson et al., 1984) of the factor IX cDNA was chosen, so as to be the most conservative change possible. The expression results suggest that this mutation is also likely to be conservative, as predicted. Perhaps significantly, neither this mutation nor the S136T mutation (also shown to be conservative previously) have been reported in the database of mutations affecting the factor IX genes of 806 haemophilia B patients (4th edition; Giannelli et al., 1993).

Although all the manipulated constructs (K10-FIX-G4, K10RB-FIX and K10RB-FIX-G4) expressed significantly higher levels of factor IX than the basic construct, the *t* test revealed that no significant difference existed amongst the three of them. The K10RB-FIX-G4 construct has the repaired K10 cap site and 5'UTR sequence in addition to the appropriate sequence context surrounding the "AUG" initiator codon. It seems likely that the sequence manipulation surrounding the "AUG" codon was responsible for the more efficient expression of K10-FIX-G4 compared to the basic construct. It was therefore expected, on theoretical grounds, that K10RB-FIX-G4 might express more factor IX than K10-FIX-G4 and perhaps even be the highest expressing construct. This was clearly not the case (table 9). The sequence that would form the 5'UTR of the K10RB-FIX-G4 construct was retrospectively analysed for potential negative features that could result in lowered expression. No upstream "AUG" sites were found. A structural analysis was performed using a computer program "STEMLOOP" (which is part of the GCG package) to predict the formation of hairpin structures in the 5'UTR. These are known to be an important cause of poor translation of mRNAs. The 5'UTRs of all the other constructs and the BKVIb gene were also simultaneously analysed. As is seen in figure 48, the K10RB-FIX and K10RB-FIX-G4 constructs are predicted to form hairpin structures in the 5'UTR. Although the sequence of the 5'UTR was designed to be identical to that of the BKVIb gene, the *Kpn*I site used for subcloning the synthetic fragment (absent in the BKVIb gene) was responsible for the stemloop formation with the complementary sequence in the BKVIb 5'UTR. In contrast, neither of the other constructs nor BKVIb are predicted to produce mRNAs with hairpin structures in their 5'UTRs.

Kozak (1991b) has summarised certain experimentally established rules regarding the theoretical prediction of the negative effects on translation attributed to secondary structures upstream from the "AUG" codon. In general, it depends on the strength and position of the hairpin. A modest amount of secondary structure within 12 nucleotides of the cap site drastically inhibits translation by preventing the binding of 40S ribosome subunits (Kozak, 1989b). On the other hand, when the secondary structure is sufficiently far from the cap site then the entry of the 40S ribosome is not affected and the stability of

Figure 48: Structural analysis of the 5'UTRs of the various K10-FIX gene expression constructs

This figure shows the result of the analysis of the 5'UTRs of the various gene constructs. The computer program "STEMLOOP" was used to perform a search for potentially deleterious hairpin structures. The sequence of the 5'UTRs are shown with the cap sites underlined. The 5'UTR of the K10 (BKVIb) gene is also analysed. The *Kpn*I recognition sequence in the K10RB-FIX and K10RB-FIX-G4 constructs is also indicated. Only the constructs with the repaired 5'UTR show the formation of a hairpin structure. The stem is 6 bp long and the loop is a symmetrical structure with 11 nucleotides. The only difference that exists between the natural K10 gene and the repaired constructs (in this region), was the presence of the *Kpn*I recognition sequence, and it can be seen from the sequence of the stemloop that this is what is responsible for the formation of the secondary structure (indicated by an arrow). The base of the stemloop is situated 4 nucleotides downstream from the cap site and 3 nucleotides upstream from the "ATG" initiation codon.

Figure 48: "STEMLOOP" Analysis

```
K10-FIX-I/II: GATCCCCGGGTACCATGC                    No stemloops
                    ↑
              putative CAP
                    ↓
K10-FIX-G4:   GATCCCCGGGTACCATGG                    No stemloops

BKVIb Gene:   CACTCCCTGAGCTAAACAGCATCACCATGC        No stemloops
                    ↑                ↑
            BKVIb/K10 CAP      5'UTR          KpnI
                    ↓                ↓            ↓
K10RB-FIX:    CACTCCCTGAGCTAAACAGCATCACCGGTACCATGC
K10RB-FIX-G4: CACTCCCTGAGCTAAACAGCATCACCGGTACCATGG
```

"STEMLOOP":

```
      C
    A   A
  A       G
  A       C
  T       A
    C   T
      G-C
      A A
      G-C
      T C
      C-G   ←KpnI
      C-G
CACTC TACCATGC/G
```

the structure determines whether or not it inhibits scanning. It has been shown that the structure has to be more stable than -30 kcal/mol in order to impair translation in COS cells (Kozak, 1986b) or cell free systems (Kozak, 1989b). As seen in figure 48, the hairpin structure predicted in the mRNAs generated from K10RB-FIX and K10RB-FIX-G4 is situated 4 nucleotides from the cap site, a distance at which even moderately stable structures can have a major effect.

The following conclusions can therefore be drawn from the above observations: (i) Sequence manipulations introduced into the basic construct (K10-FIX-II) to create K10-FIX-G4, K10RB-FIX and K10RB-FIX-G4, resulted in a significant improvement of gene expression. (ii) The sequence at the two positions, -3 (A/G) and +4 (G) are indeed important for efficient translation initiation, possibly resulting in enhanced message stability. (iii) Secondary structures near the cap site possibly inhibit translation. A preliminary analysis for secondary structures should be performed prior to the construction of gene expression vectors. As illustrated in this case, simply a choice of a different cloning site in the polylinker may have averted the generation of these potentially inhibitory structures. (iv) Both S136T and Q-45V are likely to be conservative mutations (a functional assay will have to be performed to be entirely sure). (v) The K10-FIX-I construct is effective *in vivo*. This conclusion was made in collaboration with Dr. Y. Alexander who made the transgenic mice and performed Northern blot and in situ hybridisation analysis to demonstrate factor IX mRNA in epidermis. As part of this project factor IX protein was assayed in the plasma of these (Y-line) mice.

4.2.2 EVALUATION OF THE STRATEGY

Many groups have achieved high levels of factor IX expression *in vitro*. Although a wide variety of cell types have been used, all those achieving high levels have used strong viral promoters to drive the expression of factor IX. The levels achieved have ranged from 70 to >3000 ng/10^6 cells/24 hours (Palmer et al., 1989; Armentano et al., 1990; Yao et al., 1991; Yao and Kurachi, 1992; Gerrard et al., 1993; Kay et al., 1993; Lozier et al., 1994). As mentioned before, to achieve long-term expression, cellular promoters were used, which resulted in sustained but low levels of factor IX expression

(Scharfmann et al., 1991; Dai et al., 1992). In this project a cellular promoter was used to drive expression of factor IX in a tissue- and differentiation-specific fashion. Using the K10 promoter to express factor IX in keratinocytes, a maximum level of 34.6 (26-45) ng/10^6 cells/24 hours of biologically active factor IX was secreted into the conditioned medium. This level is perhaps as much as ten to fifteen fold less than what most investigators have achieved using strong viral promoters.

It is however very likely that these relatively low levels are not due to the inefficient expression of the factor IX gene construct *per se*. The results most probably are a reflection of the limitations of the experimental system. The K10 gene though highly expressed *in vivo*, is not expressed in conventional keratinocyte cultures and so the only way to test the efficacy of the K10 promoter *in vitro*, was to use the HaCaT cell line that expresses an unusually wide variety of keratins, including K10. The expression of K10 in HaCaT cells is, however, very low and even though it increases with increasing cell densities, it is still expressed at much lower levels than the other keratins (Boukamp et al., 1988). Moreover, using immunofluoresence it was shown that K10 was expressed in only a few cells in the total culture and that the increased expression with increasing cell density was due to the increase in staining observed in the few K10-positive cells in the culture (Ryle et al., 1989). By extension, this indicates that the K10 promoter will be expressed in only a fraction of the cells transfected. Furthermore, even after enriching the cultures for cells containing the transgene (by G418 selection) and incubating cells at high density, the factor IX secretion is likely to originate from far fewer cells than in comparable studies using viral promoters, where all cells in the culture are expected to be efficient expressers. It is therefore not entirely surprising that a ten-fold lower level of factor IX was detected in the present study. Indeed this might represent a reasonably high level of expression assuming that less than 10% of HaCaT cells actually express K10. An estimate of the relative efficiencies of factor IX expression can be made by a comparison with the results of two recent studies that employed retroviral-mediated transfer for the expression of factor IX (Kay et al., 1993; Gerrard et al., 1993). Kay et al. (1993) obtained expression of factor IX at levels of 225 ng/10^6 cells/24 hours and Gerrard et al. (1993) obtained expression of 580 ng/10^6 cells/24 hours (the latter group used primary

human keratinocytes). The former group did not selectively enrich for transduced cells and estimated (using a *lacZ* reporter construct) a transfection efficiency of 30-40%. Gerrard et al. (1993) had at least twice as many cells expressing factor IX as they selectively enriched the transduced cell population (using G418). If it is assumed that 5-10% of the transfected HaCaT cells expressed factor IX from the K10 promoter, then it would seem that the efficiency of expression was comparable to that achieved by the retroviral LTR, which is indeed high. Demonstration of the ability of the K10-FIX-I construct to express biologically active factor IX *in vivo* (in the Y-line mice) at a level of 28 ng/ml also adds credence to this. Dai et al. (1992) who injected mice with retrovirally transduced primary mouse myoblasts, achieved long-term expression of factor IX *in vivo* at a maximum level of 20 ng/ml (after two consecutive administrations of 1-2 x 10^7 cells). Also Kay et al. (1993) who achieved between 0.3-1.0% hepatic transduction efficiency by direct intra-portal delivery of recombinant retroviral vectors into (2/3 partial) hepatectomised canine haemophilia B dogs, achieved sustained plasma factor IX levels of between 2-10 ng/ml. Moreover, judging from the *in vitro* results presented in this project, it could be predicted that the expression of factor IX from the manipulated constructs *in vivo* might be even higher than that from K10-FIX-I (perhaps as much as two-fold).

There are however other reasons to indicate that this experimental system may not be optimum: (i) Due to the high sequence identity observed in the non-coding sequences (including extensive intronic regions) of the human, bovine and murine counterparts of the K10 gene, Reiger et al. (1988) inferred that there may be *cis*-acting elements for transcriptional enhancement in regions other than the (5 kb) 5' flanking sequence that was used in this project. (ii) Jallat et al. (1990) showed that the factor IX cDNA (i.e. without any introns) may not be an ideal sequence for the expression of factor IX. They observed a 40-200 fold increased expression with the inclusion of introns in transgenic mice (constructs containing only the first intron or all introns gave similar results). These results are nevertheless likely to pertain to the observations of Brinster et al. (1988) who demonstrated the necessity for any intron for efficient expression in transgenic mice. The heterologous SV40 small t-intron (used in this project) may serve this non-specific function. (iii) HaCaT cells are not 'normal' keratinocytes and may differ

from the primary human keratinocytes (in secretory function and γ-carboxylation) that were successfully used by Gerrard et al. (1993). This is however not likely to be a major problem judging from the expression levels (in the presence of vitamin K) detected using the A7 antibody, that detects γ-carboxylated factor IX. (iv) The ELISAs in this project had a high background reading (that had to be subtracted from the expression results) which made the assay less sensitive and therefore the results might represent underestimates of the true expression level. Other investigators have reported similar sensitivities where they could detect up to 1 ng/ml (Smith et al., 1993) and 2.5 ng/ml (Jallat et al., 1990). Gerrard et al. (1993) who used the same A5/A7 antibodies, however, reported a much higher sensitivity of <0.1 ng/ml human recombinant factor IX in mouse plasma. (v) Other mundane possibilities such as a mutation in the factor IX coding sequence and disruption of the transcriptional unit during random integration were also considered. A comprehensive mutation analysis, linearisation of the constructs by cutting in a nonessential region and an expected sized transcript detected by Northern blots, served to adequately rule out these possibilities.

4.2.3 FUTURE PERSPECTIVES

Keratinocytes are an attractive cell type for somatic gene therapy. Fenjves et al. (1989) demonstrated that the dermo-epidermal basement membrane can allow ectopically synthesised proteins access to the systemic circulation. In addition, Gerrard et al. (1993) also demonstrated that keratinocytes are capable of carrying out the required post-translational modifications to result in biologically active factor IX. In an attempt to exploit this cell type and to achieve long-term expression *in vivo*, this project has demonstrated the use of a cytokeratin gene promoter to direct the expression of (γ-carboxylated) factor IX in a keratinocyte-specific fashion. It was shown that a gene construct produced in this project was also effective *in vivo*. As has been discussed, the levels of expression attained are considered comparable to those achieved by other groups who used strong viral promoters.

The epidermis is unique in that it can be completely regenerated from cultured somatic cells. Rheinwald and Green (1975) showed that it was possible to serially

cultivate human diploid keratinocytes on lethally irradiated fibroblasts and produce sufficient coherent epithelial sheets to cover the entire body surface (Green et al., 1979; Gallico et al., 1984). The basal layer of the epidermis is the proliferative compartment, of which up to 10% is comprised of epidermal stem cells. These cells give rise to the suprabasal keratinocytes, which migrate outwards and eventually get cornified and are shed by desquamation. To transfer gene constructs into keratinocytes, sufficient numbers of stem cells would have to be transfected in order to continuously replenish the recombinant keratinocytes lost by desquamation. From the experience gained in the use of cultured epithelial autografts for the healing of full-thickness burn wounds, it is now known that epidermal stem cells survive in culture (Gallico et al., 1984; Compton et al., 1989). Recently Jones and Watt (1993) discovered that epidermal stem cells could be purified on the basis of high surface expression of β_1 integrins and rapid adhesion to type IV collagen, fibronectin or keratinocyte extracellular matrix. This should greatly aid in the targeting of sufficient stem cell numbers *in vitro* and achieve the required long term expression from genetically manipulated autologous grafts. Keratinocytes from neonates or paediatric age groups will be more suitable for this form of therapy due to the higher plating efficiency and growth potential of these cells (Green, 1989). In this respect, allogenic grafts will not be useful because it is known that they repair epidermis by stimulating stem cells in the surrounding normal keratinocytes and do not therefore integrate into the regenerated epidermis (Green, 1989). Besides the *ex vivo* approach, gene constructs could be transferred to keratinocytes directly *in vivo*. This could be achieved by the use of liposomes or recombinant viral vectors (with a keratin promoter driving expression), either via topical application or intradermal injection. To target sufficient stem cells by this method however, repeated administrations might be required. The possibility of stimulating the stem cell compartment to become more receptive to transfected DNA *in vivo* has not yet been investigated by anyone. This is an area of biology which deserves more attention if keratinocyte gene therapy was to be furthered. The first clinical trial using a cationic-liposome delivery system for treatment of subcutaneous lesions of metastatic melanoma has been approved (January, 1992) and

amongst the main aims of this trial are to assess the safety and efficacy of this delivery system (Nabel and Felgner, 1993).

To estimate the number of keratinocytes required to actively express factor IX *in vivo,* the method described by Teumer et al. (1990) was used. The formula used was as follows: equilibrium concentration (ng / ml) = [factor IX production from the graft (ng / h) / equilibrium volume (ml)] / K, where K = In2 / $T_{1/2}$. The equilibrium volume (extracellular volume) is calculated as 14,000 ml (20% of body weight in a 70 kg human male) and $T_{1/2}$ (half life of factor IX in human plasma) = 25 h (Biggs and Denson, 1963). To maintain a plasma factor IX level of 50 ng / ml (i.e., 1% of normal plasma factor IX concentration), it is calculated that a graft will have to synthesise 19.4 μg / h or 465.6 μg / 24 h. Judging from the level of expression achieved by the HaCaT cells transfected with K10-FIX-G4 (35 ng / 10^6 cells / 24 h), there would have to be 1.3 X 10^{10} cells in the graft, to achieve therapeutic expression levels. If it is assumed that only 10% of the recombinant HaCaT cells were expressing factor IX in the *in vitro* model used in this project (section 4.2.2), whereas *in vivo* one would hope to obtain ≥95% expression, then 1.3 X 10^9 cells would be required. It is roughly estimated that the number of suprabasal keratinocytes in human epidermis (12 - 15 cell layers) is about 1.8 X 10^8 cells / 100 cm^2. This means that at presently observed expression levels, an approximately 722 cm^2 graft (4% of the adult body surface area) would be required to achieve and maintain a steady state level of 1% of normal plasma factor IX. These measurements are based on the assumption that every suprabasal keratinocyte in the patch would express the foreign gene. This is thought to be possible, provided sufficient stem cells are targeted. The targeting of sufficient stem cells *in vitro* may be possible by using a combination of selection for epidermal stem cells and using a very efficient transfection protocol. Levels of factor IX greater than 1% with grafts of a smaller size could therefore be achieved by selecting for stem cells, improving the efficiency of transfection and gene expression. It remains to be seen if the enriched stem cell population can successfully generate epidermis with a normal architecture. Due to the easy accessibility of the graft, it may also be possible to apply an inductive agent topically and thus stimulate an inducible

enhancer element to upregulate heterologous gene expression, for e.g., topical steroids to stimulate a steroid responsive element (Vogel et al., 1993).

The results obtained from the Y-line transgenic mice provide a means of assessing the efficacy of the gene constructs *in vivo* also taking into account the delivery of factor IX from suprabasal keratinocytes into the systemic circulation by transport across the dermo-epidermal basement membrane. The Y-line mice were created using K10-FIX-I, which has been shown to be a less effective gene construct than K10-FIX-G4 *in vitro*. On the basis of the *in vitro* factor IX expression obtained from HaCaT cells transfected with K10-FIX-I (22 ng / 10^6 cells / 24 h), it is estimated that a 100 cm^2 graft of human epidermis will express about 40 μg / 24 h (instead of 64.5 μg / 24 h as for K10-FIX-G4). Using the formula described by Teumer et al. (1990) it is estimated that the suprabasal keratinocytes of a Y-line mouse must produce 242.4 ng / 24 h to maintain the circulating factor IX level of 28 ng/ml in plasma. For these calculations the following values were used: $T_{1/2}$ of human factor IX in mouse plasma = 12 h (Gerrard et al., 1992) and equilibrium volume = 6.6 ml (22% of total body weight in a 30 g mouse; Teumer et al., 1990). In order to estimate the amount of factor IX that would be produced by the suprabasal keratinocytes in a similar area (~100 cm^2) of human epidermis, the structural differences between mouse and human epidermis and the results obtained by in situ hybridisation on the Y-line epidermis (performed by Dr. Y. Alexander) were considered. The mouse epidermis has only 2 layers of suprabasal keratinocytes compared to 12 - 15 in human, and in situ hybridisation showed about 25% of each layer expressed the transgene at high levels. By correcting for this, it is estimated that the suprabasal keratinocytes in a 100 cm^2 graft of human epidermis will express factor IX at about 6.8 μg / 24 h. This 6 fold difference between the expected and observed levels of factor IX expression *in vivo* can be due to one or a combination of the following reasons: (i) The effect of the difference in the cellular environments *in vivo* and *in vitro*. It is known that gene expression *in vivo* does not always correlate with levels observed *in vitro* (Kriegler, 1990a). (ii) The difference between HaCaT cells (which is a spontaneously transformed cell line) and normal keratinocytes *in vivo*. Even though the calculations have accounted for the fewer number of HaCaT cells expressing K10 *in vitro*, it is also known that these

cells express K10 at lower levels than suprabasal keratinocytes *in vivo* (Boukamp et al., 1988; Ryle et al., 1989), thus further distorting the correlation. (iii) The observed suppression of ELISA readings (by up to 40%) by plasma, leading to an underestimation of the plasma factor IX levels in the Y-line mice (Smith et al., 1993). (iv) The possible instability of factor IX with the S136T missense mutation *in vivo* (even though no such effect was seen *in vitro*) and (v) The inefficient transport of factor IX (57 kD) across the dermo-epidermal basement membrane. Gerrard et al. (1993) transplanted genetically modified primary keratinocytes at a sub-dermal site and therefore the demonstration of factor IX in the systemic circulation of the recipient athymic mice in their experiments, does not necessarily indicate that the basement membrane is permeable to factor IX. However, the demonstration of the transport of apolipoprotein E (30 kD) from epidermis to blood (Fenjves et al., 1989), and the movement of large plasma proteins (albumin and immunoglobulins) from blood to epidermis (Coruh and Mason, 1980), indicate that the last reason is not likely to be the major cause. With regard to this problem, it may be beneficial to attempt keratinocyte-specific expression using the cytokeratin 5 and 6 (K5 and K6) gene promoters. Expression of K5 and K6 are specific to the basal layer of the epidermis and hair follicles, respectively, both of which are situated closer to the systemic circulation than suprabasal keratinocytes. The K5 promoter will however drive expression in only a single (basal) cell layer and will therefore result in a lot fewer expressing cells in a graft.

Dai et al. (1992) were among the first to achieve sustained (but low) levels of factor IX expression *in vivo* (via intramuscular injections of genetically modified myoblasts) by the incorporation of a cellular enhancer element in their recombinant factor IX expressing retroviral vector. Clearly, a better understanding of the various sequence elements in the promoter/enhancers will be required to achieve high level expression using cellular promoters. A considerable amount of research is being performed in order to determine the various modulators of gene expression in keratins (Fuchs et al., 1994). Modifications that could be attempted include for example, deletion of negative regulatory elements identified in promoter/enhancers (e.g. in the HPRT gene, Rincon-Limas et al., 1991) or identification and use of positive elements present in other regions

of genes, that are not normally used for directing heterologous gene expression (e.g. in the second intron and 3' to the structural gene for human β-globin; Behringer et al., 1987). Smith et al. (1993), using adenoviral vectors *in vivo*, achieved very high but transient factor IX expression due to the host immune response. Development of less immunogenic adenoviral vectors would greatly improve the outlook for this convenient method of gene delivery (Yang et al., 1994). Low but sustained levels of factor IX expression were attained by an *in vivo* approach used by Kay et al. (1993), who administered recombinant retroviral vectors directly into the portal circulation to transduce hepatocytes *in vivo* (in partially-hepatectomised haemophilia B dogs). Improvement in the level of expression would be possible if higher titres and transduction efficiencies (they had between 0.3-1.0% efficiency) could be achieved.

The results obtained by Kay et al. (1993) are especially encouraging because they demonstrate that a phenotypic improvement is indeed possible by the secretion of relatively low levels of factor IX *in vivo*. With factor IX antigen levels after gene transfer being only 0.1% of the endogenous normal level, they achieved a sustained (up to nine months in dog 1) reduction in the WBCT (whole blood clotting time) from 45-55 min to 15-20 min (normal = 6-8 min) and a shortening of the PTT (partial thromboplastin time) from 322 to 174 seconds.

In conclusion, the results obtained in this study show that keratinocytes appear to be effective target cells for somatic gene therapy of haemophilia B and in addition, indicate the suitability of using tissue-specific cellular promoters to achieve sustained and high level expression of factor IX. However, optimisation of the regulatory elements, and investigation of a means to efficiently target epidermal stem cells, may have to be performed before this approach can be considered suitable for formal preclinical studies.

APPENDIX I

Primers for amplification of the factor VIII cDNA *

Primer [1]	Position	Sequence 5' to 3'
R1-A	-69	GGG AGC TAA AGA TAT TTT AGA GAA G
R1-B	1322	TTC CTA CCA ATC CGC TGA GG
R1-C	-50	GAG AAG AAT TAA CCT TTT GCT TCT C
R1-D	1223	CAG CAG CAA TGT AAT GTA CC
R2-A	1080	AGA AGC GGA AGA CTA TGA TG
R2-B	2316	TTG CCT AGT GCT AGG GAG TC
R2-C	1114	TCT GAA ATG GAT GTG GTC AG
R2-D	2282	AAG CTT CTT GGT TCA ATG GC
D1-A	2214	GTC TAT GGA TTC TGG GGT GC
D1-B	3692	GGC AAA ACT ACA TTC TCT TG
D2-A	3537	CGT AGG ACT CAA AGA GAT GG
D2-B	5219	CTG TTT CTT AGA ACA TGT GG
R4-A	5924	TTC ATT TCA GTG GAC ATG TG
R4-B	7164	TGT CTG CTA GGA TTT AGC AC
R4-C	6071	ATC TAC ATG CTG GGA TGA GC
R4-D	7143	AAG GTA GAA GGC AAG CCA GG

* Sequence information was obtained from Professor F. Giannelli; [1]Primers that end in A/C and B/D are forward and reverse primers, respectively; Those that end in A/B and C/D were used for the initial PCR and nested PCR amplifications, respectively. See figure 3 for the various PCRs. mRNA positions are according to Wood et al. (1984).

Primers for the amplification of individual exons of the factor VIII gene *

Primer [1]	Sequence 5' to 3'
2X	TTG AAG TGT CCA CCA AAA TGA ACG ACT
2Y	GAT ACC CAA TTT CAT AAA TAG CAT TCA
4X	TAC AGT GGA TAT AGA AAG GAC AAT TTT
4Y	TGC TTA TTT CAT CTC AAT CCT ACG CTT
5X	CCT CCT AGT GAC AAT TTC CTA CAA TGA
5Y	AGC AGA GGA TTT CTT TCA GGA ATC CAA
6X	CAT GAG ACA CCA TGC TTA GCT GAC TCT A
6Y	AAC TCT GGT GCT GAA TTT GGA AGA CCC T
7X	TCA GAT TCT CTA CTT CAT AGC CAT AGG T
7Y	GTA GGA CTG GAT ATT TAT AAT ATT CAT T
8X	ATA TAG CAA GAC ACT CTG ACA TTG
8Y	AGA GAG TAC CAA TAG TCA AA
11X	ATG GTT TTG CTT GTG GGT AG
11Y	GGA TCC GAC ATA CAC TGA GAA TGA A
12X	CTA GCT CCT ACC TGA CAA CAT CAG TAG C
12Y	CTC AAG CTG CAC TCC AGC TG
14X	GAG AAC CTC TAA CAG AAC GT
14Y	AGC AGA GCA AAG GAA TAA CC
16X	GTC GTT ATT GTT CTA CAG GTA
16Y	GTG GTC AAG CAC AAT AGA CAC CTG C
17X	TGA TGA GAA ATC CAC TCT GG
17Y	GTG CAA TCT GCA TTT CAC AG
20X	ATT TTG TGC ACT CTA GTT ACT GTG T
20Y	TAT AAT CAG CCC AGG TTC TTG GAG C
21X	ATA TAG CAA GAC ACT CTG ACA TTG
21Y	GAG TGA ATG TGA TAC ATT TCC CAT CA
22X	AAA TAG GTT AAA ATA AAG TGT TAT
22Y	TGA CTA ATT ACA TAC CAT TAA GGT T
23X	CTC TGT ATT CAC TTT CCA TG
23Y	AAG GAT ATG GGA TGA CTT GGC ACT

* Sequence information was obtained from Higuchi et al. (1991a); [1]Primers that end with an "X" and "Y" are forward and reverse primers, respectively.

Appendix II

Student's *t* test for significance of the difference between the means of two samples

The *t* value is calculated as the ratio of the observed difference between the two means to the calculated standard error of that difference.

$$t = \frac{M1 - M2}{\sqrt{s^2/n_1 + s^2/n_2}}$$

M1 and M2 are the means of the observations in the two samples; "s" is the standard deviation in the universe and n_1 and n_2 are the number of observations made in the same two samples.

To calculate s^2, the sum of the squared deviations of the observations from their means is calculated for each separate group. Then a combined estimate of the standard deviation is made as follows:

$$s^2 = \frac{\text{sum } (X1 - M1)^2 + \text{sum } (X2 - M2)^2}{(n_1 - 1) + (n_2 - 1)}$$

The probability for a given value of *t* is then looked up in statistical tables with $(n_1 - 1) + (n_2 - 1)$ degrees of freedom.

REFERENCES

- Akhurst, R.J. (1989) Prospects for Gene Therapy Now and in the future. J. Inher. Metab. Dis. 12, 191-201.

- Akli, S., Caillaud, C., Vigne, E., Stratford-Perricaudet, L.D., Poenaru, L., Perricaudet, M., Kahn, A., Peschanski, M.R. (1993) Transfer of a foreign gene into the brain using adenovirus vectors. Nature Genet. 3, 224-228.

- Alber, T. (1989) Mutational effects of protein stability. Ann. Rev. Biochem. 58, 765-98.

- Alwine, J.D., Kemp, D.J., Stark, G.R. (1977) Method for detection of specific RNAs in gels by transfer to diazobezyloxymethyl paper and hybridisation with DNA probes. Proc. Natl. Acad. Sci. USA 74, 5350-5254.

- Anderson, W.F. (1984) Prospects for human gene therapy. Science 226, 401-409.

- Anderson, W.F. (1992) Human gene therapy. Science 256, 808-813.

- Andersson, L.O., Forsman, N., Huang, K. (1986) Isolation and characterisation of human factor VIII: Molecular forms in commercial factor VIII concentrate, cryoprecipitate and plasma. Proc. Natl. Acad. Sci. USA 83, 2979-2983.

- Anson, D.S., Choo, K.H., Rees, D.J.G., Giannelli, R.F., Gould,K., Huddleston, J.A. Brownlee, G.G. (1984). The gene structure of human anti-hemophilic factor IX. EMBO J. 3, 1053-1060.

- Anson, D.S., Austen, D.E.G., Brownlee, G.G. (1985) Expression of active clotting factor IX from recombinant DNA clones in mammalian cells. Nature, 315, 683-685.

- Antonarakis, S.E., Waber, P.G., Kittur, S.D., Patel, A.S., Kazazian, H.H., Mellis, M.A., Counts, R.B., Stamatoyannopoulous, G., Bowie, E.J.W., Fass, D.N., Pitmann, D.D., Wozney, J.M., Toole, J.J. (1985) Hemophilia A: Detection of molecular defects and of carriers by DNA analysis. New Engl. J. Med. 313, 842-848.

- Antonarakis, S.E., Kazazian, H.H. (1988) The molecular basis of hemophilia A in man. Trends in Genetics 4, 233-237.

- Arai, M., Inaba,H., Higuchi,M., Antonarakis, S.E., Kazazian, H.H., Fujimaki, M., Hoyer, L.W. (1989a) Direct charcterization of factor VIII in plasma: Detection of a mutation altering a thrombin cleavage site (arginine 372→histidine). Proc. Natl. Acad. Sci. USA 86, 4277-4281.

- Arai, M., Scandella, D., Hoyer, L.W. (1989b) Molecular basis of factor VIII inhibition by human antibodies. Antibodies that bind the factor VIII light chain prevent the interaction of factor VIII with phospholipid. J. Clin. Invest. 83, 1978-1984.

- Arai, M., Higuchi, M., Antonarakis, S.E., Kazazian, H.H., Phillips, J.A., Janco, R.L., Hoyer, L.W. (1990) Characterization of a thrombin cleavage site mutation (Arg 1689 to Cys) in the factor VIII gene of two unrelated patients with cross-reacting material positive hemophila A. Blood 75, 384-389.

- Armentano, D., Thompson, A.R., Darlington, G., Woo, S.L.C. (1990) Expression of human factor IX in rabbit hepatocytes by retrovirus mediated gene transfer: Potential for gene therapy of Haemophila B. Proc. Natl. Acad. Sci. USA 87, 6141-6145.

- Asselineau, D., Bernard, B.A., Bailly, C., Darmon, M., Prunieras, M. (1986) Human epidermis reconstructed by culture: is it normal? J. Invest. Dermatol. 86, 181-186.

- Atweh, G.F., Wong, C., Reed, R., Antonarakis, S.E., Zhu, D., Ghosh, P.K., Maniatis, T., Forget, B., Kazazian, H.H. (1987) A new mutation in IVS-1 of the human β-globin gene causing β-thalassaemia due to abnormal splicing. Blood 70, 147-151.

- Axelrod, J.H., Read, M.S., Brinkhous, K.M., Verma, I.M. (1990) Phenotypic correction of factor IX deficiency in skin fibroblasts of hemophilic dogs. Proc. Natl. Acad. Sci. USA 87, 5173-5177.

- Bairoch, A. (1993) The PROSITE dictionary of sites and patterns in proteins, its current status. Nucleic Acids Res. 21, 3097-3103.

- Bajocchi, G., Feldman, S.H., Crystal, R.G., Mastrangeli, A. (1993) Direct in vivo gene transfer to ependymal cells in the central nervous system using recombinant adenovirus vectors. Nature Genet. 3, 229-234.

- Ballabio, A., Carrozo, R., Parenti, G., Gil, A., Zollo, M., Persico, M.G., Gillard, E., Affara, N., Yates, J., Ferguson-Smith M.A., Frants, R.R., Ericksson, A.W., Andria, G. (1989) Molecular heterogeneity of steroid-sulphatase deficiency: a multicenter study on 57 unrelated patients at DNA and protein levels. Genomics 4, 36-40.

- Balland, A., Faure, T., Carvallo, D., Cordier, P., Ulrich, P., Fournet, B., de la Salle, H., Lecoq, J.P. (1988) Eur. J. Biochem. 172, 565-572.

- Ballieul, B., Surani, A.M., White, S., Barton, S., Brown, K., Blessing, M., Jorcano, J. Balmain, A. (1990) Skin hyperkeratosis and papilloma formation in transgenic mice expressing a ras oncogene from a suprabasal keratin promoter. Cell 62, 697-708.

- Bebeneck, K., Joyce, C.M., Fitzgerald, M.P., Kunkel, T.A. (1990) The fidelity of DNA synthesis catalysed by derivatives of Escherichia coli DNA polymerase I. J. Biol. Chem. 265, 13878-13887.

- Beddington, R.S., Morgernstern, J., Land, H., Hogan, A. (1989) An *in situ* transgenic enzyme marker for the midgestation mouse embryo and the visualisation of inner cell mass clones during early organogenesis. Development 106, 37-46.

- Behringer, R.R., Hammer, R.E., Brinster, R.L., Palmiter, R.D., Townes, T.M. (1987) Two 3' sequences direct adult erythroid-specific expression of human beta-globin genes in transgenic mice. Proc. Natl. Acad. Sci. USA 84, 7056-7060.

- Bender, M.A., Gelinas, R.E., Miller, A.D. (1989) A majority of mice show long-term expression of a human beta-globin gene after retrovirus transfer into hematopoietic stem cells. Mol. Cell. Biol. 9, 1426-1434.

- Berg, L.P., Wieland, K., Millar, D.S., Schlosser, M., Wagner, M., Kakkar, V.V., ReissJ., Cooper, D.N. (1990) Detection of a novel point mutation causing hemophilia A by PCR/direct sequencing of ectopically transcribed factor VIII mRNA. Hum Genet. 85, 655-658.

- Bertina, R.M., Veltkamp, J.J. (1978) The abnormal factor IX of hemophilia B+ variants. Thrombosis and Haemostasis 40, 335-349.

- Bidichandani, S.I. (1991) MSc thesis, Faculty of Medicine, University of Glasgow.

- Biggs, R., Denson, K.W.E. (1963) The fate of prothrombin and factors VIII, IX, and X transfused to patients deficient in these factors. Br. J. Haematol. 9, 532-547.

- Bird, A.P. (1986) CpG rich islands and the function of DNA methylation. Nature 321, 209-213.

- Blessing, M., Zentgraf, H., Jorcano, J. (1987) Differentially expressed bovine cytokeratin genes. Analysis of gene linkage and evolutionary conservation of 5'-upstream sequences. EMBO J. 6, 567-575.

- Blessing, M., Jorcano, J., Franke, W. (1989) Enhancer elements directing cell-type-specific expression of cytokeratin genes and changes of the epithelial cytoskeleton by transfections of hybrid cytokeratin genes. The EMBO J. 8, 117-126.

- Blessing, M., Nanney, L.B., King, L.E., Jones, C.M., Hogan, B.L.M. (1993) Transgenic mice as a model to study the role of TGF-β-related molecules in hair follicles. Genes and Development 7, 204-215.

- Bodine, D.M., Karlsson, S., Nienhuis, A.W. (1989) Combination of interleukins 3 and 6 preserves stem cell function in culture and enhances retrovirus-mediated gene transfer into hematopoietic stem cells. Proc. Natl. Acad. Sci. USA 86, 8897-8901.

- Bornstein, P., McKay, J., Morishima, J., Devarayalu, S., Gelinas, R.E. (1987) Regulatory elements in the first intron contribute to transcriptional control of the human α-1 (I) collagen gene. Proc. Natl. Acad. Sci. USA. 84, 8869-8873.

- Bos, J.L. (1989) Ras oncogenes in human cancer: A review. Cancer Res. 49, 4682-89.

- Bottema, C.D.K., Ketterling, R.P., Li, S., Yoon, H.P., Phillips, J.A., Sommer, S.S. (1991) Missense mutations and evolutionary conservation of amino acids: evidence that many of the amino acids in factor IX function as 'spacer' elements. Am. J. Hum. Genet. 49, 820-838.

- Boukamp, P., Petrussevska, R.T., Breitkreutz, D., Hornung, J., Markham, A., Fusenig, N. (1988) Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. J. Cell Biol. 106, 761-771.

- Brawerman, G. (1989) mRNA decay: finding the right target. Cell 57, 9-10.

- Bray, G.L. (1992) Current status of clinical studies of recombinant factor VIII (recombinate) in patients with hemophila A . Transfus. Med. Rev. 6, 137-45.

- Breakefield, X.O., DeLuca, N.A. (1991) Herpes simplex virus for gene delivery to neurons. New Biol. 3, 203-218.

- Breathnach R., Chambon, P. (1981) Organisation and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

- Brettler, D.B., Levine, P. (1989) Factor concentrates for treatment of Hemophilia: Which one to choose? Blood 73, 2067-2073.

- Brinkhous, K., Davis, P.D., Graham, J.B., Dodds, W.J. (1973) Expression and linkage of genes for X-linked hemophilias A and B in the dog. Blood, 41, 577-585.

- Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E. Palmiter, R.D. (1988) Introns increase transcriptional efficiency in transgenic mice. Proc. Natl. Acad. Sci. USA 85, 836-840.

- Bristol, J.A., Furie, B.C., Furie,B. (1993) Propeptide processing during factor IX biosynthesis. J. Biol. Chem. 268, 7577-7584.

- Bullock, P., Champoux, J.J., Botchan, M. (1985) Association of cross-over points with topoisomerase 1 cleavage sites: a model for non-homologous recombination. Science 230, 954-957.

- Burns, P.A., Neilson, F., Jack, A., Haddow, S., Balmain, A. (1991) Transformation of mouse skin endothelial cells in vivo by direct application of plasmid DNA encoding the human T-24 ras oncogene. Oncogene 6, 1973-1978.

- Buxton, J., Shelbourne, P., Davies, J., Jones, C., van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B., Williamson, R., Johnson, K. (1992) Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. Nature 355, 547-548.

- Caskey, C.T., Mitani, K. (1993) Delivering therapeutic genes- matching approach and application. TIBTECH 11, 162-166.

- Chalkey, G. Harris, A. (1991) Lymphocyte mRNA as a resource for detection of mutations and polymorphisms in the CF gene. J. Med. Genet. 28, 777-780.

- Chamberlain, J.S., Gibbs, R.A., Rainer, J.E., Nguyen, P.N., Caskey, C.T. (1988) Deletion screening of muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res. 16, 11141-11156.

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

- Chelly, J., Concordet, J.P., Kaplan, J.C., Kahn, A. (1989) Illegitimate transcription: Transcription of any gene in any cell type. Proc. Natl. Acad. Sci. USA, 86, 2617-2621.

- Chelly, J., Gilgenkrantz, H., Hugnot, J.P., Hamard, G., Lambert, M., Recan, D., Akli, S., Cometto, M., Kahn, A., Kaplan, J.C. (1991) Illegitimate transcription. Application to the analysis of truncated transcripts of the dystrophin gene in non-muscle cultured cells from Duchenne and Becker patients. J. Clin. Invest. 88, 1161-1166.

- Chen, S.H., Zhang, M., Thompson, A.R., Bray, G.L., Scott, C.R. (1991) Splice junction mutations in factor IX gene resulting in severe haemophilia B. Nucleic Acids Res. 19, 1172.

- Chomczynski, P., Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem., 162, 156-159.

- Choo, K.H., Gould, K.G., Rees, D.J.G., Brownlee, G.G. (1982) Molecular cloning of the gene for human anti-hemophilic factor IX. Nature 299, 178-180.

- Chowdhury, J.R., Grossman, M., Gupta, S., Chowdhury, N.R., Baker, J.R., Wilson, J.M. (1991) Long-term improvement of hypercholesterolaemia after ex vivo gene therapy in LDLR-deficient rabbits. Science 254, 1802-1805.

- Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalysed by prokaryotic and eukaryotic DNA polymerases. Nucleic Acids Res. 16, 9677-9686.

- Collins, P.W., Jenkins, P.V., Goldman, E., Lee, C.A., Pasi, K.J. (1994) Intron 22 inversions and haemophilia. Lancet, 343, 791.

- Compton, C., Gill, J.M., Bradford, D.A., Regauer, S., Gallico, G.G., O'Connor, N. (1989) Skin regenerated from cultured epithelial autografts on full-thickness burn wounds from 6 days to 5 years after grafting: A light, electron microscopy and immunohistochemical study. Lab. Invest. 60, 600-612.

- Connor, B.J., Reyes, A.A., Morin, C., Hakura, K., Teplitz, R.L., Wallace, R.B. (1983) Detection of sickle cell β S-globin allele by hybridisation with synthetic oligonucleotides. Proc. Natl. Acad. Sci. USA 80, 278-282.

- Connor, J.M., Pettigrew, A.F., Hann, I.M., Forbes, C.D., Lowe, G.D.O., Affara, N. (1985) Application of an intragenic genomic probe to genetic counselling for hemophilia B in the West of Scotland. J. Med. Genet. 22, 441-446.

- Connor, J.M., Pettigrew, A.F., Schiach, C., Hann, I.M., Lowe, G.D.O., Forbes, C.D. (1986) Application of three intragenic DNA polymorphisms for carrier detection in hemophila B. J. Med. Genet. 23, 300-309.

- Cooper, D.N., Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease Hum. Genet. 78, 151-155.
- Cooper, D.N., Krawczak, M. (1993) Human Gene Mutation, Bios Scientific Publishers.

- Cotten, M., Wagner, E., Zatloukal, K., Phillips, S., Curiel, D.T., Birnsteil, M.L. (1992) Proc. Natl. Acad. Sci. 89, 6094-6098.

- Cotton, R.G.H., Rodrigues, N.R., Campbell, R.D. (1988) Reactivity of cytosine, thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc.Natl.Acad. Sci. USA, 85, 4397-4401.

- Coulondre, J., Miller, J.H., Farabaugh, P.J., Gilbert, W. (1978) Molecular basis of the substitution hotspots in Eschericia Coli. Nature 274, 775-780.

- Cui, W., Fowlis, D.J., Cousins, F.M., Duffie, E., Bryson, S., Purchio, A.F., Balmain, A., Akhurst, R.J. (1994) Concerted action of TGF1 and its type II receptor in control of epidermal homeostasis in transgenic mice. In Press.

- Culver, K.W., Anderson, W.F., Blaese, R.M. (1991) Lymphocyte gene therapy. Hum. Gene Ther. 2, 107-109.

- Cutting, G.R., Kasch, L.M., Rosenstein, B.J., Zielenski, J., Tsui, L.C., Antonarakis, S.E., Kazazian, H.H. (1990) A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis conductance regulator protein. Nature 346, 366-369.

- Dai, Y., Roman, M., Naviaux, R.K., Verma, I.M. (1992) Gene therapy via primary myoblasts: Long term expression of factor IX protein following transplantation in vivo. Proc. Natl. Acad. Sci. USA 89, 10892-10895.

- Darby, S.C., Rizza, C.R., Doll, R., Spooner, R.J.R., Stratton, I.M., Thakrar, B. (1989) Incidence of AIDS and excess of mortality with HIV in haemophiliacs in the United Kingdom: report on behalf of the directors of haemophilia centres in the United Kingdom. Br. Med. J. 298, 1064-1068.

- Davidson, B.L., Allen, E.D., Kozarsky, K.F., Wilson, J.M., Roessler, B.J. (1993) A model system for in vivo gene transfer into the central nervous system using an adenoviral vector. Nature Genet. 3, 219-223.

- Davie, E.W. Ratnoff, O.D. (1964) Waterfall sequence for intrinsic blood clotting. Science 145, 1310-1311.

- Devereux, J., Haeberli, P., Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res., 12, 387-395.

- Dhawan, J., Pan, L.C., Pavlath, G.K., Travis, M.A., Lanctot, A.M., Blau, H.M. (1991) Systemic delivery of human growth hormone by injection of genetically engineered myoblasts. Science 254, 1509-1512.

- Diamond, C., Kogan, S., Levinson, B., Gitschier, J. (1992) Amino acid substitutions in conserved domains of factor VIII and related proteins: study of patients with mild and moderately severe hemophilia A. Hum. Mut. 1, 248-257.

- Dichek, D.A., Neville, R.F., Zwiebel, J.A., Freeman, S.M., Leon, M.B., Anderson, W.F. (1989) Seeding of intravascular stents with genetically engineered endothelial cells. Circulation 80, 1347-1353.

- DiScipio, R.G., Kurachi, K., Davie, E.W. (1978) Activation of human factor IX (Christmas factor) J. Clin. Invest. 61, 1528-1538.

- Dominski, Z., Kole, R. (1991) Selection of splice sites in pre-mRNAs with short internal exons. Mol. Cell. Biol. 11, 6075-6083.

- Dunn, J.M., Phillips, R.A., Zhu, X., Becker, A., Gallie, B.L. (1989) Mutations in the RB1 gene and their effects on transcription. Mol. Cell. Biol. 9, 4596-4604.

- Dunnen, J.T., den Bakker, E., Klein Breteler, E.G., Pearson, P.L., van Ommen J.B. (1987) Direct detection of more than 50% of the Duchenne muscular dystrophy mutations by field inversion gels. Nature 329, 640-642.

- Dzau, V.J., Morishita, R., Gibbons, G. (1993) Gene therapy for cardiovascular disease. TIBTECH 11, 205-210.

- Dzierzak, E.A., Papayannopulou, T., Mulligan, R.C. (1988) Lineage-specific expression of a human beta-globin gene in murine bone marrow transplant recipients reconstituted with retrovirus-transduced stem cells. Nature 331, 35-41.

- Eaton, D., Rodriguez, H., Vehar, G.A. (1986) Proteolytic processing of human factor VIII: correlation of specific cleavages by thrombin, factor Xa and activated protein C with activation and inactivation of factor VIII:C coagulant activity. Biochemistry 25, 505-512.

- Eaton, D.L., Hass, P.E., Riddle, L. (1987) Characterisation of human recombinant factor VIII. J. Biol. Chem. 262, 3285-3290.

- Eckert, R.L., Green, H. (1986) Structure and evolution of the human involucrin gene. Cell 46, 583-589.

- Eckert, K.A., Kunkel, T.A. (1991) The fidelity of DNA polymerases used in PCR. In: PCR A practical approach. Eds. M.J. McPherson, P.Quirke, G.R.Taylor. IRL Press Oxford, pp 227-246.

- Elder, B., Lakich, D., Gitschier, J. (1993) Sequence of the murine factor VIII cDNA. Genomics, 16, 374-379.

- European Working Group on CF Genetics (1990) Gradient of distribution in Europe of the major CF mutation and of its associated haplotype. Hum. Genet. 85, 436-445.

- Evans, J.P., Brinkhous, K.M., Brayer, G.D., Reisner, H.M., High, K.A. (1989) Canine hemophilia B resulting from a point mutation with unusual consequences. Proc. Natl. Acad. Sci. USA 86, 10095-10099.

- Fass, D.N., Knutson, G.J., Katzmann, J.A (1982) Monoclonal antibodies to porcine factor VIII and their use in the isolation of active coagulant protein. Blood 59, 594-600.

- Fay, P.J., Smudzin, T.M., Walker, F.J. (1991) Activated protein-C catalysed inactivation of human factor VIII and factor VIIIa: identification of cleavage sites and correlation of proteolysis with co-factor activity. J. Biol Chem. 266, 20139-20145.

- Fay, P.J., Haidari, P.J., Smudzin, T.M. (1991) Human factor VIIIa subunit structure. Reconstitution of factor VIIIa from the isolated A1/A3-C1-C2 dimer and A2 subunit. J. Biol. Chem 266, 8957-8962.

- Felgner, P.L., Ringold, G.M. (1989) Cationic liposome-mediated transfection. Nature 337, 387-388.

- Fenjves, E.S., Gordon, D.J., Pershing, L.K., Williams, D.L., Taichman, L.B. (1989) Systemic distribution of apolipoprotein E secreted by grafts of epidermal keratinocytes: Implications for epidermal function and gene therapy. Proc. Natl. Acad. Sci. USA 86, 8803-8807.

- Ferry, N., Duplessis, O., Houssin, D., Danos, O., Heard, J.M. (1991) Retroviral-mediated gene transfer into hepatocytes in vivo. Proc. Natl. Acad. Sci. USA 88, 8377-8381.

- Fischer, S.G., Lerman, L.S. (1983) DNA fragments differing by single base substitutions are separated in denaturing gradient gels. Correspondence with melting theory. Proc. Natl. Acad. Sci. USA 80, 1579-1583.

- Fisher, C.W., Fisher, C.F., Chuang, J.L., Lau, K.S., Chuang, D.T., Cox, R.P. (1993) Occurrence of a 2-bp (AT) deletion allele and a nonsense (G-to-T) mutant allele at the E2 (DBT) locus of six patients with Maple Syrup urine disease: Multiple-exon skipping as a secondary effect of the mutations. Am. J. Hum. Genet. 52, 414-424.

- Flowers, M.E.D., Stockschlaeder, M.A.R., Schuening, F.G., Niederwieser, D., Hackman, R., Miller, A.D., Storb, R. (1990) Long-term transplantation of canine keratinocytes made resistant to G418 through retrovirus mediated gene transfer. Proc. Natl. Acad. Sci. USA 87, 2349-2353.

- Fonknechten, N., Chomel, J., Kitzis, C., Kahn, A., Kaplan, J.C. (1992) Skipping of exon 5 as a consequence of the 711+1G to T mutation in the CFTR gene. Hum. Mol. Genet. 1, 281-282.

- Forrest, S.M., Cross, G.S., Speer, A., Gardner-Medwin, D., Burn, J., Davies, K.E. (1987) Preferential deletion of exons in Duchenne and Becker muscular dystrophies. Nature 329, 638-640.

- Forrest, S.M., Cross, G.S., Flint, T., Speer, A.., Robson, K.J.H., Davies, K.E. (1988) Further studies of deletions that cause Duchenne and Becker muscular dystrophies. Genomics 2, 109-114.

- Foster, P.A., Fulcher, C.A., Houghton, R.A., Zimmerman, T.S. (1988) An immunogenic reaction within residues Val1670-Glu1684 of the factor VIII light chain induces antibodies which inhibit binding of factor VIII to von Willebrand factor. J. Biol. Chem. 263, 5230-5234.

- Foster, P.A., Fulcher, C.A., Houghton, R.A., Zimmerman, T.S. (1990) synthetic factor VIII peptides with amino acid sequences contained within the C2 domain of factor VIII inhibit factor VIII binding to phosphatidylserine. Blood, 75, 1999-2004.

- Fowler, W.E., Fay, P.J., Arvan, D.S., Marder, V. (1990) Electron microscopy of human factor V and factor VIII: Correlation of morphology with domain structure and localisation of factor V activation fragments. Proc. Natl. Acad. Sci. USA 87, 7648-7652.

- Freidmann, T., Jinnah, H. (1993) Gene therapy for disorders of the nervous system. TIBTECH 11, 192-196.

- Freije, D., Schlessinger, D.A. (1992) A 1.6-Mb contig of yeast artificial chromosome around the human factor VIII gene reveals three regions homologous to probes for the DXS115 locus and two for the DXYS64 locus. Am. J. Hum. Genet. 51, 66-80.

- Friedmann, T. (1989) Progress toward human gene therapy. Science 244, 1275-1281.

- Fu, Y.H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, R.G., Warren, S.T., Ostra, B.A., Nelson, D.L., Caskey, C.T. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67, 1047-1058.

- Fuchs, E. (1988) Keratins as biochemical markers of epithelial differentiation. Trends in Genetics 4, 277-281.

- Fuchs, E., Esteves, R.A., Coulombe, P.A. (1988) Transgenic mice expressing a mutant keratin 10 gene reveal the likely genetic basis for epidermolytic hyperkeratosis. Proc. Natl. Acad. Sci. USA 89, 6906-6910.

- Fuchs, E. (1990) Epidermal differentiation: The bare essentials. J. Cell Biol. 111, 2807-2814.

- Fuchs, E. (1991) Keratin genes, epidermal differentiation and animal models for the study of human skin diseases. Biochem. Soc. Trans. 1112-1115.

- Fuchs, E., Green, H. (1991) Regulation of terminal differentiation of cultured human keratinocytes by vitamin A. Cell 25, 617-625.

- Fuchs, E., Aneskievich, B., Bryne, C., Chan, Y.M., Cheng, J., Faus,I., Hsu, H., Hutton, M.E. Syder,A., Turksen,K., Yang, Z., Zhou, P., Zinkel, S. (1994) Mechanisms controlling keratinocyte specific gene expression: Implications for gene therapy and for elucidating the genetic basis of skin diseases. J. Cell. Biochem. Suppl 18A, 220.

- Fulcher, C.A., Zimmerman, T.S. (1982) Characterisation of the human factor VIII procoagulant protein with a heterologous precipitating antibody. Proc. Natl. Acad. Sci. USA 79, 1648-1652.

- Fulcher, C.A., Roberts, J.R., Zimmerman, T.S. (1983) Thrombin proteolysis of purified factor VIII procoagulant protein. Correlation of activation with generation of specific polypeptide. Blood 61, 807-811.

- Fulcher, C.A., Gardiner, J.E., Griffin, J.H., Zimmerman, T.S. (1984) Proteolytic inactivation of human factor VIII procoagulant protein by activated human protein C and its analogy with factor V. Blood 63, 486-489.

- Furie, B, Furie, B.C. (1988) The Molecular basis of blood coagulation. Cell 53, 505-518.

- Gallico, G.G., O'Connor, N.E., Compton, C.C., Kehnide, O., Green, H. (1984) Permanent coverage of large burn wounds with autologous cultured human epithelium. New Engl. J. Med. 311, 448-451.

- Gaucher, C., Jorieux, S., Mercier, B., Oufkir, D., Mazurier, C. (1991) The Normandy variant of von Willebrand disease: Characterisation of a point mutation in the von Willebrand factor gene. Blood 77, 1937-1941.

- Gerrard, A.J., Austen, D.E. Brownlee, G.G. (1992) Subcutaneous injection of factor IX for the treatment of hemophilia B. British J. Hematol. 81, 610-613.

- Gerrard, A.J., Hudson, D.L., Brownlee, G.G., Watt, F.M. (1993) Towards gene therapy for hemophilia B using primary human keratinocytes. Nature Genet. 3, 180-193.

- Giannelli, F., Anson, D.S., Choo, K.H. (1984) Characterisation and use of an intragenic polymorphic marker for the detection of carriers of heamophila B (factor IX deficiency). Lancet 1, 239-241.

- Giannelli, F., Green, P.M., High, K.A., Sommer, S., Poon, M.C., Ludwig, M., Schwaab, R., Reitsma, P.H., Goosens, M., Yoshoika, A., Brownlee, G.G. (1993) Haemophilia B: database of point mutations and short additions and deletions- fourth edition. Nucleic Acids Res. 21, 3075-3087.

- Gitschier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.E., Vehar, G.A., Capon, D.J., Lawn, R.M. (1984) Characterization of the human factor VIII gene. Nature, 312, 326-330.

- Gitschier, J., Drayna, D., Tuddenham, E.G.D., White, R.L., Lawn, R.M. (1985a) Genetic mapping and diagnosis of haemophila A achieved through a *Bcl* I polymorphism in the factor VIII gene. Nature 314, 738-740.

- Gitschier, J., Lawn, R.M., Rotblat, F., Goldman, E., Tuddenham, E.G.D. (1985b) Antenatal diagnosis and carrier detection of hemophilia using a factor VIII gene probe. Lancet 1, 1093-1094.

- Gitschier, J., Wood, W.I., Tuddenham, E.G.D., Shuman, M.A., Goralka, T.M., Chen, E.Y., Lawn, R.M. (1985c) Detection and sequence of mutations in the factor VIII gene of haemophiliacs. Nature, 315, 427-430.

- Gitschier, J., Kogan, S., Levinson, B., Tuddenham, E.G.D. (1988) Mutations of factor VIII cleavage sites in hemophilia A. Blood 72, 1022-1028.

- Goldberg, A.L., St. John, A.C. (1976) Intracellular protein degradation in mammalian and bacterial cells: Part 2 Ann. Rev. Biochem. 45, 747-803,

- Goodeve, A.C., Preston, F.E., Peake, I.R. (1994) Factor VIII gene rearrangements in patients with severe haemophilia A. Lancet, 343, 329-330.

- Graves, R.A., Pandey, N.B., Chodchoy, N. Marzluff, W.F. (1987) Translation is required for regulation of histone mRNA degradation. Cell 48, 615-626.

- Green, H. (1989) Regeneration of the skin after grafting of epidermal cultures. Lab. Invest. 60, 583-584.

- Greenhalgh, D.A., Rothnagel, J.A., Quintanilla, M.I., Orengo, C.C., Gagne, T.A., Bundman, D.S., Longley, M.A. Roop, D.R. (1993) Induction of epidermal hyperplasia, hyperkeratosis, and papillomas in transgenic mice by a targeted v-Ha-ras oncogene. Molecular carcinogenesis 7, 99-110.

- Grompe, M. (1993) The rapid detection of unknown mutations in nucleic acids. Nature Genet., 5, 111-117.

- Grunert, S., Jackson, R.J. (1994) The immediate downstream codon strongly influences the efficiency of utilisation of eukaryotic translation initiation codons. The EMBO J. 13, 3618-3630.

- Gruss, P., Khoury, G. Rescue of a splicing defective mutant by insertion of a heterologous intron. Nature 286:634-637, 1980.

- Gyllensten, U.B. (1989) Direct sequencing of in vitro amplified DNA. In: PCR Technology Principles and applications for DNA amplification Ed. H. Erlich, Stockton Press, pp 45-60.

- Gyllensten, U.B., Erlich, H.A. (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc. Natl. Acad. Sci. USA, 85, 7652-765

- Hamer, D., and Leder, P. Splicing and the formation of stable RNA. (1979) Cell 18, 1249-1302.

- Handford, P.A., Baron, M., Mayhew, H., Willis, A., Beesley, T., Brownlee, G.G. Campbell, I.D. (1990) The first EGF-like domain from human factor IX contains a high-affinity calcium site. EMBO J. 9, 475-480.

- Handin, R.I., Wagner, D.D. (1989) Molecular and cellular biology of von Willebrand factor. Prog. Hemost. Thromb. 9, 233-259.

- Harley, H.G., Brook, J.D., Rundle, S.A., Crow, S., Reardon, W., Buckler, A.J., Harper, P.S., Houseman, D.E., Shaw, D.J. (1992) Expansion of an unstable DNA region and phenotype variation in myotonic dystrophy. Nature 355, 545-546.

- Hase, S., Nishimura, H, Kawabata, S.I., Iwanaga, S., Ikenaka, T. (1990) The structure of (xylose)2 Glucose-O-Serine 53 found in the first epidermal growth factor-like domain of bovine blood clotting factor IX. J. Biol.Chem. 265, 1858-1861.

- Hawkins, J.D. (1988) A survey of intron and exon lengths. Nucleic Acids Res. 16, 9893-9908.

- Hay, C.R.M., Preston, F.E., Trigger, D.R., Underwood, J.C.F. (1985) Progressive liver disease in haemophila A: An understated problem. Lancet 1, 1495-1498.

- Hayashi, K. (1991) PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. PCR Methods Appl. 1, 34-38.

- Higuchi, R. (1989) Using PCR to engineer DNA. In: PCR Technology Principles and applications for DNA amplification Ed. H. Erlich Stockton Press, pp 61-70.

- Higuchi, M., Wong, C. Kochlan, L., Olek, K., Aronis, S., Kasper, C.K., Kazazian, H.H., Antonarakis, S.E. (1990) Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA. Genomics, 6, 65-71.

- Higuchi, M., Antonarakis, S.E., Kasch, L., Oldenburg, J., Economou, E.P., Olek, K., Arai, M., Inaba, H., Kazazian, H.H., (1991a) Molecular characterization of mild-to-moderate hemophilia A: Detection of the mutation in 25 of 29 patients by denaturing gradient gel electrophoresis. Proc. Natl. Acad. Sci. USA., 88, 8307-8311.

- Higuchi, M., Kazazian, H.H., Kasch, L., Warren, T.C., McGinniss, M.J., Phillips, J.A., Kasper, C., Janco, R., Antonarakis, S.E. (1991b) Molecular characterization of severe hemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the factor VIII gene. Proc. Natl. Acad. Sci. USA., 88, 7405-7409.

- Hill-Eubanks, D.C., Lollar, P. von Willebrand factor is a co-factor for thrombin catalysed cleavage of the factor VIII light chain. J. Biol.Chem. 265, 17854-17858.

- Hoeben, R.C., van der Jagt, R.C.M., Schoute, F., van Tilburg, N.H., Verbee, M.P., Briet, E., van Ormondt, H., van der Eb, A.J. (1990) Expression of functional factor VIII in primary human skin fibroblasts after retrovirus mediated gene transfer. J. Biol. Chem. 256, 7318-7323.

- Hoeben, R.C., Einerhand, M.P.W., Briet, E., van Ormondt H., Valerio, D., van der Eb, A.J. (1992) Toward gene therapy in hemophila A: retrovirus-mediated transfer of a factor VIII gene into murine hematopoetic progenitor cells. Thrombosis and Haemostasis 67, 341-345.

- Holton, T.A., Graham, M.W. (1990) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. 19, 1156.

- Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., Pease, L.R. (1989) Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. Gene 77, 61-68.

- Hoyer, L.W. (1981) The factor VIII complex: Structure and function. Blood 58, 1-13.
- Hoyer, L.W. (1994) Hemophilia A. New Engl. J. Med. 330, 38-47.
- Huang, M.T., Gorman, C.M. (1990) The Simian virus 40 small-t intron, present in many common expression vectors, leads to aberrant splicing. Mol. Cell. Biol. 10, 1805-1810.
- Hutter, P., Antonarkis, S.E., Delozier-Blanchet, C., Morris, M.A. (1994) Exon skipping associated with a A to G transition at +4 of the IVS 33 splice donor site of the neurofibromatosis type 1 (NF-1) gene. Hum. Mol. Genet., 3, 663-665.
- Israel, D.I., Kaufman, R.J. (1989) Retroviral mediated transfer and amplification of a functional human factor VIII gene. Blood 75, 1074-1080.
- Jackson, R.J. (1993) Cytoplasmic regulation of mRNA function: The importance of the 3' untranslated region. Cell 74, 9-14.
- Jagadeeswaran, P., Lavelle, D.E., Kaul, R., Mondas, T., Warren, S.T. (1984) Isolation and characterisation of human factor IX cDNA: Identification of Taq I polymorphism and regional assignement. Som. Cell mol. Genet. 10, 465-473.
- Jallat, S., Perraud,F., Dalemans, W., Balland, A., Dieterle, A., Faure, T., Meullien, P., Pavirani, A. (1990) Characterisation of recombinant human Factor IX expressed in transgenic mice and in derived transimmortalised hepatic cell lines. The EMBO J. 9, 3295-3301.
- Jaye, M., de la Salle, H., Schamber, F., Balland, A., Kohli, V., Findeli, A., Tolstoshev, P., Lecocq, J.P. (1983) Isolation of a human anti-hemophilic factor IX cDNA clone using a unique 52-base synthetic oligonucleotide probe deduced from the amino acid sequence of bovine factor IX. Nucleic Acids Res. 11, 2325-2335.
- Jenny, R.J., Pittman, D.D., Toole, J.J., Kriz, R.W., Aldape, R.A., Hewick, R.M., Kaufmann, R.J., Mann, K.G. (1987) Complete cDNA and derived amino acid sequences of human factor V. Proc. Natl. Acad. Sci. USA 84, 4846-4850.
- Jensen, P.K.A., Bolund, L. (1991) Tissue culture of human epidermal keratinocytes:a differentiating model system for gene testing and somatic gene therapy. J. Cell Sci. 100, 255-259.
- Jiang, C.K., Epstein, H.S., Tomic, M., Freedberg, I.M., Blumenberg, M. (1991) Functional comparison of the upstream regulatory DNA sequences of four human epidermal keratin genes. J. Invest. Dermatol. 96, 162-167.
- Jones, P.H., Watt, F.M. (1993) Separation of human epidermal stem cells from transit amplifying cells on the basis of differences in integrin function and expression. Cell 73, 713-724.
- Jones, P.K., Ratnoff, O.D. (1991) The changing prognosis of classical hemophilia (factor VIII deficiency) Ann. Int. Med. 114, 641-648.

- Kaleko, M., Garcia, J.V., Osborne, W.R.A., Miller, A.D. (1990) Expression of human adenosine deaminase in mice after transplantation of gentically modified bone marrow. Blood 75, 1733.

- Kane, W.H., Davie, E.W.(1986) Cloning of a cDNA for coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin. Proc. Natl. Acad. Sci. USA 83, 6800-6804.

- Kane, W.H., Ichinose, A., Hagan, F.S., Davie, E.W. (1987) Cloning of cDNAs coding for the heavy chain region and connecting region of human factor V, a blood coagulation factor with four types of internal repeats. Biochemistry, 26, 6508-6514.

- Kane, W.H., Davie, E.W. (1988) Blood coagulation factors V and VIII: Structural and functional similarities and their relationship to hemorrhagic and thrombotic disorders. Blood, 71, 539-555.

- Kaneda, Y., Iwai, K., Uchida, T. (1989) Increased expression of DNA cointroduced with nuclear protein in adult rat liver. Science 243, 375-378.

- Kasper, C.K., Sterud, B., Minami, J.Y., Shonick, W., Rapaport, S.I. (1977) Hemophilia B: Characterisation of variants and detection of carriers. Blood 50, 351-366.

- Katayama, K., Ericsson, L.H., Enfield, D.L. (1979) Comparison of amino acid sequence of bovine coagulation factor IX (Christmas factor) with that of other vitamin-K dependent plasma proteins. Proc. Natl. Acad. Sci. USA 76, 4990-4994.

- Kaufman, R.J., Wasley, L.C., Dorner, A.J. (1988) Synthesis, Processing and Secretion of recombinant human VIII expressed in mammalian cells. J. Biol. Chem. 263, 6352-6362.

- Kaufman, R.J., Wasley, L.C., Davies, M.V., Wise, R.J., Israel, D.I. Dorner, A.J. (1989) Effect of von Willebrand factor co-expression on the synthesis and secretion of factor VIII in Chinese hamster ovary cells. Mol. Cell. Biol. 9, 1233-1242.

- Kay, M.A., Rothenberg, S., Landen, C.N., Bellinger, D.A., Leland, F., Toman, C., Finegold, M., Thompson, A.R., Read, M.S., Brinkhous, K., Woo, S.L. (1993) In vivo gene therapy of Hemophilia B: Sustained partial correction in Factor IX-deficient dogs. Science, 262, 117-119.

- Klein, H.G., Aledort, L.M., Bouma, B.N., Hoyer, L.W., Zimmerman, T.S., DeMets, D.L. (1977) A co-operative study for the detection of the carrier state of classical hemophilia. New Engl. J. Med. 296, 959-962.

- Kochinsky, M.L., Funk, W.D., van Oost, V.A., MacGillivray, R.T.A. (1986) Complex cDNA sequence of human preceruloplasmin. Proc. Natl. Acad. Sci. USA., 83, 5086-5090.

- Kogan, S.C., Doherty, M.B.S., Gitschier, J.D. (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences: Application to hemophilia A. New Engl. J. Med. 317, 985-90.

- Kogan, S.C., Gitschier, J. (1990) Genetic prediction of hemophilia A, in PCR Protocols: A guide to methods and applications Academic press, Inc. pp 289-299.
- Konieczny, S.F., Emersen, C.P. (1987) Complex regulation of the muscle-specifc contractile protein (troponin-I) gene. Mol. Cell. Biol. 7, 3065-3075.
- Kopan, R., Traska, G., Fuchs, E. (1987) Retinoids as important regulators of terminal differentiation: examining keratin expression in individual epidermal cells at various stages of keratinisation. J. Cell. Biol 105, 427-440.
- Kozak, M. (1986a) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283-292.
- Kozak, M. (1986b) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes Proc. Natl. Acad. Sci. USA. 83, 2850-2854.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucl. Acids Res. 15, 8125-8148.
- Kozak, M. (1989a) Context effects and inefficient initiation at non-AUG codons in eukaryotic cell-free translation systems. Mol. Cell. Biol. 9, 5073-5080.
- Kozak, M. (1989b) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. Mol. Cell. Biol. 9, 5134-5142.
- Kozak,M. (1989c) The scanning model for translation: An update. J. Cell Biol. 108, 229-241.
- Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribososmes. Proc. Natl. Acad. Sci. USA 87, 8301-8305.
- Kozak, M. (1991a) An analysis of vertebrate mRNA sequences: Intimations of translational control. J. Cell Biol. 115, 887-903.
- Kozak, M. (1991b) Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. 266, 19867-19870.
- Kozak, M. (1991c) A short leader sequence impairs the fidelity of inititation by eukaryotic ribosomes. Gene Expression 1, 111-115.
- Kozak, M. (1991d) Effects of long 5' leader sequences on initiation by eukaryotic ribosomes in vitro. Gene Expression 1, 117-125.
- Kozak, M. (1992) Regulation of translation in eukaryotic systems Annu. Rev. Cell Biol. 8, 197-225.
- Kozak, M. (1994) Features in the 5' non-coding sequences of rabbit α and β-globin mRNAs that affect translational efficiency. J. Mol. Biol. 235, 95-110.

- Krawczak, M., Cooper, D.N. (1991) Gene deletions causing human genetic disease: Mechanisms of mutagenesis and the role of the local DNA sequence environment. Human Genet. 86, 425-441.

- Krawczak, M., Reiss, J., Cooper, D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum. Genet. 90, 41-54.

- Kriegler, M. (1990a) Gene transfer and expression. A laboratory manual. Ed. Kriegler, M. Stockton Press.

- Kriegler, M. (1990b) Assembly of enhancers, promoters, and splice signals to control expression of transferred genes. Meth. Enzymol. 185, 512-527.

- Kunkel, L.M., Smith, K.D., Boyer, S.H. and Borgaonkar, D.S. (1977) Analysis of human Y chromosome specific reiterated DNA in chromosome variants. Proc. Natl Acad Sci USA 68, 120-124.

- Kunkel, T.A., Soni, A. (1988) Mutagenesis by transient misalignement. J. Biol. Chem. 263, 14784-14789.

- Kurachi, K., Davie, E.W. (1982) Isolation and characterisation of a cDNA coding for human factor IX. Proc. Natl. Acad. Sci. USA 79, 6461-6464.

- Kwok, S., Kellog, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C. Sninsky, J.J. (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type-1 model studies. Nucleic Acids Res. 18, 999-1005.

- Laird, P.W, Zijderveld, A., Linders, K., Rudnicki, A., Jaenisch, R. and Berns, A. (1991) Simplified mammalian DNA isolation procedure. Nucl. Acids Res. 19, 4293.

- Lakich, D., Kazazian, H., Antonarakis, S.E., Gitschier, J. (1993) Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nature Genetics, 5, 236-241.

- Lalloz, M.R., McVey, J.H., Pattinson, J.K., Tuddenham, E.G.D. (1991) Heamophilia A diagnosis by analysis of a hypervariable dinucleotide repeat within the factor VIII gene. Lancet, 338, 207-211.

- Langlois, S., Deeb, S., Brunzell, J.D., Kastelein, J.J., Hayden, M.R. (1989) A major insertion accounts for a significant proportion of mutations underlying human lipoprotein lipase deficiency. Proc. Natl. Acad. Sci. USA. 86, 948-952.

- Larsson, S.A. (1985) Life expectancy of Swedish hemophiliacs 1831-1980. British J. Hematol. 59, 593-602.

- Leask, A., Bryne, C., Fuchs, E. (1991) Transcription factor AP2 and its role in epidermal-specific gene expression. Proc. Natl. Acad. Sci. USA 88, 7948-7952.

- Leclerc, G., Gal, D., Takeshita, S., Nikol, S., Weir, L., Isner, J.M. (1992) Percutaneous arterial gene transfer in a rabbit model. Efficiency in normal and balloon-dilated atherosclerotic arteries. J. Clin. Invest. 90, 936-944.

- Lee, J.I., Taichman, L.B. (1989) Transient expression of a transfected gene in cultured epidermal keratinocytes: Implications for future studies. J. Invest. Dermatol. 92, 267-271.

- Lenhert, M.E., Jorcano, J., Zentgraf, H., Blessing, M., Franz, J.K., Franke, W. (1984) Characterisation of bovine keratin genes: similarities of exon patterns in genes coding for different keratins. The EMBO J. 3, 3279-3287.

- Lersch, R., Stellmach, V., Stocks, C., Giudice, G., Fuchs, E. (1989) Isolation, sequence, and expression of a human keratin K5 gene: transcriptional regulation of keratins and insights into pairwise control. Mol. Cell. Biol. 9, 3685-3697.

- Levinson, B., Kenwrick,S., Lakich, D., Hammonds, G., Gitschier, J. (1990) A transcribed gene in an intron of the factor VIII gene. Genomics 7, 1-11.

- Levinson, B., Kenwrick, S., Gamel, P., Fisher, K., Gitschier, J. (1992) Evidence for a third transcript from the human factor VIII gene. Genomics 14, 585-589.

- Lewin, B. (1994) Genes V, Oxford University Press.

- Li, Y., Bollag, G., Clark, R., Stevens, J., Conroy, L., Fults, D., Ward, K., Freidman, E., Samowitz, W., Robertson, M., Bradley, P., McCormick, F., White, R. and Cawthon, R. (1992) Somatic mutations in the neurofibromatosis-1 gene in human tumours. Cell 69, 275-281.

- Lim, S.K., Sigmund, C.D., Gross, K.W., Maquat, L.E. (1992) Nonsense codons in human β-globin mRNA result in the production of mRNA degradation products. Mol. Cell. Biol. 12, 1149-1161.

- Lin, S.W., Smith, K.J., Welsch, D., Stafford, D.W. (1990) Expression and characterisation of human factor IX and factor IX-factor X chimaeras in mouse C127 cells. J. Biol. Chem. 265, 144-150.

- Lin, H., Parmacek, M.S., Morle, G., Bolling, S., Leiden, J.M. (1990) Expression of recombinant genes in myocardium in vivo after direct injection of DNA. Circulation 82, 2217-2221.

- Lin, S.W., Lin, S.R., Shen, M.C. (1993) Characterization of genetic defects of hemophilia A in patients of Chinese origin. Genomics, 18, 496-504.

- Ljung, R., Petrini, P., Lindgren, A.C., Tengborn, L., Nilsson, I.M. (1992) Factor VIII and factor IX inhibitors in hemophiliacs. Lancet 339, 1550.

- Ljung, R.C.R. (1994) Intron 22 inversions and haemophilia. Lancet, 343, 791.

- Loeb, L.A., Kunkel, T.A. (1982) Fidelity of DNA synthesis. Ann. Rev. Biochem. 52, 429-457.

- Loeb, L.A., Preston, B.D. (1986) Mutagenesis by apurinic/apyrimidinic sites. Ann. Rev. Genet. 20, 201-230.

- Loeffler, J.P., Behr, J.P. (1993) Gene transfer into primary and established mammalian cell lines with lipopolyamine coated DNA. Meth. Enzymol. 217, 599-617.

- Lollar, P., Parker, C.G. (1990) pH-dependent denaturation of thrombin activated porcine factor VIII. J. Biol. Chem. 265, 1688-1692.

- Lorenzo, J.I., Garcia, R., Molina, R. (1992) Factor VIII and factor IX inhibitors in hemophiliacs. Lancet 339, 1551.

- Lozier, J.N., Thompson, A.R., Hu, C-P., Read, M., Brinkhous, K.M., High, K.A., Curiel, D.T. (1994) Efficient transfection of primary cells in a canine hemophilia model using adenovirus-polylysine-DNA complexes. Hum. Gene Ther. 5, 313-322.

- Lupski, J.R., Montes de Oca-Luna, R., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Chakravarti, A., Patel, P.I. (1991) DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell 66, 219-232.

- Lusher, J.M., Arkin, K.S., Abildgaard, C.F., Schwartz, R.S. (1993) (Kogenate previously unrelated patient study group). Recombinant factor VIII for the treatment of previously untreated patients with hemophilia A: safety, efficacy, and development of inhibitors. New Engl. J. Med. 328, 453-459.

- Lynch, C.M., Clowes, M.M., Osborne, W.R.A., Clowes, A.W., Miller, A.D. (1992) Long-term expression of human adenosine deaminase in vascular smooth muscle cells of rats: A model for gene therapy. Proc. Natl. Acad. Sci. USA 89, 1138-1142.

- Lynch, C.M., Israel, D.I., Kaufman,R.J., Miller, D.A. (1993) Sequences in the coding region of clotting factor VIII act as dominant inhibitors of RNA accumulation and protein production. Hum. Gene Ther. 4, 259-272.

- MacFarlane, R.G. (1964) An enzyme cascade in the blood clotting mechanism and its function as as a biochemical amplifier. Nature 202, 498-499.

- Malcolm, S., Cowell, J.K. Young, B.D. (1986) Specialist techniques in research and diagnostic clinical cytogenetics. In: Human Cytogenetics A Practical approach Eds. D.E. Rooney, B.H. Czepulowski IRL Press Oxford, pp 197-219.

- Mann, K.G., Jenny, R.J., Krishnaswamy, S. (1988) Cofactor proteins in the assembly and expression of blood clotting enzyme complexes. Ann. Rev. Biochem. 57, 915-956.

- Mann, K.G., Nesheim, M.E., Church,W.R., Haley, P., Krishnaswamy, S. (1990) Surface dependent reactions of the vitamin-K dependent enzyme-complexes. Blood 76, 1-16.

- Marchuk, D., Drumm, M., Saulino, A., Collins, F.S. (1990) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. 19, 1154.

- Marzluff, W.F. Pandey, N.B. (1988) Multiple regulatory steps control histone mRNA concentration. TIBS 13, 49-52.

- Mather, J.P. (1990) Optimizing cell and culture environment for production of recombinant proteins. Meth. Enzymol. 185, 567-577.

- Matsuhita, T., Tanimoto, M., Yamamoto, K., Suguira, I., Hamaguchi, M., Takamatsu, J., Saito, H. (1989) Nucleotide sequence analysis of hemophilia B with the inhibitor phenotype. Blood Suppl. 74, 251a.

- Mazurier, C., Dieval, J., Jorieux, S., Delobel, J., Goudemand, M. (1990) A new von Willebrand factor (vWF) defect in a patient with factor VIII (FVIII) deficiency but with normal levels and multimeric patterns of both plasma and platelet vWF: Characterisation of abnormal vWF/FVIII interaction. Blood 75, 20-26.

- McGinniss, M.J., Kazazian, H.H., Hoyer, L.W., Bi, L., Inaba, H., Antonarakis, S.E. (1993) Spectrum of mutations in CRM +ve and CRM -ve hemophilia A. Genomics 15, 392-398.

- Messerschmidt, A., Huber, R. (1990) The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin. Modelling and structural relationships. Eur. J. Biochem. 187, 341-352.

- Mgone, C.S., Lanyon, W.G., Moore, M.R., Connor, J.M. (1992), Detection of seven point mutations in the porphobilinogen deaminase gene in patients with acute intermittent porphyria by direct sequencing of in vitro amplified cDNA. Hum. Genet., 90:12-16.

- Michaud, J., Brody, L.C., Steel, G., Fontaine, G., Martin, L.S., Valle, D., Mitchell, G. (1992) Strand-separating conformational polymorphism analysis: Efficacy of detection of point mutations in the human ornithine delta amino transferase gene. Genomics 13, 389-394.

- Millar, D.S., Steinbrecher, R.A., Wieland, K., Grundy, C.B., Martinowitz, U., Krawczak, M., Zoll, B., Whitmore, D., Stephenson, J., Mibashan, R.S., Kakkar, V.V., Cooper, D.N. (1990) The molecular genetic analysis of hemophilia A; characterisation of six partial deletions in the factor VIII gene. Hum. Genet. 86, 219-227.

- Miller, A.D., Rosman, G.J. (1989) Improved retroviral vectors for gene transfer and expression. BioTechniques 7, 980-990.

- Miller, A.D. (1990) Progress toward Hum. Gene Ther. Blood 76, 271-278.

- Miller, A.D. (1992) Human gene therapy comes of age. Nature 357, 455-460.

- Milstone, L.M., Edelson, R.L. (1988) eds. Endocrine, metabolic and immunologic functions of keratinocytes. Ann. N. Y. Acad. Sci. 548, 1-366.

- Modrich, P. (1987) DNA mismatch correction. Ann. Rev. Biochem. 56, 435-466.
- Moll, R., Franke, W.W., Schiller, D.L., Geiger, R., Krepler, R. (1982) The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumours and cultured cells. Cell 31, 11-14.
- Moore, D.D., Marks, A., Buckley, D., Kapler, G., Payvar, F., Goodman, H.M. (1985) The first intron of the human growth hormone gene contains a binding site for glucocorticoid receptor. Proc. Natl. Acad. Sci. USA. 82, 699-702.
- Morgan, J.R., Barrandon, Y., Green, H., Mulligan, R. (1987) Expression of an exogenous growth hormone gene by transplantable human epidermal cells. Science 237, 1476-1479.
- Mosesson, M., Fass, D.N., Lollar, P., DiOrio, J.P., Parker, C.G., Knutson, G.J., Hainfield, J.F., Wall, J.S. (1983) Structural model of porcine Factor VIIIa molecules based on scanning transmission electron microscope (STEM) images and STEM mass analysis. J. Clin. Invest. 85, 1983-1990.
- Moullier, P., Bohl, D., Heard, J.M., Danos, O. (1993) Correction of lysosomal storage in the liver and spleen of MPS VII mice by implantation of genetically modified skin fibroblasts. Nature Genet. 4, 154-159.
- Mulligan, R.C., Howard, B.H., Berg, P. (1979) Synthesis of rabbit β-globin in cultured monkey kidney cells following infection with a SV40 β-globin recombinant genome. Nature 277, 108-111.
- Mulligan, R.C., Berg, P. (1980) Expression of a bacterial gene in mammalian cells. Science 209:1422-1427.
- Mulligan, R.C. (1993) The basic science of gene therapy. Science 260, 926-932.
- Myers, R.M., Fischer, S.G., Lerman, L.S., Maniatis, T. (1985a) Nearly all single base substitutions in DNA fragments joined to a GC clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res. 13, 3131-3145.
- Myers, R.M., Larin, Z., Maniatis, T. (1985b) Detection of single base substitution by ribonuclease A cleavage at mismatches in RNA:DNA duplexes. Science 230, 1242-1246.
- Nabel, E.G., Plautz, G., Boyce, F.M., Stanley, J.C., Nabel, G.J. (1989) Recombinant gene expression in vivo within endothelial cells of the arterial wall. Science 244, 1342-1344.
- Nabel, E.G., Plautz, G., Nabel, G.J. (1990) Site-specific gene expression in vivo by direct gene transfer into the arterial wall. Science 249, 1285-1288.

- Nabel, E.G., Plautz, G., Nabel, G.J. (1992) Transduction of a foreign histocompatibility gene into the arterial wall induces vasculitis. Proc. Natl. Acad. Sci. USA 89, 5157-5161.

- Nabel, G.J., Felgner,P.L. (1993) Direct gene transfer for immunotherapy and immunisation. TIBTECH 11, 211-215.

- Nakano, Y., Nishihara, T., Sasayama, S., Miwa, T., Kamada, S., Kakunaga, T. (1991) Transcriptional regulatory elements in the 5' upstream and first intron regions of the human smooth muscle (aortic type) α-actin-encoding gene. Gene 99, 285-289.

- Naylor, J.A., Green, P.M., Montandon, A.J., Rizza, C.R., Giannelli, F. (1991) Detection of three novel mutations in two haemophilia A patients by rapidly screening of whole essential region of factor VIII gene. Lancet, 337, 635-639.

- Naylor, J.A. Green, P.M. Rizza, C.R., Giannelli, F. (1992) Factor VIII gene explains all cases of hemophila A. Lancet 340, 1066-1067.

- Naylor, J.A., Green, P.M., Rizza, C.R., Gianelli, F. (1993a) Analysis of factor VIII mRNA reveals defects in everyone of 28 haemophilia A patients. Hum. Mol. Genet., 2, 11-17.

- Naylor, J., Brinke, A., Hasscock, S., Green, P.M., Gianelli, F. (1993b) Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions. Hum. Mol. Genet., 2, 1773-1778.

- Nilsson, I.M., Berntorp, E., Lofqvist, T. Pettersson, H. (1992) Twenty-five years experience of prophylactic treatment in severe hemophilia A and B. J. Intern Med. 232, 25-32.

- Nisen, P.D., Stamberg, J., Ehrenpreis R., Velasco, S., Shende, A., Engelberg, J., Kararyalcin, G., Waber, L. (1986) The molecular basis of hemophilia B in a girl. New Engl. J. Med. 315, 1139-1142.

- Nisen, P.D., Waber, P.G., (1989) Non-random X chromosome DNA methylation patterns in hemophiliac females. J. Clin. Invest. 83, 1400-1403.

- Nishimura, H., Takao, T., Hase, S., Shimonishi, Y., Iwanaga, S. (1992) J. Biol. Chem. 267, 17520-17525.

- Nishino, M., Girma, J.P., Rothschild, C., Fressinaud, E., Meyer, D. (1989) New variant of von Willebrand disease with defective binding to factor VIII. Blood, 74, 151-159.

- Novack, D.F., Casna, N.J., Fischer, S.G., Ford, J.P. (1986) Detection of single base pair mismatches in DNA: by chemical modification followed by electrophoresis in a 15% polyacrylamide gel. Proc. Natl. Acad. Sci. USA 84, 3339-3343.

- O'Brien, D.P., Pattinson, J.K., Tuddenham, E.G.D. (1990) Purification and characterisation of factor VIII 372-Cys: A hypofunctional co-factor from a patient with moderately severe hemophilia A. Blood 75, 1664-1672.

- O'Brien, D.P., Johnson, D., Byfield, P., Tuddenham, E.G.D. (1992) Inactivation of factor VIII by factor IXa. Biochemistry 31, 2805-2812.

- Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F., Mandel, J.L. (1991) Instability of a 550 base pair DNA segment and abnormal methylation in Fragile X syndrome. Science 252, 1097-1102.

- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA 86, 2766-2770.

- Osborne, W.R.A., Hock, R.A., Kaleko, M., Miller, A.D. (1990) Long term expression of human adenosine deaminase in mice after transplantation of bone marrow infected with amphotrophic retroviral vectors. Hum. Gene Ther. 1, 31.

- Ouzounis, C., Sander, C. (1991) A structure-derived sequence pattern for the detection of type I copper binding domains in distantly related proteins. FEBS Letters, 279, 73-78.

- Pakula, A.A. Sauer, R.T. (1989) Genetic analysis of protein stability and function. Ann. Rev. Genet. 23, 289-310.

- Palmer, T.D., Thompson,A.R., Miller,A.D. (1989) Production of human factor IX in animals by genetically modified skin fibroblasts: Potential therapy for hemophilia B. Blood 73, 438-445.

- Palmer, T.D., Rosman, G.J., Osborne, W.R.A., Miller, A.D. (1991) Genetically modified skin fibroblasts persist long after transplantation but gradually inactivate introduced genes. Proc. Natl. Acad. Sci. USA 88, 1330-1334.

- Patterson, M.N., Bell, M.V., Bloomfield, J., Flint, T., Dorkins, H., FrosterIskenius, U., Sommer, S., Sobell, J., Sciad, D., Thibodeau, S., Davies, K.E. (1989) An intronic region within the factor VIII gene is duplicated within Xq28 and is homologous to the polymorphic locus DXS115 (767). Am. J. Hum. Genet. 44, 679-688.

- Pattinson, J.K., Millar, D.S., Grundy, C.B., Wieland, K., Mibashan, R.S., Martinowitz, U., McVey, J., Tan-Un, K., Vidaud, M., Goossens, M., Sampietro, M., Camerino, G., Krawcak, M., Reiss, J., Zoll, B., Whitmore, D., Bradshaw, A., Wensley, R., Ajani, A., Mitchell, V., Rizza, C., Maia, R., Winter, P., Mayne, E.E., Schwartz, M., Kakkar, V.V., Tuddenham, E.G.D., Cooper, D.N. (1990a) The molecuar genetic analysis of hemophilia A; a directed search strategy for the detection of point mutations in the human factor VIII gene. Blood 76, 2242-2248.

- Pattinson,. J.K., McVey, J.H., Boon, M., Ajani, A., Tuddenham, E.G.D. (1990b) CRM+ haemophila A due to a missense mutation (372 Cys) at the internal heavy chain thrombin cleavage site. Br. J. Hematol. 75, 73-77.

- Pentao, L., Wise, C.A., Chinault, C., Patel, P.I., Lupski, J.R. Charcot-Marie-Tooth disease type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. (1992) Nature Genet. 2, 292-300.

- Pinna, L.A. (1990) Casein kinase 2: An "eminence grise" in cellular regulation? Biochim. Biophys. Acta. 1054, 267-284.

- Pittman, D.D., Kaufmann, R.J. (1988) Proteolytic requirements for thrombin activation of antihemophilic factor (factor VIII). Proc. Natl. Acad. Sci. USA 85, 2429-2433.

- Pittman, D.D., Kaufmann, R.J. (1989) Structure-function relationships of factor VIII elucidated through recombinant DNA technology. Thrombosis and Haemostasis 61, 161-165.

- Pittman, D.D., Millenson, M., Marquette, K., Bauer, K., Kaufmann, R.J. (1992a) A2 Domain of human recombinant derived factor VIII is required for procoagulant activity but not for thrombin cleavage. Blood, 79, 389-397.

- Pittman, D.D., Wang, J.H., Kaufmann, R.J. (1992b) Identification and functional importance of tyrosine sulfate residues within recombinant factor VIII. Biochemistry 3315-3325.

- Poole, S., Firtel, R.A., Lawar, E., Rowekamp, W. (1981) Sequence and expression of the discoidin gene family in Dictyostelium discoideium. J. Mol. Biol. 153, 273-289.

- Potten, C.S. Morriss, R.J. (1988) Epithelial stem cells in vivo. J. Cell. Sci. (Suppl.) 10, 45-62.

- Purrello, M., Alhhadeff, B., Esposito, D., Szabo, P., Rocchi, M., Truett, M., Masiarz, F., Siniscalo, M. (1985) The human genes for hemophila A and B flank the X chromosome fragile site at Xq27.3. EMBO J. 4, 725-729.

- Ragot, T., Vincent, N., Chafey, P., Vigne, E., Gilgenkrantz, H., Couto, D., Cartaud, J., Briand, P., Kaplan, J.C., Perricaudet, M., Kahn, A. (1993) Efficient adenovirus-mediated transfer of a human minidystrophin gene into skeletal muscle of mdx mice. Nature 361, 647-650.

- Recombinant DNA Advisory Committee (RAC) (1990) Guidelines for human gene transfer. Hum. Gene Ther. 1, 93-103.

- Reiner, A.P., Thompson, A.R. (1992) Screening for nonsense mutations in patients with severe hemophilia A can provide rapid, direct carrier detection. Hum. Genet. 89, 88-94.

- Rheinwald, J.G., Green, H. (1975) Serial cultivation of strains of human epidermal keratinocytes: the formation of keratininzing colonies from single cells. Cell 6, 331-334.

- Rieger, M., Jorcano, J., Franke, W. (1985) Complete sequence of a bovine type I cytokeratin gene: conserevd and variable intron positions in genes of polypeptides of the same cytokeratin subfamily. EMBO J. 4, 2261-2267.

- Rieger, M., Franke, W. (1988) Identification of an orthologous mammalian cytokeratin gene J. Mol. Biol. 204, 841-856.

- Roa, B.B., Garcia, C.A., Suter, U., Kulpa, D.A., Wise, C.A., Mueller, J., Welcher, A.A., Snipes, J.G., Shooter, E.M., Patel, P.I., Lupski, J.R. (1993) Charcot-Marie-Tooth disease type 1A: Association with a spontaneous point mutation in the PMP22 gene. New Engl. J. Med. 329, 96-101.

- Robberson, B.L., Cote, G.J., Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell Biol. 10, 84-94.

- Roberts, R. G., Bentley, D.R., Barby, T.F.M., Manners, E., Bobrow, M (1990) Direct diagnosis of carriers of Duchenne and Becker muscular dystrophy by amplification of lymphocyte mRNA. Lancet 336, 1523-1526.

- Roberts, R., Bobrow, M., Bentley, D. (1992) Point mutations in the dystrophin gene. Proc. Natl. Acad. Sci. USA 89, 2331-2335.

- Rosenberg, S.A., Aebersold, P., Cornetta, K., Kasid, A., Morgan, R.A., Moen, R., Karson, E.M., Lotze, M.T., Yang, J.C., Topalian, S.L., Merino, M.J., Culver, K., Miller, A.D., Blaese, R.M., Anderson, W.F. (1990) N. Engl. J. Med. 323, 570.

- Rosenfeld, M.A., Siegfried, W., Yoshimura, K., Yoneyama, K., Fukayama, M., Stier, L.E., Paakko, P.K., Gilardi, P., Stratford-Perricaudet, L.D., Perricaudet, M., Jallat, S., Pavirani, A., Lecocq, J.P., Crystal, R.G. (1991) Adenovirus-mediated transfer of a recombinant alpha 1 antitrypsin gene to the lung epithelium in vivo. Science 252, 431-434.

- Rosenfeld, M.A., Yoshimura, K., Trapnell, B.C., Yoneyama, K., Rosenthal, E.R., Dalemans, W., Fukayama, M., Bargon, J., Stier, L.E., Stratford-Perricaudet, L., Perricaudet, M., Guggino, W.B., Pavirani, A., Lecocq, J.P., Crystal, R.G. (1992) In vivo transfer of the human cystic fibrosis transmembrane conductance regulator gene to the airway epithelium. Cell 68, 143-155.

- Rosenthal, D.S., Chung, S., Huff, C.A., Johnson, J., Yuspa,S.H., Roop, D.R. (1991) A human epidermal differentiation-specific keratin gene is regulated by calcium but not negative modulators of differentiaton in transgenic mouse keratinocytes. Cell Growth and Differentiation 2, 107-113.

- Ross, J. (1988) Messenger RNA turnover in eukaryotic cells. Mol. Biol. Med. 5, 1-14.

- Rossiter, J.P., Young, M., Kimberland, M.L., Hutter, P., Ketterling, R.P., Gitschier, J., Horst, J., Morris, M.A., Schaid, D.J., deMoerloose, P., Sommer, S., Kazazian, H.H., Antonarakis, S.E. (1994) Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. Hum. Mol. Genet. 3, 1035-1039.

- Rotblat, F., Goodall, A.H., O'Brien,D.P.O., Rawlings,E., Middleton, S., Tuddenham, E.G.D. (1983) Monoclonal antibodies to human procoagulant factor VIII. J. Lab. Clin. Med. 101, 736-746.

- Rotblat, F., O'Brien, D.P.O., O'Brien, F., Goodall, A.H., Tuddenham, E.G.D. (1985) Purification of human factor VIII:C and its characterisation by Western blotting using monoclonal antibodies. Biochemistry 24, 4294-4300.

- Ryle, C.M., Breikreutz, D., Stark, H.J., Leigh, I.M., Steinert, P.M., Roop, D. Fusenig, N.E. (1989) Density-dependent modulation of synthesis of keratins 1 and 10 in the human keratinocyte line HaCaT and in ras-transfected tumorigenic clones. Differentiation 40, 42-54.

- Saiki, R.K., Gelfand, S.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Ehrlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, 239, 487-491.

- Salier, J.P., Hiroshiwa, S., Kurachi, K., (1990) Functional characterisation of the 5' regulating region of human factor IX gene. J.Biol. Chem. 265, 7062-7068.

- Sambrook, J., Fritsch, E.F., Manniatis, T. (1989) Molecular Cloning: A laboratory manual, 2nd edition, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

- Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc.Natl. Acad.Sci. USA, 74, 5463-5467.

- Sarkar, G., Sommer, S.S. (1989) Access to messenger RNA sequence or its protein product is not limited by tissue or species specificity. Science 244, 331-334.

- Sarkar, G., Cassady, J., Bottema, C.D.K. Sommer, S.S. (1990) Characterization of polymerase chain reaction amplification of specific alleles. Anal. Biochem. 186, 64-68.

- Sawadago, M., VanDyke, M.W. (1991) A rapid method for purification of deprotected oligonucleotides. Nucleic Acids Res. 19, 674.

- Scharf, S.J., Horn, G.T., Erlich, H.A. (1990) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 239, 487-491.

- Scharfmann, R., Axelrod, J., Verma, I.M. (1991) Long-term in vivo expression of retrovirus-mediated gene transfer in mouse fibroblast implants. Proc. Natl. Acad. Sci. USA 88, 4626-4630.

- Schwaab, R., Ludwig, M., Kocchan, L., Oldenburg, J., McVey, J.H., Egli, H., Brackmann, H.H., Olek, K. (1991) Detection and characterisation of two missense mutations at a cleavage site in the factor VIII light chain. Thrombosis Res. 61, 225-234.

- Sellheyer, K., Bickenbach, J.R., Rothnagel, J.A., Bundman, D., Longley, M.A., Krieg, T., Roche, N.S., Roberts, A., Roop, D.R. (1993) Inhibition of skin development by overexpression of transforming growth factor β1 in the epidermis of transgenic mice. Proc. Natl. Acad. Sci. USA 90, 5237-5241.

- Sellner, L.N. Coelen, M., McKenzie, I. (1992) Reverse transcriptase inhibits Taq polymersase activity. Nucleic Acids Res. 20, 1487-1490.

- Shapiro, M.B., Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15, 7155-7174.

- Shenk, T.E., Rhodes, C., Rigby, P.W.J., Berg, P. (1975) Proc. Natl. Acad. Sci. USA 72, 989-993.

- Shima, M., Ware, J., Yoshioka, A., Fukui, H., Fulcher, C.A. (1989) An arginine to cysteine amino acid substitution at a critical thrombin cleavage site in a dysfunctional factor VIII molecule. Blood 74, 1612-1617.

- Smit, C., Rosendaal, F.R., Varekamp, I. (1989) Physical condition, longevity, and social performance of Dutch hemophiliacs, 1972-1985. Br. Med. J. 298, 235-238.

- Smith, K., Ono, K. (1984) Monoclonal antibodies to factor IX: characterisation and use in immunoassays for factor IX. Thromb. Res. 33, 211-224.

- Smith, K.J., Singaraju, C., Smith, L.F. (1986) Factor IX metal ion-dependent antigen assays for measurement of warfarin effect. Am J. Clin. Pathol. 87, 370.

- Smith, K.J. (1988) Immunoaffinity purification of factor IX from commercial concentrates and infusion studies in animals. Blood 72, 1269-1277.

- Smith, C.L., Kleo, S.R., Cantor, C.R (1988) Pulsed-field gel electrophoresis and the technology of large DNA molecules. In: Genome analysis: A practical approach. ed. K.E. Davies IRL press, Oxford, pp 41-72.

- Smith, H.O., Annau, T.M., Chandrasegaran, S. (1990) Finding sequence motifs in groups of functionally related proteins. Proc. Natl. Acad. Sci. USA 87, 826-830.

- Smith, T.A.G., Mehaffey, M.G., Kayda, D.B., Saunders, J.M., Yei, S., Trapnell, B.C., McClelland, A., Kaleko, M. (1993) Adenovirus mediated expression of therapeutic plasma levels of human factor IX in mice. Nature Genet. 5, 397-402.

- Smithies, O., Gregg, R.G., Biggs, S.S., Koralewski, M.A. Kucherlapati, R.S. (1985) Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature 317, 230-234.

- Sommer, S., Bowie, E.J.W., Ketterling, R.P., Bottema, C.D.K. (1992) Missense mutations and the magnitude of functional deficit: the sample of factor IX. Hum. Genet. 89, 295-297.

- Southern, E.M. (1975) Detection of specifc sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

- Southern, P.J., Berg, P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. 1, 327-341.

- St. Louis, D., Verma, I.M. (1988) An alternative approach to somatic cell gene therapy. Proc. Natl. Acad. Sci. USA 85, 3150-3154.

- Staff of the Center for Biologics evaluation and research, FDA (1991) Points to consider in Human Somatic Cell Therapy and Gene Therapy. Hum. Gene Ther. 2, 251-256.

- Steinert, P.M., Roop, D. (1988) Molecular and cellular biology of intermediate filaments. Ann. Rev. Biochem. 57, 593-625.

- Stellmach, V., Leask, A., Fuchs, E. (1991) Retinoid-mediated transcriptional regulation of keratin genes in human epidermal and squamous cell carcinoma cells. Proc. Natl. Acad. Sci. USA 88, 4582-4586.

- Stenflo, J. (1991) Structure function relationships of epidermal growth factor modules in vitamin-K dependent clotting factors. Blood 78, 1637-1651.

- Sternberg, E.A., Spizz, G., Perry, W.M., Vizard, D., Weil, T. Olson, E. (1988) Identification of upstream and intragenic regulatory elemnts that confer cell-type restricted and differentiation specific expression on the muscle creatine kinase gene. Mol. Cell. Biol. 8, 2896-2909.

- Sterner, D.A., Berget, S.M. (1993) In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. Mol. Cell Biol. 13, 2677-2687.

- Stratford-Perricaudet, L.D., Makeh, I., Perricaudet, M., Briand, P. (1992) Widespread long-term gene transfer to mouse skeletal muscles and heart. J. Clin. Invest. 90, 626-630.

- Stribling, R., Brunette, E., Liggitt, D., Gaensler, K., Debs, R. (1992) Aerosol gene delivery in vivo. Proc. Natl. Acad. Sci. USA 89, 11277-11281.

- Stubbs, J.D., Lekutis, C., Singer, K.L., Bui, A., Yuzuki, D., Srinivasan, U., Parry, G. (1990) cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like sequences. Proc. Natl. Acad. Sci. USA., 87, 8417-8421.

- Study group of the UK hemophila Centre directors (1988) Effect of dry heating of coagulation factor concentrates at 80°C for 72 hours on transmission of non-A, non-B hepatitis. Lancet 1, 814.

- Takahashi, N., Ortel, T.L., Putnam, F.W. (1984) Single-chain structure of human ceruloplasmin: The complete amino acid sequence of the whole molecule. Proc. Natl. Acad. Sci. USA 81, 390-394.

- Tantravahi, U., Murty, V.V.W.S., Jhanwar, S.C., Toole, J.J., Woozney, J.M., Chaganti, R.S.K., Latt, S.A. (1986) Physical mapping of the factor VIII gene proximal to two polymorphic DNA probes in human chromosome band Xq28: Implications for factor VIII gene segregation analysis. Cytogenet. Cell Genet. 42, 75-79.

- Teumer, J., Lindahl, A., Green, H. (1990) Human growth hormone in the blood of athymic mice grafted with cultures of hormone secreting human keratinocytes. FASEB J. 4, 3245-3250.

- Thompson, A.R. (1991) Status of gene transfer for hemophilia A and B. Thrombosis and Haemostasis 66, 119-122.

- Tindall, K.R., Kunkel,T.A. (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry 27, 6008-6013.

- Tizzano, E.F., Altisent, C., Tusell, J., Domenech, M., Baiget, M. (1994) Intron 22 inversions and haemophilia. Lancet, 343, 792.

- Toole, J.J., Knopf, J.L., Wozney, J.M., Sultzman, L.A., Buecker, J.L., Pittman, D.D., Kaufman, R.J., Brown, E., Shoemaker, C., Off, E.C., Amphlett, G.W., Foster, W.B., Coe, M.L., Knutson, G.J., Fass, D.N., Hewick, R.M. (1984) Molecular cloning of a cDNA encoding human antihaemophilic factor. Nature, 312, 342-347.

- Toole, J.J., Pittman, D.D., Orr, E.C., Murtha, P., Wasley, L.C., Kaufman, R.J. (1986) A large region (~95kDa) of human factor VIII is dispensable for in vitro procoagulant activity. Proc. Natl. Acad. Sci. USA., 83, 5939-5942.

- Traystman, M.D., Higuchi, M., Kasper, C.K., Antonarakis, S.E., Kazazian, H.H., (1990) Use of denaturing gradient gel electrophoresis to detect point mutations in the factor VIII gene. Genomics, 6, 293-301.

- Tuddenham, E.G.D. (1989) von Willebrand factor and its disorders: an overview of recent molecular studies. Blood Rev. 3, 251-262.

- Tuddenham, E.G.D., Cooper, D.N., Gitschier, J., Higuchi, M., Hoyer, L.W., Yoshioka, A., Pewake, I.R., Schwabb, R., Olek, K., Kazazian, H.H., Lavergne, J.M., Giannelli, F., Antonarakis, S.E. (1991) Haemophilia A: database of nucleotide substitutions, deletions, insertions, and rearrangements of the factor VIII gene. Nucleic Acids Res. 19, 4821-4833.

- van Diejen, G., Tans, G., Rosing, J., Hemker, H.C. (1981) The role of phospholipid and factor VIIIa in the activation of bovine factor X. J. Biol. Chem. 256, 3433-3442.

- Vasan, N.S., Kuivaniemi, H., Vogel, B.E., Minor, R.R., Wooton, J.A.M., Tromp, G, Weksberg, R., Prockop, D.J. (1991) A mutation in the pro α2(I) gene (COL1A2) for type I procollagen in Ehler-Danlos syndrome type VII: evidence suggesting that skipping of exon 6 in RNA splicing may be a common cause of the phenotype. Am. J. Hum. Genet. 48, 305-317.

- Vassar, R., Rosenberg, M., Ross, S., Tyner, A., Fuchs, E. (1989) Tissue-specific and differentiation-specific expression of a human K14 keratin gene in transgenic mice. Proc. Natl. Acad. Sci. USA 86, 1563-1567.

- Vassar, R., Fuchs, E. (1991) Transgenic mice provide new insights into the role of TGF-α during epidermal development and differentiation. Genes & Dev. 5, 714-727.

- Vehar, G.A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D., Roblat, F., Oppermann, H., Keck, R., Wood, W.I., Harkins, R.N., Tuddenham, E.G.D., Lawn, R.M., Capon, D.J. (1984) Structure of human factor VIII. Nature, 312, 337-342.

- Veltkamp, J.J., Meilof, J., Remmelts, H.G., Van der Vlerk, D., and Loeliger, E.A. (1970) Another genetic variant of hemophila B: Hemophila B Leyden. Scand. J. Haemat. 7, 82-90.

- Verkerk, A.J.M.H., Piereti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P.A, Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., Eussen, B.E., van Ommen, GJB, Blonden, L.A.J. Riggins, G.J. Chastain, J.L., Kunst, CB., Galjaard, H., Caskey, C.T., Nelson, D.L., Ostra, B.A. Warren, S.T. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exibiting length variation in the fragile X syndrome. Cell 65, 905-914.

- Vermeer, C. (1990) γ-carboxyglutamate-containing proteins and the vitamin K dependent carboxylase. Biochem. J. 266, 625-636.

- Villarreal, L.P., White, R.T. (1983) A splice junction deletion deficient in th transport of RNA does not polyadenylate nuclear RNA. Mol. Cell. Biol. 3, 1381-1388.

- Vogel, J. (1993) Keratinocyte gene therapy. Arch. Dermatol. 129, 1478-1483.

- Wagner, E., Zatloukal, K., Cotten, M., Kirlappos, H., Mechtler, K., Curiel, D.T., Birnsteil, M.L. (1992) Coupling of adenovirus to transferrin-polylysine/DNA complexes greatly enhances receptor-mediated gene delivery and expression of transfected genes. Proc. Natl. Acad. Sci. USA 89, 6099-6103.

- Walker, F.J., Fay, P.J. (1992) Regulation of blood coagulation by the protein C system. FASEB J. 6, 2561-2567.

- Wapenaar, M.C., Kievits, T., Hart, K.A., Abbs, S., Blonden, L.A.J., Den Dunnen J.T., Grootscholten, P.M., Bakker, E., Verellen-Dumoulin, Bobrow, M., van Ommen, G.J.B., Pearson, P.L. (1988) A deletion hot-spot in the Duchenne muscular dystrophy gene. Genomics 2, 101-108.

- Weil, D., D'Alessio, M., Ramirez, F., Steinmann, B., Wirtz, M.K., Glanville, R.W., Hollister, D.W. (1990) Temperature-dependent expression of collagen splicing defect in fibroblasts of a patient with Ehler-Danlos syndrome type VII. J. Biol. Chem. 264, 16804-16809.

- White, G.C., McMillan, C.W., Kingdon, H.S., Shoemaker, C.B. (1989) Use of recombinant anti-hemophilic factor in the treatment of two patients with classical hemophilia. New Engl. J. Med. 320, 166-170.

- White, M.B., Carvalho, M., Derse, D., O'Brien, S.J., Dean, M. (1992) Detecting single base substitutions as heteroduplex polymorphisms. Genomics 12, 301-306.

- Williams, R.S., Johnston, S.A., Reidy, M., DeVit, M.J., McElligott, S.G., Sanford, J.C. (1991) Introduction of foreign genes into tissues of living mice by DNA-coated microprojectiles. Proc. Natl. Acad. Sci. USA 88, 2726-2730.

- Wilson, J.M., Birinyi, L.K., Salomon, R.N., Libby, P., Callow, A.D., Mulligan, R.C. (1989) Implantation of vascular grafts lined with genetically modified endothelial cells. Science 244, 1344-1346.

- Winship, P.R., Anson, D.S., Rizza, C.R., Brownlee, G.G. (1984) Carrier detection in hemophilia B using two further intragenic restriction fragment length polymorphisms. Nucleic Acids Res. 12, 8861-8872.

- Wise, C.A., Garcia, C.A., Davis, S.N., Heju, Z., Pentao, L., Patel, P.I., Lupski, J.R. (1993) Molecular analyses of unrelated Charcot-Marie-Tooth (CMT) disease patients suggests a high frequency of the CMT1A duplication. Am. J. Hum. Genet. 53:853-863.

- Wolff, J.A., Malone, R.W., Williams, P., Wang, C., Ascadi, G., Jani, A., Felgner, P.L. (1990) Direct gene transfer into mouse muscle in vivo. Science 247, 1465-1468.

- Wong, C., Antonarakis, S.E., Goff, S.C., Orkin, S.H., Forget, B.G., Nathan, D.G., Giardina, P.J.V., Kazazian, H.H. (1989) β-thalassaemia due to two novel nucleotide substitutions in consensus acceptor splice sequences of the β-globin gene. Blood 73, 914-918.

- Wood, W.I., Capon, D.J., Simonsen, C.C., Eaton, D.L., Gitschier, J., Keyt, B., Seeburg, P.H., Smith, D.H., Hollingshead, P., Wion, K.L., Delwart, E., Tuddenham, E.G.D., Vehar, G.A., Lawn, R.M. (1984) Expression of active factor VIII from recombinant DNA clones. Nature, 312, 330-337.

- Wu, G.Y., Wu, C.H. (1987) Receptor-mediated in vitro gene transformation by a soluble carrier system. J. Biol. Chem. 262, 4429-4432.

- Wu, D.Y., Ugozolli, L., Pal, B.K., Wallace, R.B. (1989) Allele-specifc enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. USA 86, 2757-2760.

- Wu, S.M., Cheung, W.F., Frazier, D., Stafford, D.W. (1991) Cloning and expression of the cDNA for human γ-glutamyl carboxylase. Science 254, 1634-1636.

- Wu, G.Y., Wilson, J.M., Shalaby, F., Grossman, M., Shafritz, D.A., Wu, C.H. (1991) Receptor-mediated gene delivery in vivo. Partial correction of genetic analbuminaemia in nagase rats. J. Biol. Chem. 266, 14338-14342.

- Yang, Y., Nunes, F.A., Berensci, K., Gonczol, E., Engelhardt, J.F., Wilson, J.M. (1994) Inactivation of *E2a* in recombinant adenoviruses improves the prospect for gene therapy in cystic fibrosis. Nature Genet. 7, 362-369.

- Yao, S.N., DeSilva, A.H., Kurachi, S., Samuelson, L.C., Kurachi, K. (1991) Characterisation of a mouse factor IX cDNA and developmental regulation of the factor IX gene expression in liver. Thrombosis and Haemostasis 65, 52-58.

- Yao, S.N., Kurachi, K. (1992) Expression of human factor IX in mice after injection of genetically modified myoblasts. Proc. Natl. Acad. Sci. USA. 89, 3357-3361.

- Yen, T.J., Machlin, P.S., Cleveland, D.W. (1988) Autoregulated instability of β-tubulin by recognition of the nascent amino acids of β-tubulin. Nature 334, 580-585.

- Yoshimura, K., Rosenfeld, M.A., Nakamura, H., Scherer, E.M., Pavirani, A., Lecocq, J.P., Crystal, R.G. (1992) Expression of the human cystic fibrosis transmembrane conductance regulator gene in the mouse lung after in vivo intratracheal plasmid-mediated gene transfer Nucleic Acids Res. 20, 3233-3240.

- Yoshitake, S., Schah, B.G., Foster, D.C., Davie, E.W., Kurachi,K. (1985) Nucleotide sequence of the gene for human factor IX (Anti-Hemophilic Factor B). Biochemistry 24, 3736-3750.

- Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J.C., Warren, S.T., Schlessinger, D., Sutherland, G.R., Richards, R.I. (1992) Fragile X genotype characterised by an unstable region of DNA. Science 252, 1179-1181.

- Yuspa, S.H., Kilkenny, A.E., Steinert, P.M., Roop, D.R. (1989) Expression of murine epidermal differentiation markers is tightly regulated by restricted extracellular calcium concentrations in vitro. J. Cell. Biol. 109, 1207-1217.