Received: 19 August 2022 | Revised: 24 January 2023 | Accepted: 6 February 2023

https://doi.org/10.1016/j.jtha.2023.02.005

ORIGINAL ARTICLE

An updated interactive database for 1692 genetic variants in coagulation factor IX provides detailed insights into hemophilia B

Ziqian Xu | Holly J. Spencer | Victoria A. Harris | Stephen J. Perkins

Research Department of Structural and Molecular Biology, University College London, London, UK

Correspondence
Stephen J. Perkins, Department of Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, UK.
Email: s.perkins@ucl.ac.uk

Funding information
Lister Institute for Preventive Medicine

Abstract

Background: Genetic variants in coagulation factor IX (FIX) are associated with hemophilia B, a rare bleeding disease. *F9* variants are widespread across the gene and were summarized in our FIX variant database introduced in 2013.

Objectives: We aimed to rationalize the molecular basis for 598 new *F9* variants and 1645 new clinical cases, totaling 1692 *F9* variants and 5358 related patient cases.

Methods: New *F9* variants were identified from publications and online resources, and compiled into a MySQL database for comparison with the human FIXa protein structure.

Results: The new total of 1692 *F9* variants correspond to 406 (88%) of the 461 FIX residues and now include 70 additional residues. They comprise 945 unique point variants, 281 deletions, 352 polymorphisms, 63 insertions, and 51 others. Most FIX variants were point variants, although their proportion (56%) has reduced compared to 2013 (73%); at the same time, the proportion of polymorphisms has increased from 5% to 21%. The 764 unique mild severity variants in the mature protein with known phenotypes include 74 (9.7%) quantitative type I variants and 116 (15.2%) predominantly qualitative type II variants. The remaining 574 variants types are unspecified. Inhibitors are associated with 152 hemophilia B cases out of 5358 patients (2.8%), an increase of 93 from the previous database.

Conclusion: The even distribution of the *F9* variants revealed few mutational hotspots, and most variants were associated with small perturbations in the FIX protein structure. The updated database will assist clinicians and researchers in assessing treatments for patients with hemophilia B.

KEYWORDS
coagulation factors, genetic variants, hemostasis, inherited coagulation disorders, protein structure/folding

Manuscript handled by: Riitta Lassila

Final decision: Riitta Lassila, 6 February 2023

1 | INTRODUCTION

Hemophilia B, originally known as Christmas disease, is an X-linked recessive blood clotting disorder that affects 1 in 25,000 to 30,000 male births worldwide. It is associated with deficiencies in coagulation factor IX (FIX) as a result of genetic variants in the 33.5 kb *F9* gene located at the distal end of the long arm of the X chromosome (Xq27.1) (Figure 1) [1]. FIX, is a vitamin K–dependent single-chain zymogen that is activated to FIXa. The cleavage of FIX to FIXa is achieved either by factor VIIa via the tissue factor pathway or by factor XIa via the contact pathway, both of which require the presence of Ca^{2+}. FIXa consists of 1 light chain with a γ-carboxyglutamic acid-rich (Gla) domain and 2 epidermal growth factor-like domains (EGF-1 and EGF-2) and 1 heavy chain with the serine protease (SP) domain with the catalytic triad His267-Asp315-Ser411 (Figure 1). The 2 chains are linked by the Cys178-Cys335 disulphide bridge. The Gla domain contains 11 γ-carboxyglutamate residues that are post-translationally modified from glutamic acid to enhance the affinity of FIXa for Ca^{2+}, and enables host-membrane interactions, while the EGF-1 and SP domains contain 2 high Ca^{2+} binding sites that are essential for FIXa function.

Genetic variants in hemophilia B are associated with a wide array of molecular defects in the *F9* gene, such as missense, nonsense, and splice site variants, or more complex deletions and insertions that may result in a reading frame shift. This differs from other coagulation factor deficiencies such as hemophilia A, where intrachromosomal recombination is associated with 45% of the severe patient cases [2]. Clinically, hemophilia B is monitored by low FIX plasma procoagulant levels (FIX:C), which are designated as mild, moderate, or severe, and abnormal levels are termed a type II qualitative disorder. The quantity of FIX protein circulating in plasma (FIX:Ag) is measured by enzyme-linked immunoassay and is often observed at low levels in patients with hemophilia B, being termed a type I quantitative disorder. The determination of those *F9* variants is important for assessing the risk of inhibitor development, as well as in clinical management of pregnancy and the early neonatal period [3]. Furthermore, therapeutic approaches for specific FIX variants are related to advances in medicinal therapy such as antibiotics [4] and gene therapy [5] involving *F9* variants.

Over 1500 distinct *F9* variants have currently been identified, and simple listings are no longer a useful tool. To address this issue, starting in 2005, we introduced powerful interactive web databases at University College London for the coagulation factors VIII, IX, X, and XI that enable rapid and comprehensive phenotype and genotype analyses of these proteins [6–10]. That for FIX and hemophilia B in 2013 enabled web searches based on inputted parameters, and this has received over 42,000 visits to date [6]. Other websites for *F9* variants exist in general mutation databases such as UniProt (http://www.uniprot.org) and the Human Gene Mutation Database (www.hgmd.cf.ac.uk). More specific variant lists exist such as those of the Hemophilia B Mutation Project [11] and the EAHAD *F9* database, which is an older outdated copy of our 2013 database [6]. These other resources offer limited interactive features. Here, we aimed to significantly increase the number of *F9* variants from 1113 previously to 1692 using a wider range of sources and upgrade our database features. We describe the distribution of *F9* variants, provide molecular insights into the occurrence of hemophilia B, and identify prospects for genetic counseling and future treatment of hemophilia B.

Essentials

- We update our widely used interactive factor IX (FIX) web database first released in 2013.
- We report 1692 *F9* variants from 5358 patient cases covering 88% of the 461 FIX residues.
- Most missense variants cause small structural perturbations in the FIXa protein structure.
- The database provides an easy-to-use resource for clinicians working with hemophilia B.

2 | METHODS

2.1 | Database design and data collection

The updated interactive FIX database is available at https://www.factorix.org/ and its architecture was described previously [6]. The database format followed the layout of our previous similar databases (https://www.factorix.org/, https://www.factorx-db.org/, and https://www.factorxi.org) [7–10]. Copyright is retained by S.J.P. and University College London, and database copying is not allowed without explicit permission. Here, data for each variant were first accumulated from relevant articles and sources for an Excel spreadsheet for import into an open-source MySQL relational database. The FIX MySQL database was established using open-source phpMyAdmin software (https://www.phpmyadmin.net/). The front end of the database is maintained on a University College London server and utilizes CSS, HTML, PHP, and JavaScript programming for database access via https://www.factorix.org/. For personal or private research only, a FIX variant listing and their associated fields can be downloaded from the "Variants" menu on the website (Supplementary Material).

The original flat listing of variants from 1997 contained 973 unique FIX variants that were incorporated, with permission, to give 1113 unique variants from 3721 individual cases in the 2013 FIX database at https://www.factorix.org/ [6,12]. Our 2021 update provided 1329 unique variants from 4815 individual patients [13], and the present 2022 study provided 1692 unique variants from 5358 individual patients with FIX deficiency and FIX Leiden (cut-off date: July 2022). The database integrity was enhanced using multiple searches. Initially, variants were compiled from online literature searches of peer-reviewed articles, predominantly PubMed (https://pubmed.ncbi.nlm.nih.gov/), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), and dbSNP (http://www.ncbi.nlm.nih.gov/SNP) [14]. Next, abstracts from international hematology and hemostasis meetings and

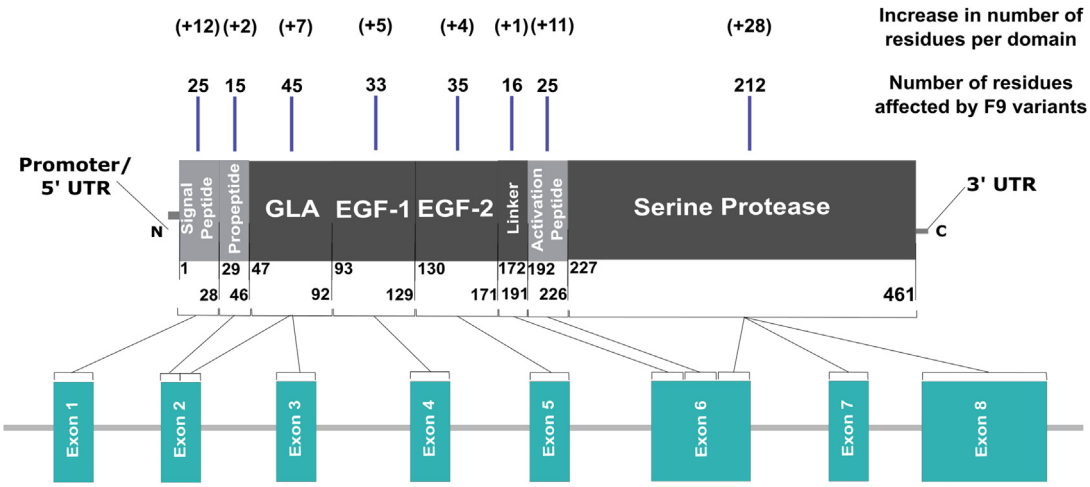

FIGURE 1 Domain cartoon of the 461 amino acids in FIX. The *F9* gene comprises 7 introns and 8 exons that are transcribed into an mRNA of 2802 nucleotides. The transcript consists of a 28-bp signal prepeptide and an 18-bp propeptide, followed by the 415-residue mature FIX protein. During activation to FIXa, the Arg191-Ala192 peptide bond is cleaved to give a FIX intermediate with two chains (residues 47-191 and 192-461), and then, the Arg226-Val227 peptide bond is cleaved to release an activation peptide (residues 192-226) and FIXa. Above the domains, the totals of the 406 FIX residues affected by variants in each domain are shown. The increases in the number of residues per domain since 2013 are shown above in parentheses. Starting from the N-terminus, the signal peptide is followed by the propeptide, the GLA domain, and two epidermal growth factor domains EGF-1 and EGF-2. A linker joins EGF-2 to the activation peptide. The serine protease domain is at the C-terminus. The domains are not drawn to scale, and the 4 folded domains in FIXa are colored in dark grey. The residue numbering corresponds to the HGVS format (starting with +1 at the beginning of the signal peptide). N and C represent the N- (amino terminal end) and C- (carboxyl terminal end) of FIX, respectively. Below the domain cartoon, the exon(s) of the *F9* gene structure are shown. FIX, factor IX.

genetic databases such as the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/ac/index.php) were searched using keywords such as "hemophilia B," "Factor IX," "F9," "mutation," or "variant." Finally, to enhance the diversity of the database, variants were incorporated for non-English non-PubMed sources, such as those in Chinese, Japanese, Spanish, or Russian. The sources were listed in the "Reference" tab of the FIX database.

2.2 | Presentation of clinical data

For each variant, clinical data were included on FIX procoagulant activity (FIX:C) and FIX antigen level (FIX:Ag). Information on hemophilia B Leiden and FIX inhibitor formation was also recorded. For the classification of severity, we clarified with the authors whether any record did not follow the conventional boundaries of FIX:C for determining the hemophilia B severity. In cases where confusion remained, the clinical severity was presented in its original form with the discrepancy stated in the database's comments section. For the reason of privacy and to prevent data misuse, patient IDs were not shown online, although such information may have been reported in the literature.

To minimize the impact of errors in the FIX database, which occur even from peer-reviewed articles, all data were curated prior to their inclusion. Any corrected entries were noted with their original format in the comments section, and logged in order to substantiate its credibility. Where possible, the published reference was linked to the PubMed database (https://pubmed.ncbi.nlm.nih.gov/). For data from unpublished sources, such as local statistics from a disease center, it is common for those data to be published later. In such scenarios, only the published source was used to prevent any data duplication.

2.3 | Sequence numbering

As described previously [6], the database was compiled using the widely accepted sequence nomenclature provided by the Human Genome Variation Society (HGVS) (http://varnomen.hgvs.org/) [15,16]. All variants were represented by systematic names. The HGVS format avoids potential data corruption and allows data merger from different sources. Because the HGVS format is universally recognized, our data can be interpreted worldwide. In older publications, when reported, the legacy numbering for amino acids was converted to the HGVS format by adding 46 to this (+1 in legacy numbering corresponded to +47 in HGVS numbering). To help the user, both the HGVS and legacy numbering were reported for each variant in the database. The first HGVS residue is the first methionine (ATG) of the signal peptide. cDNA changes were indicated by "c" and followed by the nucleotide number, and change (eg, c.38G>A), whereas amino acid changes were denoted by "p" (eg, p.Ile7Phe). The chymotrypsin numbering system for the SP domain is often used because the numbering is equivalent across all SPs, but was not used here because the N-terminal Gla and EGF regions were not represented in this numbering.

2.4 | Protein structural analyses

The database permitted 3 types of searches: (i) causative genetic variants of hemophilia B in the *F9* gene; (ii) their related patient

entries, which include any subfeatures and clinical results; and (iii) experimental protein structures of FIXa in the Protein Data Bank. The human FIXa structure (UNIPROT accession number: P00740) was unchanged from 2013, in which 3 partial structures for human FIXa were merged with a crystal structure for full-length porcine FIXa [6]. The uncleaved full FIX structure was taken from AlphaFold (https://alphafold.ebi.ac.uk/entry/P00740). AlphaFold is an artificial intelligence system based on neural networks that predicts protein structures based on their amino acid sequence [17]. The database also included a pipeline that predicts structure alterations based on the existing FIXa models.

The protein backbone secondary structure of each residue in the FIXa structure was identified using the Definition of Secondary Structure of Proteins (DSSP) tool (https://www3.cmbi.umcn.nl/xssp/) [18,19]. Residues were individually assigned to be one of either H (α-helix), B (β-bridge), E (extended β-strand), G (3_{10} helix), I (π-helix), T (hydrogen-bonded turn), S (bend), or C (undefined coil region). DSSP also gave the exposed surface area of each residue in the FIXa structure, which were converted into percent accessibility by dividing this by the theoretical solvent accessible surface area of the amino acid in question [18–20]. The results were simplified by assigning percentage accessibilities of 0% to 9% as 0, 10% to 19% as 1, 20% to 29% as 2, and so on. Residues with accessibilities of 0 or 1 were considered to be solvent buried, and accessibilities of 2 to 9 were classified as solvent exposed.

The disruptive effect of a given missense variant on the FIXa protein structure was assessed by 4 independent substitution analyses. Bar charts below were used in order to place each individual prediction in the context of all the variants. They were Polymorphism Phenotyping v2 (PolyPhen2) [21–23], Sorting Intolerant From Tolerant (SIFT) [24], Protein Variation Effect Analyzer (PROVEAN) [25,26], and the Grantham matrix from 1974 [27–30]. Both PolyPhen2 and SIFT gave scores ranging from 0 to 1, where PolyPhen2 scores close to 1 indicated the more damaging variants and SIFT scores close to 0 indicated the more damaging variants. Unlike SIFT and PolyPhen2, PROVEAN generated scores that are compared with a threshold; scores above −2.5 were neutral and those below −2.5 were deleterious. The Grantham analysis differed in that this was not sequence-specific but based on the substituted amino acids. Grantham scores ranged from 0 to 215, with larger scores indicating more damaging changes, and a score of 0 had no effect. While more accurate, the American College of Medical Genetics and Genomics presents a five-level classification of variants (pathogenic, likely pathogenic, variant of unknown significance, likely benign, and benign) [31], but this was not used here for reason of the lack of specific data descriptions in our sources.

3 | RESULTS

3.1 | Distribution of factor IX variants in the updated database

The current version of the FIX database (https://www.factorix.org) incorporates 598 newly reported *F9* variants in addition to the total of 1113 variants in the 2013 version of our database. The variants were presented within the myPHPadmin database based on 5358 patient records (Figure 2A). The variants were accumulated from 355 research publications and abstracts, the PubMed literature resource (https://www.ncbi.nlm.nih.gov/pmc/), and the ClinVar variants resource (https://www.ncbi.nlm.nih.gov/clinvar/). The variants were further trimmed during curation to eliminate redundant or incorrect entries, thus making an overall total of 1692 *F9* genetic variants (Figure 2B). In the database, 1645 entries were newly reported, representing database growth by 52% in *F9* variants and 44% in patient entries. Forty-eight novel experimentally determined structures of human FIXa were also incorporated into the database for reference, as well as an AlphaFold structural model for full-length uncleaved FIX that revealed a disordered structure for the activation peptide.

Users can perform basic searches based on amino acid or nucleotide number or advanced searches based on specified criteria, including domain location, reference, disease severity, and phenotype. The database has interactive functions derived from our upgraded FX website (https://www.factorx_db.org) [9–10]. This includes a site map that facilitates user navigation across the website pages. The website homepage features 2 FIXa movies that displayed its main secondary structures. A variant map of the FIX amino acid sequence is color-coded to mark the domain boundaries (Figure 1). On this, variants are shown in red and are accessible by clicking on these.

The 1692 *F9* genetic variants include 945 point variants (individual nucleotide base change), 281 deletions (loss of at least one nucleotide), 63 insertions (gain of at least one nucleotide), 23 indels (combination of insertions and deletions), 22 duplications (copying of at least one nucleotide), 6 complex (combination of insertions, deletions, and substitutions), and 352 polymorphisms (genetic variants occurring in >1% of the population). Point variants and polymorphisms are dominant in the updated database together, accounting for 77% of *F9* variants (Figure 2B), although the proportion of point variants has reduced from 73% in 2013 to 56% here. In terms of the 598 new variants, 50% are single-nucleotide polymorphisms reported from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP). Deletions comprise 102 new variants, which is comparable to the previous one. Interestingly, 22 duplications are reported, which is much increased from the previous one. Three duplications from 4 isolated patients with hemophilia replicated specific sections of the *F9* gene. Two of those 4 patients are associated with severe hemophilia B through the duplication of exon b-f and d-f, respectively [32,33].

In terms of the variant locations, 1419 are from exons, 174 are from introns, 62 are from one of the untranslated regions, and 37 are larger variants that span across exon and introns (Figure 2C). The 945 missense point variants are distributed across the *F9* gene, most being missense ones (Figure 2D). The majority of the new variants in this update occur in the exons, especially exon 8 (Figure 2E), which is in consistency with the previous analysis. Importantly, exon 8 contains part of the SP domain that contains the catalytic triad residues. The least new variants occurred in intron 5, which displayed only 16 variants overall (novel and previously reported). Out of the 1419 *F9* exon variants (Figure 2C), 54.7% of them (776) are within the SP

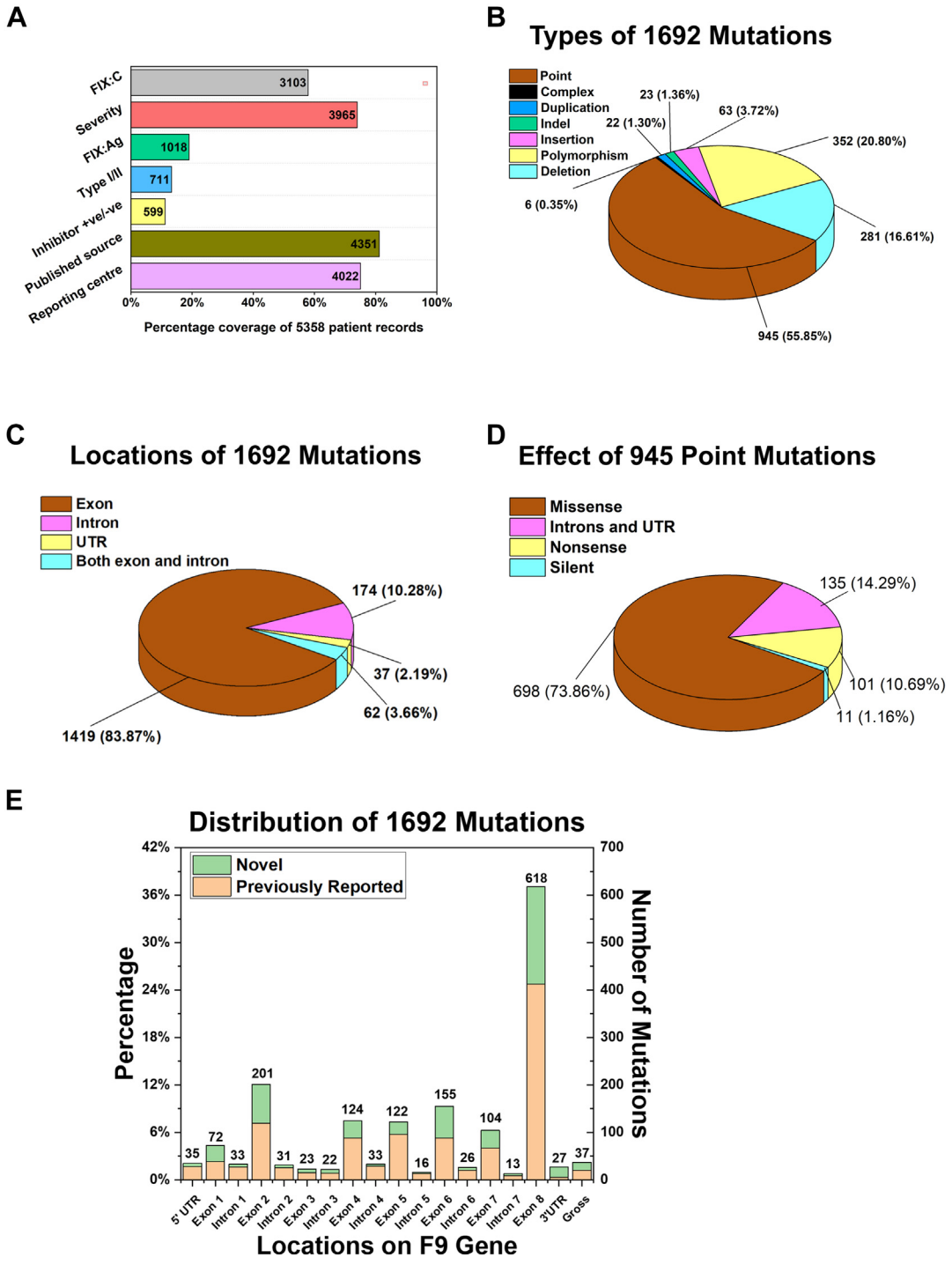

FIGURE 2 Statistical analysis of the 1692 variants from 5358 patients in the FIX database. (A) Percentage coverage of clinical data reported in entries of 5358 patients with hemophilia B. The number of each type of data presented is shown within each bar. (B) Distributions of the 1692 variants according to their occurrence as 7 types of variants. (C) The distribution on the 1692 variants between the exons, introns, and UTR regions in the *F9* gene structure. (D) The effect of the 945 point variants on the gene sequence. (E) Gene distribution of the new variants from this update (green) and the variants from 2013 (orange). The total number of variants in each gene segment is shown above each bar. FIX, factor IX; UTR, untranslated region.

domain, whereas 12.4% (176), 8.5% (121), and 8.0% (114) are within the Gla, EGF-1, and EGF-2 domains, respectively. Those figures are in proportion to the domain sizes in FIX, ie, 235 residues out of 461 (51%) for SP; 46 residues (10%) for Gla, 37 residues (8%) for EGF-1, and 42 residues (9%) for EGF-2. This shows that there are no variant hotspots in FIXa. Overall, variants are reported in 70 new FIX residues, which increases the total number of affected FIX residues to 406, accounting for 88% of the total 461 residues (Figure 3).

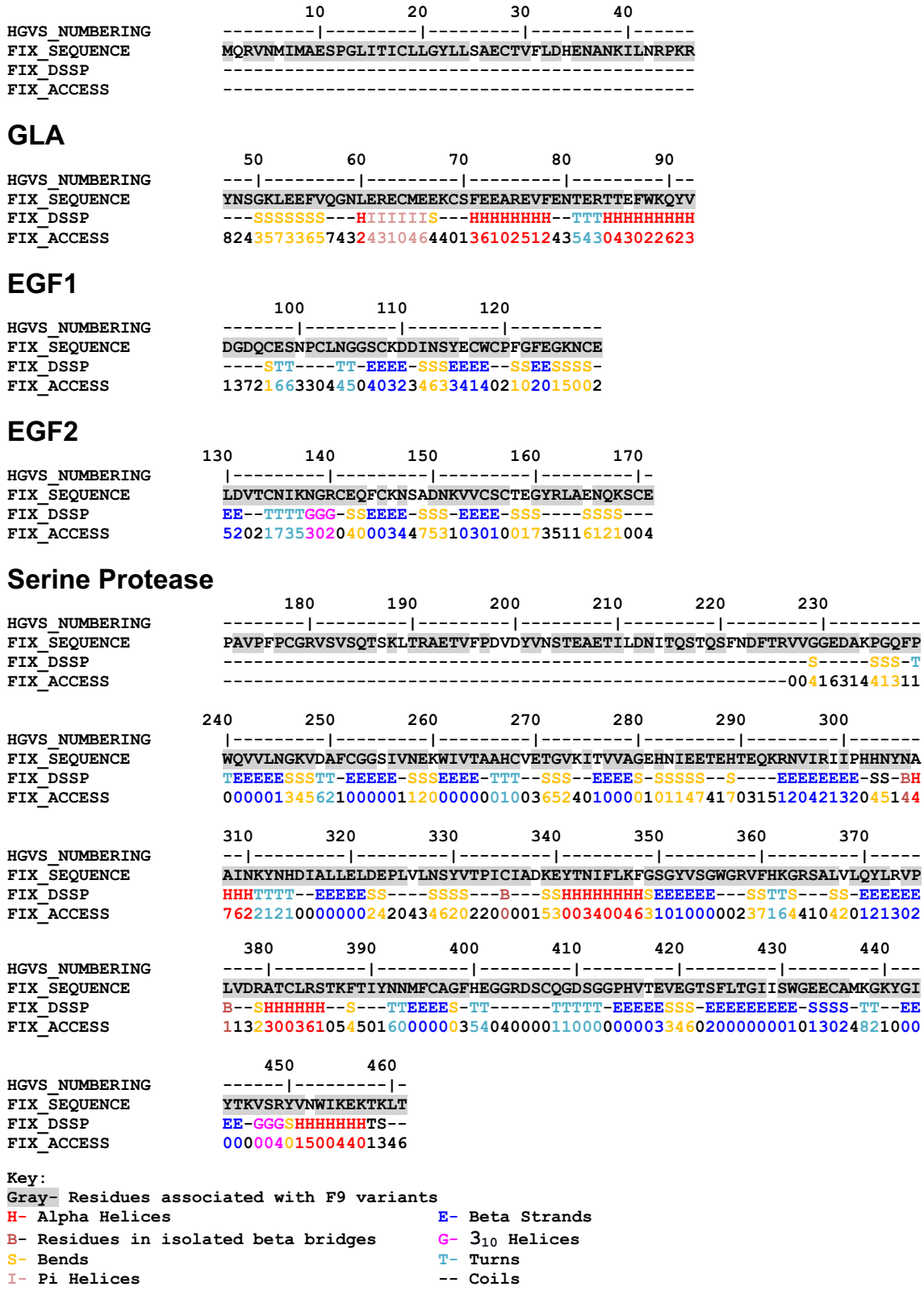

FIGURE 3 Protein sequence of human FIXa and its variants, including its secondary structures and solvent accessibilities. The 406 protein residues out of 461 in total associated with the *F9* variants are highlighted in gray. The catalytic triad is located at His267-Asp315-Ser411. The DSSP secondary structure assignments are reported as follows: H, alpha helices (red); E, beta strands (blue); B, residues in isolated beta bridges (brown); G, 3_{10} helices (magenta); I, pi helices (pink); S, bends (yellow); T, turns (cyan); and -, coils (uncolored). The relative side chain solvent accessibilities (ACCESS) are converted to values ranging from 0 (0%-9% accessibility; completely buried) to 9 (90%-100% accessibility, fully solvent exposed).

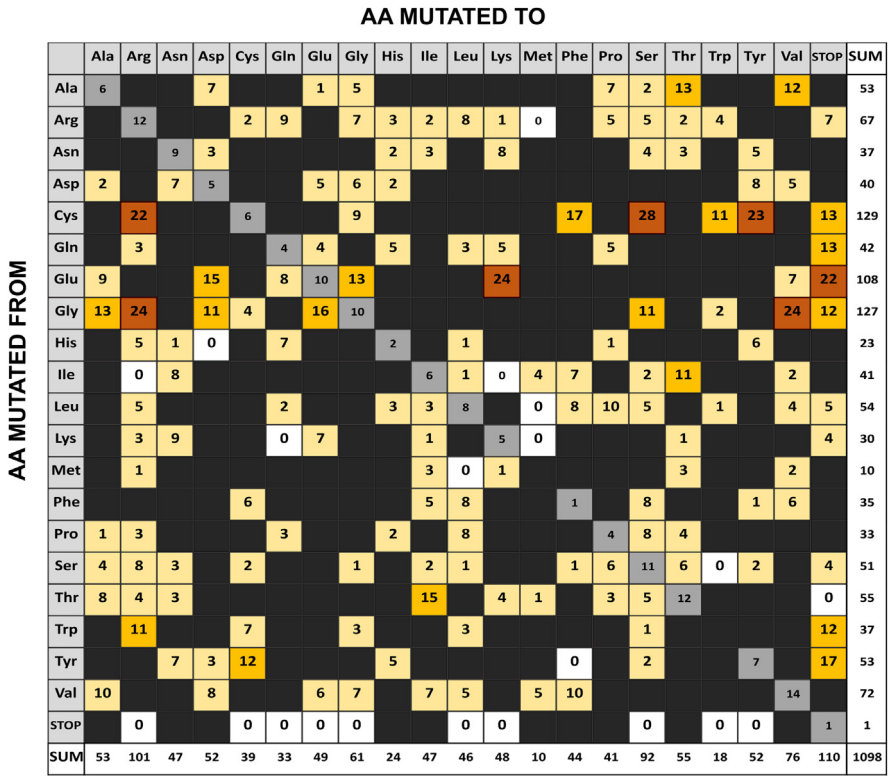

FIGURE 4 Substitution grid representing 1098 point missense variants in the *F9* gene. The grid presents the number of missense variants that occurs for each defined amino acid change. The totals are summed at the right or at the bottom. All the substitutions are the result of a single-nucleotide change. Any grid substitutions that would require more than a single-nucleotide change are shown in black, however none were seen. White boxes represent possible substitutions that do not occur in the *F9* gene. Yellow boxes represent substitutions that occur between one and 9 times, orange represents substitutions that occur between ten and 19 times, and red represent substitutions that occur over 20 times.

A substitution grid assessed the 1098 missense and nonsense variants in FIX. There, Cys (129), Gly (127), Glu (108), Val (72), and Arg (67) were the 5 most commonly affected residues in the wildtype protein (right column, Figure 4). The prominence of Cys residues is attributed to disulphide bridge disruption in the 4 FIX domains, resulting in misfolded protein. The most frequent change in FIX was the 28 variants where wildtype Cys residues became Ser residues. The most common outcome of the variants was 110 occurrences of the stop codon that replaced Glu, Tyr and other residues, indicating that about 10% of the reported FIX variants corresponded to the truncated protein. Silent variants (gray diagonal in Figure 4) were most commonly found in Val (14) or Arg (12) residues. The black boxes of Figure 4 signify 2 nucleotide changes in a given variant, of which there were none observed.

3.2 | Hemophilia B phenotypes

The clinical parameters were the FIX procoagulant activity (FIX:C) and the FIX antigen level (FIX:Ag). The clinical severity in the patient was determined by FIX:C. The 3 phenotypes were classified conventionally as severe (<1% of normal FIX level; <0.01 IU mL^{-1}), moderate (1%-5% FIX; 0.01–0.05 IU mL^{-1}), or mild (>5% to <40% FIX; >0.05–<0.40 IU mL^{-1}). This severity information was reported for 74% of patients with hemophilia B, with 3103 cases using their FIX:C levels (Figure 2A). Thus, 2060 patients have severe hemophilia B, whereas 1141 and 764 patients have moderate and mild phenotypes, respectively. The remaining 1393 cases either did not report any phenotype or corresponded to patients with hemophilia B Leiden.

The database showed that, while *F9* variants were dispersed across the *F9* gene (Figure 2E), patients with hemophilia B can be clustered on certain variants. Of the 1692 variants, 102 showed at least 8 clinical cases, and these clusters constituted over 50% of the 5358 cases. 977 variants occurred only with single cases. The most frequent *F9* variant was c.1025C>T on exon 8 at a CpG site with 175 identified cases from 58 reports or unpublished studies. This was Thr342Met that replaced a polar residue with a hydrophobic one for which 103 showed moderate or mild phenotypes, 7 cases were severe, and 65 had no reported phenotype. Over 100 cases have been reported for 6 other variants. These comprise 1 nonsense variant (Arg75*), 4 missense variants (Arg43Gln, Gly106Ser, Arg191His, and Arg294Gln), and 1 gross deletion of the entire *F9* gene.

For the 3965 cases of hemophilia B with a defined phenotype, the phenotype distribution depended on the variant type. For the missense variants, the distribution of phenotypes was relatively balanced, where 18.8%, 25.7%, and 32.3% of the patients had mild, moderate, and severe phenotypes, respectively. The remaining 23.2% of missense variants had unreported severities. For the silent variants, 69.8% of the cases did not report severities, while the remaining 13.2%, 6.9% and 10.1% of cases showed mild, moderate, and severe phenotypes, respectively. Most of the patients with frameshift (53.6%) or nonsense (68.4%) variants have severe phenotypes. Unsurprisingly, 80 of 102 patients (78.4%) who have lost the entire *F9* gene suffer from severe phenotypes. There, 2 moderate-classified phenotypes showed low FIX:C levels of 1%–2%, one mild phenotype corresponded to a female carrier, and 19 had unreported phenotypes.

3.3 | Hemophilia B Leiden

Hemophilia B Leiden (sometimes spelled Leyden) is an atypical subtype of hemophilia B caused by variants that affect the *F9* promotor region. Patients show severe or sometimes moderate/mild phenotypes from birth, but their FIX:C levels gradually increase with age after puberty. The amelioration was initially associated with androgen response element binding sites located between nucleotides −65 and −51 in the promotor [34]. These binding sites were thought to be responsive to androgens such as testosterone, which would recompense the nonfunctioning transcription factor binding sites as hormone levels increase with age. More recently, because many FIX variants that do not have these androgen sites occur in the protein, growth hormone may be directly responsible for puberty-onset recovery of FIX production [35].

The database showed that 38 point variants of 94 downstream of the androgen response element in the promotor have been associated with hemophilia B Leiden. The most frequent is c.-35G>A, which affects 50% of the reported individuals. The Leiden cases account for 40.4% of the patients with variants between nucleotides −50 and 1, and 0.71% of the total hemophilia B patients in the database. Interestingly, in Argentina, 1.9% of patients with hemophilia B belong to the hemophilia B Leiden subtype [36]. It should be noted that the recording of FIX:C levels over time is not always performed, making any hemophilia B Leiden cases difficult to identify.

3.4 | Anti–factor IX inhibitors

Anti-FIX inhibitors are typically IgG4 or IgG1 antibodies in patients with hemophilia B as a result of environmental and genetic risk factors. These inhibitors make FIX replacement therapy ineffective and can lead to allergic/anaphylactic reaction or nephrotic syndrome that increase the morbidity of patients with hemophilia B [37]. Inhibitor titers in assays are measured in Nijmegen-Bethesda units (NBU) or Bethesda units (BU). Patients with high titer inhibitors with >5.0 NBU/BU are more challenging to treat compared with patients with low titer inhibitors (≤5.0 NBU/BU).

In the database, 152 of the 5358 hemophilia B cases were inhibitor–positive, which was almost doubled compared to 59 reports out of 3721 in 2013. This incidence remained within the 1.5%-3% range of hemophilia B inhibitors in populations worldwide [38]. The 40 inhibitor–positive variants in 152 patients corresponded to 10 deletions, 3 indels, 10 nonsense variants, 12 missense variants, 2 silent variants, and 3 other point variants at splice sites or untranslated regions. Only 15 of those 40 variants were new to the database. Although most variants lead to a null allele, 29 inhibitor–positive cases were caused by missense variants, such as the 8 cases out of 11 for Glu66Asp. For these 152 patients, 107 have a variant which leads to a null allele (deletion, indel or nonsense). Looking at all the database clinical records, gross variants remain the riskiest type associated with inhibitor formation, with 51 of 191 gross variant cases (27%) being associated with positive inhibitors. In comparison, 7% of nonsense variants (46 of 639 cases) are inhibitor-positive. On the other hand, 447 database entries were inhibitor-negative. The remaining 4759 hemophilia B cases did not mention inhibitors, and it was not known whether inhibitors were screened for. Because the database includes a bias toward severe phenotypes, and because inhibitors are rare in nonsevere hemophilia B, it was not feasible to give a representative estimate of the risk for inhibitors. A more accurate risk of inhibitor development is given by the frequency of inhibitors in severe hemophilia B patients, where 99 severe cases were found with inhibitors out of 2060 cases (4.8%).

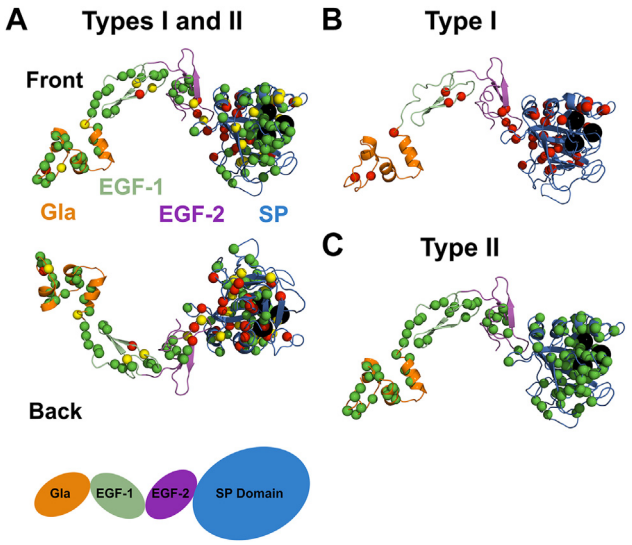

FIGURE 5 Distribution of the type I and type II variants in the FIXa structure. (A) Residues with type I variants are represented by red spheres and those with type II variants are represented by green spheres. Residues associated with both types I and II are represented by yellow spheres. The catalytic triad His267-Asp315-Ser411 at the right is shown as 3 black spheres. The Gla, EGF-1, EGF-2, and SP domains are colored in orange, light green, violet, and blue, respectively, to correspond to the colors used in the website. (B, C) For clarity, the individual type I and type II residues are also shown separately in the FIXa structure. Of the 181 type I variants, totals of 7, 13, 12, and 110 are localized to the Gla, EGF-1, EGF-2, and SP domains in that order. Of the 530 type I variants, totals of 43, 59, 21, and 262 are localized to the Gla, EGF-1, EGF-2, and SP domains in that order. Several variants in the same domain can occur on the same residue, meaning that only one sphere is shown at that location.

3.5 | Single-nucleotide polymorphisms

Single-nucleotide polymorphisms or SNPs are defined as non–disease-associated substitutions with frequencies higher than 1% in populations. Similar to point variants, SNPs can be synonymous or non-synonymous, the latter resulting in an amino acid change. The updated database contained 352 SNPs, compared to 54 reported previously. Of these, 122 were silent and 158 resulted in missense variants.

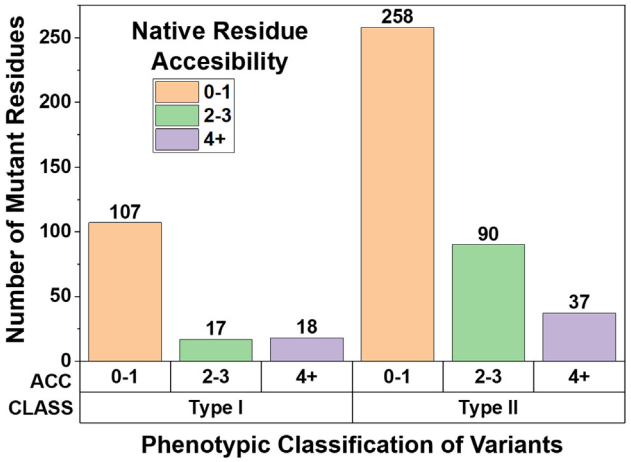

FIGURE 6 Solvent exposure of the type I and type II residues in the FIXa protein structure. The residue accessibilities for 142 type I and 385 type II variants (CLASS) are grouped according to their relative solvent accessibilities (ACC), where the values 0 to 1 correspond to buried sidechains, values 2 to 3 indicate slightly exposed sidechains, and values of ≥4 indicate fully exposed sidechains. The accessibilities are taken from Figure 3.

3.6 | Protein structural analysis of type I and type II variants

F9 variants are classified as type I (quantitative) or type II (qualitative) depending on the phenotypic FIX:C/FIX:Ag ratio. A type I defect corresponds to a ratio above 0.7, and type II if this ratio is below 0.7 [39]. In the database, only 711 cases have been assigned in this way (Figure 2A). Of these, 181 cases were type I, whereas 530 cases were type II. Twenty-two *F9* variants were associated with both type I and type II. Both phenotypes were distributed throughout the 4 FIX domains, indicating that the whole FIXa structure is affected by variant-mediated damage. This distribution differs from other coagulation factor deficiencies, where type I is predominant in factor XI, or type I/II were more evenly distributed in factor VIII [7,8,10]. The greater abundance of type II variants is seen in the FIXa structure (Figure 5B, C). Almost one fifth of the type II variants (99/530) were related to Arg residues in the linker or propeptide regions, which are not part of functional FIXa. It is likely that the FIX-FIXa cleavage was disrupted in those mutants, resulting in little FIXa and a type II phenotype.

To clarify the causative effects of the types I and II variants, the accessible surface area of each residue was computed from the FIXa crystal structure model using DSSP and normalized (Methods) [18,20,40]. The resulting relative solvent accessibilities (Figure 3) showed that variants at buried locations in FIXa dominated the types I and II phenotypes (Figure 6). The involvement of 75% type I and 67% type II variants at buried positions linked perturbations of the FIXa structure with their disease-associated variants. Five missense variants were reported over 10 times: Ala279Thr (types I/II 14/4), Arg294Gln (types I/II 18/3), Gly106Ser (types I/II 10/23), Thr342Met (types I/II 16/27), and Ile443Thr (type II; 27 times). All those residues except Arg294 are completely buried (Figure 3), being in or proximal to α-helices or β-strands in the core secondary structure. These observations were thus reproducible and show that amino acid packing changes within FIXa result in either misfolding or functional damage.

The identification of a missense FIX variant cannot be assumed to cause hemophilia B. The database enables the 856 missense variants to be assessed using 4 independent prediction tools (Methods). Overall, the PolyPhen-2 analysis showed that 598 (70%) gave scores between 0.9 to 1.0 and were assigned to be damaging, but that 142 (17%) were predicted to be benign (Figure 7A). The SIFT analysis predicted that 683 (80%) of the missense variants were damaging (Figure 7B). Therefore, the majority of FIXa variants are predicted to be linked with hemophilia B. The PROVEAN and Grantham analyses were less clear with a wider distribution of scores, showing that 636 (75%) of the PROVEAN scores and 844 (99%) of the Grantham scores were deleterious. Regardless, both the PROVEAN and Grantham analyses also demonstrated that many FIXa variants were damaging (Figure 7C, D). The FIXa database gives all prediction scores for the 856 missense variants in order to enable an informed decision about the significance of a score for a variant.

4 | DISCUSSION

The upgraded FIX database with 1692 variants has significantly improved its quality with a wider data set compared to our original study with the database of 1113 variants in 2013 [6]. The website presented genetic and structural information on the FIX as 2 distinct but parallel themes, similar to our earlier databases on factors V, VIII, X, and XI [7–10,41]. This is illustrated using genetic and structural outputs for the Arg294Gln variant, for which 113 patient records exist (Supplementary Figure S1). On the left, insight into the conservation of Arg294 from the AA alignments tab shows that Phe301 aligned with 10 other mammalian species to show that this residue is almost fully conserved and therefore important for FIX function. On the right, the structural analysis shows that Arg294 is exposed at the surface of the SP domain in FIXa. Arg is the most basic of the 20 common amino acids and is significant in stabilizing ionic interactions at protein surfaces or those buried within protein structures. Although no gross changes occurred in the distribution of the F9 variants, new findings include the large growth of polymorphisms. At least one variant has now been identified for 406 (88%) of the 461 FIX residues in the sequence, thus providing a significantly broader coverage of FIX. The number of null variants has grown compared to point variants, further complexing the diversity of variants. The interactive interface of the database makes it easily accessible for clinicians and researchers. This allows the database to provide further clinical insights into the occurrence of hemophilia B.

The present FIX database upgrade was focused on the extraction of *F9* variants from the PubMed, ClinVar, and dbSNP resources; meeting abstracts; and non-English sources (Methods). These sources resulted in an additional 579 unique variants here to total 1692. Other *F9* databases have been presented. The CHBMP compilation of 1133 *F9* mutations up to 2014 (https://www.cdc.gov/ncbddd/

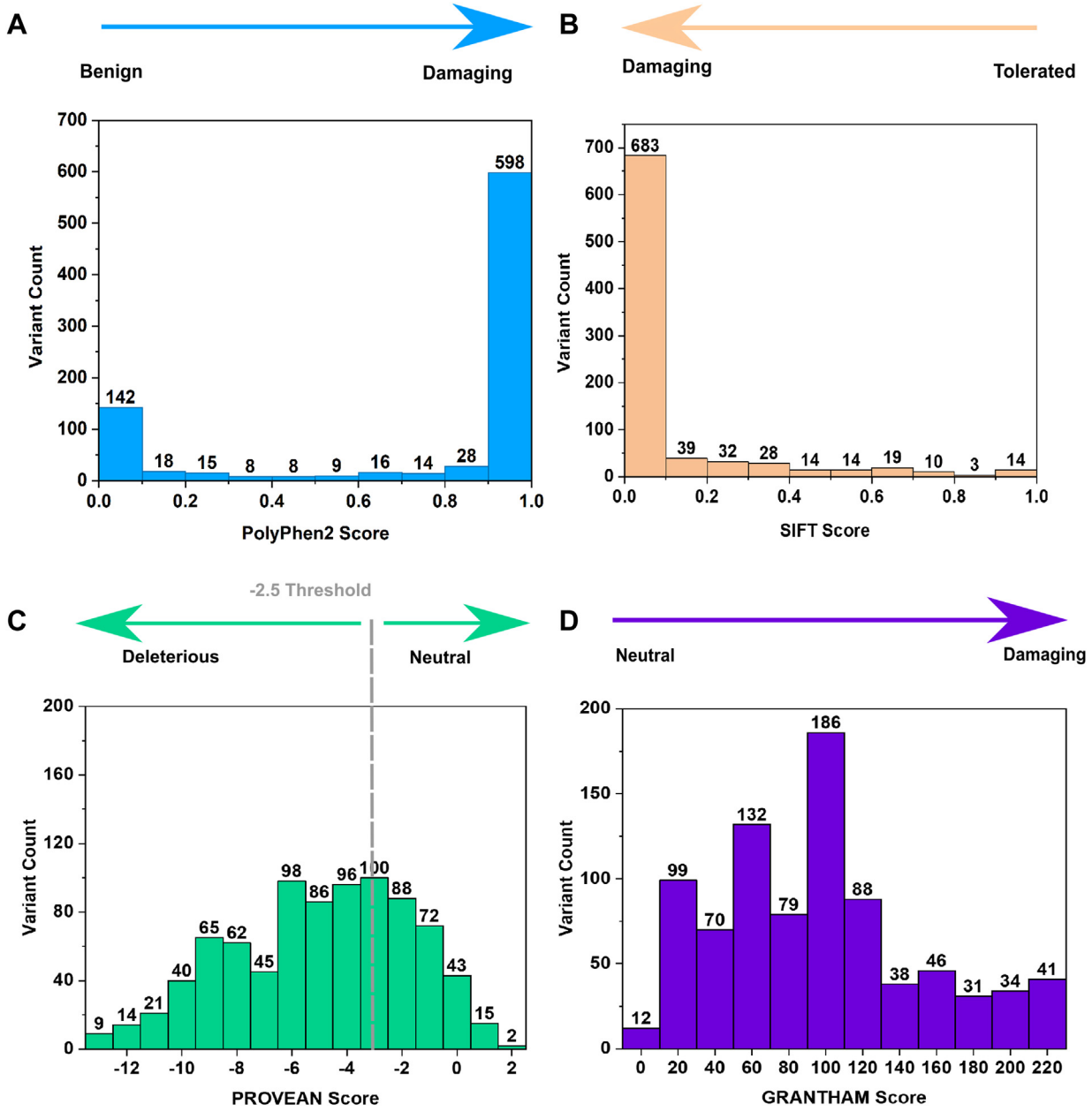

FIGURE 7 Substitution analysis of 856 missense variants in the F9 gene. The 4 substitution analyses predict the damaging effects of substitution variants on the protein structures. The 856 variants have been scored using 4 different prediction analyses, namely, (A) PolyPhen2, (B) SIFT, (C) PROVEAN, and (D) Grantham (Methods). In the PROVEAN analysis, the threshold was set as −2.5, below which the variant was considered to be damaging, and 5 variants were not included for reason of the nonavailability of the PROVEAN web site. Five PROVEAN scores for the variants E63D, V92F, G230A, P239L, and C407Y were unavailable for the reason of a PROVEAN web-service closure.

hemophilia/champs.html) is a downloadable Excel file that reports phenotypes and references. However, CHBMP does not present three-dimensional structural analyses of the FIX protein. A study of 599 patients with hemophilia with *F9* genetic variation has been reported but is not online [42]. A Spanish listing from 2010 with 82 variants and their phenotypes has been presented [43]. The EAHAD *F9* mutation list (https://f9-db.eahad.org/) is an outdated copy of our database and was last updated in February 2020 to 1244 unique variants, which is an increase of 111 variants compared to our increase of 579 variants presented here, with the increase arising mostly from the 2017 listing [42]. Our updated website is enhanced (Methods) by including the source of inclusions, nomenclature, curation of the variants, avoidance of duplicates, worldwide distribution of variants, and newer FIX structures.

The utility of variant databases is affected by 3 factors. First, they become outdated, most often caused by the departure of the original database curator. The upgraded FIX database takes advantage of our multiple similar databases in a standard format for the coagulation and complement genes, facilitating easier maintenance by a senior curator with oversight [6–10]. We observed that approximately 30 novel *F9*

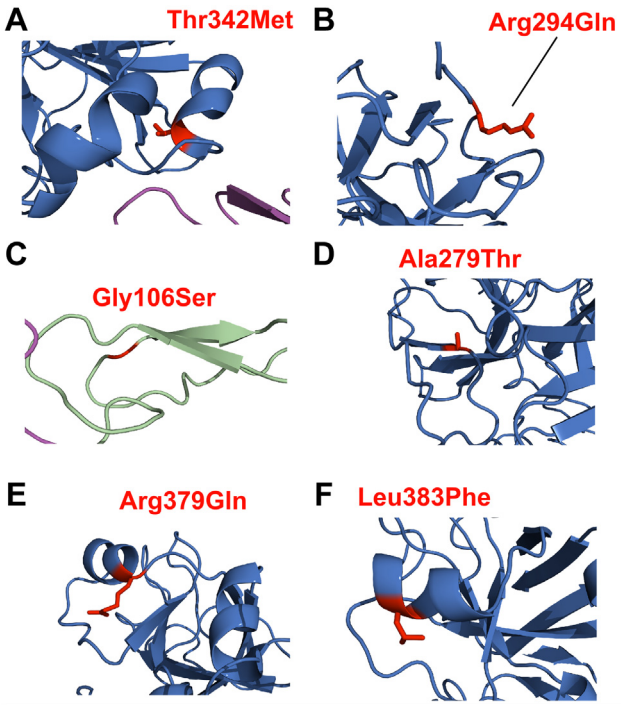

FIGURE 8 Molecular graphic representation of 6 residues within the FIX protein structure. Molecular views of the most commonly occurring FIX variants found in patients. Only those missense variants that are visible in the FIXa crystal structure are shown here. For each, the residue of interest in the native FIX protein is shown. Panels A–F highlight the substitutions Thr342Met (SP domain; 175 patients), Arg294Gln (SP domain; 114 patients), Gly106Ser (EGF-1 domain; 105 patients), Ala279Thr (SP domain; 87 patients), Arg379Gln (SP domain; 75 patients), and Leu383Phe (SP domain; 38 patients). All residue numbering is given in HGVS format.

variants and 120 individual cases were reported annually. The magnitude of the present upgrade should have captured the major features of the *F9* variants, as it is now unlikely that enough novel variants will be found to reveal changes in variant patterns. Second, while the identification of *F9* variants was prioritized, their relevant patient details are also important for understanding the prevalence and severity of the specific variant. Unfortunately, clinical centers do not always report phenotypes, weakening the aim of characterizing FIX variants. Sometimes, it was also unclear whether the patients were isolated sporadic cases or from the same family, which is important to appreciate variant prevalence and severity [44]. Third, the locations of the clinical data are important. Most database cases are from developed countries, which are more complete in their details, whereas data from some developing countries, such as China, India, or Brazil, may be limited by economic factors [45]. In the USA, only ~20% of hemophiliacs have been genotyped due to the lack of medical insurance and limited local testing availability. The prevalence of hemophilia B (per 10,000 males) varies significantly in countries with different economic backgrounds [46]. Despite several regional-based patient reports, it is unlikely that analyses will be available for every country [47–49]. To assist with this issue, the upgraded FIX database has the facility to sort reports by country and display these as a map, and may be used as an alternative resource when clinical summaries are not available for a specific region.

Inhibitor developments in hemophilia B are rare but associated with serious outcomes. The low occurrence of hemophilia B inhibitors also means that current knowledge and treatment relies on more frequent hemophilia A inhibitors, which differ in disease outcome and pathology [50]. In particular, the characterization of patients with severe hemophilia B in the database associated with inhibitor formation may allow better estimates of inhibitor prevalence and more specific anti-inhibitor therapies to be developed. Nevertheless, the majority of the literature reports does not refer to inhibitor formation, and hence, the frequency we observed is not comparable to more specific inhibitor incidence analyses [51]. Furthermore, because inhibitor development depends on multiple factors, those insights should be viewed with environmental factors such as age and immune-related genetic polymorphisms outside *F9* to assess their clinical consequences and decisions.

Our database clarifies the damaging protein structural effects of the variants from the use of accessibility calculations and substitution predictions such as Polyphen2 and SIFT. Molecular views of the most common 6 FIXa variants (Figure 8) show that the local molecular environment is most variable. These analyses benefit from well-defined protein structures, as well as additional insights from AlphaFold modeling [17]. As for FX and FXI, analyses of the type I and type II variants show their association with small but significant structural perturbations of FIXa [9,10]. Buried protein residues are frequently associated with the type I/II classifications, and this is reinforced by the SIFT and PolyPhen2 analyses. The modification of Arg residues in the disordered activation peptide also correlates with type II variants. In the absence of functional studies of recombinant mutant FIXa, these structural analyses provide a reassurance that a given variant is associated with hemophilia B.

ACKNOWLEDGMENTS

The authors were supported by grants from the Lister Institute for Preventive Medicine for this work.

AUTHOR CONTRIBUTIONS

Z.X. and H.J.S. updated the variant database and analyzed the data. H.J.S. and V.A.H. updated and maintained the website. Z.X., H.J.S., and S.J.P. wrote the paper. All authors contributed to the preparation of the manuscript and read and approved its final version.

DECLARATION OF COMPETING INTEREST

The authors state that they have no conflicts of interest.

ORCID

Ziqian Xu https://orcid.org/0000-0002-0295-3054
Holly J. Spencer https://orcid.org/0000-0003-2243-2130
Victoria A. Harris https://orcid.org/0000-0002-4240-0008
Stephen J. Perkins https://orcid.org/0000-0001-9218-9805

REFERENCES

[1] Anson DS, Choo KH, Rees DJG, Giannelli F, Gould K, Huddleston JA, Brownlee GG. The gene structure of human anti-haemophilic factor IX. *EMBO J*. 1984;3:1053–60.

[2] Lakich D, Kazazian HH, Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet*. 1993;5:236–41.

[3] Chalmers EA, Williams M, Brennand J, Liesner R, Collins P, Richards M. Guideline on the management of haemophilia in the fetus and neonate. *Br J Haematol*. 2011;154:208–15.

[4] James PD, Raut S, Rivard GE, Poon MC, Warner M, McKenna S, Leggo J, Lillicrap D. Aminoglycoside suppression of nonsense mutations in severe hemophilia. *Blood*. 2005;106:3043–8.

[5] Chowdary P, Shapiro S, Makris M, Evans G, Boyce S, Talks K, Dolan G, Reiss U, Phillips M, Riddell A, Peralta MR, Quaye M, Patch DW, Tuddenham E, Dane A, Watissée M, Long A, Nathwani A. Phase 1-2 trial of AAVS3 gene therapy in patients with hemophilia B. *N Engl J Med*. 2022;387:237–47.

[6] Rallapalli PM, Kemball-Cook G, Tuddenham EG, Gomez K, Perkins SJ. An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *J Thromb Haemost*. 2013;11:1329–40.

[7] Rallapalli PM. *Interactive location-specific databases and evolutionary aspects of the mutations in coagulation proteins*. PhD thesis. University College London; 2014.

[8] Saunders RE, O'Connell NM, Lee CA, Perry DJ, Perkins SJ. Factor XI deficiency database: an interactive web database of mutations, phenotypes, and structural analysis tools. *Hum Mutat*. 2005;26:192–8.

[9] Harris VA, Lin W, Perkins SJ. Analysis of 180 genetic variants in a new interactive FX variant database reveals novel insights into FX deficiency. *TH Open*. 2021;5:e557–69.

[10] Harris VA, Lin W, Perkins SJ. Analysis of 272 genetic variants in the upgraded interactive FXI web database reveals new insights into FXI deficiency. *TH Open*. 2021;5:e543–56.

[11] Li T, Miller CH, Payne AB, Hooper WC. The CDC hemophilia B mutation project mutation list: a new online resource. *Mol Genet Genomic Med*. 2013;1:238–45.

[12] Giannelli F, Green PM, Sommer SS, Poon M-C, Kudwig M, Schwaab R, Reitsma PH, Goossens M, Yoshioka A, Figueiredo MS, Brownlee GG. Haemophilia B: database of point mutations and short additions and deletions, 7th edition. *Nucleic Acids Res*. 1997;25:133–5.

[13] Xu Z, Harris VA, Perkins SJ. *The analysis of 1329 genetic variants in the updated coagulation factor IX database for haemophilia B. MSc Thesis*. London: University College London; 2021.

[14] Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res*. 2000;28:352–5.

[15] den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, Mcgowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PEM. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat*. 2016;37:564–9.

[16] den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat*. 2000;15:7–12.

[17] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–9.

[18] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22:2577–637.

[19] Touw WG, Baakman C, Black J, te Beek TA, Krieger E, Joosten RP, Vriend G. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res*. 2015;43:D364–8.

[20] Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum allowed solvent accessibilites of residues in proteins. *PLOS ONE*. 2013;8:e80635.

[21] Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;76:7.20.1–41.

[22] Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30:3894–900.

[23] Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN. PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng*. 1999;12:387–94.

[24] Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–4.

[25] Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31:2745–7.

[26] Sandell L, Sharp NP. Fitness effects of mutations: an assessment of PROVEAN predictions using mutation accumulation data. *Genome Biol Evol*. 2022;14:evac004.

[27] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–9.

[28] Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40:W452–7.

[29] Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLOS ONE*. 2012;7:e46688.

[30] Grantham R. Amino acid difference formula to help explain protein evolution. *Science*. 1974;185:862–4.

[31] Ellard S, Baple EL, Callaway A, Berry I, Forrester N, Turnbull C, et al. ACGS best practice guidelines for variant classification in rare disease 2020. *Assoc Clin Genom Sci*. 2019:1–32.

[32] Xie X, Chen C, Liang Q, Wu X, Wang X, Wu W, Ding Q. Characterization of two large duplications of F9 associated with mild and severe haemophilia B, respectively. *Haemophilia*. 2019;25:475–83.

[33] Wheeler RB, Cutler JA, Alamelu J, Mitchell MJ. The first report of a multi-exon duplication in the F9 gene causative of severe haemophilia B. *Haemophilia*. 2015;21:e433–5.

[34] Crossley M, Ludwig M, Stowell KM, De Vos P, Olek K, Brownlee GG. Recovery from hemophilia B Leyden: an androgen responsive element in the factor IX promoter. *Science*. 1992;257:377–9.

[35] Ahmed SZ, O'Rourke M, Jenkins V, Regan I, Nolan B. Progressive increase in FIX level in males with haemophilia B Leyden and c.35G>A mutation in early childhood not related to androgen effect. *Br J Haematol*. 2020;189:e262–5.

[36] Radic CP, Rossetti LC, Abelleyro MM, Candela M, Bianco RP, Pinto M, de T, Larripa IB, Goodeve A, De Brasi CD. Assessment of the F9 genotype-specific FIX inhibitor risks and characterisation of 10 novel severe F9 defects in the first molecular series of Argentinian patients with haemophilia B. *Thromb Haemost*. 2013;109:24–33.

[37] Santoro C, Quintavalle G, Castaman G, Baldacci E, Ferretti A, Riccardi F, Tagliaferri A. Inhibitors in hemophilia B. *Semin Thromb Hemost*. 2018;44:578–89.

[38] Di Michele D. Inhibitor development in haemophilia B: an orphan disease in need of attention. *Br J Haematol*. 2007;138:305–15.

[39] Laffan M, Brown SA, Collins PW, Cumming AM, Hill FGH, Keeling D, Peake IR, Pasi KJ. The diagnosis of von Willebrand disease: a guideline from the UK Haemophilia Centre Doctors' Organization. *Haemophilia*. 2004;10:199–217.

[40] Joosten RP, Te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, Schneider R, Sander C, Vriend G. A series of PDB related databases for everyday needs. *Nucleic Acids Res*. 2011;39:411–9.
[41] Efthymiou C, Print EHT, Simmons A, Perkins SJ. Analysis of 363 genetic variants in F5 via an interactive web database reveals new insights into FV deficiency and FV Leiden. *Thromb Haemost Open*. 2023;7:e31–40.
[42] Johnsen JM, Fletcher SN, Huston H, Roberge S, Martin BK, Kircher M, Josephson NC, Shendure J, Ruuska S, Koerper MA, Morales J, Pierce GF, Aschman DJ, Konkle BA. Novel approach to genetic analysis and results in 3000 hemophilia patients enrolled in the My Life, Our Future initiative. *Blood Adv*. 2017;1:824–34.
[43] Vidal F, Gallardo D. Hemobase, http://www.hemobase.com; 2010 [accessed January 2023].
[44] Kasper CK, Lin JC. Prevalence of sporadic and familial haemophilia. *Haemophilia*. 2007;13:90–2.
[45] Diop S, Haffar A, Mahlangu J, Chami I, Kitchen S, Pierce G. Improving access to hemophilia care in sub-Saharan Africa by capacity building. *Blood Adv*. 2019;3:1–4.
[46] Stonebraker JS, Bolton-Maggs PHB, Michael Soucie J, Walker I, Brooker M. A study of variations in the reported haemophilia B prevalence around the world. *Haemophilia*. 2012;18:e91–4.
[47] Jenkins VV, Egan H, Keenan C, O'Shea E, Smith OP, Nolan B, White B, O'Donnell J. Mutation analysis of haemophilia B in the Irish population: Increased prevalence caused by founder effect. *Haemophilia*. 2008;14:717–22.
[48] Li T, Miller CH, Driggers J, Payne AB, Ellingsen D, Hooper WC. Mutation analysis of a cohort of US patients with hemophilia B. *Am J Hematol*. 2014;89:375–9.
[49] Mårtensson A, Letelier A, Halldén C, Ljung R. Mutation analysis of Swedish haemophilia B families - high frequency of unique mutations. *Haemophilia*. 2016;22:440–5.
[50] Ljung R, Auerswald G, Benson G, Dolan G, Duffy A, Hermans C, Jiménez-Yuste V, Lambert T, Morfini M, Zupančić-Šalek S, Santagostino E. Inhibitors in haemophilia A and B: management of bleeds, inhibitor eradication and strategies for difficult-to-treat patients. *Eur J Haematol*. 2019;102:111–22.
[51] Male C, Andersson NG, Rafowicz A, Liesner R, Kurnik K, Fischer K, Platokouki H, Santagostino E, Chambost H, Nolan B, Königs C, Kenet G, Ljung R, van den Berg HM. Inhibitor incidence in an unselected cohort of previously untreated patients with severe haemophilia B: a PedNet study. *Haematologica*. 2021;106:123–9.

SUPPLEMENTARY MATERIAL

The online version contains supple mentary material available at https://doi.org/10.1016/j.jtha.2023.02.005