

Haemophilia B: database of point mutations and short additions and deletions, 7th edition

F. Giannelli, P. M. Green*, S. S. Sommer¹, M-C. Poon², M. Ludwig³, R. Schwaab³, P. H. Reitsma⁴, M. Goossens⁵, A. Yoshioka⁶, M. S. Figueiredo⁷ and G. G. Brownlee⁸

Paediatric Research Unit, Guy's Tower, London Bridge, London SE1 9RT, UK, ¹Mayo Clinic, 200 Southwest First Street, Rochester, MN 55905, USA, ²Division of Hematology and Hematologic Malignancies, Foothills Hospital, University of Calgary, 1403, 29th Street, NW Calgary, AB, Canada, ³Institute of Clinical Biochemistry, Sigmund-Freud-Strasse 25, 5300 Bonn 1, Germany and Institute of Experimental Haematology and Blood Transfusion, University of Bonn, Bonn, Germany, ⁴Laboratory of Experimental Internal Medicine, Academic Medical Centre, Amsterdam, The Netherlands, ⁵National Institute of Health and Medical Research, 51 Av. du Maréchal de Lattre de Tassigny, 94010 Creteil, France, ⁶Department of Pediatrics, Nara Medical College, Kashihara City, Nara, Japan, ⁷Department of Clinical Medicine, School of Medicine of Ribeirão Preto, University of São Paulo, 14048-900 Ribeirão Preto, São Paulo, Brazil and ⁸Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received October 7, 1996; Accepted October 8, 1996

ABSTRACT

The seventh edition of the haemophilia B database lists in easily accessible form all known factor IX mutations due to small changes (base substitutions and short additions and/or deletions of <30 bp) identified in haemophilia B patients. The 1535 patient entries are ordered by the nucleotide number of their mutation. Where known, details are given on: factor IX activity, factor IX antigen in circulation, presence of inhibitor and origin of mutation. References to published mutations are given and the laboratories generating the data are indicated.

Haemophilia B or Christmas disease is an X-linked recessive disorder due to mutations causing a marked deficit of coagulation factor IX—a glycoprotein of 415 amino acid residues normally present in plasma and an essential component of the clotting cascade. The disease affects 1 in ~30 000 males and is only very rarely found in females. The reduced genetic fitness of affected individuals has resulted in the fairly rapid loss of mutations from the population and continued renewal of the population pool of haemophilia B genes. Thus unrelated patients usually carry mutations of independent origin, although clusters of defects identical by descent do occur. For general reviews see refs (1,2).

The purpose of this database is to assemble in a yearly updated accessible, summary form, molecular data on the causative mutations of haemophilia B patients worldwide. It is not intended to replace primary publications, although it does contain a significant amount of unpublished work. We have continued our database numbering system (3) giving all patients a unique Patient Identity Number (PIN or ID number). As in previous years, we have included repeat observations of the same mutation, as well as molecularly unique mutations.

The factor IX gene lies on the long arm of the X chromosome at Xq27.1 and its entire sequence of 33 kb is known (4). It contains eight exons (a–h) encoding six major domains of factor IX. These are: (i) exon a, a hydrophobic *signal* peptide which targets the protein for secretion from the hepatocyte into the blood stream; (ii) exons b and c, a propeptide and *gla* domain—the latter containing 12 γ -carboxyglutamyl residues. This post-translational modification is required for the correct folding and calcium binding of factor IX. (iii) Exon d, a type B, or first epidermal growth factor (EGF)-like domain, which shows homology to EGF and, in addition, contains conserved carboxylate residues including a β -hydroxyaspartate at amino acid 64. This domain binds an additional Ca^{2+} with high affinity (5). (iv) Exon e, a type A, or second EGF-like domain which lacks the conserved carboxylate residues of the EGF type B domain. (v) Exon f, an activation domain, within which factor XIa, and VIIa plus tissue factor, cleave twice, converting factor IX to IXa; (vi) exons g and h, the serine protease or catalytic domain, responsible for the proteolysis of factor X to Xa. This region is homologous to other well studied serine proteases (e.g. chymotrypsin) and it is thought likely that his (221), asp (269) and ser (365), all participate in the classical catalytic mechanism.

Factor IX is initially synthesized in the liver as a precursor molecule, either 46, 41 or 39 amino acids [it is not known which, although 39 is probable (6)] longer at its N-terminus than the 415-long mature factor IX found in plasma. Processing steps occur in the hepatocyte prior to secretion and sequentially remove the hydrophobic signal peptide and the propeptide. In addition to the γ -carboxylation of the 12 N-terminal glutamyl residues carried out by a vitamin K-dependent carboxylase, and the partial β -hydroxylation of aspartate 64, N-linked carbohydrate side chains are added at residues 157 and 167 and an O-linked carbohydrate at serine 53 and 61 and threonine 159 and 169 (7–9). The crystal structure of factor IX has been determined recently (10).

* To whom correspondence should be addressed. Tel: +44 171 955 4645; Fax: +44 171 955 4644; Email: pgreen@umds.ac.uk

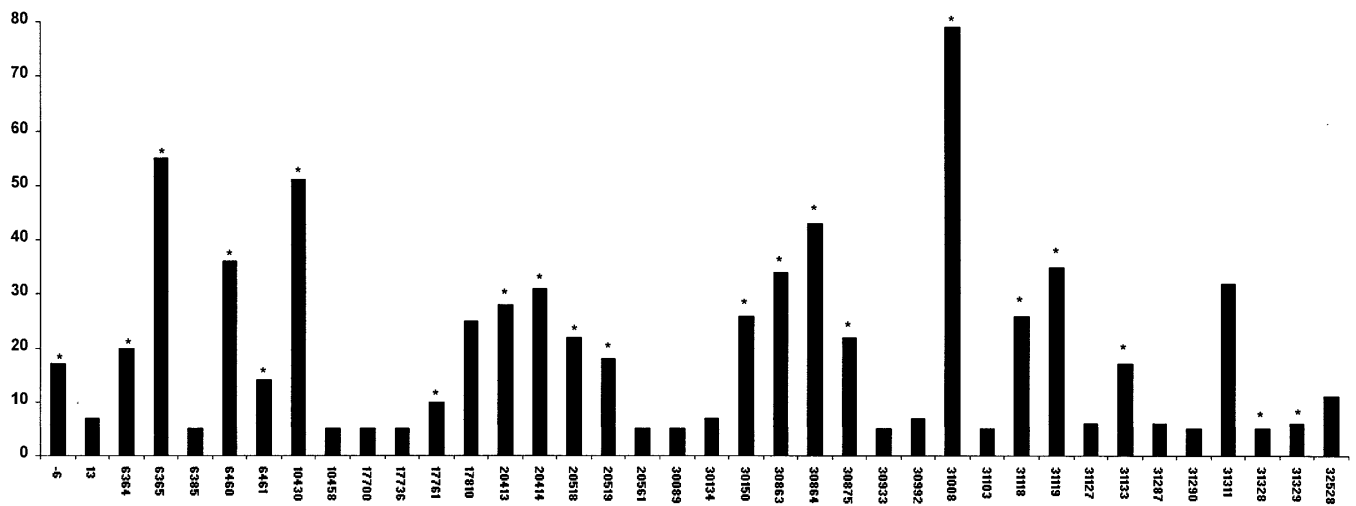


Figure 1. Positions and number of repeat mutations observed more than four times. Vertical axis: number of repeats of each mutation; horizontal axis: nucleotide position. Transitions at CpG sites are indicated by *.

There are 1535 patient entries in this seventh edition of the database compared with 1380 patients last year (11). Besides point mutations, these show 119 short (defined as <30 nucleotides) deletions or additions or both, made up from 91 deletions, 22 additions and six examples involving both additions and deletions. There are also 21 double mutations and one triple mutation. Twenty patients are known to have developed inhibitors, two are somatic mosaics and 25 are females (either affected or nonsymptomatic carriers). The list excludes the 29 patients with partial or complete gene deletions or more complex rearrangements quoted by Thompson, 1990 (12) and others observed since (13,14). Complete population studies suggest an incidence of 2–5% for these gross defects in haemophilia B patients. Of the 1535 patients listed, 597 show unique molecular events probably causing the disease, while the remainder are repeats. Many of these repeats occur at CG doublets and involve a CG→TG or CA change. As discussed before (3), such sites are believed to be genuine ‘hotspots’ for mutation. However, it is now becoming clear that the high number of repeat observations at some CG doublets, particularly those causing mild disease (e.g. 80 examples at 31 008) are caused, at least in part, by founder effects. A founder effect is responsible for the many repeats of a mutation at residue 31 311 that is not part of a CG (12,15). There are many examples in the database of repeating mutations; probably some but not all of these will have a common origin. The number of repeat observations (more than four) at each nucleotide site is shown in Figure 1. The database attempts to offer a view of the spectrum of mutations causing haemophilia B that is as accurate as possible and this is helped by the fact that about one-third of all mutations have been detected as a result of full population studies. However, some bias cannot be completely avoided. Obviously there is an over-representation of severe haemophilia-causing mutations as these tend to be the first analysed and the most likely to come to notice. We also expect under-representation of double mutants as not all laboratories have done ‘complete’ gene screens.

The distribution of mutants according to protein domains and control regions within the gene shows that mutations have been detected in all regions except the poly(A) site. The contribution of different base substitutions to the total of point mutations is

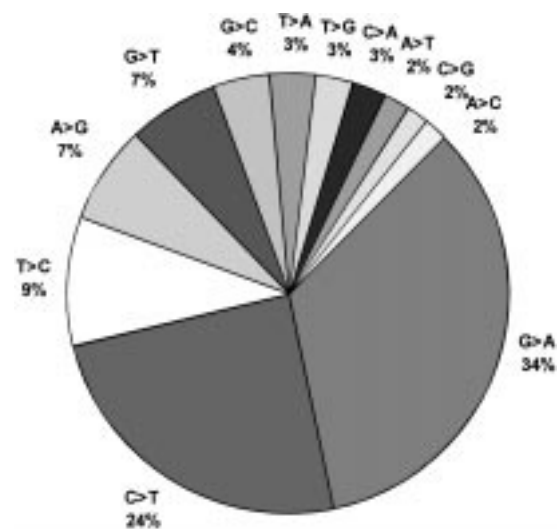


Figure 2. Pie chart showing all single base substitutions recorded in the database.

shown in Figure 2, while the breakdown of mutations according to class is shown in Figure 3. Remarkably, there are now 18 molecularly unique mutants occurring within a short region of the promoter, and these are invaluable in studying gene regulation (16–19). Missense mutations within exons give valuable information as to the importance of particular amino acid residues. The present list contains 360 different amino acid substitutions (probably detrimental). One hundred and three residues of factor IX show two or more amino acid substitutions. These include the active site serine (amino acid 365), and the proposed active site aspartate (amino acid 269), but none is yet known for the proposed active site histidine at amino acid 221. Mutations at nine of the 12 carboxylglutamyl residues have now been detected, confirming their critical role for the function of factor IX. Amino acid substitutions have also been found for every one of the 22 cysteines of circulating factor IX and therefore

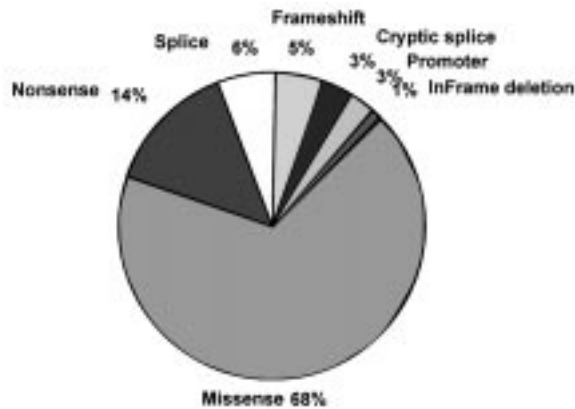


Figure 3. Pie chart showing different classes of mutation recorded in the database.

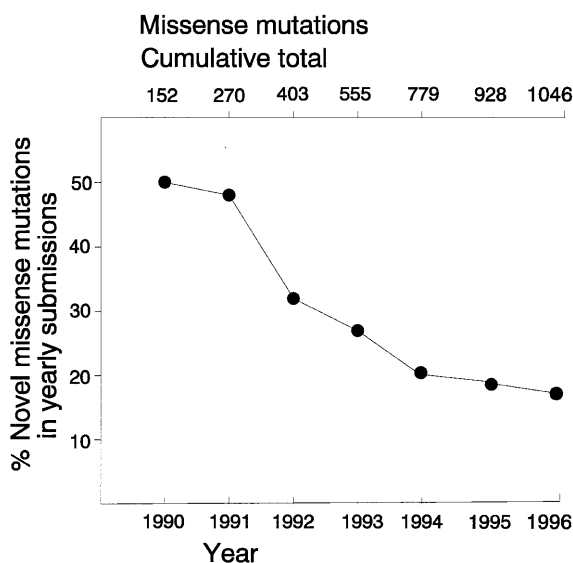


Figure 4. Graph of percentage of novel missense mutations reported in successive years.

mutations have been found that compromise each one of the disulphide bridges of the mature protein, thus confirming the importance of such structures. This year 118 missense mutations were added, of which 27 are new. The percentage of new missense mutations in successive yearly contributions is illustrated in Figure 4.

This year's database is now available on the World Wide Web (<http://www.umds.ac.uk/molgen/haemBdatabase>), but is too large to publish as a printed table. Copies on disk (in both text and database format) can be obtained from the nearest country co-ordinator if internet access is not possible. The data are arranged in 10 fields. The 'Patient' field gives the code for the patient provided by the contributor: this is usually unique except when the contributors have not given separate codes to patients with the same mutation (e.g. Vancouver, Fr) or a code for their patient ('unnamed'). The PIN field distinguishes all patient entries. Thus a patient with a double mutation has two database entries but with the same name and PIN. Occasionally, PINs are removed (and not replaced) when two or more patients previously thought to be unrelated and uniquely reported are subsequently

shown to be related or reported more than once. Factor IX coagulation (FIX:C) and antigen assays are given (as iu/dl or %) in 'Clotting' and 'Antigen' fields. The nucleotide numbering is as in ref. (4), and the amino acid numbering as in ref. (20). The base change is given in the 'mutation' field: a minus or plus sign signifies deletion and addition, respectively. The amino acid change is given where relevant or '-X' to signify an amino acid deleted. A yes/no 'CpG' field replaces previous years' footnote 4 indicating transitions at CpG sites. The comments field contains further assorted information such as Frameshift, Double (meaning double mutant, cross referenced), N (normal variant), Gla (γ -carboxy glutamic acid), Bm (prolonged bovine prothrombin time) (21), Inhibitors (declared presence of), de novo [mutation originating in mother, maternal grandmother (MGM) or maternal grandfather (MGF)]. The 'reference' field indicates contributors quoted in detail in a separate file.

The database was compiled by the central co-ordinators (Giannelli and Green) from separate lists updating the previous year's list prepared by co-ordinators for the different countries. This year these were as follows: Giannelli and Green representing the UK, Sweden and Iceland; Sommer representing USA; Poon representing Canada; Ludwig and Schwaab representing Germany; Reitsma representing the Netherlands; Goossens representing France; Yoshioka representing Japan; Figueiredo representing South America and Brownlee, the rest of the world. New data or notification of errors or omissions should be sent to the individual country co-ordinators.

REFERENCES

- Roberts H.R. (1993) *Thrombosis Haemostasis* **70**, 1-9.
- Green P.M., Naylor J. and Giannelli F. (1995) *Adv. Genet.* **32**, 99-139.
- Giannelli F., Green P.M., High K.A. *et al* (1991) *Nucleic Acids Res.* **19** (Suppl.): 2193-2219.
- Yoshitake S., Schach B.G., Foster D.C., Davie E.W. and Kurachi K. (1985) *Biochemistry* **24**, 3736-3750.
- Handford P.A., Mayhew M., Baron M., Winship P.R., Campbell I.D. and Brownlee G.G. (1991) *Nature* **351**, 164-167.
- Pang C.P., Crossley M., Kent G. and Brownlee G.G. (1990) *Nucleic Acids Res.* **18**, 6731-6732.
- Hase S., Kawabata S., Nishimura H., Takeya H., Sueyoshi T., Miyata T., Iwanaga S., Takao T., Shimonishi Y. and Ikenaka T. (1988) *J. Biochem.* **104**, 867-868.
- Nishimura H., Takao T., Hase S., Shimonishi Y. and Iwanaga S. (1992) *J. Biol. Chem.* **267**, 17520-17525.
- Agarwala K.L., Kawabata S., Takao T., Murata H., Shimonishi Y., Nishimura H. and Iwanaga S. (1994) *Biochemistry* **33**, 5167-5171.
- Brandstetter H., Bauer M., Huber R., Lollar P. and Bode W. (1995) *Proc. Natl Acad. Sci. USA* **92**, 9796-9800.
- Giannelli F., Green P.M., Sommer S.S. *et al.* (1996) *Nucleic Acids Res.* **24**, 103-118.
- Thompson A.R. (1990) *Prog. Hemost. Thromb.* **10**, 175-214.
- Ketterling R.P., Vielhaber E. and Sommer S.S. (1994) *Am. J. Hum. Genet.* **54**, 831-835.
- Saad S., Rowley G., Tagliavacca L., Green P.M. and Giannelli F. (1994) *Thrombosis and Haemostasis* **71**, 563-570.
- Bottema C.D.K., Koeberl D.D., Ketterling R.P., Bowie E.J.W., Taylor S.A.M., Lillicrap D., Shapiro A., Gilchrist G. and Sommer S.S. (1990) *Br. J. Haematol.* **75**, 212-216.
- Crossley M., Ludwig M., Stowell K.M., De Vos P., Olek K. and Brownlee G.G. (1992) *Science* **257**, 277-279.
- Picketts D.J., Lillicrap D.P. and Mueller C.R. (1993) *Nature Genet.* **3**, 175-179.
- Reijnen M.J., Peerlinck K., Maasdam D., Bertina R.M. and Reitsma P.H. (1993) *Blood* **82**, 151-158.
- Naka H. and Brownlee G.G. (1996) *Br. J. Haematol.* **92**, 231-240.
- Anson D.S., Choo K.H., Rees D.J.G. *et al.* (1984) *EMBO J.* **3**, 1054-1064.
- Hougie C. and Twomey J.J. (1967) *Lancet* **i**, 699.