UPTEC X 02 035
JUN 2002

ISSN 1401-2138

ANN-CHARLOTTE RÖNN

Analysis of sequence variations in the factor VIII gene in haemophilia A patients

Master's degree project

Molecular Biotechnology Programme
Uppsala University School of Engineering

UPTEC X 02 035	**Date of issue 2002-06**

Author

Ann-Charlotte Rönn

Title (English)

Analysis of sequence variations in the factor VIII gene in haemophilia A patients

Title (Swedish)

Abstract

A 23kb region upstream of the factor VIII gene was screened for mutations in haemophilia A patients using solid phase fluorescent chemical cleavage of mismatch. The haplotypes found could not explain the difference in phenotype between patients sharing the same causative mutation. A putatively functional mutation was found in the promoter region. To further investigate the ancestry of the causative mutations in this study, the patients were haplotyped for several known intragenic and extragenic polymorphisms. This study shows that some of the causative mutations investigated probably have arisen on more than one occasion. No conclusive evidence can be given in this study regarding the functional role of the promoter mutation.

Keywords

Haemophilia A, factor VIII gene, polymorphism, haplotype, chemical cleavage of mismatch

Supervisors

Prof. Francesco Giannelli & Dr. Peter Green
King's College, London, UK

Examiner

Dr. Matti Nikkola
Uppsala University, Uppsala, Sweden

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **26**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Analysis of sequence variations in the factor VIII gene in haemophilia A patients

Ann-Charlotte Rönn

Sammanfattning

Hemofilia A, den vanligaste av blödarsjukdomarna, drabbar 1 av 5000 män. Sjukdomen beror på en mutation i genen för faktor VIII, ett protein nödvändigt för koaguleringen av blodet. Eftersom faktor VIII-genen sitter på X-kromosomen ärvs den på mödernet och kvinnor är oftast symptomfria. Även om hemofilia A oftast ärvs, uppkommer ständigt nya mutationer. Beroende på var mutationen sitter är symptomen olika svåra, alltifrån spontana blödningar i lederna till nästintill symptomfria. Det finns ett behov av att kartlägga de mutationer som finns. Genetiska metoder skulle möjliggöra en pålitligare diagnosställning och behandling på ett tidigt stadium, vilket skulle leda till färre följdsjukdomar.

I detta examensarbetet har en DNA-sekvens av faktor VIII-genen undersökts för att finna oupptäckta mutationer hos hemofilia A-patienter. Eventuellt skulle dessa mutationer kunna förklara varför en del patienter med samma mutation trots det kan ha olika svåra symptom. Flera kända ofarliga DNA-variationer omkring faktor VIII-genen har också undersökts hos patienter med samma mutation. Detta för att kunna fastställa om exakt samma mutation ofta uppkommer flera gånger under människans utveckling, eller om det är troligare att patienter med samma mutation alltid har en gemensam anfader.

Examensarbete 20 p i Molekylär bioteknikprogrammet

Uppsala universitet Juni 2002

Table of contents

1 Introduction

Factor VIII (f8) is a glycoprotein that is essential for blood coagulation. Deficiency or functional abnormality of this protein result in haemophilia A, the most common of the bleeding disorders. Because it is an X-linked recessively inherited disorder, patients are almost exclusively males, and its incidence is of 1 in 5,000 male live birth in all ethnic populations. Females with one abnormal f8 gene are usually asymptomatic carriers. Haemophilia A has been maintained in the population by an equilibrium between new mutations and selection. It therefore has a very high level of mutational heterogeneity and about 30% of the cases arise from new mutations (1).

Haemophilia A occurs in mild, moderate or severe form. In its severe form, with a circulating f8 activity level of less than 1% of the normal value, it causes spontaneous bleeding, particularly to the joints, and excessive bleeding after even minor traumas. Patients having the mild form, with a circulating f8 activity level of 5-30% of the normal value, go undiagnosed often for most part of their lives, until exposed to major trauma involving surgery (1).

F8 participates in the intrinsic pathway of blood coagulation. It circulates in blood plasma associated with the von Willebrand factor in a noncovalent complex. There, it is activated by thrombin and acts as a cofactor for factor IXa, which in the presence of Ca^{2+} and phospholipids activates factor X. Factor Xa is essential for the coagulation cascade to continue into the common pathway, in which thrombin cleaves fibrinogen into fibrin monomers that polymerise into a stable clot. The presence of f8 and factor IX has proved to be crucial for the intrinsic pathway as defects in these genes lead to haemophilia A and B respectively (figure 1) (2,3).

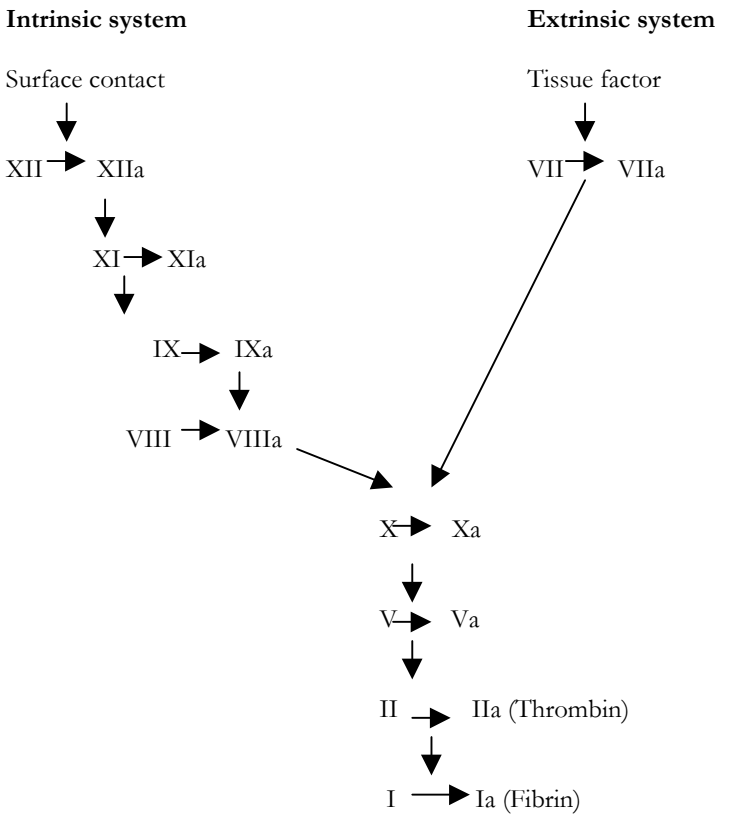


Figure 1. The coagulation cascade, in which f8 plays an important role.

The 186kb long f8 gene is located about 1000kb from the telomeric end of the long arm of the X chromosome, within band Xq28. The gene is transcribed in a telomere to centromere direction and its 26 exons encode a 9kb long messenger RNA. The RNA translates into a single polypeptide chain of 2351 amino acids that undergoes multiple cleavage events. A large intron, IVS22, contains two genes f8a and f8b, the latter using exon 23 of f8 as its second exon (figure 2) (4, 5).

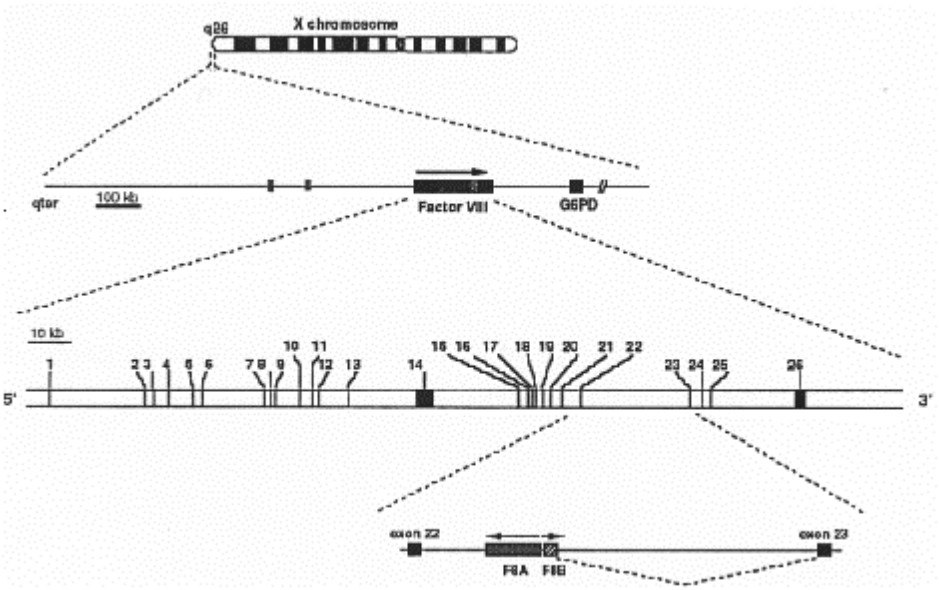


Figure 2. Schematic representation of the chromosomal localisation and structure of the f8 gene, located approx. 1000kb from the telomeric end of the long arm of the X chromosome (Xqter) (16).

1.1 Intragenic polymorphisms

Some intronic polymorphic sites are known within the f8 gene. These DNA markers are used for haemophilia A carrier detection and prenatal diagnosis within families, as they are in linkage disequilibrium with the defect in the f8 gene (6, 7, 8).

Two different dinucleotide repeats located in intron 13 and intron 22 have previously been studied. These multi-allelic length polymorphisms show X-linked mendelian inheritance and are highly informative as markers for pedigree analysis (6, 7). The $(CA)_n$ repeat within intron 13 has earlier been shown to have eight alleles ranging from 16 to 24 dinucleotide repeats with allelic frequencies ranging from 1 to 45% (6). The $(GT)_n(AG)_n$ repeat within intron 22 has been shown to have four alleles ranging from 23 to 28 repeats with allelic frequencies ranging from 1.3 to 66.7% (7).

Previously, a T to C polymorphism has been found within intron 19. The C allele of this single nucleotide polymorphism (SNP) is recognised and cleaved by the restriction enzyme *Hind* III (8). Another SNP has been found in exon 26, in the 3'UTR of the f8 gene. Neither of the alleles, A or G, are known to be recognised by any enzyme, and the SNP has to be genotyped by sequencing (9).

1.2 Extragenic polymorphisms

100-150kb 5' of the f8 gene is the VBP1 (also known as VHL binding protein in the literature) gene, transcribed in a telomere to centromere direction (figure 3) (10). This gene codes for subunit 3 of prefoldin, a chaperone protein that binds to cytosolic chaperonin and transfers target proteins to it, and thereby promotes folding of the target proteins (11).

In the 3'UTR of the VBP1 gene a C to T polymorphism has previously been studied. The transition creates a *Tsp* 5091 restriction endonuclease site that can be used for genotyping the SNP (10).

The dyskeratosis congenita 1 (DKC1) gene is situated 60kb 3' of the f8 gene (figure 3). This gene, transcribed in a centromere to telomere direction, is a member of the small nucleolar ribonucleoproteins (snoRNPs) gene family. These are involved in various aspects of rRNA processing and modification, and are also components of the telomerase complex. A defect gene causes X-linked dyskeratosis congenita (12, 13).

Two previously studied transitions are situated in exon 14 and 15 respectively of the DKC1 gene. The C to T transition in exon 14 at nucleotide (nt) 1557 represents a synonymous mutation at codon 487. The G to A polymorphism in exon 15 is situated at nt 1647 in the 3'UTR. The SNPs are detectable by digestion with the restriction enzymes *Blp* I and *Hae* III respectively (12).

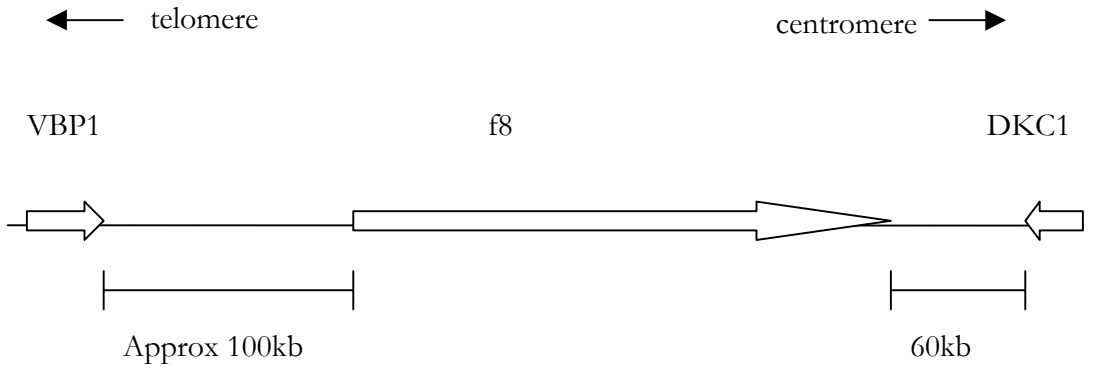


Figure 3. Map of the physical distance between f8 and VBP1 and CDK1. Genes are not drawn to scale. Block arrows indicate direction of transcription.

1.3 Chemical cleavage of mismatch

Chemical cleavage of mismatch is a technique for detecting and localising mismatches in heteroduplex DNA molecules. It was originally described in 1988 (14). Since then the technique has been improved, and the variant currently used by this laboratory was first described in 1995 (15). The modified method of chemical cleavage of mismatch, solid phase fluorescent chemical cleavage of mismatch (SPFCCM), has replaced the original radioactive labelling with fluorescently labelled nucleotides, and removed the time consuming precipitation steps by the use of biotinylated primers, streptavidin-magnetic beads and magnets (solid phase).

The method uses the chemical reactivity of T when mismatched with G, C, and T, and that of C when mismatched with A, T and C. When heteroduplex DNA, formed by annealed DNA strands of a patient and a control individual, are treated with hydroxylamine or osmium tetroxide mismatched C or T respectively are chemically modified. These modified bases are recognised and cleaved by piperidine and the lengths of the heteroduplexes can be determined by electrophoretically analysing the resulting products on an ABI 377. The four fluorescent colours detected by this system allow the analysis of three DNA segments per lane as well as the use of internal size markers.

2 Aim of project

Previously, in this laboratory, the entire coding region of the f8 gene has been screened for mutations in ca 900 haemophilia A patients, using SPFCCM. Over 600 different mutations have been found, including missense, nonsense, frameshift, deletion, inversion and defects in splice signal sites (9).

In some cases patients sharing the same functional mutation have varying levels of circulating f8 activity. The aim of this project is therefore to investigate this further by screening the 5' upstream region of f8, including the promoter region, for polymorphic changes in order to see whether alleles at these polymorphic sites are correlated with the phenotypic difference among the patients. Haplotype analysis of the polymorphisms found, together with known intragenic and close extragenic polymorphisms, may also show whether the known identical haemophilia A mutations have arisen as separate events or have a common ancestral origin.

3 Material and methods

The solutions were prepared according to table 1.

Table 1. Recipes for the solutions used

TE	10mM Tris-HCl, pH 8.0 0.1mM Na_2EDTA
10x TBE	0.89M Trizma-base 0.89M Boric acid 25mM Na_2EDTA
10x PCR Buffer	670mM Tris-HCl, pH 8.8 166mM $(NH_4)_2SO_4$ 67mM $MgCl_2$ added just prior to use: 1.7mg/ml Bovine Serum Albumin 100mM ß-mercaptoethanol
Glycerol loading dyes	30% glycerol 0.1% Bromophenol Blue 0.1% Xylene Cyanol 5mM Na_2EDTA filter sterilised
Formamide loading dyes	Deionised Formamide 0.1% Dexy Blue
10x Hybridisation Buffer	3M NaCl 1M Tris-HCl, pH 8.0
2x Binding Buffer	2M NaCl 0.4% Tween20
Polyacrylamide gel	20ml Sequencing gel, 6% acrylamide (Severn Biotech) 47µl 10% $(NH_4)_2S_2O_8$ (APS) 33µl Tetramethylethylenediamine (TEMED)
Hydroxylamine solution	0.23mg/µl Hydroxylamine 0.2% Diethylamine (DEA) dH_2O
Osmium solution	10% Osmium 0.02% Pyridine TE
Piperidine solution	20% Piperidine 4% GeneScan™-2500 ROX™ Size Marker (Applied Biosystems) Formamide loading dye

3.1 DNA extraction
DNA was obtained from peripheral lymphocytes by separating the RNA and DNA fractions extracted using Purescript® RNA Isolation Kits (Gentra Systems).

3.2 SPFCCM
Biotinylated primers were designed using the programme Primer3 (17). If possible, primers with little probability to form secondary structures (e.g. hairpins), or primer dimers due to complementary 3' ends of the primers, were chosen. The melting temperatures for the primer annealing varied between 58°C and 61°C, and the GC-content of the primers varied between 34% and 55% for all designed primers.

DNA segments were amplified using polymerase chain reaction (PCR) in a Perkin Elmer DNA Thermal Cycler. The same conditions were used for all reactions (table 2).

Table 2. The PCR cycles.

1. 94°C for 5min: denaturing of all double stranded DNA.
2. 93°C for 1min: denaturing of double stranded DNA formed during PCR.
3. 61°C for 30s: annealing of primers to single stranded DNA.
4. 72°C for 2min: elongation of double stranded DNA.
5. 72°C for 5min: final elongation.

Steps 2-4 is repeated for 35-38 cycles, to get a sufficiently high yield of product. The high annealing temperature ensures a high specificity of the products (i.e. no undesired binding of the primers).

The same recipe for all reactions was used: 1x PCR Buffer, 0.5mM dNTP's (Promega), 2.5U *Taq* DNA polymerase (Promega), 2.5U *Taq* Extender™ PCR Additive (Stratagene), 37.5ng forward primer (MWG Biotech), 37.5ng reverse primer, 50-100ng genomic DNA, made up to 25µl with TE.

In addition, all DNA segments were amplified in one control individual, not diagnosed with haemophilia A. Each of these segments were internally labelled using one of three different fluorescent dUTP's (Applied Biosystems). 4µM [R110]dUTP's, 4µM [R6G]dUTP's, or 16µM [TAMRA]dUTP's were added for that purpose to the standard recipe.

All PCR products were checked on a mini 1% agarose gel, made up with 1x TBE.

For the SPFCCM reactions, 4µl of patient product (target) and 1µl of fluorescent labelled control product (probe), from the corresponding segment, were annealed in 1x Hybridisation Buffer made up to 30 µl with TE. In most reactions three segments labelled with different dyes were mixed in one tube. The solution was heated to 95°C for 5min followed by hybridisation for one hour in 65°C.

25µl magnetic beads per SPFCCM reaction tube were washed in 50µl 2x Binding Buffer per reaction. Binding Buffer was removed, using a magnet to separate the beads, and the beads were resuspended in 30µl 2x Binding Buffer per reaction. 30µl bead mix were added to each reaction tube and left at room temperature for 15min. The supernatant was removed, using a magnet. 20µl hydroxylamine solution or osmium tetroxide solution were added to the beads in each reaction tube and left to incubate for 90min in 37°C or 15min in room temperature respectively. The supernatant was removed, using a magnet. The beads were washed in 30µl TE per reaction and all supernatant was removed, using a magnet. The beads were resuspended in 5µl piperidine solution per reaction tube and heated for 30min in 90ºC. The bead mix was chilled in -20ºC for at least one hour, before loading 1µl of the supernatant on a 6% polyacrylamide gel on an ABI Prism® 377 DNA Sequencer (Applied Biosystems) (1x TBE buffer). Cleaved products were sized by GeneScan® Analysis Software (Applied Biosystems).

3.3 Intragenic and extragenic polymorphism detection

All segments for detection of known polymorphisms were amplified using the following recipe and under the following touchdown conditions, if not stated otherwise:

1 x PCR Gold Buffer (Applied Biosystems), 0.1mM dNTP's (Promega), 1.53mM $MgCl_2$ (Applied Biosystems), 150ng forward and 150ng reverse primer, 0.18µl Ampli *Taq* Gold® DNA polymerase (Applied Biosystems) and 50-100ng genomic DNA made up to 22µl with TE. For 13 cycles: 45s at 94°C, 45s at 65°C (-5C for each cycle) 1.15min at 72°C. For an additional 30 cycles: 45s at 94°C, 45s at 56°C, 1.15min at 72C°. For all amplifying the primers described in literature were used (7, 10, 12, 18), except for amplification of exon 26 where primers 26.7F and 26.7R designed at this laboratory were used (9).

3.3.1 Digestions with restriction enzymes

To amplify the segment in exon 15 of the DKC1 gene the following recipe and conditions were used: 1 x PCR buffer, 0.2mM dNTP's (Promega), 2ng forward and 2ng reverse primer, 0.1U *Taq* polymerase (Promega) and 50-100ng genomic DNA made up to 10µl with TE. For 35 cycles: 30s at 93°C, 30s at 58°C, and 1min at 72C°.

Digestions with the restriction enzymes *Hind* III (Promega), *Hae* III (Promega) and *Blp* I (New England Biolabs) was performed using the buffers supplied by the manufacturer for each enzyme. For 5µl of medium strong PCR product: 10U enzyme and 1 x buffer was made up to 20µl with dH_2O and digested for 1h at 37°C. Cleaved/uncleaved products were checked on a 2.5% agarose gel using a 50bp Step Ladder (Promega) for size determining.

3.3.2 Dinucleotide repeat analysis

To amplify the repeat in intron 13 of the f8 gene the following recipe and conditions were used: 1 x PCR Buffer (Applied Biosystems), 0.1mM dNTP's, 1.5mM $MgCl_2$, 200ng forward and 200ng reverse primer (flourescently labelled with [HEX]) and 50-100ng genomic DNA made up to 25µl with TE. For 32 cycles: 30s at 94°C, 1min at 52C°, 1min at 72°C. The repeat in intron 22 of the f8 gene were amplified using [FAM] labelled primers.

A mix of 0.5µl PCR product, 0.45µl formamide loading dye and 0.05µl GeneScan™-500 TAMRA™ Size Marker (Applied Biosystems) were loaded on a 6% polyacrylamide gel and run on an ABI Prism® 377 DNA Sequencer (Applied Biosystems) (1x TBE buffer). Repeat lengths were sized by GeneScan® Analysis Software (Applied Biosystems).

3.3.3 Sequencing

The VBP1 segment was amplified using the standard recipe for SPFCCM-PCR. Annealing temperature was 53°C and elongation time shortened to 30s.

Sequencing was carried out by means of the dideoxy procedure (19). The primer closest to the SNP was chosen for the sequencing reaction. Amplifying primers were removed from 1µl PCR product by adding 1µl ExoSAP-IT (Amersham Pharmacia) and 4µl dH_2O, incubating at 37°C for 15min and heating at 80°C for 15min. For the sequencing reaction 100ng primer and 3µl ABI Prism® Big Dye™Terminators (Applied Biosystems) were added to the PCR products. The sequencing programme included 30 cycles of 96°C for 30s, 50°C for 15s and 60°C for 4min. DNA purification was performed by ethanol precipitation, as specified by Applied Biosystems. DNA was resuspended in 1µl formamide loading dye, loaded on a 6% polyacrylamide gel and run on an ABI Prism® 377 DNA Sequencer (Applied Biosystems) (1x TBE buffer). Analysis was performed by Sequencing Analysis Software™ (Applied Biosystems).

4 Results

4.1 Patients

From male patients previously scanned for mutations in the coding region of f8 by this laboratory, 20 were chosen. Three of them, patients 4865, 5019, and 4453, have a base change (G to A) at nt 541, causing an amino acid change from valine (V) to methionine (M) at codon 162. Patients 4607, 4613 and 4530 all have a base change (C to T) at nt 6532, causing an amino acid change from arginine (R) to cysteine (C) at codon 2159. Four patients, 4998, 4579, 4702 and 4260, have a base change at nt 2149 (C to T), causing an amino acid change from R to tryptophan (W) at codon 698. Three patients have a base change at nt 6744 (G to T), causing an amino acid change from W to C at codon 2229. Two patients have a base change (A to T) at nt 1213, causing an amino acid change from isoleucine (I) to phenylalanine (F) at codon 386. There is also one patient (number 4401) with the base change G to A at nt 6545, causing an amino acid change from R to histidine (H) at codon 2163 (table 3).

Each group of mutations are named after the number of the mutant codon. Four patients (no. 5043, 4594, 4743 and 4937) had no known mutations, even though they had been previously scanned over the same sequences as the other patients. This group is referred to as unknown. In each mutation group patients with very different circulating f8 levels have been chosen. Patient 4401 was randomly chosen. In addition to the 20 patients originally chosen, two patients, 4920 and 5241, with no known mutations in the coding region of f8 were added in later experiments. The reason for this will be discussed later (table 3).

4.2 DNA region for SPFCCM experiments

A 23kb sequence 5' of exon 1 at the f8 gene was divided into 20 segments of approximately 1.5kb each, overlapping by approximately 200bp (figure 4).

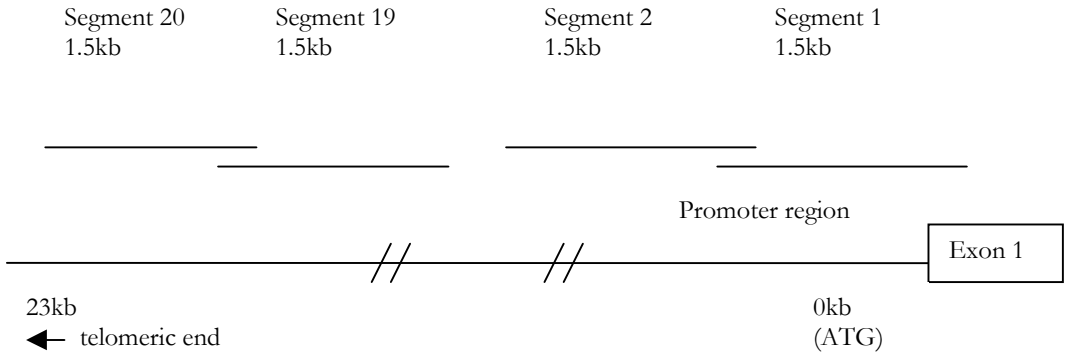


Figure 4. Schematic picture of how the 23kb upstream region of exon1 in the f8 gene was divided into 1.5kb segments, overlapping by 200bp.

Of the 20 sequence segments six proved very difficult to amplify and have been excluded from further analysis (segments 3, 5, 6, 7, 16 and 17). The reason for the failure of the amplification of these segments has not been fully investigated, but it was probably due to repetitive sequences within the segments. It is also possible that the primer sequences could have contained one or more incorrect base due to discrepancies with the contig sequence used for primer design.

Table 3. Haemophilia A patients chosen for the experiments. 16 had previously recorded mutations, causing haemophilia A. In six of the patients no mutations had been found in the coding regions of f8 in previous scanning by this lab. Unfortunately the circulating level of f8 was not known for patient 4920.

Patient no.	F8 activity level (% of normal)	Nucleotide no.	Base change from	Base change to	Amino acid change fraom	Amino acid change to	Codon no. (group no.)
4865	16	541	G	A	V	M	162
5019	10	541	G	A	V	M	162
4453	4	541	G	A	V	M	162
4607	30	6532	C	T	R	C	2159
4613	4	6532	C	T	R	C	2159
4530	3	6532	C	T	R	C	2159
4998	36	2149	C	T	R	W	698
4579	22	2149	C	T	R	W	698
4702	8	2149	C	T	R	W	698
4260	5	2149	C	T	R	W	698
5065	12	6744	G	T	W	C	2229
4883	2	6744	G	T	W	C	2229
4328	0	6744	G	T	W	C	2229
4576	13	1213	A	T	I	F	386
4575	0	1213	A	T	I	F	386
4401	17	6545	G	A	R	H	2163
5043	30						
4594	0						
4743	0						
4937	0						
4920	unknown						
5241	30						

4.3 SPFCCM results

In the remaining 14 segments at least 11 mismatches were found. Five small cleavage bands (i.e. less than half the size of the full product size) were found within the same segment, 19, making it impossible to establish exactly how many mismatches were present. Two of these were visible after both the hydroxylamine and the osmium tetroxide SPFCCM reactions. The smallest cleavage band was exclusive to the hydroxylamine reaction and two cleavage bands were exclusive to the osmium tetroxide reaction. Segments 1, 9, 12, 13, 14, 18 each contained one mismatch. In three of these, segments 1, 13 and 18, the cleavage occurred in both the hydroxylamine and the osmium tetroxide SPFCCM reactions. The cleavage bands in segment 9 could only be seen in the hydroxylamine reaction, and the cleavage bands in segments 12 and 14 were exclusive to the osmium tetroxide reaction. Almost all of the bands could be seen in the same patients, dividing the patient material into two distinct haplotypes (figure 5). The only exceptions from these haplotypes were the extra band present in patient 4743 (unknown) in segment 19, and the exclusive mismatch in patient 5043 (unknown) in segment 1. The haplotype containing no mismatches, i.e. the haplotype of the control individual, is referred to as haplotype A. The haplotype containing all the mismatches, except for the two exceptions mentioned above, is referred to as haplotype B. Haplotype B includes patients 4702 (mutation group 698), 4401 (mutation group 2163) and all patients in mutation groups 2159 and 386. Patients 4743 and 5043 both had all the mismatches in haplotype B as well as the above mentioned extra cleavage bands (table 4).

The estimated positions of the polymorphisms were based on the position of the primers used for each amplified segment and their distance from the ATG codon in exon 1 in the f8 gene, and the size of the small cleavage band. Since the cleavage band size does not reveal whether the polymorphism is in the centromeric or telomeric part of the segment, there are still two possibilities for the estimated position of each polymorphism (table 4).

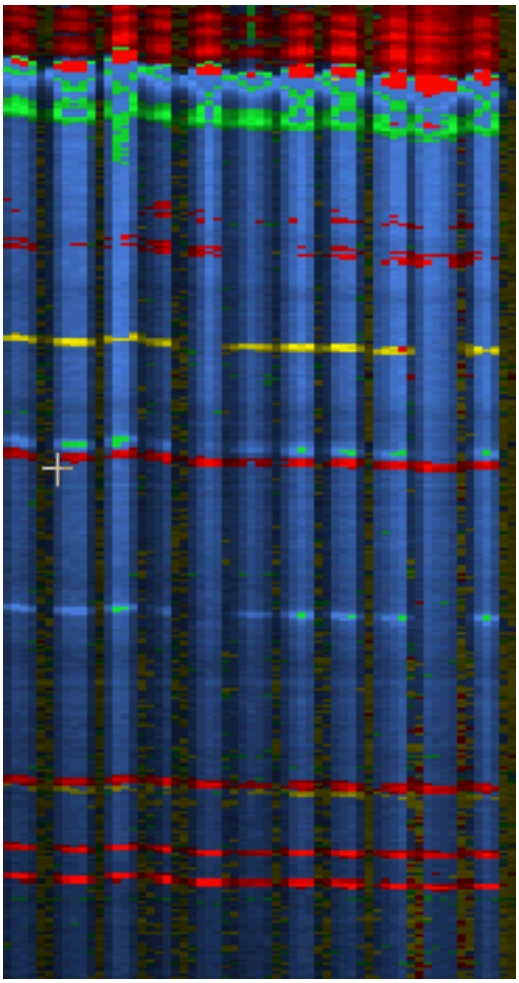


Figure 5. Mismatch gel from a hydroxylamine reaction on 11 patient samples. The clear yellow band of 975 bases represents one of the many cleavages in segment 19 (full gel is not shown). The two blue bands of 621 and 842 bases are the products of the mismatch in segment 13. No mismatch was seen in segment 2, labelled green. Later experiments proved the faint green bands, shining through the blue, to be false (data not shown). Red bands are size marker and top bands are full-length products. The three cleavage bands are here seen in, from left to right, patient samples: 4401, 4530, 4575, 4576, 4607, 4613, 4702, 4743 and 5043. No bands are present in patient samples 4594 and 4865, lane five and ten respectively from left to right.

Table 4. Mismatches found in the investigated region 5' of exon 1 in the f8 gene. Only the smaller of the two cleavage bands, resulting from one mismatch, was considered as it was usually easier to score. The band size was set to be the average band size obtained from all patients with that same mismatch. The estimated positions of the polymorphisms refer to the physical distance upstream from the ATG codon in exon 1 of the f8 gene. Haplotype B includes patients 4702, 4401 and all patients in mutation groups 2159 and 386. The "x" marks in what patients the mismatches were found.

Segment	Small cleavage band size (bp)	Estimated positions of polymorphism (bp)	SPFCCM reaction (hydroxyl-amine/os-mium)	Haplotype B	Patient 5043	Patient 4743
1	316	219/1116	h/o		x	
9	395	9642/10382	h	x	x	x
12	168	13463/14656	o	x	x	x
13	671	15207/15390	h/o	x	x	x
14	319	16104/16925	o	x	x	x
18	370	20885/21663	h/o	x	x	x
19	317	22166/23044	h	x	x	x
19	338	22187/23023	h/o	x	x	x
19	434	22283/22927	o			x
19	485	22334/22876	o	x	x	x
19	531	22380/22830	h/o	x	x	x

4.3.1 The segment 1 mismatch

The seemingly rare mismatch found in segment 1 in patient 5043, needed to be further investigated. In part because of its uniqueness but also due to the fact that segment 1 covered the promoter region (20). The position of the mutation was established by sequencing segment 1, using both 3' and 5' primers for that segment. Dr Naushin Waseem, member of the laboratory, performed the sequencing. The mismatch turned out to be a base change from C to T in nt -218 (considering ATG in exon 1 to be nt +1), a base change previously encountered by Dr Waseem in two patients originally not chosen for this study. The two patients, 4920 and 5241, were now included in the rest of the project to investigate whether patients 4920, 5241 and 5043 shared the same haplotype, not only in the 23kb upstream region of the f8 gene, but also when considering known intragenic and extragenic polymorphic sites of the f8 gene.

To quickly establish whether patients 4920 and 5241 belonged to haplotype A or B, one of the mismatches was chosen to represent the haplotype. This could be done because of the lack of deviation from these two haplotypes. The mismatch in segment 12 was chosen because of its small cleavage band and thereby closeness to one of the primers. It is usually easier to get good sequence for shorter distances. The segment was sequenced in both directions for patients 4920 and 5241 as well as two patients belonging to haplotype A and B respectively. The sequences showed that both patients 4920 and 5241 carried the same base change as the patient belonging to haplotype B. Thus it was concluded that they both shared the same haplotype as patient 5043 in the 23kb upstream region of the f8 gene.

To further investigate the rare mutation in segment 1, 46 female control individuals, with no family history of haemophilia A, were sequenced using a reverse primer 246bp downstream of the mutation. All control individuals were homozygous for the C allele.

A χ^2 test for association between the nt –218 mutation and haemophilia A patients with no known causative mutation (study group) was performed, using all patient data from this study and adding available data from this laboratory. The control group consisted of patients with known causative mutations and the 46 phenotypically normal females (table 5).

Table 5. χ^2 test for association between the promoter mutation and haemophilia A patients with no known causative mutation (study group), using all available data from this laboratory.

	No. of chromosomes	No. of T alleles
Study group	103	3
Control group	167	0
$\chi^2 = 4.92$		

The test showed a significant ($P < 0.05$) difference in no. of individuals with T allele between study group and control group.

4.4 Investigation of known intragenic and extragenic polymorphic sites

4.4.1 Dinucleotide repeats

In the patients the investigated $(CA)_n$ repeat within intron 13 of the f8 gene showed 5 alleles, from 24 to 28 dinucleotides (figure 6). The estimated no. of repeats was based on the assumption that a dinucleotide repeat of 20 gives a fragment of 141bp (6). Fragment lengths obtained from a second gel run was also taken into account where data were uncertain (data not shown)(table 3).

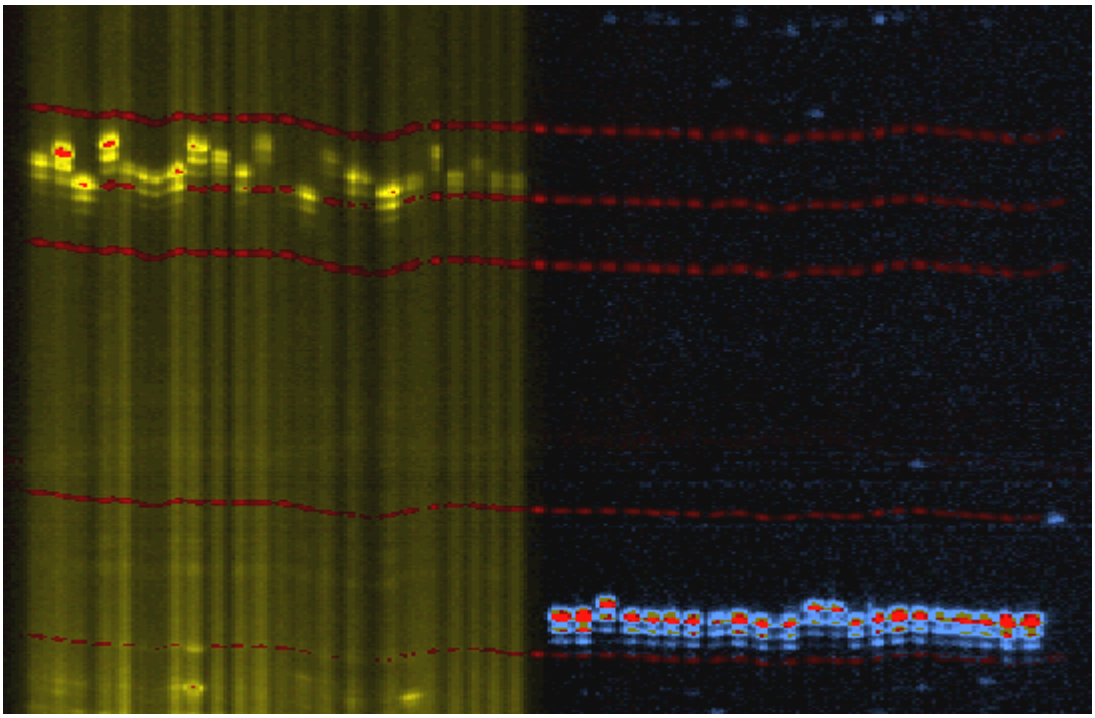


Figure 6. Dinucleotide repeat length analysis. $(CA)_n$ repeats within intron 13 are labelled yellow and $(GT)(AG)_n$ repeats within intron 22 are labelled blue. Red bands are size markers. All 22 patient samples are loaded in numerical order according to patient number, from left to right, except the 21st and 22nd sample which are patient samples 4920 and 5241 respectively.

The $(GT)_n(AG)_n$ repeat within intron 22 of the f8 gene showed little variation. Notably patients 4702, 4401 and 4743 had a 26 dinucleotide repeat, while the rest of the patients had 25. The assumption that 26 dinucleotides gives a fragment of 83bp was used to estimate the no. of repeats (7) (table 6).

Table 6. No. of (CA) repeats in the investigated region of intron 13 and no. of (GT)(AG) repeats in the investigated region of intron 22, in the f8 gene.

Patient no. (mutation group no.)	Length of fragment (bp) Intron 13	Estimated no. of (CA) repeats	Length of fragment (bp) Intron 22	Estimated no. of (GT)(AG) repeats
4865 (162)	155.91	27	81.57	25
5019 (162)	156.00	27-28	81.68	25
4453 (162)	155.98	27	81.59	25
4607 (2159)	152.03	26	81.65	25
4613 (2159)	156.15	28	81.79	25
4530 (2159)	152.18	26	81.79	25
4998 (698)	151.71	25-26	81.48	25
4579 (698)	156.06	28	81.66	25
4702 (698)	150.00	24-25	83.70	26
4260 (698)	152.18	26	81.79	25
5065 (2229)	154.63	27	81.70	25
4883 (2229)	154.02	27	81.64	25
4328 (2229)	154.12	27	81.79	25
4576 (386)	152.00	25-26	81.59	25
4575 (386)	152.02	26	81.74	25
4401 (2163)	150.00	24-25	83.97	26
5043 (unknown)	152.95	26	81.86	25
4594 (unknown)	153.97	26	81.70	25
4743 (unknown)	149.83	24	83.74	26
4937 (unknown)	152.00	25-26	81.55	25
4920 (unknown)	152.00	25-26	81.71	25
5241 (unknown)	152.02	26	81.71	25

4.4.2 Restriction enzyme digestions

In exon 14 of the DKC gene, a 360bp long segment was digested by the restriction enzyme *Blp* I. Cleavage of the segment into smaller segments of 215bp and 145bp occurred in patients having a T at the C/T polymorphic site at nt 1557 (figure 7). Seven of the 22 patients (32 %) had a T at the site, including patients 5043, 4920, 5241, 4530, 4607 and the two patients of mutation group 386. All of these patients also belonged to haplotype B in the 23kb 5' region of the f8 gene (table 7).

A 184bp long segment located in exon 15 of the DKC gene, within the 3' untranslated region, was digested by the restriction enzyme *Hae* III. The digestion resulted in cleavage into a 157bp long segment and a 27bp short (therefore not visible) segment, for the patients having a G at the G/A polymorphic site of nt 1647 (figure 8). Only four of the patients (18 %) had an A at the polymorphic site; patients 4998 and 4260 of the 698 mutation group, 4594 and 4937 with unknown mutations, all four having haplotype A in the 23kb 5' region of the f8 gene (table 7)

A 437bp long segment located in intron 19 of the f8 gene was digested by the restriction enzyme *Hind* III. Cleavage, resulting in a 380bp long fragment and a 57bp long fragment (not visible) or non cleavage, indicates the base C or T respectively at the polymorphic site (figure 9). Ten of the patients (45 %) had a T at the polymorphic site. These patients, 4607, 4530, 4702, 4576, 4575, 4401, 5043, 4743, 4920 and 5241, also showed the haplotype B pattern in the 23kb 5' region of f8 (table 7).

Table 7. The PCR products cleaved (+) or not cleaved (-) by *Blp* I, indicating T or C respectively, at the polymorphic site of exon 14 in the DKC gene. The PCR products cleaved (+) or not cleaved (-) by *Hae* III, indicating G or A respectively, at the polymorphic site of exon 15 in the DKC gene. The PCR products cleaved (+) or not cleaved (-) by *Hind* III, indicating C or T respectively, at the polymorphic site of intron 19 of the f8 gene.

Patient no. (mutation group no.)	Cleaved/not cleaved (+/-) by Blp I	Cleaved/not cleaved (+/-) by Hae III	Cleaved/not cleaved (+/-) by Hind III
4865 (162)	-	+	+
5019 (162)	-	+	+
4453 (162)	-	+	+
4607 (2159)	+	+	-
4613 (2159)	-	+	+
4530 (2159)	+	+	-
4998 (698)	-	-	+
4579 (698)	-	+	+
4702 (698)	-	+	-
4260 (698)	-	-	+
5065 (2229)	-	+	+
4883 (2229)	-	+	+
4328 (2229)	-	+	+
4576 (386)	+	+	-
4575 (386)	+	+	-
4401 (2163)	-	+	-
5043 (unknown)	+	+	-
4594 (unknown)	-	-	+
4743 (unknown)	-	+	-
4937 (unknown)	-	-	+
4920 (unknown)	+	+	-
5241 (unknown)	+	+	-

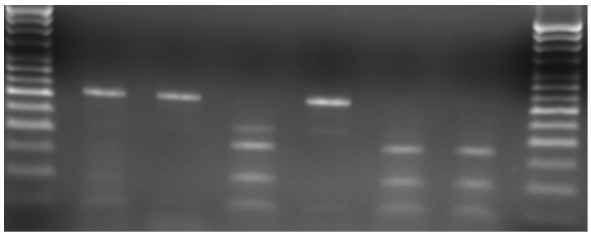
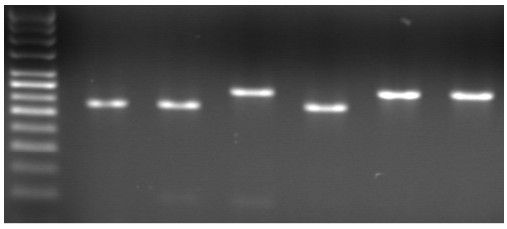


Figure 7. Restriction enzyme digestion of fragment in exon 14. The patient samples are from left to right: 4998, 5019, 5043, 5065, 4920 and 5241. Uncleaved products are 360bp. Cleaved productsare 215bp and 145bp (50bp ladder).



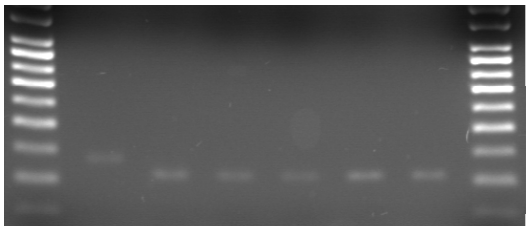


Figure 8. Restriction enzyme digestion of fragment in exon 15. The patient samples are from left to right: 4998, 5019, 5043, 5065, 4920 and 5241. Uncleaved products are 184bp. Cleaved products are 157bp and 27bp (not visible) (50bp ladder).

Figure 9 (left). Restriction enzyme digestion of fragment in intron 19. The patient samples are from left to right: 4998, 5019, 5043, 5065, 4920 and 5241. Uncleaved products are 437bp. Cleaved products are 380bp and 57bp (not visible) (50bp ladder).

4.4.3 Sequenced polymorphisms

A C to T transition in the 3' UTR of the VBP1 gene was investigated by sequencing a 933bp long PCR product. The C/T polymorphism was found 263bp from the sequencing primer. Twelve of the patients had the T (54 %), including patients 5043 and 5241, but notably patient 4920 had the C. In one of the patients, 4865, no information was obtained (table 8).

The A/G polymorphism of exon 26 of the f8 gene was also investigated by sequencing. The PCR product was 355bp long, with the polymorphic site 125bp from the sequencing primer. Ten of the patients (45 %) had the A, all of them belonging to haplotype B in the 23kb 5' region of f8, while the haplotype A patients had the G. The only discrepancy from this pattern is found in patient 4613, belonging to haplotype B, where a G was found (table 8).

Table 8. The C/T polymorphism in the 3' UTR of the VBP1 gene. No information was obtained for patient 4865. The A/G polymorphism in exon 26 of the f8 gene.

Patient no. (mutation group no.)	3' UTR VBP1 polymorphism	Exon 26 f8 polymorphism
4865 (162)	Unknown	G
5019 (162)	C	G
4453 (162)	T	G
4607 (2159)	T	A
4613 (2159)	C	G
4530 (2159)	T	A
4998 (698)	T	G
4579 (698)	C	G
4702 (698)	T	A
4260 (698)	T	G
5065 (2229)	C	G
4883 (2229)	C	G
4328 (2229)	C	G
4576 (386)	T	A
4575 (386)	T	A
4401 (2163)	T	A
5043 (unknown)	T	A
4594 (unknown)	C	G
4743 (unknown)	T	A
4937 (unknown)	C	G
4920 (unknown)	C	A
5241 (unknown)	T	A

4.5 Haplotypes

All extragenic and intragenic polymorphisms detected, including the haplotype in the promoter region, was summarised and arranged according to their position in the chromosome and overall haplotype. Eleven haplotypes, not considering the causative mutations in the coding regions, could be distinguished among the 22 patients. Only one haplotype was found for mutation groups 2229 and 386 respectively. Mutation groups 162 and 2159 each had two different haplotypes and mutation group 698 had three. All haplotypes within a mutation group differed by more than one polymorphism. However the three patients sharing the nt –218 mutation differed at only one polymorphic site (table 8).

Table 8. All detected polymorphisms arranged according to their position in the chromosome and haplotype. As described above patient 4743 had an additional mutation in segment 19 (+°), and the patients 5043, 5241 and 4920 had the nt –218 mutation (+*).

Patient no. (mutation group no.)	VBP1 3' UTR polymor-phism	F8 promoter Haplotype B	F8 intron13 Estimated no. of (CA) repeats	F8 intron19 Cleaved/ not cleaved (+/-) by Hind III	F8 intron22 Estimated no. of (GT)(AG) repeats	F8 exon 26 polymor-phism	DKC1 exon15 Cleaved/ not cleaved (+/-) by Hae III	DKC1 exon14 Cleaved/ not cleaved (+/-) by Blp I
5019 (162)	C		27-28	+	25	G	+	-
4579 (698)	C		28	+	25	G	+	-
5065 (2229)	C		27	+	25	G	+	-
4883 (2229)	C		27	+	25	G	+	-
4328 (2229)	C		27	+	25	G	+	-
4594 (unknown)	C		26	+	25	G	-	-
4937 (unknown)	C		25-26	+	25	G	-	-
4453 (162)	T		27	+	25	G	+	-
4865 (162)	Unknown		27	+	25	G	+	-
4260 (698)	T		26	+	25	G	-	-
4998 (698)	T		25-26	+	25	G	-	-
4613 (2159)	C	+	28	+	25	G	+	-
4743 (unknown)	T	+°	24	-	26	A	+	-
4702 (698)	T	+	24-25	-	26	A	+	-
4401 (2163)	T	+	24-25	-	26	A	+	-
4607 (2159)	T	+	26	-	25	A	+	+
4530 (2159)	T	+	26	-	25	A	+	+
4576 (386)	T	+	25-26	-	25	A	+	+
4575 (386)	T	+	26	-	25	A	+	+
5043 (unknown)	T	+*	26	-	25	A	+	+
5241 (unknown)	T	+*	26	-	25	A	+	+
4920 (unknown)	C	+*	25-26	-	25	A	+	+

5 Discussion

The SPFCCM is a very fast and reliable method for the detection and analysis of polymorphic variation in the human genome. DNA segments of 1.5kb in length are usually appropriate to screen X-chromosome sequences. Larger segments can contain multiple mismatches that are difficult to analyse.

By the use of three different fluorescent colours three segments can be chemically cleaved in the same tube and loaded in the same well on the ABI gel. If 64 lanes are used per ABI gel run, a screening rate of 144kb per ABI gel run can be accomplished. All mismatches are detected except the T-G mismatch, in which the T is in the probe strand and lies 3' to a G.

Mismatches at 50bp and less from each end of a segment are also not detectable because the small cleavage bands produced are too weakly labelled and the large ones too similar in size to the full segment. This can be avoided however by designing segments that overlap, as has been done in this study where most segments overlap by 200bp.

The chemistry of SPFCCM has also proven to be very reliable with reproducible results. The chemicals used are cheap and stable for up to three months. They are very toxic, and great care must be taken to handle them correctly. The amounts used for the reactions are very small however, and the risks small when carrying out the procedures in a fume cupboard.

The method is also very reliable in estimating the exact position of a polymorphism, within a margin of error of ±2bp. Although this has not been investigated in this study, it is worth mentioning that the position of the segment 1 polymorphism was estimated from the mismatch results with only 1bp error, and thus coincide with earlier studies concerning the reliability of SPFCCM (21).

5.1 F8 activity level discrepancies

No conclusions can be drawn concerning the differences in f8 activity levels within mutation groups. The two haplotypes found in the 23kb region upstream do not discriminate between f8 activity levels.

One explanation concerns the two basic methods of measuring the circulating f8 activity level. In short, the 1-stage assay is based on measuring the clotting time of the patient blood plasma The 2-stage assay has a different approach. In the first stage clotting is prevented by binding the patient blood plasma to $Al(OH)_3$. The plasma is then mixed with human serum that contains factor X. The rate of generation of activated factor X is dependent on the f8 activity. In the second stage normal blood plasma is added and the time to clot formation is dependent on the amount of factor X generated in the first stage (22, 23).

Interestingly, several studies have shown a discrepancy in the measured f8 activity level when using different methods. Even though both assays have been available for four decades it is still not known which method more accurately represent the f8 activity level in vivo, and the reason for the differences. The possible explanations all concern the site of the specific mutation and how it can affect the rate of f8 activation and deactivation. In this study it is not known which of the assays have been used to measure the f8 activity levels in the 22 patients, and it is possible that some of the stated levels should not be compared (22, 23).

Other possibilities to be considered, when searching for the explanation to phenotypic differences among patients with the same deleterious mutation, include polymorphic variation in genes that affect the secretion, transport, activation and degradation of f8.

5.2 The haplotypes

In all cases but one, the haplotype of the 23kb region 5' of the f8 gene is shared by all patients from the same mutation group. When considering all extragenic and intragenic polymorphic sites investigated, more diversity within mutation groups are seen. As haemophilia A mutations are fairly rapidly eliminated from the population, recombination is unlikely to be an explanation for the diversity within mutation groups unless the haplotypes were to differ by only one polymorphic site, or several adjacent sites.

However, for mutation groups 162, 698 and 2159 the haplotypes differ between patients within a group by more than one polymorphism that could not be due to a single recombination event. All three mutations are C to T transitions on either sense or antisense DNA strand, the most common point mutation, which could imply a hotspot for mutations.

Thus, it is very probable that the mutations at codon 162 and 2159 have arisen on at least two separate occasions respectively, as two haplotypes are seen within both these groups, and the mutation at codon 698 has arisen on at least three occasions, as three haplotypes are seen within this group. No conclusions can be drawn regarding common ancestry for mutations at codon 386 and 2229, as the patients included in this study show no diversity at the polymorphic sites investigated.

5.3 The segment 1 mutation

No conclusive evidence can be given in this study regarding the significance of the mutation at nt -218 observed in three patients with mild haemophilia A. The arguments in favour of its potential functional relevance are the following. No other mutations have been found in these patients' DNA when screening the coding regions of the f8 gene. Two of them share the same f8 activity level, 30%. Unfortunately the level of the third patient is unknown. A χ^2 test shows significant association between the mutation and patients with unknown causative mutations, and the mutation is not found in the control group.

Critical evidence in favour of the causative role of the mutation would be obtained if either it could be shown that the mutation has arisen anew in one of the patients' family or that it has occurred independently in two different haplotypes, that could not be thought to derive from one another by a single recombination event.

The polymorphism in the VBP1 gene is not shared by all three patients. This polymorphism is less than 100kb from the f8 gene, which should correspond to less than 0.1cM. However the likelihood for a recombination event is dependent on time and the age of the mutation is unknown, therefore the possibility that the haplotype difference is due to a single recombination event that has occurred since the appearance of the mutation at nt –218 cannot be excluded.

6 Acknowledgements
Firstly, I wish to thank Professor Francesco Giannelli for welcoming me as a guest student, and for his much appreciated guidance and advice, especially regarding the writing up of this project; Dr. Peter Green for supervision, support, trust and always being available. I also wish to thank all people working in the laboratory for their help, patience and nice company; especially Dr. Naushin Waseem for her time and invaluable advice, and Theo Anagnostopoulos for enjoyable demonstrations and aid.

7 References

1. Hoyer LW. Haemophilia A.*N Engl J Med* 330 (1): 38-47 (1994)
2. Lenting PJ, van Mourik JA, Mertens K. The life cycle of coagulation factor VIII in view of its structure and function. *Blood* 92 (11): 3983-6 (1998)
3. Mann KG, Nesheim ME, Church WR, Haley P, Krishnaswamy S. Surface-dependant reactions of the vitamin K-dependant enzyme complexes. *Blood* 76 (1): 1-16 (1990)
4. Gitschier J, Wood WI, Goralka TM, Wion KL, Chen EY, Eaton DH, Vehar GA, Capon DJ, Lawn RM. Characterization of the human factor VIII gene. *Nature* 312 (5992): 326-30 (1984)
5. Levinson B, Kenwrick S, Lakich D, Hammonds G Jr, Gitschier J. A transcribed gene in an intron of the human factor VIII gene. *Genomics* 7 (1): 1-11 (1990)
6. Lalloz MRA, McVey JH, Pattinson JK, Tuddenham EGD. Haemophilia A diagnosis by analysis of a hypervariable dinucleotide repeat within the factor VIII gene. *Lancet* 338: 207-11 (1991)
7. Lalloz MRA, Schwaab R, McVey JH, Michaelides K, Tuddenham EGD. Haemophilia A diagnosis by simultaneous analysis of two variable dinucleotide tandem repeats within the factor VIII gene. *British J Haematology* 86: 804-9 (1994)
8. Ahrens P, Kruse TA, Schwartz M, Rasmussen PB, Din N. A new HindIII restriction fragment length polymorphism in the hemophilia A locus. *Hum Genet* 76 (2): 127-8 (1987)
9. Personal contact: Waseem N. Department of medical and molecular genetics, GKT school of medicine, King's College, Guy's Hospital, London
10. Brinke A, Green PM, Giannelli F. Characterization of the gene (VBP1) and transcript for the von Hippel-Lindau binding protein and isolation of the highly conserved murine homologue. *Genomics* 45: 105-12 (1997)
11. Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL, Cowan NJ. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93 (5): 863-73 (1998)
12. Hassock S, Vetrie D, Giannelli F. Mapping and characterization of the X-linked dyskeratosis congenita (DKC) gene. *Genomics* 55: 21-7 (1999)
13. Knight SW, Heiss NS, Vulliamy TJ, Greschner S, Stavrides G, Pai GS, Lestrigant G, Varma N, Mason PJ, Dokal I, Poustka A. X-linked dyskeratosis congenita is predominantly cause by missense mutations in the DKC1 gene. *Am J Hum Genet* 65 (1): 50-8 (1999)
14. Cotton RG, Rodrigues NR, Campbell RD. Reactivity of cytosine and 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 (12): 4397-401 (1988)
15. Rowley G, Saad S, Giannelli F, Green PM. Ultrarapid mutation detection by multiplex, solid-phase chemical cleavage. *Genomics* 30 (3): 574-82 (1995)
16. Antorakis SE, Kazazian HH, Tuddenham EGD. Molecular etiology of factor VIII deficiency in hemophilia A. *Hum Mut* 5: 1-22 (1995)
17. Primer3 Old Version. http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi/ (Oct. 2001).
18. Surin VL, Zhukova EL, Krutov AA, Solovyev GY, Grineva NI. Simple and convenient detection of a HindIII polymorphic site in intron 19 of factor VIII using PCR. *Nucleic Acids Research* 18 (11): 3432 (1990)

19. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74 (12): 5463-7 (1977)
20. Figueiredo MS, Brownlee GG. Cis-acting elements and transcription factors involved in the promoter activity of the human factor VIII gene. *J Biol Chem* 270 (20): 11828-38 (1995)
21. Anagnostopoulos T, Green PM, Rowley G, Lewis CM, Giannelli F. DNA variation in a 5-Mb region of the X chromosome and estimates of sex-specific/type-specific mutation rates. *Am J Hum Genet* 64: 508-17 (1999)
22. Keeling DM, Sukhu K, Kemball-Cook G, Waseem N, Bagnall R, Lloyd JV. Diagnostic importance of the two-stage factor VIII:C assay demonstrated by a case of mild haemophilia associated with His 1954 – Leu substitution in the factor VIII A3 domain. *British J Haematology* 105: 1123-26 (1999)
23. Pipe SW, Saenko EL, Eickhorst AN, Kemball-Cook G, Kaufman RJ. Hemophilia A mutations associated with 1-stage/2-stage activity discrepancy disrupt protein-protein interactions within the triplicated A domains of thrombin-activated factor VIIIa. *Blood* 97 (3): 685-91 (2001)