Analysis of 180 genetic variants in a new interactive FX variant database reveals novel insights into FX deficiency

Victoria A Harris, Weining Lin, Stephen J Perkins.

Affiliations below.

DOI: 10.1055/a-1704-0841

Please cite this article as: Harris V A, Lin W, Perkins S J. Analysis of 180 genetic variants in a new interactive FX variant database reveals novel insights into FX deficiency. TH Open 2021. doi: 10.1055/a-1704-0841

Conflict of Interest: The authors declare that they have no conflict of interest.

This study was supported by Lister Institute of Preventive Medicine (http://dx.doi.org/10.13039/501100001255), Not available

Abstract:
Coagulation Factor X (FX), often termed Stuart-Prower Factor, is a plasma glycoprotein composed of the γ-carboxyglutamic acid (Gla) domain, two epidermal growth factor domains (EGF-1, EGF-2) and the serine protease (SP) domain. FX plays a pivotal role in the coagulation cascade, activating thrombin to promote platelet plug formation and prevent excess blood loss. Genetic variants in FX disrupt coagulation and lead to FX or Stuart-Prower Factor deficiency. To better understand the relationship between FX deficiency and disease severity, an interactive FX variant database has been set up at https://www.factorx-db.org, based on earlier websites for the Factor XI and IX coagulation proteins. To date (April 2021), we report 427 case reports on FX deficiency corresponding to 180 distinct F10 genetic variants. Of these, 149 are point variants (of which 128 are missense), 22 are deletions, three are insertions and six are polymorphisms. FX variants are phenotypically classified as being Type I or Type II. Type I variants involve the simultaneous reduction of FX coagulant activity (FX:C) and FX antigen levels (FX:Ag), whereas Type II variants involve a reduction in FX:C with normal FX:Ag plasma levels. Both types of variants were distributed throughout the FXa protein structure. Analyses based on residue surface accessibilities showed the most damaging variants to occur at residues with low accessibilities. The interactive FX web database provides a novel easy-to-use resource for clinicians and scientists to improve the understanding of FX deficiency. Guidelines are provided for clinicians who wish to use the database for diagnostic purposes.

Corresponding Author:
DPhil Stephen J Perkins, University College London, Structural and Molecular Biology, Darwin Building, WC1E 6BT London, United Kingdom of Great Britain and Northern Ireland, s.perkins@ucl.ac.uk

Affiliations:
Victoria A Harris, University College London, Structural and Molecular Biology, London, United Kingdom of Great Britain and Northern Ireland
Weining Lin, University College London, Structural and Molecular Biology, London, United Kingdom of Great Britain and Northern Ireland
Stephen J Perkins, University College London, Structural and Molecular Biology, London, United Kingdom of Great Britain and Northern Ireland
Analysis of 180 genetic variants in a new interactive FX variant database reveals novel insights into FX deficiency

Victoria A. Harris†, Weining Lin† and Stephen J. Perkins†

† Research Department of Structural and Molecular Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, United Kingdom.

Correspondence: Professor Stephen J. Perkins, Department of Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, U.K. e-mail: s.perkins@ucl.ac.uk This manuscript contains 27 pages, including 9 Figures.

ABSTRACT

Coagulation Factor X (FX), often termed Stuart-Prower Factor, is a plasma glycoprotein composed of the γ-carboxyglutamic acid (Gla) domain, two epidermal growth factor domains (EGF-1, EGF-2) and the serine protease (SP) domain. FX plays a pivotal role in the coagulation cascade, activating thrombin to promote platelet plug formation and prevent excess blood loss. Genetic variants in FX disrupt coagulation and lead to FX or Stuart-Prower Factor deficiency. To better understand the relationship between FX deficiency and disease severity, an interactive FX variant database has been set up at https://www.factorx-db.org, based on earlier websites for the Factor XI and IX coagulation proteins. To date (April 2021), we report 427 case reports on FX deficiency corresponding to 180 distinct F10 genetic variants. Of these, 149 are point variants (of which 128 are missense), 22 are deletions, three are insertions and six are polymorphisms. FX variants are phenotypically classified as being Type I or Type II. Type I variants involve the simultaneous reduction of FX coagulant activity (FX:C) and FX antigen levels (FX:Ag), whereas Type II variants involve a reduction in FX:C with normal FX:Ag plasma levels. Both types of variants were distributed throughout the FXa protein structure. Analyses based on residue
surface accessibilities showed the most damaging variants to occur at residues with low accessibilities. The interactive FX web database provides a novel easy-to-use resource for clinicians and scientists to improve the understanding of FX deficiency. Guidelines are provided for clinicians who wish to use the database for diagnostic purposes.

248 words

**Key words:** Coagulation factors, haemostasis, protein structure/folding, gene mutations, inherited coagulation disorders

**Introduction**

Factor X (FX) is a vitamin K-dependent serine protease encoded by the F10 gene, located on chromosome 13. F10 is composed of eight exons and is 27kb in length. The gene encodes a pre-peptide and a pro-peptide, cumulatively referred to as the pre-pro leader, a γ-carboxyglutamic acid (Gla) domain, two epidermal growth factor domains (EGF-1 and EGF-2), a 52-amino acid activation peptide (AP) and a catalytically active serine protease (SP) domain (Figure 1). In the coagulation cascade, FX is responsible for the conversion of prothrombin to thrombin to drive the formation of the platelet plug. FX is synthesized as a zymogen and is activated via the extrinsic or intrinsic coagulation pathways, involving the Tissue Factor (TF) / Factor VIIa (FVIIa) complex or the Factor VIIa (FVIIa) / Factor IXa (FIXa) complex, respectively. Activation involves cleavage of theArg234-Ile235 bond in the FX heavy chain, releasing the 52-amino acid AP and activated FXa (Figure 1). Ile235 then forms a salt bridge with Asp418 in the substrate binding pocket. Additional FX cleavage at Lys475-Ser476 is often observed but has no major functional implications. FXa typically complexes with cofactor FVa to form the prothrombinase complex, enhancing prothrombin affinity and catalysis by $10^5$-fold. This complex interacts with membranous-phosphatidyserine in a Ca$^{2+}$ dependent manner. Following membrane assembly, prothrombinase cleaves prothrombin at two distinct sites, Arg271 and Arg320, yielding the N-terminal peptide and disulphide linked thrombin. Thrombin subsequently promotes clot formation. In addition to playing a direct role in the coagulation cascade, FXa is implicated in conditions such as tissue fibrosis, atherosclerosis and angiogenesis. FXa binds the protease-activated receptors PAR-1 and PAR-2 in vascular, immune and epithelial cells. FXa
cleaves the N-termini of PAR-1 and PAR-2, creating a novel N-terminus that functions as a tethered ligand that leads to the triggering of intracellular signalling cascades.\textsuperscript{1,7}

Experimentation involving \textit{F10} KO mice emphasizes the importance of FX in physiological function. The absence of FX results in partial embryonic lethality and fatal neonatal bleeding.\textsuperscript{1} Similarly, in humans, variants that impede FX activity lead to disease states, commonly termed Congenital Factor X deficiency. Such states can also be referred to as Stuart-Prower Factor deficiency, named after the first recognised deficient patients.\textsuperscript{8-9} FX deficiency occurs at a frequency of 1:500,000 – 1:1,000,000 across the globe, with 1 in 500 individuals believed to be heterozygous carriers.\textsuperscript{9-10} FX deficiency is more prevalent in regions where consanguineous marriage remains common, such as in the Middle East.\textsuperscript{9,11-12} Deficient patients experience variable bleeding symptoms, ranging from easy bruising, gum bleeding, nose bleeds and haematuria in mild cases to haemarthrosis, hematomas, intramuscular bleeds, gastro-intestinal tract bleeds and intracranial haemorrhage in more severe cases. All FX deficient patients are at greater risk of prolonged post-operative bleeding.\textsuperscript{8-13} The predominant treatments for FX deficiency include FX replacement and the use of antifibrinolytics.\textsuperscript{13-16} Antifibrinolytics such as \textit{ε}-aminocaproic and tranexamic acid or fibrin glue are often used for the treatment of minor bleeds.\textsuperscript{15-17} Variants causing FX deficiency are characterised as having cross-reacting material negative (CRM\textsuperscript{-}) or cross-reacting material positive (CRM\textsuperscript{+}) phenotypes. CRM\textsuperscript{-} or ‘Type I’ variants are those that result in the parallel reduction of FX coagulant activity (FX:C) and FX plasma antigen levels (FX:Ag). Contrastingly, CRM\textsuperscript{+} or ‘Type II’ variants reduce FX:C without impacting the FX:Ag level. Such variants are functional variants that go undetected by cellular quality control mechanisms.\textsuperscript{8,13} FX:C in healthy human plasma ranges from 50-150 IU/dL.\textsuperscript{18} Mildly deficient patients have 6-10\% of normal FX:C, moderately deficient patients 1-5\% and severely deficient patients < 1\%.\textsuperscript{9,14,16}

The diagnosis and treatment of FX deficiency will benefit from a more complete understanding of the correlation between the \textit{F10} genotype and the bleeding severity.\textsuperscript{19} In order to better understand this relationship, we have developed the first interactive database at https://www.factorx-db.org for FX variants. The database contains genetic information from literature sources in relation to 180 unique variants from 427 patients, as well as structural information from FXa protein crystallography. The new FX database was adapted from our previously developed interactive coagulation factor databases for Factor XI (FXI) and Factor
IX (FIX) at [https://www.factorxi.org](https://www.factorxi.org) and [https://www.factorix.org](https://www.factorix.org), which underwent initial development in 2003 and 2013 respectively. In this new FX database, we show how the extended domain structure of FXa is correlated with the distribution of Type I and Type II variants where the phenotype is reported. We compare this distribution with that in FXI when our recently updated variant and structural analyses indicate a predominance of Type I variants in FXI. We show that changes in the solvent surface accessibility of individual residues are well correlated with the occurrence of genetic variants in the FX structure, regardless of the phenotypic classification. We offer guidelines on the interpretation of deleterious variants in the database. The availability of this new database will help clarify the molecular basis of FX deficiency for clinicians and scientists.

**Methods**

**Source of the FX database**

The new interactive FX web database at [https://www.factorx-db.org](https://www.factorx-db.org) carries 180 distinct F10 genetic variants that are associated with FX deficiency. The database was created at University College London and the website copyright is retained by S. J. Perkins and University College London. Database copying is not permitted without explicit permission. The FX database was derived from the newly upgraded FXI database at [https://www.factorxi.org](https://www.factorxi.org), and the FIX database at [https://www.factorix.org](https://www.factorix.org). The variants in the FX database were populated through online literature searches of peer-reviewed articles, primarily those found on PubMed at [https://pubmed.ncbi.nlm.nih.gov/](https://pubmed.ncbi.nlm.nih.gov/). The literature cut-off date was April 2021, from which a total of 180 unique FX variants and 427 patients was extracted. The variants were compiled into a spreadsheet in order to establish the FX MySQL database, using phpMyAdmin software ([https://www.phpmyadmin.net/](https://www.phpmyadmin.net/)) as an intermediary platform. The database is maintained on the University College London server and utilises HTML programming, PHP and JavaScript to enable public access via the web. If desired for personal or private research use only, a list of the FX variants and their associated fields can be downloaded from the Variants menu on the website (Supplementary Materials).

**Analysis of FX variants**

The FX database records DNA changes in HGVS format, where +1 refers to the A of the ATG initiation codon, at the start of the 40-residue pre-pro leader peptide. When publications have used legacy numbering for DNA changes, the origin of the residue
numbering is stated clearly in the variant comments section on the website. Protein changes are recorded in HGVS format, with codon +1 referring to the ATG initiation codon. To facilitate comparison with older publications, legacy numbering is presented also for protein changes on the database, with codon +1 referring to the first codon of the mature unactivated FX protein (+41 in HGVS numbering).

While 136 FXa crystal structures are available on the PDB database (https://www.rcsb.org), many structures do not show the full-length FXa protein, with many lacking the Gla and EGF-1 domains entirely. Of the 136 FXa structures, PDB ID: 1P0S was alone in successfully determining the Gla domain, yet this crystal structure did not model the EGF-1 domain. Five FXa crystal structures successfully determined the EGF-1 domain, with PDB ID: 1XKA encompassing the most light- and heavy-chain residues at the highest resolution, showing the EGF-1, EGF-2 and SP domains in full. The 1P0S crystal structure showed a structural resolution of 0.280 nm and a goodness-of-fit R-value of 0.207. The 1XKA model showed a higher resolution of 0.230 nm and an improved R-value of 0.196. Thus, the 1P0S structure was superimposed in full onto the 1XKA structure to result in a full-length merged FXa model with all four domains visible in an extended conformation. The 1P0S structure was then edited out, leaving only its Gla domain along with the full 1XKA structural model. The complete merged FXa model consisted of a light chain ranging from Ser43 to Arg179 and a heavy chain ranging from Ile235 to Arg469 (Figure 1). The 1XKA structure was retained as much as possible owing to its improved resolution compared to 1P0S. Ramachandran analysis at http://molprobity.biochem.duke.edu was used to confirm the accuracy of the model. Our final native FXa model was not refined by energy minimisation, nor was the structure refined after making the amino acid substitutions.

In order to interpret the effect of the FX missense variants, four different substitution analyses were carried out on them, namely Polymorphism Phenotyping v2 (PolyPhen-2) at http://genetics.bwh.harvard.edu/pph2/, Sorting Intolerant From Tolerant (SIFT) at https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html, Protein Variation Effect Analyzer (PROVEAN) at http://provean.jcvi.org/seq_submit.php and Grantham analysis, using the Grantham matrix proposed in 1974. Both the PolyPhen-2 and SIFT algorithms give variants scores ranging from 0.0 to 1.0. However, in the PolyPhen-2 algorithm, the closer to 1.0 the prediction score is, the more damaging the variant is likely to be. The SIFT scores are...
reversed where values closer to 0.0 predict more damaging variants and those closer to 1.0 are more tolerated. The PROVEAN threshold used was -2.5, thus variants with scores below the threshold were considered deleterious and those above -2.5 were neutral. The Grantham analysis differs from the other three in that it is not sequence specific and is purely based on the amino acids undergoing substitution. Grantham scores range from 0 (silent variants) to 215, with larger scores indicating more radical and likely more damaging substitutions.

The Definition of Secondary Structure of Proteins (DSSP) tool at https://www3.cmbi.umcn.nl/xssp/ was used to identify the secondary structure of each residue in the merged FXa structure. DSSP was applied both to the merged FXa structure as well as to the four separate FXa domains. Residues were individually assigned secondary structures to be one of 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 was also used to determine the relative surface accessibility of each residue in the FXa crystal structure in Å². The accessibilities were converted into % accessibility by dividing the DSSP output by the theoretical solvent accessible surface area of the amino acid sidechain in question. The results were simplified as follows. Percentage accessibilities of 0-9% were given the value 0, 10-19% the value 1, 20-29% the value 2, and so on. Residues with accessibilities of 0 or 1 were classified as buried and those with accessibilities of 2-9 were classified as solvent-exposed.

Results

The new FX interactive web database

The interactive FX database (https://www.factorx-db.org) presents genetic and structural information in relation to 180 FX variants and 427 patients, sourced from 118 research articles (Figure 1). Specific genetic features in the database include basic and advanced search tools, an interactive clickable variant map, and interspecies F10 multiple sequence alignments, permitting phylogenetic analysis of the F10 gene (left, Figure 2). Additionally, the database offers structural analyses of the identified variants, a list of published FX crystal structures and links to further FX resources (right, Figure 2). These features are accessed from the home page, which features movies of FXa with and without its associated variants. From the home page, the basic search tool enables users to search for variants of specific type or with defined cDNA or amino acid numbers. The advanced search
tool allows for more explicit searches, including options to search by variant effect, location and/or reference. The genetic searches report on variant information, patient information and relevant comments. Variant information includes variant type, effect, location and allelic specific information. Allelic frequencies (AFs) are provided for variants wherever possible using data supplied by the genome aggregation database (gnomAD) version 2.1.1 at https://gnomad.broadinstitute.org/.

The gnomAD v2.1.1 data set spans 125,748 exome sequences and 15,708 whole-genome sequences and contains 51 (28%) of the 180 identified FX variants. The AF is used as an indication of the relative frequency of a given variant at a specific genetic locus in a genome. The AF cut-off used was 0.01, thus those AF > 0.01 indicated a common variant. Of the 51 variants present in the gnomAD dataset, only two had AF > 0.01, both of which corresponded to known polymorphisms (Thr264Thr and c.-40C>T). The predominance of rare FX variants in the gnomAD population highlights the importance of FX gene conservation. The anonymised patient information reports on patient age, gender, race, assay results and disease severity.

The database provides structural analyses for the missense variants which can be mapped onto an interactive FXa structure for inspection. A molecular Jmol viewer for each variant is provided through Java software. It should be noted that Java applets need to be enabled within the browser to permit these features to be seen. Users are able to apply prediction software to better understand the structural effects of any potential missense variant (right, Figure 2). Users have the option to submit novel variants and to contact the site administrator with additional queries. The database site map and help pages make for easy navigation and comprehension of the site.

**Classification of FX variants and polymorphisms**

Of the 180 variants presented in the FX database, 82.78% are point variants, 12.22% deletions, 3.33% polymorphisms and 1.67% insertions (Figure 3(a)). The 149 point variants are further subdivided into missense (86.58%), nonsense (6.71%) and splice site (6.71%) variants (Figure 3(b)). Of the 180 genetic variants, 174 are disease-associated alterations and six are non-disease associated polymorphisms, including the above pair (Thr264Thr and c.-40C>T). The 180 variants are spread across the F10 gene, with 51.11% clustered in the large SP domain (Figure 3(c)). As illustrated in Figure 4, 46.55% of the disease-causing variants are phenotypically classified as Type I, 16.09% as Type II and 37.36 % have an unknown phenotype. Here, Type II variants are characterised by a FX:C to FX:Ag ratio < 0.7. Closer
phenotypic analysis revealed that 64.29% of the Type II variants were in the SP domain, while 14.29% occurred in the Gla domain, 7.14% in the EGF-1 domain, 7.14% in the EGF-2 linker region and less than 5.00% in each of the pre-pro leader and AP sequences. The abundance of Type II variants in the SP and Gla domains highlights their functional activity in terms of peptide hydrolysis and in binding to host cell membranes respectively. The reduced abundance of Type II variants in the EGF-1 and EGF-2 domains and the abundance of Type I variants in these indicates their importance in the formation of a correctly-folded protein structure. Of the 10 most commonly occurring FX variants in patients (Pro383Ser, 25 occurrences; Gly21Arg, 21; Gly262Asp, 20; Glu54Lys, 16; Glu91Lys, 15; Asp322Asn, 15; Ser374Pro, 14; Gly406Ser, 14; Gly421Asp, 13; Glu142Lys, 12; in order of frequency as shown), six are Type II and four are Type I. Interestingly the Asp322Asn variant is a Type II variant involving the catalytic triad residue Asp322, meaning that this variant loses FXa catalytic activity. The occurrence of this disease-causing variant (15 case reports to date) highlights the importance of the FX catalytic triad in FXa proteolytic function.

Substitution analysis was carried out on 128 FX missense variants to predict the damaging effects of each amino acid change. All the substitutions involved single nucleotide changes; none appeared in the greyed boxes of Figure 5. As illustrated in Figure 5, the total count for each missense change is shown, with Gly, Glu, Arg and Cys being the most commonly mutated residues (23, 18, 14 and 12 cases respectively). For these four residues, a change involving a charged residue is often observed, such as Gly>Arg, Glu>Lys and Cys>Arg. Such substitutions are readily seen to be protein damaging. Of interest, the Gly residues (23) were the most commonly affected. This concurs with Gly being the smallest residue and thus its replacement with a bulkier residue is likely to cause protein structural rearrangements. Charged residues are also more often varied than others; Glu (18) and Arg (14) are the second and third most commonly substituted residues. The removal of a charged group in a protein can often destabilise its structure. Cys residues are the fourth most commonly mutated residue. The formation of Cys-Cys disulphide bridges is key to protein folding stabilisation, so the loss of one bridge can be disruptive to the overall three-dimensional structure.

Phenotype and accessibility analysis of FX missense variants
The merged FXa crystal structure (Methods) resulted in a full-length FXa model that permitted structural analyses of the variants. When subjected to Ramachandran stereochemical analysis, the 1XKA crystal structure (used for the EGF-1, EGF-2 and SP domains) revealed 92.8% of all residues to be in the most favoured conformational regions, with a further 6.6% of residues in the additional allowed regions. The 1P0S crystal structure (used for the Gla domain) revealed that 90.3% of all residues were in the most favoured conformational regions, with a further 8.2% of residues residing in additional allowed regions. The 1XKA structure showed two Ramachandran outliers (Leu113 and Arg126), while the 1P0S structure showed seven (Leu131, Gln144, Asn145, Asp159, Asn160, Gln356, Glu369). The merged FXa model revealed that 99.4% of all residues were in favoured regions with 91.8% of these residing in the most favoured regions. The only two Ramachandran outliers found in the merged model were Leu113 and Arg126, present also in the EGF domains in the 1XKA crystal structure. The merged FXa structure was concluded to be of high enough quality for an accurate structural assessment of the FX variants.

The merged FXa structure was used to show the distribution of 113 missense variants and 4 non-disease associated polymorphisms in mature FXa (Figures 6(a) and 6(b)). The Gla domain contained three α-helices, A1-A3 and a single π-helix, P1. The EGF-1 and EGF-2 domains were primarily β-sheet structures, with EGF-1 forming two β-strands A-B and EGF-2 forming four β-strands A-D. The SP domain contained 13 β-strands A-M, as well as three minor α-helices, three 310-helices and one π-helix (Figure 4). A schematic representation of the FXa domain structure showed that the four FXa domains were in an extended arrangement (Figure 6(c)). In the merged FXa structure, 186 out of 488 residues were assigned accessibilities of 0 or 1, meaning that their sidechains were buried, and a further 186 residues were assigned accessibilities of 2 or more, indicating that their sidechains were solvent exposed (Figure 4). The remaining 116 residues were not present in the FXa protein structure and thus were unavailable for accessibility analysis. These 116 residues included the 40-residue pre-pro leader peptide, the 52 amino acid AP removed on FX activation, and several others including those at the end of the SP domain. Of the 127 variants with defined accessibilities, several of which occurred at the same location in the FX sequence, 49 were classified as surface exposed with accessibilities of 2 or more and 78 were classified as buried with accessibilities of 0 or 1 (Figure 4).
When observing the full-length FXa structure, the ratio of Type II to Type I variants stood at 0.35, while this ratio increased to 0.44 when looking solely at the SP domain. Given the catalytic function of the SP domain, unsurprisingly Type II variants occurred more frequently in the SP domain compared to the rest of FXa. The clustering of the Type II variants (green) was visible in the SP domain when mapped onto the FX model (Figure 6(a)). Interestingly, of the 17 Type II point variants in the SP domain, three had surface accessibilities of 2 or more, indicating sidechain exposure, while the remaining 14 were classified as buried. Such findings indicated that, regardless of surface exposure, the SP variants were able to impede native FXa function without disrupting the overall FXa structure, resulting in Type II defects. In contrast, the Type II variants in the Gla, EGF-1 and EGF-2 domains were all at exposed locations, suggesting that these may affect interdomain or intermolecular interactions in the functioning of FX. The predominance of Type I variants in the Gla, EGF-1 and EGF-2 domains suggested that these domains were more important in maintaining an extended FXa structure rather than having a functional role.

The observation of a missense FX variant in the gene of a patient is not necessarily causative of a bleeding disorder. To assist interpretations, the FX website offered guidelines to clarify whether a given variant is likely to disturb the overall protein structure and function. The PolyPhen-2 algorithm predicted that 109 (85%) of the 128 variants were damaging, with scores of 0.9 to 1.0 (Figure 7(a)). The SIFT analysis predicted that 110 (86%) of variants were damaging with scores of 0.0 to 0.2 (Figure 7(b)). This outcome indicated that most FX variants were causative for FX deficiency as expected. The PROVEAN and Grantham analyses demonstrated less clear trends, although the majority of variants showed damaging scores (Figure 7(c,d)). The interactive FX database provides variant-specific PolyPhen-2, SIFT, PROVEAN and Grantham scores for the 128 missense variants, thus providing an easy-to-use clinical support tool to clarify the significance of a variant (Figure 2). The significance of the four scores for a given variant is clarified from Figure 7. A high damaging score would highlight that the substitution in question is likely to explain FX deficiency in the patient.

In order to further examine the relationship between FX phenotypes and surface accessibilities, the residue surface accessibilities were displayed as a function of phenotypes for (a) the intact FXa protein and (b) the individually separated FXa domains (Figure 8). Notably, a high 68% of Type I variants (36 of 53) showed accessibilities of 0 or 1,
highlighting their predisposition to be buried within the intact FXa protein structure (Figure 8(a)). In contrast, Type II variants and polymorphisms appeared in both the surface-exposed and buried regions of the FXa structure, with no clear preference for either location. Many variant residues were of unknown phenotype, however interestingly the majority of these showed low accessibilities. Following the separation of the four domain structures in FX, the resulting changes in surface accessibility compared to the intact protein enabled the identification of residues that made interdomain contacts. An even higher proportion of 91% of Type I variants (48 of 53) was located to residues that showed no accessibility change after domain separation (Figure 8(b)). The same outcome was also seen for Type II variants and unassigned phenotypes. The predominance of variants at buried locations within FX showed that small perturbations in the FXa structure, through the introduction of variants that lead to slight changes in surface accessibility, are sufficient to impair the protein and lead to disease states.

To illustrate the utility of the website, four individual variant residues with low surface accessibilities were visually highlighted on the FXa model (Figure 9), namely (a) Asp135Glu in EGF-2 (Type I, accessibility 2), (b) Gly244Arg/Glu in SP (Type I, accessibility 0), (c) Gly262Asp in SP (Type I, accessibility 0) and (d) Asp322Asn in SP (Type II, accessibility 0). The separation of the domain structures alters the accessibility of Asp135 from 2 to 3, and that of Gly244 from 0 to 3. These views explain the visual context of residue substitutions by three-dimensional inspections of the native FXa structure. It should be noted that these changes (particularly to buried residues) may be causing local structural changes in the protein. For the first variant, the replacement of an Asp with a Glu residue seen at the EGF-1 and EGF-2 interface was sufficient to destabilise the FXa protein folding, presumably by altering the interdomain interface to form a less stable one. For the second and third variant residues, the replacement of a buried Gly residue by a much bulkier charged Arg, Glu or Asp residue was likely to destabilise the protein folding of the SP domain, resulting in the observed Type I phenotype. The third variant is exemplified in Figure 2. The fourth example is the substitution of Asp322 by Asn322 within the catalytic triad His276-Asp322-Ser419. The change impacts only on FXa function, not on the overall protein structure, thus explaining why this is reported as a Type II defect. Missense variants at such locations have perturbed these inter- and intradomain interactions, disrupting the native FXa structure and resulting in protein degradation or dysfunction. Thus, the website provides molecular explanations for the relative abundance of Type I and II variants in FX in terms of
small but significant disruptions to the intricately-folded domain structure, even though the FXa domain structure is an extended one and relatively solvent exposed.

DISCUSSION

The compilation of 180 genetic variants for FX in this study represents to our knowledge the first interactive web database for FX that summarises documented patient records, their genetic variants, and protein structural analyses, all assembled as one convenient resource. The new FX website represents the continuation of our other related interactive websites for FVII, FVIII, FIX and FXI (https://www.factorvii.org, https://www.factorviii-db.org, https://www.factorix.org and https://www.factorxi.org).20-23,37 For FX, this website will significantly improve the quality of variant analyses and will lead to further insights into the occurrence of FX disease states. Most notably, we show that these variants are observed across the entire FXa protein structure (Figure 6), and that accessibility changes in the amino acid packing arrangement in the folded FXa structure by residue substitution is a major cause of FX deficiency (Figure 8). Accessibility changes may be a good predictor for the changes associated with a variant. The disruption scores from Polyphen-2, SIFT, PROVEAN and Grantham offer useful insight into the impact of the variants and will provide the clinician with guidelines on the significance of a variant. Such an analysis of the FX variants has resulted from three main advances: (a) the availability of sufficient literature on FX genetic variants as the result of the increasing use of patient genotyping (Figure 1); (b) the availability of a merged FXa crystal structure of a sufficient high quality for analyses (Methods); (c) our increasing expertise in developing interactive databases related to haemophilia.23

Many multidomain plasma proteins present extended domain arrangements. Examples include the three-dimensional structures for FXa, activated FVII and activated FIX, all three of which have the same four domains termed Gla-EGF1-EGF2-SP. Other plasma proteins present more compact domain arrangements such as that of FXI.23 Both FX and FXI present variants that are distributed throughout their protein structures, implying that the extended or compact nature of their domain structures do not influence variant distributions. However, where the phenotypes are known, FVII, FIX and FX variants show a higher proportion of Type II phenotypes which are associated with functional defects, rather than Type I phenotypes which correspond to misfolding defects.20-23,37 This outcome is as expected given
that FVII, FIX and FX all have extended domain arrangements. FXI shows a higher proportion of Type I defects, and this outcome is attributed to its compact domain arrangement which is more easily damaged by disease-associated variants. Heparin and Factor Va exosites are known for FX and both are shown on the SP domain of Figure 6b, together with the catalytic triad. Of interest is that many Type II variants in the SP domain (green) are spatially close to these exosites.

Our interactive FX database serves as a useful resource for clinicians and scientists to diagnose FXI deficiency and provide insight into the effect of a variant. Database technology becomes required in the light of large increases in the known genetic variants in FX, when a simple spreadsheet listing of mutations is no longer adequate to monitor these. The website layout is designed to present genetic and structural information on FX as two distinct but parallel themes, similar to that for our original FVII, FVIII, FIX and FXI databases. This is illustrated using genetic and structural outputs for the commonly-reported Gly262Asp variant, for which 20 patient records exist (Figure 2). On the left, further insight into the conservation of Gly262 is obtained from the AA Alignments tab which shows Gly262 aligned with six other mammalian species. This alignment reveals that the Gly262 residue is conserved in only three other sequences and therefore may be able to accommodate a bulkier sidechain. On the right, the structural analysis shows that Gly262 is a buried residue on a β-strand, and the JMol viewer shows that this is located inside the SP domain. Further research into FX, including experimental studies, will be key to a better understanding of the relationship between FX deficiency and disease severity.

What is known on this topic:

(1) Many FX genetic variants have been published in the literature yet there is no interactive database which compiles both variant and patient information and offers sequence and structural variant analysis.

(2) A total of 136 FXa crystal structures are available on the Protein Databank website (https://www.rcsb.org), yet there are no available structural models that show all four FXa domains.

What this paper adds:
(1) Our new interactive FX database enables sequence and structural analysis of F10 genetic variants identified prior to April 2021 and serves as a literature pool of an up-to-date FX resource.

(2) Information regarding 180 unique FX variants from 427 case reports is presented in our interactive database, of which 174 variants are disease-associated and 6 are non-disease-associated polymorphisms.

(3) We present a full-length structural model of FXa created from two previously published FXa crystal structures, and present guidelines to assess the likely impact of newly-reported missense variants.

Research grants and financial support:
The authors V.A.H. and W.L. were supported by studentships from the Lister Institute of Preventive Medicine for this project. The authors declare that they have no conflicts of interest with this research.

REFERENCES


23. Harris VA, Lin W, Perkins SJ. Analysis of 272 genetic variants in the upgraded interactive FXI web database reveals new insights into FXI deficiency. 2021; Submitted


Figure 1. Distribution of the 180 variants identified within the F10 gene. The six protein segments are not drawn to scale. The number of variants in each of the respective domains is shown in large font above or below the variant lists. Intronic variants are included in their respective domains according to sequence numbering. The residue numbering under the domains marks the amino acids framing each domain, given in HGVS format. N and C represent the N- and C-termini respectively of FX. Only 177 variants are displayed in this schematic as the remaining variants are undefined insertions and deletions.

Figure 2: Screenshots of the new FX interactive web database to illustrate the genetic and structural analysis made for the Gly262Asp variant. The upper panel displays the output when residue 262 is inputted on the FX website home page. By clicking “Show” next to the patient information heading, the lower left panel lists genetic information for the 20 patients reported with the Gly262Asp variant, of which the first 12 records are visible here, together with the source of the patient record. The multiple sequence alignment is shown underneath with Gly262 highlighted by a red arrow. Clicking “HERE” under the structural interpretation heading gives the image shown on the bottom right panel. This assesses the buried or exposed accessibility of the variant and its location in the FXa protein structure. Four different substitution analyses (Grantham, PolyPhen-2, SIFT and PROVEAN) are provided in order to help predict the likely effects of each missense variant on FXa. A Jmol view of the FXa structure is displayed that can be rotated and zoomed into as desired. Additional information regarding the amino acid side chain properties of both the wild type and mutant residues is also displayed on this page.

Figure 3. Distribution of the 180 variants identified within the F10 gene. The panels (a-c) illustrate the breakdown of the 180 FX variants by variant type, effect and location within the F10 gene sequence. The charts illustrated here distinguish the non-disease associated polymorphisms from the disease-causing genetic variants.

(a) Relative frequency of four different types of genetic variants in the F10 gene.

(b) Genetic classification of 149 point variants in the F10 gene.

(c) Relative distribution of the 180 FX variants across the F10 gene and within the protein domains. The six undefined variants include two unique deletions that stretch across exons 1-6 and 1-8 respectively, one undefined deletion and three missense variants preceding the F10 coding sequence.
Figure 4. Secondary structure and accessibility analysis of variants in the merged FXa crystal structure. The FX sequence is shown with the major secondary structure assignments highlighted in grey boxes. Residues are denoted 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). The SP catalytic triad His276-Asp322-Ser419 is highlighted in black. β-strands are labelled alphabetically in the order in which they occur. The β-strands in the EGF-1, EGF-2 and SP domains are denoted A-B, A-D and A-M respectively. α-helices are denoted A1-A3, 3_10-helices G1-G2 and π-helices P1. The positions of 136 point variants found in the F10 gene are highlighted in boxes, where 57 red boxes denote Type I mutations, 26 green boxes denote Type II mutations, and 33 orange boxes denote mutations with unknown phenotype. Several boxes correspond to multiple variants at one residue position. The 5 purple boxes denote non-disease associated polymorphisms. All numbering is in HGVS format.

Figure 5: Substitution grid representing 128 missense variants in the F10 gene. The grid illustrates the number of variants of each defined amino acid change. All substitutions shown are the result of a single nucleotide change. Any substitutions that require more than a single nucleotide change are shown in dark grey. Silent variants are excluded from the grid and shown in pale grey. The white boxes represent possible substitutions that do not occur in the F10 gene. The yellow boxes represent substitutions that occur once or twice, the orange boxes represent substitutions that occur three or four times and the red boxes represent substitutions that occur five or more times.

Figure 6: Structural view of the distribution of variants within FXa
(a) and (b) illustrate the front and back views (rotated by 180°) of the distribution of 113 missense variants and 4 polymorphisms within mature FX. The merged FXa crystal structure is shown in ribbon format, with the Cα positions of the variants and polymorphisms highlighted as spheres. The variant locations are coloured according to their phenotype. Type I (CRM-) variants are in red, Type II (CRM+) variants in green, and those with unknown phenotype in yellow. The non-disease associated polymorphisms are depicted in purple. The catalytic triad His276-Asp322-Ser419 (HGVS numbering) is shown in (b) in the SP domain as black spheres. In (b) also, orange spheres denote the residues Arg313, Lys316, Arg346, Arg 387, Lys391, Lys460 and Arg469 (HGVS numbering) correspond to a heparin binding site in FX (chymotrypsin numbering Arg93, Lys96, Arg125, Arg165, Lys169, Lys236 and
Arg240). These heparin binding residues together with Asp413-Asp418 and Val455-Arg469 (HGVS numbering) are involved with Factor Va binding (chymotrypsin numbering Asp185-Asp189 and Val231-Arg245). The ribbon colours correspond to those used on our website https://www.factorx-db.org.

(c) Schematic representation of the four domains in FX. The four domains are aligned and depicted in the same orientation and colours as in the ribbons in (a) and (b) above.

Figure 7: Substitution analysis of 128 missense variants in the F10 gene. Substitution analyses predict the damaging effects of substitution variants.

(a) Graphical distribution of variants determined by their PolyPhen-2 score.
(b) Graphical distribution of variants determined by their SIFT score.
(c) Graphical distribution of variants determined by their PROVEAN score. The PROVEAN threshold used was -2.5.
(d) Graphical distribution of variants determined by their Grantham score.

Figure 8: Accessibility analyses of 108 FX missense variants and four polymorphisms in the FX protein structure.

(a) FX variants in the FXa structure are grouped by phenotypic classification (CLASS). The variants are further subdivided according to the native residue accessibility (ACC) of the intact protein. Accessibility was determined using DSSP and is explained in detail in the Methods section. Accessibilities of 0 or 1 indicate sidechain burial and values >1 indicate sidechain exposure to solvent.

(b) FX variants are again grouped by phenotypic classification (CLASS) and accessibility (ACC). Here, accessibility refers to the change in residue accessibility when the intact FX protein is separated into its four distinct domains, namely the Gla, EGF-1, EGF-2 and SP domains.

Figure 9: Molecular graphic representation of four selected variants in the FXa protein structure. The views are taken from the unmodified FXa structural model (Methods). The left-hand panel highlights the residue of interest and its calculated solvent accessibility in the intact FXa protein. The right-hand side panel shows the same residue in its respective isolated domain, with its corresponding accessibility. Panels (a), (b), (c) and (d) highlight the residues Asp135, Gly244, Gly262 (Figure 2) and Asp322 (catalytic triad) respectively. Accessibility
calculations are detailed in the Methods section. The domain colouring corresponds to that in Figure 5. All residue numbering is given in HGVS format.
Factor X Gene (F10)

Variant Database

Search Results: 1 unique variant retrieved.

Terms with a "*" next to them are explained on the Help Page:

c.786G>A p.Gly262Asp (Legacy AA No. 222)

Variant Type: Point
Variant Effect: Missense
Phenotype: 1
Allele Count*: 3

Domain: Serine protease
Codon Change: G:A
Location: Exon 7
No. of Patients Reported: 20
Allele Number*: 202606
Allele Frequency*: 0.000011

References and Comments:
Payvandi et al 2002; Herrmann et al 2006; Epacceen et al 2012
Variant disrupts protein folding and may perturb FX catalytic activity due to proximity to catalytic triad.

Patient Information: Show

Structural Interpretation:
Please click HERE for in-depth variant analysis.

© Copyright 2003-2021, Structural Immunology Group, University College London, Gower Street, London WC1E 6BT
“Any part of this article may be copied or used in any way without permission.”

Residue Information:

Substitution Analysis:
- Grantham Score: 59
- PolyPhen-2 Prediction: Probably Damaging (Saroo: 1.005)
- SIFT Prediction: Probably Damaging (Saroo: 0.09)
- PROVEAN Prediction: Tolerated (Saroo: 2.02 E-3)

Structural Information:
- Gly262 is a buried residue (surface accessibility value < 0).
- Gly262 is in a region of secondary structure within the FX domains.

The DSSP assignment for this residue is ... E.

Controls:
The molecule can be rotated with the mouse holding the left button and the image can be magnified using the middle mouse button.
Right-click on the Jmol screenshots for more options.

Multiple Sequence Alignment of F10's Protein Sequences from 5 Other Species

Sequence identity: 95.3% 94.8% 99.9% 99.9% 99.9% 99.9% 99.9% 99.9% 99.9% 99.9%
<table>
<thead>
<tr>
<th></th>
<th>Ala</th>
<th>Arg</th>
<th>Asn</th>
<th>Asp</th>
<th>Cys</th>
<th>Gln</th>
<th>Glu</th>
<th>Gly</th>
<th>His</th>
<th>Ile</th>
<th>Leu</th>
<th>Lys</th>
<th>Met</th>
<th>Phe</th>
<th>Pro</th>
<th>Ser</th>
<th>Thr</th>
<th>Trp</th>
<th>Tyr</th>
<th>Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Arg</td>
<td>0</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Asp</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Cys</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Gln</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Glu</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Gly</td>
<td>0</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>His</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>Ile</td>
<td>0</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Leu</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Lys</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Met</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Phe</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Pro</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Ser</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Thr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Trp</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Tyr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Val</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>SUM</td>
<td>4</td>
<td>22</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>128</td>
</tr>
</tbody>
</table>

**AA MUTATED TO**

**AA MUTATED FROM**