The Molecular Genetics, Prevention and Control of Haemophilia

Ian Peake

The cloning and isolation of the human factor VIII (FVIII) and factor IX (FIX) genes in the 1980s has lead to an increased understanding of the genetic and hence the molecular basis of haemophilia A and B respectively. These studies are of considerable benefit for accurate carrier detection and prenatal diagnosis in families with haemophilia, and provide insights into the relationships between genetic defects and clinical manifestations.

Haemophilia A

The Factor VIII Gene

The FVIII gene is 186 kb in length situated on the long arm of the X chromosome at Xq28. It consists of 26 exons and 25 introns. The processed mRNA is about 9 kb in length and predicts a precursor protein of 2351 amino acids, which in turn leads to a mature FVIII protein of 2332 amino acids. The largest intron, intron or IVS 22, is unusual in that it contains a CpG island associated with two further genes called F8A and F8B[1,2]. The latter is transcribed in the same direction as FVIII from an exon in intron 22 but then including exons 23-26 of the FVIII gene. The function of its transcribed product is unknown. F8A consists of an intronless gene encoded entirely from within intron 22 in the opposite direction to FVIII. The presence of two further copies of this gene some 500 kb telomeric to the intron 22 copy[1] has considerable significance since intrachromosomal recombination between these homologous sequences is the cause of almost 50% of cases of severe haemophilia A. The function of the F8A gene transcript is at present unknown.

The FVIII Protein

FVIII is a large plasma glycoprotein, stabilised in plasma by its non-covalent association with von Willebrand factor (VWF). Where this association is disturbed, as in type 2N von Willebrand's disease (VWD) by mutations in the VWF gene, plasma FVIII levels are reduced, presumably through its decreased half-life. This condition can be confused in males with mild haemophilia A[3].

The mature FVIII protein of 2332 amino acids consists of a series of repeating domains in the order A1,A2,B, A3,C1,C2[4]. There is close homology between FVIII and FV except in the B domain (encoded by exon 14), which shows no significant homology to any known protein sequence. In fact, B-domainless FVIII, produced by genetic engineering, has normal FVIII activity and is being studied as a treatment of haemophilia[5]. Certain sequences within the FVIII molecule are important for its role in blood coagulation and include proteolytic cleavage sites at amino acids 372 and 1689 [6], the VWF-binding region of about 40 amino acids within domain A2, and the phospholipid-binding sites within the C1 and C2 domains.

The Genetic Basis of Haemophilia A

An increasing number of nucleotide changes within the gene in individuals with haemophilia have been reported. Some of the changes are polymorphisms and, while also occurring in the normal non-haemophilia population, can be of considerable use in family studies to track the haemophilic FVIII gene in affected families. Ethnic variations in allele frequencies mean that the usefulness of these polymorphisms varies throughout the world. The highly informative CA repeat polymorphism in intron 13 and the (GT)n(AG)n repeat in intron 22 are apparently equally informative world wide, though there are variations in the frequency of the different alleles[7].

The exponential increase in the number of reported causative mutations in haemophilia A (and haemophilia B) since the mid 1980s is directly related to developments in DNA and RNA analysis technology. Southern blotting has been generally replaced by PCR-based DNA amplification methods starting from both DNA itself and following reverse transcription from mRNA. This latter procedure utilises the small amount of ectopically expressed FVIII mRNA found in circulating leukocytes. Screening for differences in DNA sequence between haemophilic and normal FVIII genes has been greatly facilitated by the rapid mismatch detection methods. These procedures allow for PCR amplified sequences to be directly compared and include methods based on chemical cleavage of the DNA and on gel electrophoresis of the sequences under specialised conditions, often as DNA heteroduplexes containing normal and putative abnormal DNA. In Sheffield the technique of CSGE (conformation sensitive gel electrophoresis) has proved to be particularly useful in the detection of FVIII gene mutations and polymorphisms[8]. When differences are found, the respective region of the "abnormal" DNA is sequenced either manually or, increasingly, by automated procedures.

Southern blotting techniques are still valuable in detecting large gene deletions or insertions and for detecting the intron 22 inversion mutation.

An international database of mutations detected in individuals with haemophilia A has been created by Tuddenham and colleagues and is now accessible on the

World Wide Web[9] at http://europium. mrc. rpms.ac.uk/. Up to 1998 536 patient entries were reported, of which 304 were unique molecular events

FVIII Gene Mutations in Haemophilia A

Gene deletions

Deletions are arbitrarily divided into large (>200 bp) or small (<200 bp) and at present there are over 80 large deletions reported on the database, ranging from less than 1 kb to complete gene deletions. They almost always result in severe haemophilia A, with over 30% developing inhibitors to FVIII after replacement therapy . The few reported cases of milder disease probably result from in-frame splicing of mRNA to yield a truncated FVIII molecule which has some biological activity[10-12].

Small deletions can vary from single bases to up to 200bp. They generally result in frameshifts and the resulting termination or STOP codon. Rarely, deletions of a codon triplet or triplets have been reported to give rise to milder forms of the disease. Of considerable interest is the report by Young and colleagues[13] of a single T deletion within exon 14 of a patient with mild haemophilia A, resulting in the sequence AAAAAAAATAA being replaced by Ax10. Theoretically this deletion should result in a frameshift mutation and severe disease. However detailed sequencing of cloned PCR sequences of DNA and mRNA from this region of the gene in both normals and from the patient clearly show that a significant number of mRNA transcripts are found in which an inframe situation exists. This is by the formation of an Ax8 or Ax11 sequence, presumably by slippage during transcription of the repetitive sequence in the patient. It is also possible that the abnormal mRNA sequence is misread at translation to produce some normal FVIII protein (ribosomal frameshifting). Although these transcripts would contain an amino acid deletion or replacement it is argued by the authors that such FVIII molecules should have FVIII activity, particularly since the region encoded by exon 14, the B domain has no known role in the biological activity of the protein and can be deleted without affecting this activity (see above). Several of the single base pair deletions reported on the database occur in nucleotide repeat sequences in exon 14 and, although there are no clear genotype/phenotype discrepancies as reported above, the level of inhibitor development in these individuals is less than 5%, compared to over 30% in patients with termination codons caused by nucleotide transitions or transversions[14]. It is therefore entirely possible that such individuals can produce small amounts of a FVIII protein, sufficient to create immunological tolerance to transfused FVIII, but insufficient to modulate the genotypically-expected phenotype .

Gene Insertions

Fewer gene insertions than deletions have been reported. The haemophilia A database contains 24 reports at present, two of which are large (2.1 and 3.8 kb) LINE sequence insertions in exon 14 of the FVIII gene[15]. LINE elements are retrotransposon sequences found throughout the genome. Most of the other insertions are single bases often within existing repeat nucleotide sequences and presumable arising from misreading of the original sequence during DNA replication.

Missense mutations

These are generally single base substitutions and as such lead to amino acid changes or termination (STOP) codons. By 1998 there were 182 such unique changes reported on the database. Changes within the exon/intron boundaries will give rise to altered or absent intron splicing. Several mutations occur on the data base many times and reflect either a founder affect or cytosine to thymine transitions at CpG sites, now recognised to be “hot spots” for mutations within the human genome..

Amino acid substitution can result in either low or undetectable FVIII biological activity in plasma. This can be accompanied by similar levels of FVIII protein (measured antigenically using specific antibodies) resulting in the “so-called” cross-reacting material reduced or negative phenotype (CRM-red or CRM-ve). This presumably results from the production of abnormal FVIII protein which is either inherently unstable, poorly expressed or has a very short plasma survival time. Cases of CRM+ve haemophilia A can arise where there are normal circulating plasma levels of FVIII antigen, but the biological activity is reduced or even absent. In cases where the mutation is known, little is known of the reason for the CRM status. In fact direct phenotype/genotype correlation’s in missense cases are rare and are restricted to mutations affecting specific cleavage sites essential for the normal activation or inactivation of FVIII or to specific binding regions of the protein, e.g. tyrosine to phenylalanine at amino acid 1680 disrupts VWF binding, which normally requires the sulphation of the tyrosine residue at this location. Two cases where mutations would affect N-glycosylation sites at amino acids 1566 and 1772 have been reported, both of which result in a CRM+ve phenotype with little or no biological activity.

FVIII gene rearrangements

Studies by Naylor et al [16,17] and those of Lakich et al [18] showed that in about 50% of patients with severe haemophilia A there was an inversion and translocation of part of the FVIII gene from exon 1 to exon 22 to a position some 500 kb telomeric to the remainder of the gene (exons 23 to 26), which remained at its usual location. This relocation and concomitant inversion is believed to be a result of homologous recombination between the regions of the X-chromosome including and surrounding the 3 copies of the F8A gene referred to above. This event occurs specifically between the copy within intron 22 (the int22 homologous region) and either the distal or proximal copies and results in inactivation of the gene, so causing severe haemophilia A.

Detection of these rearrangements generally relies on the use of Southern blotting using a specific probe for the

int22 homologous region, which of course also binds to the other homologous F8A-containing regions[19]. Distal rearrangements are the most common (80%), with proximal rearrangements representing about 15%, and a smaller group of rearrangements (< 5%) involving apparent extra and/or partial copies of the homologous region outside the FVIII gene. A recently described PCR based technique for the detection of the FVIII gene inversions requires the amplification of large >10 kb DNA fragments[20].

The result of an international consortium[21] have shown that the int 22 inversion (as these rearrangements are generally called) occurs throughout the world in haemophilic populations in 40-50% of severe haemophiliacs. Interestingly, in sporadic cases of haemophilia A caused by this inversion, over 90% of mothers carry the defect, and almost exclusively the affected gametes are supplied by a normal male, i.e. the patient's normal grandfather[22]. Overall the distal inversion is responsible for about 35% of all severe haemophilia A and the proximal inversion for 7%.

Haemophilia B

The FIX Gene and Protein

The factor IX gene was cloned in the early 1980s is located at Xq27 and spans 34 kb of DNA and has 8 exons. The complete gene has been sequenced[23]. The primary FIX gene translation product contains 461 amino acids, from which the prepro 46-amino acid leader sequence is cleaved to allow release of the mature 415 amino acid single chain peptide from the hepatocyte into the blood. The functional domains of the FIX protein are to a great extent encoded by specific exons. Thus exons 1 and 2 encode the signal peptide and most of the propeptide and glutamic acid (gla) domain. The propeptide also provides a recognition site for the vitamin K-dependent gamma carboxylase which is responsible for gamma carboxylation of the first 12 gla residues of the mature FIX protein[24]. Exon 3 encodes the final part of the propeptide and a hydrophobic stack domain. Exons 4 and 5 encode the epidermal growth factor-like domains which are involved in calcium, platelet and activated factor VIII binding[25,26]. The FIX activation peptide is encoded by exons 7 and 8 and includes the catalytic domain of His 221, Asp 269 and Ser 365.

The Genetic Basis of Haemophilia B

Analysis of the FIX gene has utilised the same methods as employed in the analysis of the FVIII gene for haemophilia A. Since the gene itself is much smaller, the coding, exon-intron boundaries and the 5′ and 3′ non-coding regions are relatively easy to analyze, and direct sequencing of PCR amplified material is often performed. However, mismatch detection methods can be successfully employed[27].

A number of polymorphic diallelic sequence changes have been reported[7] and have proved very useful in linkage studies for carrier detection and prenatal diagnosis. Significant ethnic variations in allelic frequencies have been observed, though the 5′ MseI and 3′ HhaI polymorphisms appear to be informative in most populations.

The relatively easy of analysis of the FIX gene in patients with haemophilia B has resulted in a large number of causative mutations being detected. In fact analysis of the FIX gene has often served as a model system for gene analysis in general. By 1998, 1713 entries were recorded on the Haemophilia B Mutation Database World Wide Web site at http://www.umds.ac.uk/molgen/haemBdatabase. These included 652 unique molecular events.

FIX Gene Mutations in Haemophilia B.

Deletions and Insertions

About 3% of haemophilia B is caused by large deletions/rearrangements of the FIX gene (29 reported at present not on the database). Interestingly about 50% of these have developed inhibitors to FIX following replacement therapy[28]. There is no rearrangement equivalent to that seen in the FVIII gene in haemophilia A in haemophilia B. One hundred thirty-two short deletions and insertions (30 bp) have been reported.

Missense mutations

In 1998 the database contained nucleotide changes which would result in 389 amino acid substitutions and 55 STOP codons, 60% and 8%, respectively, of all unique reported mutations. Fifty-eight percent of these changes involve CpG dinucleotides, a recognised hot-spot for mutations within the human genome. In all, amino acid substitutions have been reported at 360 of the 461 amino acid residues in the primary FIX translation product. Many of these changes have confirmed the functional importance of particular amino acids, e.g. the carboxyglutamyl residues in the gla domain.

It is now clear that a significant proportion of individuals with mild haemophilia B have the same mutation. Indeed in North America four mutations (Gly60Ser, Arg248Gln, Thr296Met and Ile397Thr) are responsible for about 30% of all cases of mild disease. This is due to a small number of founder mutations[29].

Promoter Region Mutations

The FIX gene promoter is probably one of the best understood promoters in the human genome. In particular 18 unique nucleotide changes have been reported in a 5′ flanking region of 40 nucleotides around the predominant transcription start site. These mutations occur in patients with the haemophilia B Leiden phenotype, in which a spontaneous steady increase in plasma FIX levels occurs after puberty with associated decrease in clinical symptoms, and have been shown to be at locations that contain binding sequences for liver-enriched transcription factors, including C/EPB and HNF4.

Carrier Detection and Prenatal Diagnosis in Haemophilia

Carrier detection and prenatal diagnosis in haemophilia is effectively performed in many laboratories throughout the

world by genetic analysis of the FVIII and FIX genes in the families of affected individuals. Gene tracking using intragenic diallelic or multiallelic linked DNA polymorphisms has formed the basis of these procedures. However, increasingly sophisticated technology now allows for the detection of the causative mutations in the FVIII and FIX genes in practically all cases of haemophilia A and B. This provides for precise carrier detection and prenatal diagnosis without the problems associated with linkage polymorphism analysis, which can include multiple family member analysis and possible non-informativeness. Mutation detection also provides a greater understanding of the haemophilia phenotype and clinical situation. While the presence of a common FVIII gene inversion and rearrangement in about 50% of patients with severe haemophilia A means that this is the first analysis performed in such patients, those without this mutation and all patients with haemophilia B require, in practice, use of procedures to analyze the complete functional gene, since it is clear that causative mutations can occur at almost any site within the FVIII or FIX genes.

References

1. Levinson B, Kenwrick S, Lakich D, Hammonds G Jr, Gitschier J. A transcribed gene in an intron of the human factor VIII gene. Genomics 1990; 7: 1-11.
2. Levinson B, Kenwrick S, Gamel P, Fisher K, Gitschier J. Evidence for a third transcript from the human factor VIII gene. Genomics 1992; 14: 585-589.
3. Nesbitt IM, Goodeve AC, Guilliatt AM, MakrisM, Preston FE, Peake IR. Characterization of type 2N von Willebrand disease using phenotypic and molecular techniques . Thromb . Haem. 1996; 75: 959-964.
4. Vehar GA, Keyt B, Eaton D, Rodriguez H, O'Brien DP, Rotblat F, Oppermann H, Keck R, Wood WI, Harkins RN, Tuddenham EGD, Lawn RM, Capon DJ. Structure of human factor VIII. Nature 1984; 312: 337-342.
5. Fijnvandraat K, Berntorp E, tenCate JW, Johnsson H, Peters M, Savidge G, Tengborn L, Spira J, Stahl C. Recombinant, B-domain deleted factor VIII (r-VIII SQ): Pharmacokinetics and initial safety aspects in hemophilia A patients. Thromb, Haem. 1997; 77: 298-302.
6. Pittman DD, Kaufman RJ. Proteolytic requirements for thrombin activation of anti-hemophilic factor (factor VIII). Proc. Natl. Acad. Sci. USA. 1988; 85: 2429-2433.
7. Goodeve AC, Peake IR. Diagnosis of haemophilia A and B carriers and prenatal diagnosis. In: Haemophilia. Ed Forbes CD, Aledort L, Madhok. Chapman and Hill, London 1998; p63-74.
8. Williams IJ, Abuzenadah A, Winship PR, Preston FE, Dolan G, Wright J, Peake IR, Goodeve AC. Precise carrier and prenatal diagnosis in families with haemophilia A: Use of conformation sensitive gel electrophoresis for mutation screening and polymorphism analysis. Thromb. Haem 1998;79:723-726
9. Kemball Cook G, Tuddenham EGD. The factor VIII mutation database on the World Wide Web: The haemophilia A mutation, search, test and resource site—HAMSTers update (version 3.0). Nucl. Acids Res. 1997; 25: 128-132.
10. Youssoufian H, Antonarakis SE, Aronis S, Tsiftis G, Phillips DG, Kazazian HH. Characterization of five partial deletions of the factor VIII gene. Proc. Natl. Acad. Sci. USA. 1987; 84: 3772-3776
11. Wehnert M, Herrmann FH, Wulff K. Partial deletions of factor VIII gene as molecular diagnostic markers in haemophilia A. Dis Markers (1989); 7: 113-117.
12. Lavergne JM, Bahnak BR, Vidaud M, Laurian Y, Meyer D. A directed search for mutations in hemophilia A using restriction enzyme analysis and denaturing gradient gel electrophoresis. A study of seven exons in the factor VIII gene of 170 cases. Nouv. Rev. Fr. Haematol. (1992); 34: 85-91.
13. Young M, Inaba H, Meyer LW, Higuchi M, Kazazian HH, Antonarakis SE. Partial correction of a severe molecular defect in haemophilia A because of errors during expression of the factor VIII gene. Am. J. Hum. Genet. (1997); 60: 565-573.
14. Schwaab R, Brackman HH, Meyer C, Seeha Fer. J, Kirchgesser M, Haack A, Olek K, Tuddenham EGD, Oldenburg J. Haemophilia A: Mutation type determines risk of inhibitor formulation. Thromb. Haem. (1995); 74: 1402-1406.
15. Kazazian HH. Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature (1988); 332: 164-166.
16. Naylor JA, Green PM, Rizza CR, Giannelli F. Factor VIII gene explains all cases of haemophilia A. Lancet (1992); 340: 1066-1067.
17. Naylor JA, Brinke A, Hassock S, Green PM, Giannelli F. Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions. Hum. Mol. Genet. (1993); 2: 1773-1778.
18. Lakich D, Kazazian HH. Jr., Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat. Genet. (1993); 5: 236-241.
19. Goodeve AC, Preston FE, Peake IR. Factor VIII rearrangements in patients with severe haemophilia A. Lancet (1994); 343: 329-330.
20. Lui Q, Nozari G, Sommer SS. Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in haemophilia A. Blood (1998); 92: 1458-1459
21. Antonarakis SE. and a consortium of more than 50 international authors. Factor VIII gene inversions in severe haemophilia A: results of an international consortium study. Blood (1995); 86: 2206-2212.
22. Rossiter JP, Young M, Kimberland ML, Hutter P, Ketterling PR, Gitschier J, Horst J, Morris MA, Schaid DJ, de Moreloose P, Sommer SS, Kazazian HH, Antonarakis SE. Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. Hum. Mol. Genet. (1994); 3: 1035-1039.
23. Yoshitake S, Schab BG, Foster DC, Davie EW, Kurachi W. Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). Biochem. J. (1995); 24: 3736-50.
24. Galeffi P, Brownlee GG. The propeptide region of clotting Factor IX is a signal for a vitamin K dependent carboxylase: evidence from protein engineering of amino acids –4. Nucleic Acids Res. (1987); 15: 9305-13.
25. Handford PA, Baron M, Mayhew M, Willis A, Beesley T, Brownlee GG, Campbell ID. The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. EMBO J. (1990); 9: 475-80.
26. Ahmad SS, Rawala R, Cheung WF, Stafford DW, Walsh PN. The role of the second growth-factor domain of human

factor IXa in a binding to platelets and in factor-X activation. Biochem. J. (1995); 310: 427-31.

27. Hinks JL, Winship PR, Makris M, Preston FE, Peake IR, Goodeve AC. A rapid method for haemophilia B mutation detection using conformation sensitive gel electrophoresis. Brit. J. Haematol. (1999); 104: 915-18
28. High KA. Factor IX: molecular structure, epitopes, and mutations associated with inhibitor formation. Adv. Exp. Med. Biol. (1995); 386: 79-86.
29. Thompson AR, Bajaj SP, Chen S-H, MacGillivray RTA. "Founder" effect in different families with haemophilia B mutation. Lancet (1990); 335: 418.