

TH Open

Analysis of 363 variants in the coagulation Factor 5 gene via an interactive web database reveals new insights into FV deficiency and FV Leiden

Christos Efthymiou, Emily H Print, Anna Simmons, Stephen J Perkins.

Affiliations below.

DOI: 10.1055/a-1987-5978

Please cite this article as: Efthymiou C, Print E H, Simmons A et al. Analysis of 363 variants in the coagulation Factor 5 gene via an interactive web database reveals new insights into FV deficiency and FV Leiden. TH Open 2022. doi: 10.1055/a-1987-5978

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 known

Abstract:

The inherited bleeding disorder FV deficiency and clotting disorder FV Leiden are caused by genetic variants in the F5 gene. Both diseases occur with mild, moderate, severe, or asymptomatic phenotypes, and either dysfunctional or reduced amounts of plasma FV protein. Here we present an interactive web database containing 363 disease-associated unique F5 variants derived from 801 patient records that rationalizes their occurrence based on the 2224 residue sequence and new FV structures. The 363 variants correspond to 38 (10.5%) mild, 14 (3.9%) moderate, 52 (14.3%) severe, 49 (13.5%) asymptomatic, and 210 (57.8%) unreported phenotypes. The variant distribution in the FV domains A1, A2, A3, B, C1 and C2 was 42 (11.9%), 61 (17.2%), 58 (16.4%), 103 (29.1%), 28 (7.9%) and 29 variants (8.2%) respectively. Both the globular (A1-A3, C1, C2) and disordered (B) regions contain similar proportions of variants. Variants associated with either FV deficiency or FV Leiden do not cluster near known protein-partner binding sites, thus the molecular mechanism leading to their phenotypes cannot be explained. However, the widespread distribution of FV variants in combination with a high proportion of buried variant residues indicated that FV is susceptible to disruption by small perturbations in its globular structure. Variants located in the disordered B domain also appear to disrupt the FV structure. We discuss how the interactive database provides an online resource that clarifies the clinical understanding of FV-associated disorders.

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:

Christos Efthymiou, University College London, Structural and Molecular Biology, London, United Kingdom of Great Britain and Northern Ireland

Emily H Print, University College London, Structural and Molecular Biology, London, United Kingdom of Great Britain and Northern Ireland

Anna Simmons, 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

Regular Article (Category - Coagulation and Fibrinolysis)

Analysis of 363 genetic variants in F5 via an interactive web database reveals new insights into FV deficiency and FV Leiden

Christos Efthymiou[†], Emily H. T. Print[†], Anna Simmons[†], and Stephen J. Perkins[†]

[†] Research Department of Structural and Molecular Biology, University College London, London, United Kingdom.

Running title: Analysis of genetic variants in FV

Correspondence: Stephen J. Perkins, Department of Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, United Kingdom (e-mail: s.perkins@ucl.ac.uk).

ORCID

C. Efthymiou, <https://orcid.org/0000-0001-5612-7084>

E.H.T. Print, <https://orcid.org/0000-0002-8784-290X>

A. Simmons, <https://orcid.org/0000-0001-9118-2644>

S. J. Perkins, <https://orcid.org/0000-0001-9218-9805>

For purposes of manuscript review, the FV web site is available at <http://www.factorv-db.org> (username Christos; password Athens23). It will be publicly released after the manuscript is accepted for publication.

Abstract

The inherited bleeding disorder FV deficiency and clotting risk factor FV Leiden are associated with genetic variants in the *F5* gene. FV deficiency occurs with mild, moderate, severe, or asymptomatic phenotypes, and either dysfunctional or reduced amounts of plasma FV protein. Here we present an interactive web database containing 363 unique *F5* variants derived from 801 patient records, with 199 FV deficiency-associated variants from 245 patient records. Their occurrence is rationalized based on the 2224 residue sequence and new FV protein structures. The 199 FV deficiency variants correspond to 26 (13%) mild, 22 (11%) moderate, 49 (25%) severe, 35 (18%) asymptomatic, and 67 (34%) unreported phenotypes. Their variant distribution in the FV domains A1, A2, A3, B, C1 and C2 was 28 (14%), 32 (16%), 34 (17%), 42 (21%), 16 (8%) and 19 variants (10%) respectively, showing that these six regions contain similar proportions of variants. Variants associated with FV deficiency do not cluster near known protein-partner binding sites, thus the molecular mechanism leading to the phenotypes cannot be explained. However, the widespread distribution of FV variants in combination with a high proportion of buried variant residues indicated that FV is susceptible to disruption by small perturbations in its globular structure. Variants located in the disordered B domain also appear to disrupt the FV structure. We discuss how the interactive database provides an online resource that clarifies the clinical understanding of FV deficiency.

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

Introduction

The major disease associated with variants in Factor V (FV) of the blood coagulation cascade is FV deficiency. FV deficiency itself is a bleeding disorder, while FV Leiden is a risk factor for clotting, both of which are associated with genetic variants in the *F5* gene which codes for FV^{1,2}. The full-length human *F5* gene has 25 exons (OMIM #612309; Gene ID #2153) spanning approximately 80 kb and is located on chromosome 1 (1q24)³. The mRNA with 25 exons (Figure 1) contains 6672 bp, corresponding to a protein with 2224 amino acids and a molecular mass of 330 kDa. A 28-residue signal peptide is removed post-translationally in the endoplasmic reticulum, leaving a 2196 residue protein. FV is comprised of six major domains and 2 linkers in the order A1-A2-a2-B-a3-A3-C1-C2 (Figure 1)⁴. FV is homologous with coagulation Factor VIII, with their A and C domains having approximately 40% sequence identity³. FV is primarily synthesised by the liver whereby it is released into the bloodstream in the inactive zymogenic form. It circulates in the blood and is composed of a heavy chain with domains A1 and A2, and a light chain with domains A3, C1 and C2⁵.

FV is a plasma glycoprotein that plays a key role in the coagulation cascade. Specifically, activated FV (FVa) is essential in the prothrombinase complex and thrombin amplification. This complex cleaves prothrombin into thrombin, aiding formation of the platelet plug that is critical in coagulation. FV can be activated either by activated Factor X (FXa) or thrombin. Activation results in the cleavage of FV at R709, R1018, and R1545 (legacy numbering R681, R990, R1517) to release the B domain from the middle of FV66. The FV heavy and light chains remain associated through a single calcium ion. Cleavage induces a conformational change, enabling FVa to be a cofactor for FXa to cleave prothrombin. FXa alone has minimal affinity for the substrate prothrombin. However, once activated, FXa complexes with its cofactor FVa to form the prothrombinase complex which has high prothrombin affinity. Prothrombin binds an exosite on the protease complex and is

cleaved at two distinct sites, Arg271 and Arg320. Thrombin is subsequently released and facilitates platelet plug formation, preventing excess blood loss. A negative feedback loop controls this process, where thrombin induces the generation of negative control proteins to prevent excess blood clotting. Activated protein C (APC) degrades FVa, thereby preventing additional conversion of prothrombin to thrombin^{3,6}.

Genetic variants within the *F5* gene are predominantly associated with FV deficiency, for which we report 199 variants. From a clinical point of view, it is considered an autosomic recessive disease. FV deficiency can be classified as Type I or Type II. Type I deficiency is quantitative and defined by low coagulant activity (FV:C) and antigen levels (FV:A). Type II deficiency is qualitative and characterized by low FV:C even with normal FV:A⁷. FV:C in normal human plasma ranges from 60-140 IU/dL⁸. The phenotypic classification is determined by the FV:C level with severely, moderately, and mildly deficient individuals having FV:C levels of <1%, 1-5%, and 6-10% of normal⁹. Patients with FV deficiency suffer from reduced coagulation and an increased risk of uncontrolled bleeding due to reduced FV:C levels. Individuals with severe FV deficiency are typically homozygous or compound heterozygous for causative mutations, whereas individuals who are partially deficient are heterozygous, with one mutated allele. In contrast, FV Leiden disorders are less frequent, for which we report 37 variants. FV Leiden thrombophilia is linked with APC resistance, whereby FV variants render it less susceptible to inactivation by APC. In particular, APC cleaves FVa at Arg residues R334, R534, R707 (legacy R306, R506, and R679) to inactivate it; mutations in these residues resist inactivation by APC.⁸ This results in higher-than-normal coagulation due to increased thrombin production.⁹

Individuals with FV deficiency experience variable bleeding symptoms, the severity correlating with the level of protein activity. Individuals with lower FV:C will experience more severe symptoms than those with higher protein activity. Symptom onset varies, though severely affected individuals suffer from an earlier age. Mildly affected individuals experience minor symptoms such as easy bruising, nosebleeds, gum bleeding, and haematuria. Such individuals are also prone to prolonged and excessive post-operative bleeding. Severely affected individuals suffer from more severe bleeds including haemarthrosis, intramuscular bleeds, and haematomas. Additionally, such individuals are more susceptible to melena, bleeding in the gastrointestinal tract, and intracranial haemorrhages. Females suffering from FV deficiency often experience menorrhagia. Such individuals are also at greater risk of pregnancy complications including miscarriage and post-partum haemorrhage. Current treatments include antifibrinolytics, fibrin glue, and fresh frozen plasma (FFP) which promote blood clotting and reduce bleeding¹². Those with FV Leiden often do not experience symptoms. However, this polymorphism is associated with a weak risk for venous thrombosis. Symptoms depend on the location of the blood clot. In contrast to FV deficiency, an FV Leiden patient who suffers from a vein thrombosis usually needs anticoagulants (more frequently now direct oral anticoagulants rather than coumarins).⁸

The development of the first interactive searchable web databases for genetic variants in the coagulation proteases Factor XI <https://www.factorxi.org/> in 2003 and Factor IX <https://www.factorix.org/> in 2013 paved the way for a more powerful clinical tool to analyse FV variants. These interactive databases allow variants to be assessed alongside the gene and protein sequences, clinical phenotypes, and known three-dimensional protein structures. The Factor XI web site was recently upgraded¹³, and a new one developed for the coagulation protease Factor X <https://www.factorx-db.org/>¹⁴. Based on these, we now report to our

knowledge the first interactive database for FV variants. An earlier FV database is no longer available^{14,13}. Genetic repositories containing lists of variants include the Genome Aggregation Database (gnomAD; <https://gnomad.broadinstitute.org/>¹⁶), the Leiden Open-source Variation database (LOVD; <http://www.lovd.nl/3.0/home>¹⁷), the Expert Protein Analysis system (ExPASy; <https://www.expasy.org/>¹⁸), the ClinVar resource (<http://www.ncbi.nlm.nih.gov/clinvar/>¹⁹), and the public release of the Human Gene Mutation Database (HGMD; <http://www.hgmd.cf.ac.uk/ac/index.php>²⁰). However, as the number of known variants has increased, simple listings of FV variants are impractical and overwhelming. Therefore, an interactive database for FV variants is needed.

For the new FV database, we identified 363 unique FV variants, 199 being associated with FV deficiency, and the remainder being associated with FV Leiden or having an unknown association (Figure 2). F5 variants suggested to have a prothrombotic phenotype are collectively referred as FV Leiden. Our database also includes recent cryo-EM structures that reveal the molecular structures of the five globular A and C domains in FV and FVa for the first time (PDB IDs: 7KVE, 7KXY)²¹. In addition, a full-length protein model for FV that includes the disordered B domain structure was predicted by the AlphaFold neural networking and machine learning method²². The assembly of these data sets provides insights on the disease mechanisms for FV deficiency. We discuss the mild, moderate, severe, and asymptomatic phenotypes, and demonstrate that disease-associated missense variants were distributed evenly across the globular/disordered and functional/non-functional regions of FV. Therefore, the majority of the variants affect the structural integrity of FV and its protein folding. Clinicians and researchers can use the database as a tool to clarify the significance of the FV variants in FV deficiency.

Methods

Source of the FV Database

A total of 363 unique *F5* variants were identified for the interactive FV web database (<https://www.factorv-db.org>) from 801 patients. There was a focus on the 199 FV deficiency-associated variants from 245 patient records. The FV database was formatted to follow the layout of our previous FIX, FX and FXI databases (<https://www.factorix.org>; <https://www.factorx-db.org/>; <https://www.factorxi.org/>)^{13,14,23}. Copyright of the FV website is retained by S. J. Perkins and University College London, and database copying is not permitted without explicit permission. The 363 FV genetic variants were derived from online literature searches of peer-reviewed articles, primarily those in PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) and the ClinVar resource (<http://www.ncbi.nlm.nih.gov/clinvar/>). The cut-off date was December 2021 for the current database, and the sources are listed in the 'Reference' tab of the FV web database. For the FV web database, data for each variant were compiled into an excel spreadsheet, which was then used to establish the FV MySQL database using phpMyAdmin software (<https://www.phpmyadmin.net/>) as an intermediary platform. The database is maintained on a University College London server and utilises HTML, PHP, and JavaScript programming to enable public access via the web. If required for personal or private research use only, a list of the FV variants and their associated fields can be downloaded from the 'Variants' menu on the website (Supplementary Materials).

Analysis of the FV variants

The FV database records DNA and protein changes in the Human Genome Variation Society (HGVS) format (<http://varnomen.hgvs.org/>)²⁴. For DNA changes, +1 refers to the A of the ATG initiation codon at the start of the 28-residue pre-leader signal peptide. For

protein changes, +1 refers to the ATG initiation codon. In older publications, legacy numbering was used. For these publications, the legacy numbering for reported amino acid changes was converted to HGVS format by adding 28 to the original legacy number (+1 in legacy numbering refers to the first codon of the mature inactivated FV protein and corresponds to +29 in HGVS numbering). Nine human FV and FVa crystal and cryo-EM structures are available in the Protein Data Bank (PDB) at <https://www.rcsb.org> and these are listed on the FV website. Of these nine structures, six crystal structures are only for partial FV structures. Only the three cryo-EM structures show all five main A1-A3 and C1-C2 domains together (PDB IDs: 7KVE (FV with a resolution of 0.33 nm), 7KVF (FV with a resolution of 0.36 nm), and 7KXY (FVa with a resolution of 0.44 nm)). The 7KVE structure was used for variant analyses in this study as it has the highest resolution. However, no FV structures exist in which the large B domain is resolved as the B domain structure is structurally disordered and cannot be observed even with cryo-EM²⁵. To perform the structural analysis of the disordered B domain, an AlphaFold-modelled FV structure was used (<https://alphafold.ebi.ac.uk/entry/P12259>). AlphaFold is an artificial intelligence system based on neural networks that predicts protein structures based on their amino acid sequence²².

The Definition of Secondary Structure of Proteins (DSSP) tool (<https://www3.cmbi.umcn.nl/xssp/>) was used to identify the secondary structure of each residue in the FV cryo-EM structure (PDB ID: 7KVE)²⁶. Residues that were not visible in the cryo-EM structure were replaced by the AlphaFold structure prediction. Residues were individually assigned secondary structures to be one of either H (α -helix), B (β -bridge), E (extended β -strand), G (310 helix), I (π -helix), T (hydrogen-bonded turn), S (bend) or C (undefined coil region). DSSP was also used to determine the exposed surface areas of each

residue of the FV structure in Å². The accessibilities were converted into percent accessibility by dividing the DSSP output by the theoretical solvent accessible surface area of the amino acid sidechain in question²⁶⁻²⁸. 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.

To assess the disruptive effect of a given variant on the FV protein structure, four independent substitution analyses were carried out on the FV missense variants. These were Polymorphism Phenotyping v2 (PolyPhen-2) <http://genetics.bwh.harvard.edu/pph2/>; Sorting Intolerant From Tolerant (SIFT) https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html; Protein Variation Effect Analyzer (PROVEAN) http://provean.jcvi.org/seq_submit.php; and Grantham analysis, using the Grantham matrix from 1974²⁹⁻³². Both the SIFT and PolyPhen-2 algorithms give variant scores ranging from 0.0 to 1.0. For the PolyPhen-2 algorithm, prediction scores closer to 1.0 indicate variants that are more likely to be damaging. The opposite is true for SIFT where prediction scores closer to 0.0 indicate variants that are more likely to be damaging. Unlike SIFT and PolyPhen-2, the PROVEAN algorithm generates prediction scores that are valued against a threshold; variants with scores above -2.5 are considered neutral and those below -2.5 deleterious. Lastly, the Grantham analysis differs from the other three because its prediction scores are not sequence specific but are instead based on the amino acids which were substituted. Grantham scores range from 0 to 215, with a larger score indicating a variant substitution that is more likely to be damaging. Variants with a Grantham score of 0 are silent and have no effect on protein structure and function.

Results

Classification of FV deficiency variants in the updated interactive web database

The interactive FV web database (<https://www.factorv-db.org>) presents information regarding 363 unique genetic variants (Figure 2) from 801 patient entries. The variants were identified from 135 research articles in PubMed (<https://www.ncbi.nlm.nih.gov/pmc/>) and the ClinVar resource (<https://www.ncbi.nlm.nih.gov/clinvar/>) that presented another 123 unique variants or 34% of the total. The database home page features two movies of the monomeric six-domain FV AlphaFold predicted structure and the five-domain FV cryo-EM structure, both of which visualised the spatial distribution of the variants. Users can perform simple searches on the website by amino acid or nucleotide number, or perform an advanced search based on several criteria including variant type, effect, domain location, reference, etc. The database has interactive capabilities derived from our recently upgraded FIX and FX websites (<https://www.factorix.org>; https://www.factorx_db.org), including a site map which facilitates user navigation^{13,14,23,33}. The site map displays the entire FV amino acid sequence, this being highlighted to mark the boundaries of domains and linkers (Figure 1). Any known missense variants are linked in the sequence and accessed by clicking on the corresponding amino acid. The genome aggregation database (gnomAD) version 2.1.1 (<https://gnomad.broadinstitute.org>) was used to provide allelic frequencies (AF) for FV variants in the database when possible. The gnomAD v2.1.1 data set included 125,748 exome sequences and 15,708 whole-genome sequences. Thus 179 (49.3%) of the 363 identified FV variants were found in gnomAD. The AF indicates the relative frequency of a variant at a specific genetic location. Using an AF cut-off of 0.01, an AF > 0.01 indicated a commonly occurring variant, and only 21 of the 179 showed an AF > 0.01. The remaining 158 variants showed an AF < 0.01, underscoring that most FV variants are rare ones. Using a stricter AF cut-off of 0.001, only 13 additional variants occurred more frequently than this, leaving 145 FV variants with known AF as rare ones from gnomAD. Additional information is provided

for missense variants through the Polyphen-2, SIFT, PROVEAN, and Grantham scores (Methods), all of which predict if the variant is damaging to FV or benign. Other features include a multiple sequence alignment of human FV with FV from other species, helping users to understand the phylogenetic history of the *F5* gene and the extent of residue conservation in related sequences for each variant. Java applets must be enabled within the browser to permit the mutation map, the multiple sequence alignment, and other related features in order to be seen. An interactive FV structure is presented in a Jmol viewer onto which missense variants can be mapped and viewed for a clearer structural and functional assessment. Users are able to switch between the three cryo-EM structures and the AlphaFold full-length protein model. Lastly, all nine currently known FV structures and their literature references are listed on the website to provide an up-to-date knowledge of FV structural research.

Compared to the FV deficiency variants in the database, there were fewer FV Leiden/thrombosis variants, thus a database comparison between these two sets of variants was not straightforward. In addition, the ClinVar variants in the database did not mention the associated FV variant. Accordingly, we focus only on the 199 FV deficiency-associated variants. These were classified using the residual FV activity (FV:C) into asymptomatic (>10%), mild (6-10%), moderate (1-5%) and severe (<1%) disease based on a normal activity range of 60-150 IU/dL. Of the 199 variants, 26 (13%) were mild, 22 (11%) were moderate, 49 (25%) were severe, 35 (18%) were asymptomatic, and 67 (34%) were unreported phenotypes. Classifying by the type of genetic change, point variants comprised 73.9%, deletions 21.1%, duplications 3.0%, insertions 1.5%, and indels 0.5% (Figure 2(a)). Point variants can be classified into missense, nonsense, silent, and intronic variants, corresponding to 68.0%, 17.7%, 2.0%, and 12.2% of point variants respectively (Figure 2(b)). The 199

unique variants were distributed throughout the *F5* gene and FV protein structure, with variants found in all six A, B, and C domains, as well as the $\alpha 2$ linker (Figure 2(c)). Point variants found in more than 5 patients occurred more often in the A and C domains compared to the disordered B domain (green, Figure S1).

The distribution of amino acid variants in FV was assessed by a substitution analysis of the 100 FV deficiency missense variants. Only single nucleotide mutations were observed. The dark greyed boxes of Figure S2 signify two nucleotide changes for a given variant, of which there were none in *F5*. The total count for each missense change showed that Arg (18), Ser (11), Cys(9), Asp(7), and Leu (6) were the five most commonly affected residues (right column, Figure S2). The most common resulting substitution was Gly (15) (bottom row, Figure S2). The predominance of positively charged Arg and negatively charged Asp missense variants suggested that ionic interactions were important for the correct function of FV. The prevalence of the polar uncharged residue Ser and the hydrophobic residues Leu and Val suggested that the internal FV globular protein structure had been disturbed by protein misfolding mutations. Cys residues are important for forming two disulphide bridges in each of the A1 and A2 domains for stabilization, and likewise one disulphide bridge in each of the A3, C1, and C2 domains, explaining the prevalence of Cys residue mutations (Figure S1).

Cryo-EM and AlphaFold structure analysis of secondary structures

The FV protein structure enabled the three-dimensional distribution of the variants to be visualised and clarified the molecular basis of FV deficiency. In our database, three molecular structures were used for structural analysis of genetic variants. These included the two new cryo-EM structures for FV and FVa (PDB IDs: 7KVE and 7KXY respectively), which both lack the disordered B domain, and the AlphaFold full-length model of FV that

includes a model of the B domain^{21,22}. When the FV structure (7KVE) was subjected to Ramachandran plot quality analysis (<http://molprobity.biochem.duke.edu>³⁴), 81.4% (1115 residues) of amino acids were categorised in the “most favoured” conformational regions, 18.3% (251) were in the “additional allowed” regions, and 0.3% (5) were conformational outliers. This outcome was superior when compared to the FVa (7KXY) structure for which the corresponding figures were 69.8%, 26% and 4.2% (805, 300, and 48 residues out of a total of 1153) respectively. The AlphaFold prediction for full-length FV was based on new structural methods for accurate predictions using artificial intelligence²¹²⁰. Although AlphaFold is an improvement on previous structural prediction methods, it is difficult to assess the quality of the outcome. However, AlphaFold provides a score for each residue called the predicted local-distance difference test (pLDDT), which measures the local reliability for each residue from 0 to 100. For FV, the majority of the residues in the five globular domains have scores above 70, indicating high confidence in the prediction. However, nearly all of the B domain residues have scores below 50, indicating a low confidence in the prediction. This is expected because the AlphaFold confidence scores are correlated with the existence of homologs in the PDB³⁴³³. Since the B domain structure has not been experimentally solved in FV, nor in its homologue human Factor VIII, the structure prediction of the B domain is essentially arbitrary, even though this accounts for the presence of the B domain in FV. The Ramachandran analysis showed that, for the full FV sequence of 2224 residues, 71.6% (1592 residues) of amino acids were categorised in the “most favoured” conformational regions, 12.1% (269) were in the “additional allowed” regions, and 16.3% (362) were conformational outliers. Overall, the quality of the AlphaFold FV structural model was comparable to those of the cryo-EM structures, with more conformational outliers present as the result of the disordered nature of the B domain. Regardless, the AlphaFold model is advantageous as it displays the whole FV protein structure. From this, the B domain

was predicted to have a disordered outermost structure that was wrapped around the central core of the five A1-A3 and C1-C2 domains (Figures 3(a-b)).

The AlphaFold FV structure displays all 100 FV deficiency missense variants found in FV, these being colour coded according to the phenotype for each variant (Figure 3(c)). Green, yellow, and red spheres corresponded to mild, moderate, and severe cases, and were distributed throughout the six FV domains, similarly to the purple asymptomatic cases. These locations suggest that the whole FV structure is susceptible to the damaging effects of genetic variants, including the disordered B domain. In fact, the four most common missense variants (Figure 3(d)), occurred in four different globular domains of FV, as well as the next 25 most common ones. Visualizing the missense variants for each individual domain demonstrated an even distribution of variants across all structures, including in the disordered B domain. This outcome underscored the effect of the variants on disrupting the overall FV protein structure rather than disturbing localised functionally important regions (Figure 4).

Comparing the actual number of missense variants in each domain to the normalised number based on the size of each domain showed that slightly more variants occurred in most of the globular A and C domains and fewer in the A3 and disordered B domain (Figure 5). The percent difference between actual and normalised variants was 57%, 25% and -68% in the A domains, and 14% and 43% in the C domains. The percent difference between actual and normalised variants in the B domain was -67%. The lower proportion of variants in the B domain was attributable to its disordered structure, where a residue change was less likely to perturb the B domain structure. In contrast, residue changes in the globular A and C domains were more likely to affect protein folding. Interestingly, the B domain showed more deletion variants occurring in this compared to the A and C domains (Figure 5(b)). This could again

be attributed to its disordered structure being more tolerant of traditionally highly damaging variants.

Comparison of the individual variants with the FV protein structure

The identification of a missense FV variant in a patient gene cannot be assumed to cause an FV disorder. Additionally, with 100 unique FV deficiency-associated missense variants, it was not feasible to experimentally test individual variants via recombinant expression to determine if the variant damages FV function or folding. To circumvent this issue, the FV website predicted whether an individual variant was likely to disrupt protein structure and function. Four independent algorithms were used. The PolyPhen-2 analysis predicted that, of the 100 missense variants, 75 (75%) gave scores between 0.9 to 1.0 and were assigned to be damaging, but that 25 (25%) were predicted to be benign variants (Figure 6(a)). The SIFT analysis predicted that 92 (92%) of the missense variants were damaging (Figures 6(b)). Therefore, most FV deficiency-associated missense variants are predicted to be damaging. The PROVEAN and Grantham analyses regarding potential damaging effects were less clear, with 95 (95%) PROVEAN and 88 (88%) GRANTHAM scores being deleterious in a wider distribution of scores. Regardless, both the PROVEAN and Grantham analyses also demonstrated that many variants were damaging (Figures 6(c,d)). The interactive FV database provided all four PolyPhen-2, SIFT, PROVEAN, and Grantham scores for each of the 100 missense variants, in order to assist clinicians to make an informed decision about the significance of a given variant.

The relationship between FV phenotype and amino acid surface solvent accessibilities in the FV protein structure for each variant provided further information on the effect of each variant. Residue surface accessibilities were calculated using the DSSP tool for the intact FV

protein and the six individually separated FV domains, then graphically compared with the reported phenotype for the variants. For the intact FV protein, 80% of the mild variants (8 of 10 cases) had low accessibilities of 0 or 1, indicating these were mostly buried within the FV structure (Figure 7(a)). Similarly, for the moderate and severe phenotypes, 86% (32 of 37) and 84% (38 of 45) of the variants also had low accessibilities of 0 or 1. For the separated FV domains, the accessibilities showed a similar trend whereby 70% (7 of 10), 86% (32 of 37), and 64% (29 of 45) of the variants for the mild, moderate, and severe phenotypes had low accessibilities of 0 or 1 (Figure 7(b)). The predominance of damaging variants at buried sites, such as those in interdomain contacts, illustrates that small perturbations in the FV structure were sufficient to reduce normal protein function and become disease-associated.

The four most common missense variants (Figure 3(d)) were visually highlighted to illustrate the effects caused by these amino acid changes (Figure S3). The patient list shows that these four possessed a mix of all phenotypes. In keeping with Arg being the most commonly mutated residue in the FV variants (Figure S2), one of the most common amino acid mutants was an Arg residue, with a Gly, Tyr, and Asp being the other three most common (Gly420, Tyr1730, Arg2102 and Asp96; HGVS), with accessibilities of 2, 0 and 1 and 0 respectively (Figure S2). These occurred in 9-21 patients and were mutated to Cys, Cys, Cys and His residues (Figure S3). The Tyr1730, Arg2102 and Asp96 (HGVS) variants were all buried within the globular FV structure as evidenced by the low accessibility scores of 0 or 1. Three variants were mutated to Cys residues, which will present an unpaired Cys residue that may interfere with the disulphide bridge pairings that stabilize the wildtype FV structure. The Asp96His (HGVS) mutation converts a negatively charged residue to a positive charge, most likely to interfere with critical internal stability contacts within FV, particularly between the A1 and C2 domains (Figure S3(d)). Gly420 (HGVS) has a higher

accessibility score of 2; the visualization of this residue in Figure S4(a) supports this finding. In this case, the conversion of a positively charged residue to a polar uncharged residue may interfere with the stability of the A2 domain or a potential binding with the disordered B domain.

Discussion

Here, our new interactive genetic and structural web database for human FV will facilitate improved analyses of FV genetic variants for clinicians, as well as further insights into the disease mechanism leading to FV deficiency. It was not feasible to analyse the FV Leiden variants here for reason of their reduced frequency. Given the high growth in newly-reported FV variants, a simple flat listing of these is now overwhelming to use, and is unable to provide clinicians with useful information. Our interactive FV database is a powerful resource and has two predominant advantages: (a) the number of unique FV variants totals 363 variants, this being accumulated from the published literature and ClinVar (Figure 1); (b) the new cryo-EM molecular structures for FV and FVa²¹ and the full-length AlphaFold prediction for FV2120, enable the FV variants to be assessed in the context of the full three-dimensional structure of FV. The variants compiled in the database are distributed throughout the six globular and disordered FV domains. For all the missense variants, the PolyPhen-2, SIFT, PROVEAN, and Grantham analyses indicate that most variants are likely to be damaging to the FV protein. Interestingly, many detected variants are located in buried regions of low solvent accessibility (Figure 7). This outcome suggests that the alteration and perturbation of residues inside each FV domain are a major cause of FV deficiency due to disruption of the FV structure. While mutational hotspots in functionally important regions of FV can be important, these are not the dominant cause of FV genetic disease.

Several regions of FV/FVa are critical to its function and should be considered in the context of identified variants. Four major FVa ligands have been identified, namely FXa, thrombin, polyanionic membranes, and APC.

(i) First, we consider the FV regions involved in binding to FXa, labelled as BS1 to BS4 in Figure S4(a) in that order below. One possible site BS1 for FVa binding to FXa is the Arg334 region (legacy Arg306) in the A2 domain that involves either residues 339-353 in FV (legacy 311-325)³⁶ or a larger section in residues 335-376 (legacy 307-348)³⁷. In our dataset of 199 unique FV deficiency variants, mutations were found in 14% of the residues in the larger proposed binding region. Another region BS2 for the FXa binding site is the Arg534 region (legacy Arg506), with a binding cluster involving the residues Arg529, Arg534, Arg538, Ala539, Asp541, Asp605, and Asp606 (HGVS)³⁸. We identified variants only in residue Arg534, although several variants in close proximity to the binding residues were found. However, there is evidence that the Arg334 and Arg534 regions may bind prothrombin rather than FXa, or both, but this remains unclear³⁹. A third FXa binding site BS3 has been suggested as the A2 domain C-terminus (A2T) from residues 711-737 (legacy 683-709)⁴⁰. We identified variants only in residue 717 in this binding region. A fourth region BS4 for FVa binding to FXa has been suggested in the A3 domain, spanning a large portion of the domain from residues 1565-1780 (legacy 1537-1752)³⁹. We identified variants in 10% of the amino acids in this binding region.

(ii) Next, we consider FVa binding to prothrombin. The first binding region on FV for prothrombin is around regions Arg334 and Arg534 in the A2 domain (Figure S4(a); sites BS1 and BS2), as mentioned above. The A2T binding site implicated in FXa binding also may play a role in binding prothrombin, with a variant at S717 being identified in our database. However, contradictory evidence exists in regards to the validity of this region in binding

prothrombin. The light chain of FV (domains A3, C1 and C2) does not appear to be involved in prothrombin binding³⁹.

(iii) The third key interaction is the binding of FVa to anionic membranes (sites BS1 and BS2 in Figure S4(a)), which is critical for formation of the ternary prothrombinase-prothrombin complex involving FVa, FXa, and prothrombin. Residues 1982-1985, 2051, and 2055 (legacy 1954-1957, 2023, and 2027) in the C1 domain form hydrophobic “spikes” which help anchor FVa to the anionic membrane³⁹. We did not identify any variants in this binding region. The C2 domain also plays a role in binding anionic membranes, with the N-terminal region from 2065-2115 (legacy 2037-2087) containing hydrophobic and electrostatic interactions which aid binding⁴¹. We identified variants in 18% of the residues in this region.

(iv) A final critical binding interaction is APC binding to FVa to inactivate it and regulate the coagulation process. The binding site for APC to FVa is not known⁴². However, it is well-established that APC cleaves FV at R334, R534, and R707 (legacy R306, R506, and R679). We identified variants in residue 534, indicating the importance of this residues for proper regulation (Figure S4).

In summary from the above, variants have been identified in up to 18% of the residues comprising each of these six functionally important regions. Overall, for the 185 unique variants found in the exonic region of *F5*, only 38 or 20% correspond to variants in any of the functional regions of FV. These functionally important regions comprise 15% of the total number of 2224 residues in FV. This outcome indicates that 80% of identified variants are found in the remaining 85% residues of FV which lie outside the functional regions. It is concluded that mutation rates were not higher in the functional portions of FV. Interestingly, for the four most commonly mutated residues according to number of patients (Asp96, Gly420, Tyr1730, and Arg2102 (HGVS)), two occur within functional regions of FV.

Overall, this distribution of variants in FV in both the functional regions and other regions distant from functional areas show that both are capable of disrupting the overall FV structure and cause disease. Therefore, FV function is susceptible to disturbances in both its globular or disordered structures.

The majority of variants in our database were associated with FV deficiency, and these variants were found in the full sequence of residues 15-2222 in FV. Therefore, FV deficiency is associated with variants found across the length of FV rather than just in functionally active regions. The changes to protein structure due to these variants inactivate FV, preventing conversion of prothrombin to thrombin and increasing the risk of bleeding. The database is able to assess whether a variant is significant in terms of its potential effect on the protein structure from the four scoring methods and its solvent accessibility (Figure S5). While not analysed in detail, FV Leiden or thrombosis was also associated with variants that occur across the full length of the FV sequence.

FV has a unique protein domain structure in comparison to most of the coagulation proteins. It features globular A and C domains in a compact arrangement, these being surrounded by a disordered B domain (Figure 3). FV shares a similar domain structure with Factor VIII. Other coagulation proteins vary in their three-dimensional structure. For example, Factor XI possesses a compact five-domain globular structure whereas Factors IX and X have extended four-domain structures. Regardless of these structural differences, variants were distributed throughout each of these proteins, suggesting these variants disrupt the protein folding, thereby damaging function. The same tendency for variants to be distributed across the domain structures rather than localized to functional hotspots was present in these different proteins. For FV, the globular (A1-A3, C1-C2) and disordered (B)

domains represent 58% and 35% of the 2224 residues in its sequence respectively, and possess 65% and 21% of FV deficiency variants in the protein. The evidence suggests FV function is susceptible to perturbations in its structure regardless of the location of a variant, even in the disordered B domain. It had been suggested that most missense variants in the B domain of Factor VIII are unlikely to be disease-associated as the B domain is unnecessary for secretion and function of the protein⁴³. Although the B domains of FV and Factor VIII do not show high sequence similarity, there is not enough evidence to show whether variants in the B domains of FV or Factor VIII are more or less likely to be causative of disease.

Following the established format of our FXI^{13,33}, FIX²³, and FX¹⁴ web databases, the FV website layout presents separate genetic and structural information in parallel with each other. The database will facilitate additional research into FV to understand better the relationship between FV deficiency and disease severity, based on the substantial number of FV variants and their new structural interpretation afforded by the cryo-EM structures and the AlphaFold prediction. Variants predicted to be damaging to the FV structure can be prioritized for experimental analysis to understand better how the structure is disrupted.

What is known on this topic:

(1) A large number of FV variants have been published in the literature and resources such as ClinVar, however no interactive web database dedicated to FV has been available.

(2) New structural models for FV now include the disordered B domain and recent cryo-EM structures for the A1-A2-A3-C1-C2 domains, thus enabling an assessment of the FV missense variants in terms of the full FV protein structure compared to earlier.

What this paper adds:

(1) As a clinical research tool, we summarise 363 unique FV variants from 801 case reports. For FV deficiency, 199 variants were associated with FV deficiency from 245 patient records, in our interactive web database, for which there are 26 mild, 22 moderate, 49 severe, 35 asymptomatic, and 67 unreported phenotypes. For the risk factor FV Leiden, 37 unique variants were reported.

(2) Variants were distributed evenly throughout the globular (A1-A3, C1 and C2) and disordered (B) domains without evidence of localised hotspots. The globular and disordered domains comprise 58% and 35% of the full FV sequence respectively, and contain 64% and 21% of the variants.

(3) The frequency of FV variants across its sequence suggests that FV deficiency results from the sensitivity of FV function to protein structural disruption, rather than the inhibition of its binding to other proteins.

Research grants and financial support:

The authors were supported by grants from the Lister Institute for Preventive Medicine for this work. The authors declare that they have no conflicts of interest with the contents of this article.

REFERENCES

1. Lippi G, Favaloro EJ, Montagnana M, Manzato F, Guidi GC, Franchini M. Inherited and acquired factor V deficiency. *Blood Coagul Fibrinolysis* 2011;22(3):160-166
2. Kujovich JL. Factor V Leiden thrombophilia. *Genet Med.* 2011;13(1):1–16
3. Duga S, Asselta R, Tenchini M L. Coagulation factor V. *Int J Biochem Cell Biol* 2004;36(8):1393–1399
4. Cripe LD, Moore KD, Kane WH. Structure of the gene for human coagulation factor V. *Biochemistry* 2002;31(15):3777–3785
5. Huang JN, Koerper MA. Factor V deficiency: a concise review. *Haemophilia* 2008;14(6):1164–1169
6. Petrillo T, van 't Veer C, Camire RM. Characterization of the interaction between tissue factor pathway inhibitor α (TFPI α) and Factor V species. *Blood* 2019;134(Supp 1):485–485
7. Abou Mourad Y, Shamseddine A, Hamdan A, Koussa S, Taher A. Report of a rare co-occurrence of congenital factor V deficiency and thalassemia intermedia in a family. *Ann Saudi Med.* 2004;24(4):301-302
8. Stefely JA, Christensen BB, Gogakos T, et al. Marked factor V activity elevation in severe COVID-19 is associated with venous thromboembolism. *Am J Hematol.* 2020;95(12):1522-1530.
9. Paraboschi EM, Menegatti M, Rimoldi V, Borhany M, Abdelwahab M, Gemmati D, Peyvandi F, Duga S. Profiling the mutational landscape of coagulation factor V deficiency. *Haematologica* 2020;105(4):e180-e185
10. van Cott EM, Khor B, Zehnder JL. Factor V Leiden. *Am J Hematol.* 2016;91(1):46-49
11. Merli, GJ, Martinez, J. Prophylaxis for deep vein thrombosis and pulmonary embolism in the surgical patient. *Med Clin North Am* 1987; 71(3):377-397

12. Tabibian S, Shiravand Y, Shams M, Safa M, Gholami MS, Heydari F, Ahmadi A, Rashidpanah J, Dorgalaleh A. A comprehensive overview of coagulation Factor V and congenital Factor V deficiency. *Semin Thromb Hemost* 2019;45(5):523-543
13. 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; 05:e543–e556
14. 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;05:e557-e569
15. Vos HL. Inherited defects of coagulation Factor V: the thrombotic side. *J Thromb Haemost* 2006;4(1):35-40
16. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581(7809):434–443

17. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 2011;32(5):557–563
18. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003;31(13):3784-3788
19. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46(D1):D1062–D1067
20. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeyasinghe S, Krawczak M, Cooper DN Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;21(6):577–581
21. Ruben EA, Rau MJ, Fitzpatrick JAJ, di Cera E. Cryo-EM structures of human coagulation factors V and Va. *Blood* 2021;137(22):3137–3144
22. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–589
23. 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(7):1329–1340

24. den Dunnen JT, Antonarakis SE. Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion. *Hum Mutat* 2000;15(1):7-12
25. Ruben EA, Rau MJ, Fitzpatrick JAJ, di Cera E. Cryo-EM structures of human coagulation factors V and Va. *Blood* 2021;137(22):3137-3144
26. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22(12):2577–2637
27. Touw WG, Baakman C, Black J, et al. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* 2015;43(Database issue):D364–D368
28. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum allowed solvent accessibilities of residues in proteins. *PLoS One* 2013;8(11):e80635
29. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7(04):248–249
30. 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(Web Server issue):W452-W457
31. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012;7(10):e46688
32. Grantham R. Amino acid difference formula to help explain protein evolution. *Science* 1974;185(4154):862–864
33. 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(03):192–198

34. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB 3rd, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* 2018;27(1):293–315
35. Jones DT, Thornton JM. The impact of AlphaFold2 one year on. *Nature Meth* 2022;19(1):15-20
36. Kojima Y, Heeb MJ, Gale AJ, Hackeng TM, Griffin JH. Binding site for blood coagulation factor Xa involving residues 311-325 in factor Va. *J Biol Chem* 1998;273(24):14900–14905
37. Kalafatis M, Mann KG. The role of the membrane in the inactivation of factor Va by plasmin. Amino acid region 307-348 of factor V plays a critical role in factor Va cofactor function. *J Biol Chem* 2001;276(21):18614–18623
38. Steen M, Tran S, Autin L, Villoutreix BO, Tholander AL, Dahlbäck B. Mapping of the factor Xa binding site on factor Va by site-directed mutagenesis. *J Biol Chem* 2008;283(30):20805–20812
39. Schreuder M, Reitsma PH, Bos MHA. Blood coagulation factor Va's key interactive residues and regions for prothrombinase assembly and prothrombin binding. *J Thromb Haemost* 2019;17(8):1229–1239
40. Bakker HM, Tans G, Thomassen MC, Yukelson LY, Ebberink R, Hemker HC, Rosing J. Functional properties of human factor Va lacking the Asp683-Arg709 domain of the heavy chain. *J Biol Chem* 1994;269(32):20662–20667
41. Ortel TL, Quinn-Allen MA, Keller FG, Peterson JA, Larocca D, Kane WH. Localization of functionally important epitopes within the second C-type domain of

coagulation factor V using recombinant chimeras. *J Biol Chem.* 1994;269(22):15898-15905.

42. Castoldi, E. FV and APC resistance: the plot thickens. *Blood* 2014;123(15):2288–2289

43. Ogata K, Selvaraj SR, Miao HZ, Pipe SW. Most factor VIII B domain missense mutations are unlikely to be causative mutations for severe hemophilia A: implications for genotyping *J Thromb Haemst* 2011;9(6):1183-1190

Figure 1: Distribution of the 363 unique variants identified within the F5 gene and FV protein. The FV protein is comprised of the A1-A2-a2-B-a3-A3-C1-C2 domains and linkers in that order, and is shown as grey boxes that are not drawn to scale. N and C represent the N- and C-termini, respectively. Residue numbering marks the first and last amino acids which frame each domain, reported in HGVS format (starting with 1 at the signal peptide). The number of variants in each domain is shown above each protein domain. The number of variants associated with FV deficiency (red), thrombosis or FV Leiden (blue), or have an unknown association (grey) are indicated. Intronic variants (19 in total), undefined or not applicable variants (9), and variants affecting multiple domains (2) are not shown. Below the protein domains, the gene arrangement of 26 exons is shown as alternating light grey and maroon boxes drawn to scale. The FV protein domain to which each exon codes for are indicated.

Figure 2: Distribution of the 199 unique FV deficiency variants found in the F5 gene.

The panels (a-c) indicate breakdowns of the 199 FV deficiency variants into variant type, effect, and location within the *F5* gene sequence.

(a) The relative frequency of six distinct types of unique variants in the *F5* gene.

- (b) Effect of the 147 point variants found in the *F5* gene sequence.
- (c) Distribution of the 199 FV variants across the *F5* gene and FV protein domains, although six undefined or not applicable variants are not shown.

Figure 3: Structural and schematic views of FV deficiency variants within the FV domains.

- (a) The full FV structure is shown in ribbon format from the AlphaFold prediction. The structure is shown in rainbow colours, with blue corresponding to the N-terminal region and red corresponding to the C-terminal region. The N-terminus and C-terminus are denoted by N and C, respectively.
- (b) The FV structure from (a) is shown schematically in cartoon form in the same orientation and colours. The globular A1, A2, A3, C1 and C2 domains are denoted by filled circles. The disordered B domain is schematically represented by a green line.
- (c) The 100 missense variants are mapped to the ribbon diagram, where the phenotype classifications of mild, moderate, and severe effects are denoted as the traffic light colours of green, yellow, and red, respectively. Missense variants with unreported phenotype are shown in grey and asymptomatic cases are shown in purple.
- (d) The 25 most commonly reported variants are shown as spheres in the ribbon structure of FV shown in (a). Blue spheres denote the fifth to twenty-fifth more common variants, and the four magenta spheres denote the four most common variants in FV (Figure S3 below).

Figure 4: The six individual FV domain structures and their 100 missense variants.

Each domain is shown as a ribbon diagram in rainbow colours from the N-terminus (blue) to the C-terminus (red) for clarity. The structurally similar A1, A2 and A3 domains are shown with their secondary structure ribbons depicted in the same orientations, and likewise the

structurally similar C1 and C2 domains. The blue spheres denote the missense mutations associated with FV deficiency in each domain: A1, 22 variants; A2, 20 variants; A3, 22 variants; B, 9 variants; C1, 8 variants; C2, 17 variants. These variants total 98, where the remaining 2 variants (not shown) occur in the signal peptide (1 variant) and the $\alpha 2$ linker (1 variant).

Figure 5: Distribution of FV deficiency variants in the six FV domains.

- (a) The number of missense variants in each of the six FV domains is shown above the red bars. If the number of the 98 missense variants is normalised in proportion to the amino acid residues present in the sequence of each domain, the outcome is shown as blue bars. The 2 missense variants that occur in the linker region and the signal peptide are not shown.
- (b) The distribution of 42 variants in the B domain (green) is compared against the five types of the 157 genetic variants that occur in FV (orange).

Figure 6: Substitution analysis of 100 FV deficiency missense variants in the *F5* gene.

The four substitution analyses predict the damaging effects of substitution variants on the protein structures.

- (a) Analysis of the variants determined by their PolyPhen-2 scores.
- (b) Analysis of the variants determined by their SIFT scores.
- (c) Analysis of the variants determined by their PROVEAN scores. The PROVEAN threshold used was -2.5.
- (d) Analysis of the variants determined by their Grantham scores.

Figure 7: Accessibility analyses of 100 missense variants associated with FV deficiency in the FV protein structure.

- (a) Accessibilities are presented for the full protein. FV variants in the intact cryo-EM structure (PDB ID: 7KVE) and the full-length AlphaFold structure (Figure S1) were grouped by their phenotypic classification (SEVERITY). The variants were further subdivided according to the native residue accessibility (ACC) of the intact protein. The accessibility was determined using DSSP (Figure S1). Accessibilities of 0 or 1 indicate sidechain burial and values of >1 indicate sidechain exposure to solvent.
- (b) Accessibilities are presented for the six separated domains. FV variants were again grouped by phenotypic classification (SEVERITY) and accessibility (ACC). Here, accessibility refers to the change in residue accessibility when the cryo-EM FV structure (PDB ID: 7KVE) or the AlphaFold structure (Figure S1) were separated into the six A1, A2, A3, B, C1, and C2 domains. Interdomain linkers are not included in this comparison.

SUPPLEMENTARY FIGURES

Figure S1: Secondary structure and accessibility analysis of variants occurring in the FV protein. The FV amino acid sequence is shown with secondary structure assignments and solvent accessibilities indicated below each residue. The residue secondary structures are denoted as 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). Secondary structures and accessibilities based on a defined cryo-EM structure are marked in black, and those without a cryo-EM structure are marked in blue. For these latter residues, secondary structure and solvent accessibility predictions were made based on the AlphaFold structure. The positions of 285 point variants that occur in the exonic region of the *F5* gene are highlighted in boxes. These include point missense, point nonsense, and point silent variants. Yellow boxes denote

point variants that occur in 4 patients or less, green boxes denote point variants that occur in 5 patients or more, and red boxes denote point variants that occur in over 50 patients. Post-transcriptional modifications are shown. These include seven Cys-Cys disulphide bridges (highlighted in blue) and seven sulphated Tyr residues (highlighted in grey).

Figure S2: Substitution grid representing 100 point missense FV deficiency variants in the *F5* gene. The grid presents the number of missense variants that occurs for each defined amino acid change. 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 dark grey, however none were seen. Silent variants are excluded from the grid and shown in pale grey. White boxes represent possible substitutions that do not occur in the *F5* gene. Yellow boxes represent substitutions that occur between one and five times.

Figure S3: Molecular graphic representation of four residues within the FVa protein structure. The four panels highlight in the FVa cryo-EM structure the four most-commonly occurring variants found in patients (Figure 3(d)). For each, the residue of interest in the native FVa protein are shown. Panels (a-d) highlight the wildtype residues Gly420 (A1 domain; 10 patients), Tyr1730 (A3 domain; 21 patients), Arg2102 (C2 domain; 9 patients), and Asp96 (A1 domain; 11 patients). All residue numbering is given in HGVS format.

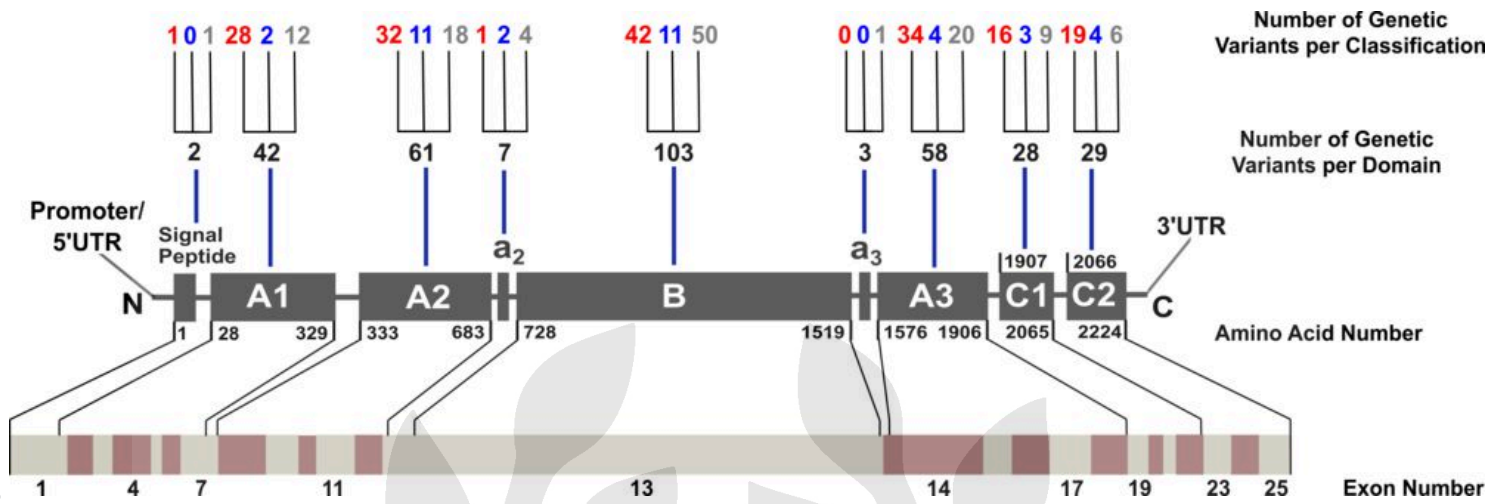
Figure S4: FV protein structure with labelled FV deficiency missense variants and known protein contact regions.

(a) The full FV structure is shown in ribbon format from the AlphaFold prediction using artificial intelligence. The structure is shown in rainbow colours, with blue corresponding to the N-terminal region and red corresponding to the C-terminal region. The N-terminus and C-

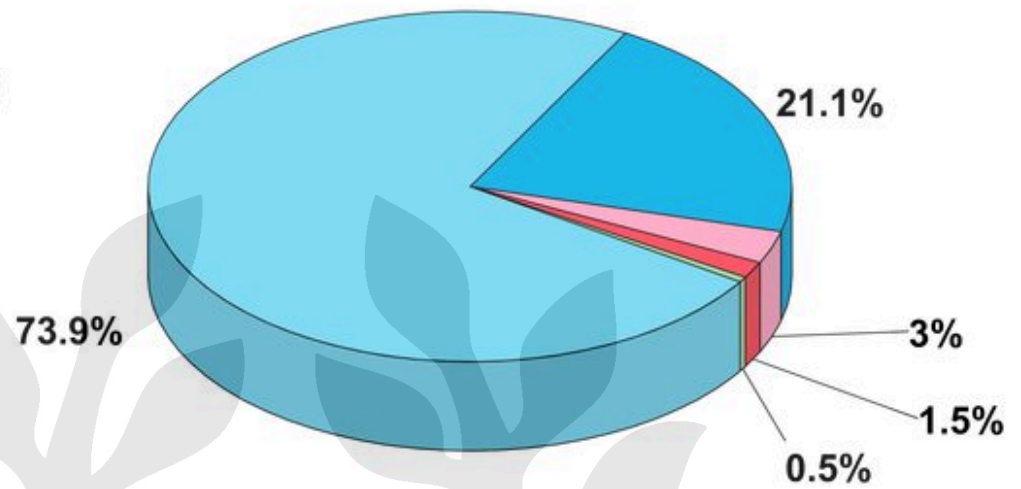
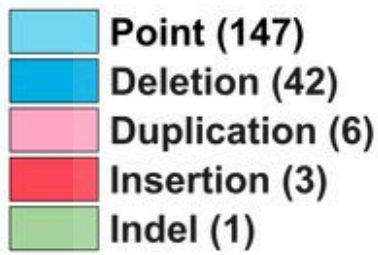
terminus are denoted by N and C respectively. The 100 missense variants are mapped to the ribbon diagram, where the phenotype classifications of mild, moderate, and severe effects are denoted as the traffic light colours of green, yellow, and red. Missense variants with unreported phenotype are shown in grey and asymptomatic cases are shown in purple. The dotted lines denote the regions of FV (FVa) that interact with coagulation proteins FXa and prothrombin, as well as the anionic phospholipid cell membrane. Multiple binding sites (BS) for the same factor are numbered accordingly. Note that the APC binding site is not indicated as it is not known.

(b) The full FV structure is shown in the same format as (a) except the main chain has been coloured grey and the binding sites were coloured red, orange, yellow, green, blue, and violet to highlight these better.

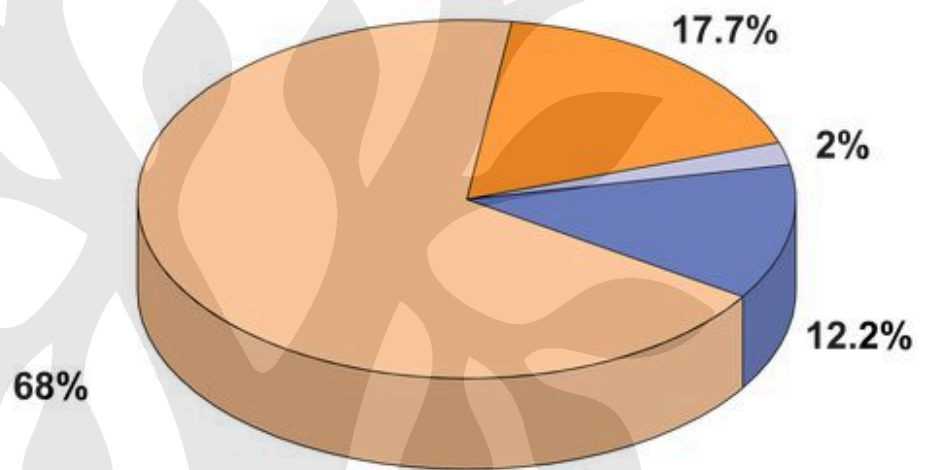
Figure S5: Screenshots of the FV web site to illustrate the analysis made for the His175Arg variant. The upper panel displays the output when the His175 residue is inputted on the home page of the interactive web site. By clicking “Show” on the patient information, the lower left panel lists genetic information for the 8 patients reported with the His175Arg variant together with the source of the patient record. Clicking “HERE” on the structural interpretation gives the image shown on the bottom right panel. This assesses the buried or exposed accessibility of the variant and its location in the FV protein structure. A JMol view of the FV structure is displayed that can be rotated and zoomed into as desired; users also can switch between FV structures. The Grantham, Polyphen-2, SIFT, and PROVEAN substitution analyses to predict the damaging effects of each missense variant are provided to facilitate clinical diagnosis.



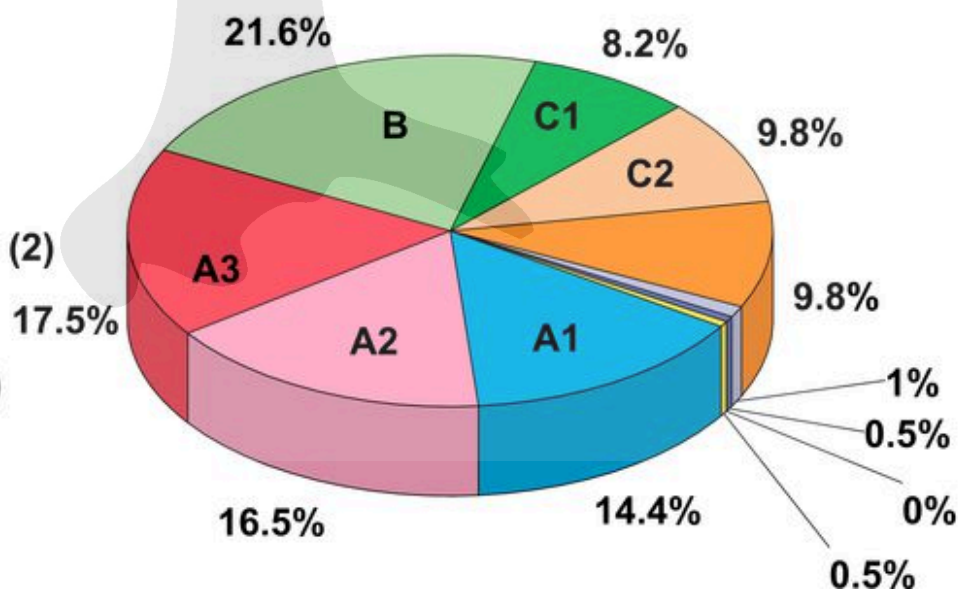
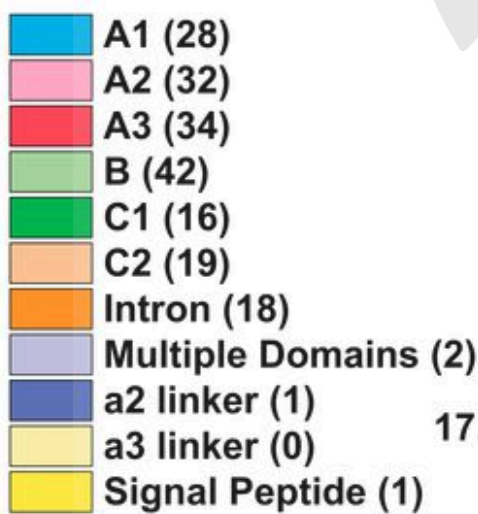
(a)

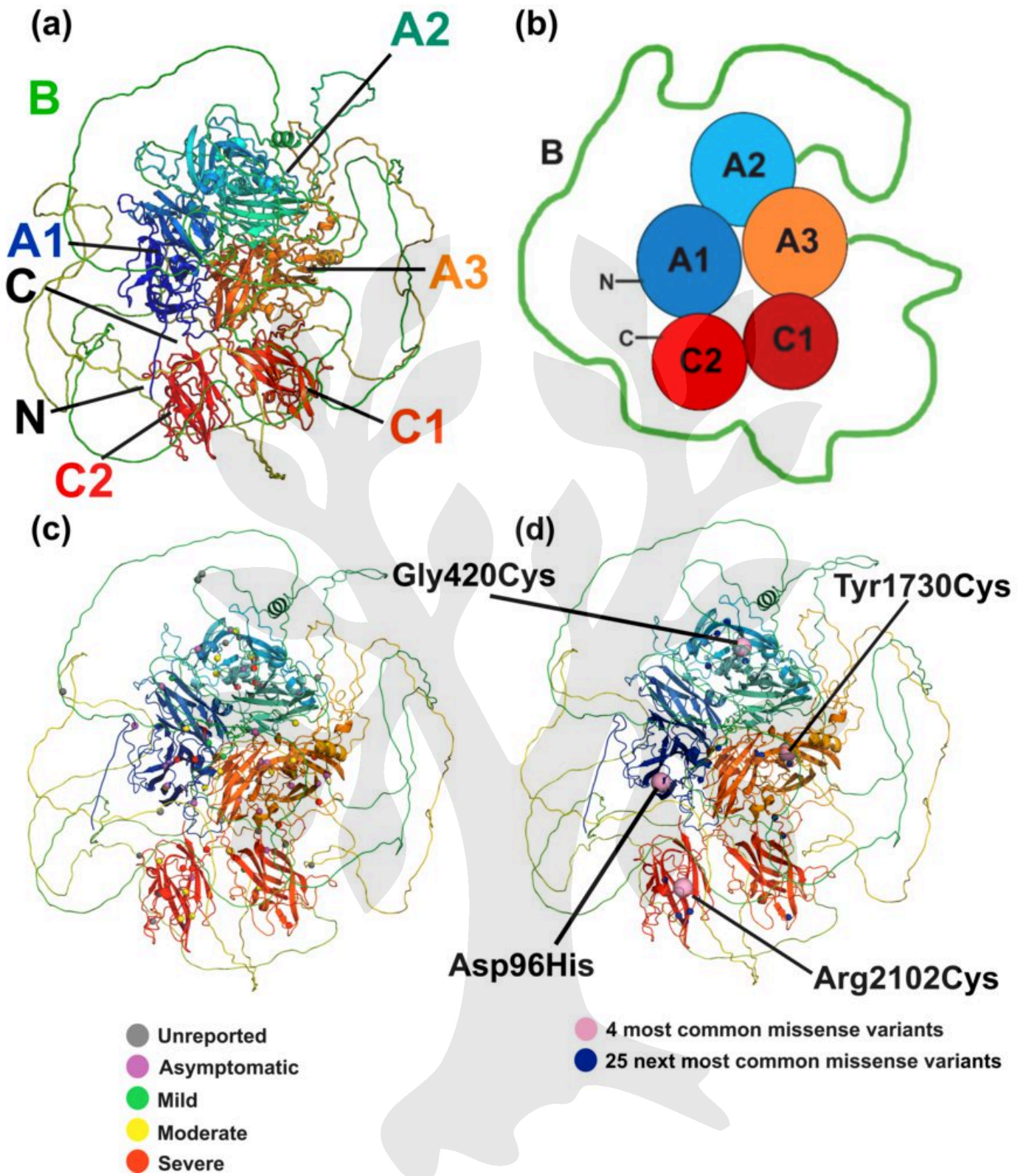


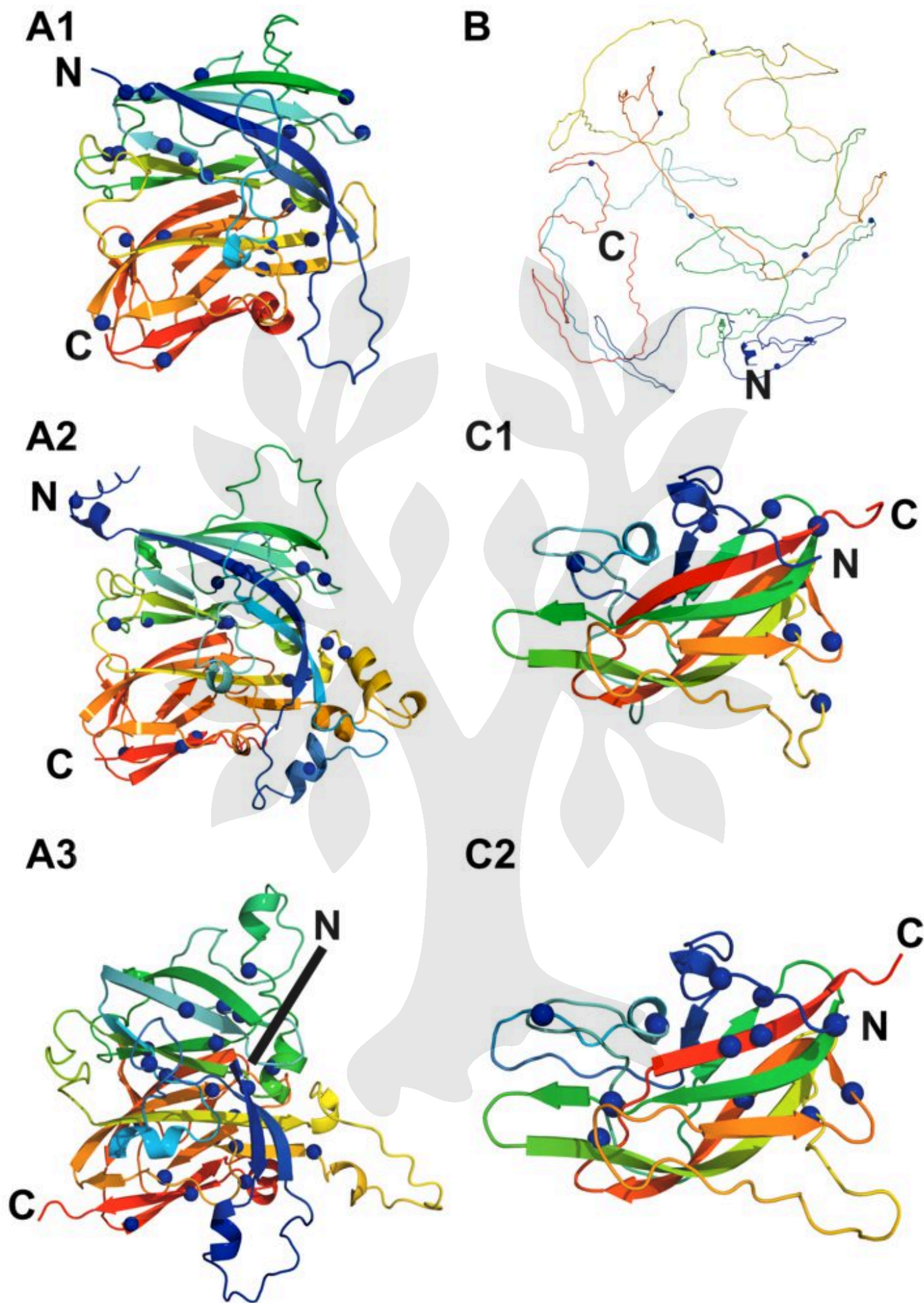
(b)

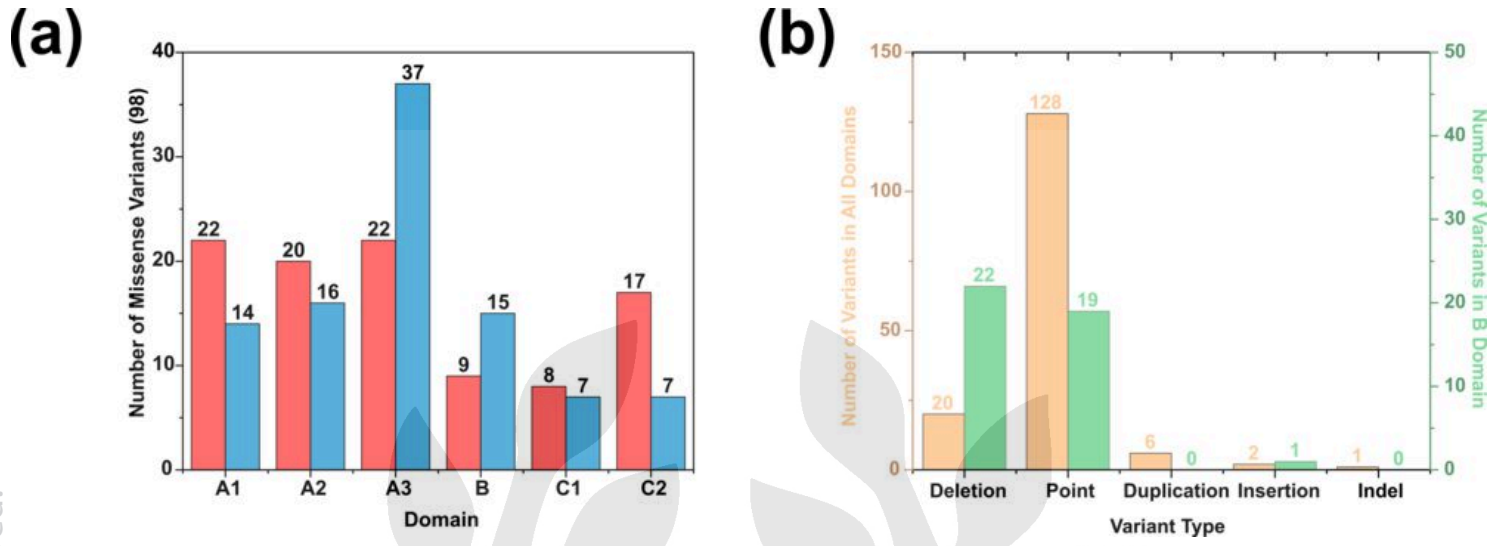


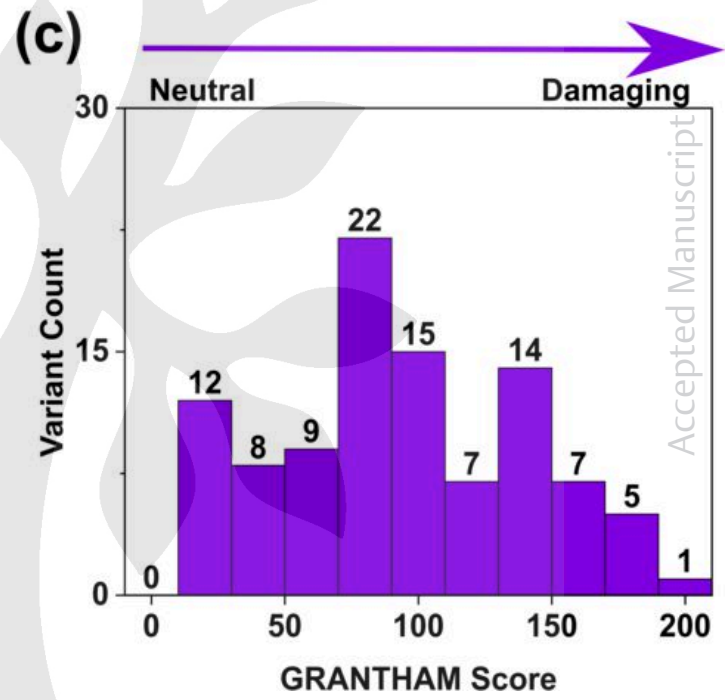
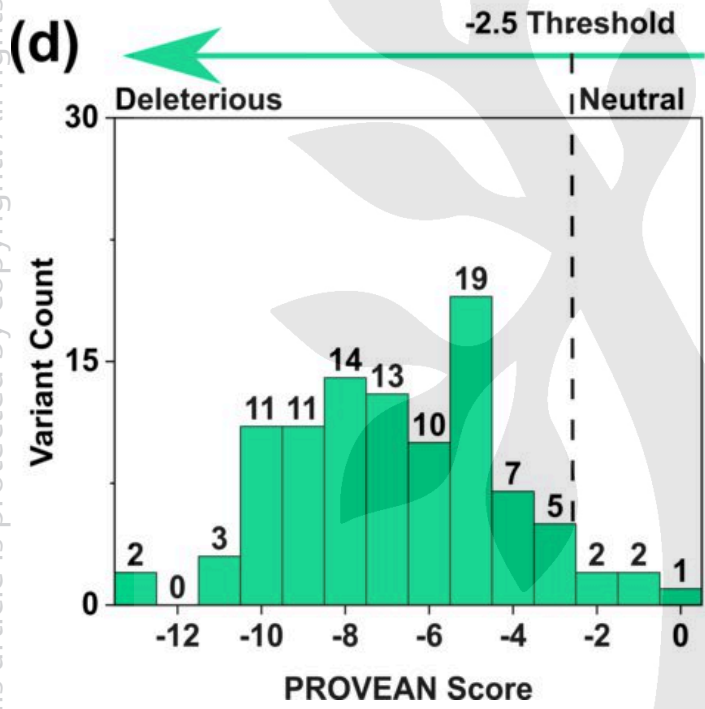
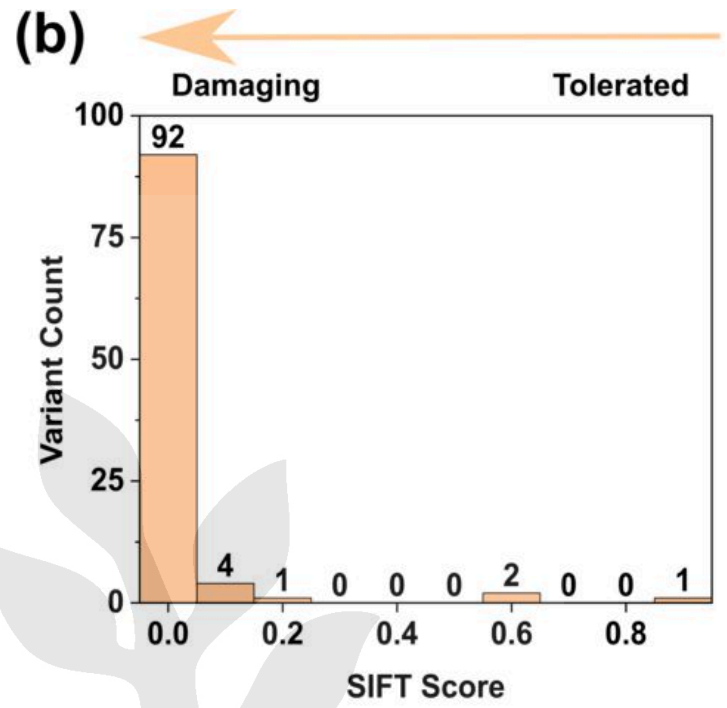
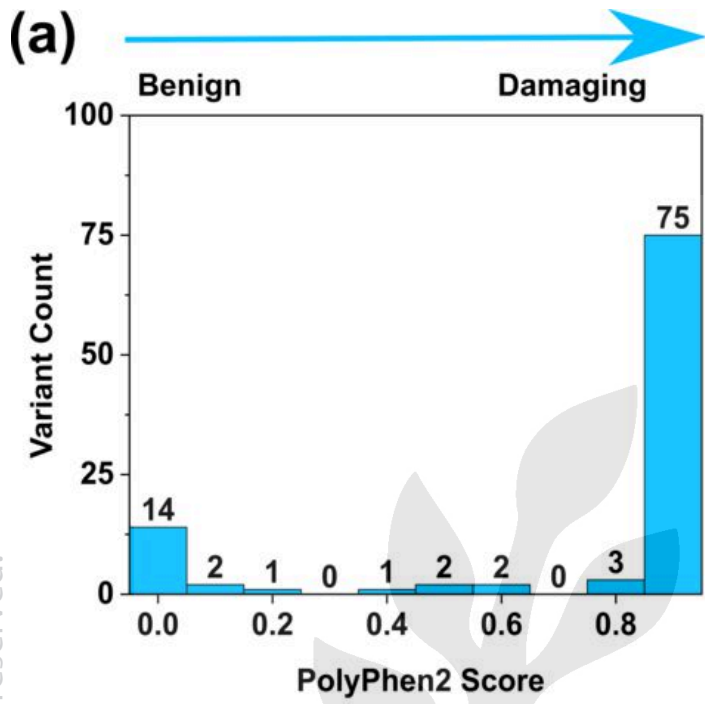
(c)

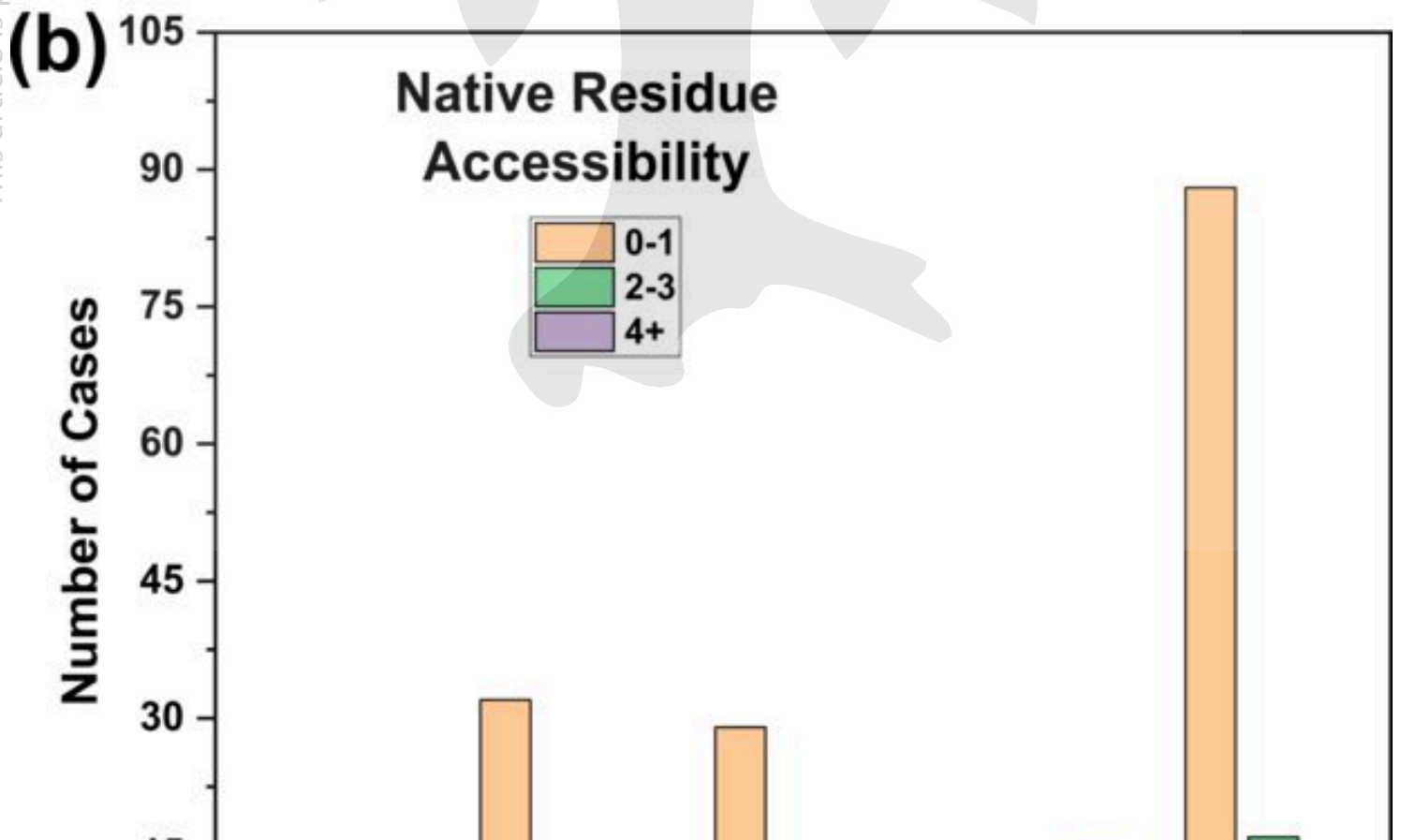
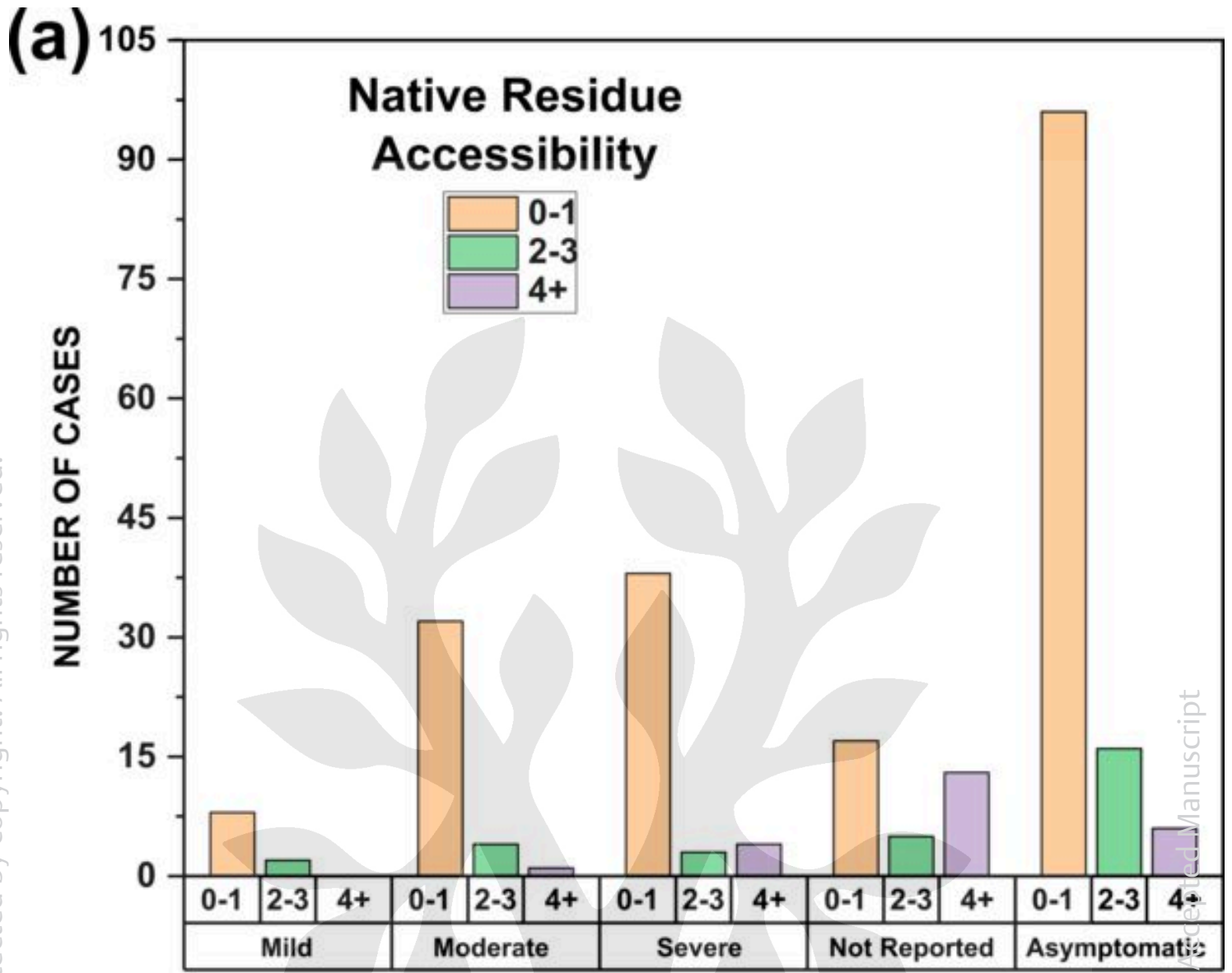












This article is protected by copyright. All rights reserved.

Accepted Manuscript

```

HGVS_NUMBERING    1140    1150    1160    1170    1180    1190    1200    1210    1220    1230
FV_SEQUENCE        IQQPDMHSTSDPSRSRSPKLESMLEYDISHKSFPTDISQMSPSSEHEVWQTVISPDLSQVTLSPQLSQTNLSPDLSSHITLSPELIQNLSPALGQMPIS
FV_DSSP_7KVE      667254452767657857778868766857345586686668767659775867776677766777677575776667476676666865777678575432
FV_ACCESS

```

```

HGVS_NUMBERING    1240    1250    1260    1270    1280    1290    1300    1310    1320    1330
FV_SEQUENCE        PDLSSHITLSPDLSSHITLSDLQLSQTNLSPQLSQTNLSPALGQMPISPDLSHITLSDLSPQLSQTNLSPQLSQTNLSPALGQMPISPDLSHITLSDL
FV_DSSP_7KVE      46547768666757667576667666677657667677978767677757667666767667777555335335665866758787677677767767576
FV_ACCESS

```

```

HGVS_NUMBERING    1340    1350    1360    1370    1380    1390    1400    1410    1420    1430
FV_SEQUENCE        SQTNLSPQLSQTNLSPALGQMPISPDLSHITLSDLQLSQTNLSPQLSQTNLSPDLSEPLFADLSQIPLTFLDLCMTLSPDLGETDLSFNFGQMSLSPDL
FV_DSSP_7KVE      54767767774766767787875777657676767766766775766766675767676775776766665676766775887756686876777876
FV_ACCESS

```

```

HGVS_NUMBERING    1440    1450    1460    1470    1480    1490    1500    1510
FV_SEQUENCE        QVTLSPDLSDITLLDLQLSPPDLDQIFYPSESLSQSLLLQEFNESPFPYDLCQMPSPSPPTLNDTFLSKFNPLVIVGLSKDG
FV_DSSP_7KVE      887775867678877767777776668666756757656576747787787877566645156677777878768686758776
FV_ACCESS

```

a3

```

HGVS_NUMBERING    1520 4    1530    515406    1550    1560    1570
FV_SEQUENCE        TDYEIIPKEVQSSEDDYAEIDYVYDDPYKTDVVRTNINSSRPDNIAAWYLRSN
FV_DSSP_7KVE      S-----EEEEEE--SB-GGGTTS--S-----S-S-EEEEEEEESSSSS-S-B-----SGGGS-TT---B---BTTEEEEEEE--B-SSS-B-----SS
FV_ACCESS        7797877788786787886747656342006207204574723934341630561

```

A3

```

HGVS_NUMBERING    1580    15907    1600    1610    1620    1630    1640    1650    1660    1670
FV_SEQUENCE        NGNRRNYIAAEEISWDYSEFVQRETDIEDSDDIPETTYKKVVFVKYLDSTFTKRDPGEYEEHLGILGPIIRAEVDIVIQVRFKNLASRPYSLHAHG
FV_DSSP_7KVE      --EEEE--EEEEEE--SB-GGGTTS--S-----S-S-EEEEEEEESSSSS-S-B-----SGGGS-TT---B---BTTEEEEEEE--B-SSS-B-----SS
FV_ACCESS        624413010001235120201335321538637542351040000220332407533310021300100001000013000000010205220000000

```

```

HGVS_NUMBERING    1680    1690    1700    1710    1720    1730    1740    1750    1760    1770
FV_SEQUENCE        LSYEKSSEGKTYEDDSEWFKENAVQPNSSYTYVWHATEISGPESPGSACRAWAYSAVNPEKDTNSGLICPLLCQKGLHKSDSNMPPDMREFVLLF
FV_DSSP_7KVE      -B---GGGT---S---SSSTGGGSS-B-TT---EEEEEE--GGGS--STT-SSE---EEE--TTSHHHHHHTT---EEEB-ES-TT---TTTT-S-SS-EEEEEE
FV_ACCESS        00010100004041416741340030636100000010160001207101000000002201200000000000000000000003441237144101421100000

```

```

HGVS_NUMBERING    1780    1790    1800    1810    1820    1830    1840    1850    1860    1870
FV_SEQUENCE        MTFDKKSWYEEKKSSWRLTSSQKKSHEPFAINGMILYSLPLKMYEQWVRLHLLNIGGSQDIHVVHFPQTLLENKQKQHLGVWPLLPQSFKTL
FV_DSSP_7KVE      EEEEESSSS--SS-SS-TTTTGGGTTTTEEEEESSSBS--TT-EE-SS---EEEE--SS-SS-EEEEET---SSS---SSEE-EE-EEEE
FV_ACCESS        00010230311454064158410264551020100001020011010223130000000001032100000000011032500030011010020000

```

```

HGVS_NUMBERING    1880    1890    1900
FV_SEQUENCE        EKASKPQWLLNTEVGNQRACMQTFPLIMDRD
FV_DSSP_7KVE      E-----EE---EEE---HHHTTT---EE---EE---S-
FV_ACCESS        1010011020000000020063001020000064

```

C1

```

HGVS_NUMBERING    1910    1920    1930    1940    1950    1960    1970    1980    1990    2000
FV_SEQUENCE        CMFAGLSTGIISDSQIKASFLQWYEPRLARLNNGGSYNAWVEKLAAEFASKPWIQVDMQKEVIITLQDGAHXYLKYTYTFYVAYSSNQINRQ
FV_DSSP_7KVE      --EE-STTSSS--GGGEEESS--TT--GGG-BTT--SS--S--SSS---EEE--SS-EEEEEEEEE---SSS---EEEEEEEEE--SSS--
FV_ACCESS        51400031220246014133101612032010524743000024617517555010001053300000010000346613020210200103437611

```

```

HGVS_NUMBERING    2010    2020    2030    2040    2050    2060
FV_SEQUENCE        IFKLNSTRNVMYFNGNSDASTIKENQFDPIVARYIRISPTRAYNRPTLLELQCEVNG
FV_DSSP_7KVE      --SSS--SSSS---SSS---EEEEEEEEEEEEEEEEEEEESSS---B-EEE--SSS
FV_ACCESS        32002013253515005324343313030002030010002524540000000000410

```

C2

```

HGVS_NUMBERING    2070    2080    2090    2100    2110    2120    2130    2140    2150    2160
FV_SEQUENCE        CSTPLGMENGIENKQITASSFKKSNGDYWEPFARLNAQSRVNAWQAKANNKQWLEIDLLKIKKIITITGCKSLSENIVKSYTIHYSEQGVV
FV_DSSP_7KVE      --EE-STTSS-S-GGEEES--B-SSS-B-TT--SS--S--SSS---TT--TT---EEEEEEEEEEEEEEE--EE-SSSS-EEEEEEEEE--SSS--
FV_ACCESS        001100034450527003030514388313000240104276501002064424502010205320000000000033855111030000120342652

```

```

HGVS_NUMBERING    2170    2180    2190    2200    2210    2220
FV_SEQUENCE        KPYRLKSSMVDIFEGNTNTGHVKNFFNPPISRFIRVIPKTWNBIALRLELFGCDIY
FV_DSSP_7KVE      EE-B-TTSSSB-----SSSS---EEEEEEEEEEEEEEEEEEEESSS---EEEE---
FV_ACCESS        23052562654150600642400211304210302001000342250000000000415

```

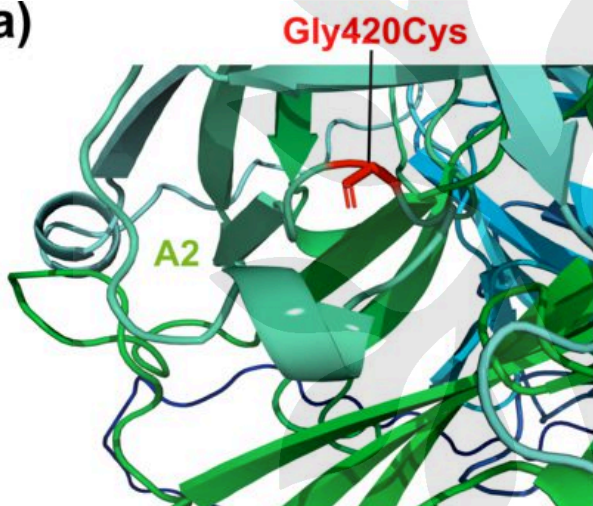
- KEY**
- Yellow: FV deficiency point variants that occur in 4 patients or less
 - Green: FV deficiency point variants that occur in 5 or more patients
 - Red: FV deficiency point variants that occur in 50 or more patients
 - Blue: Point variants associated with thrombosis
 - Purple: Point variants with unknown disease association
 - Grey: Point variants with multiple associations
 - : Cys-Cys Disulphide Bridge
 - : Sulphated Tyr Residue

AA MUTATED TO

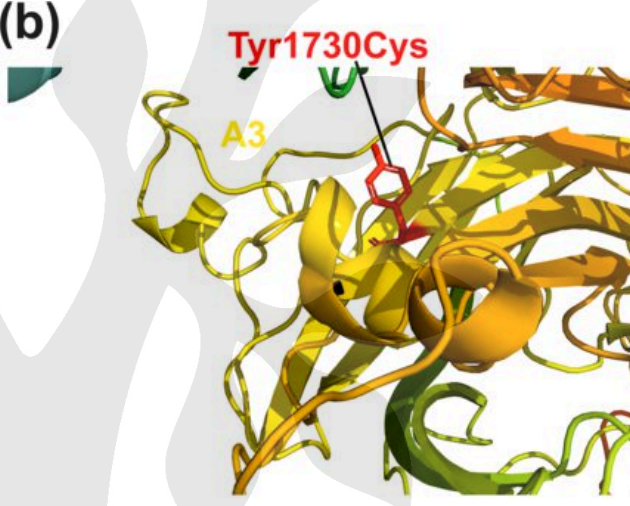
AA MUTATED FROM

| | Ala | Arg | Asn | Asp | Cys | Gln | Glu | Gly | His | Ile | Leu | Lys | Met | Phe | Pro | Ser | Thr | Trp | Tyr | Val | SUM |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | | | | 0 | | | 0 | 0 | | | | | | | 1 | 0 | 1 | | | 0 | 2 |
| Arg | | | | | 1 | 1 | | 4 | 2 | 0 | 2 | 0 | 1 | | 4 | 0 | 1 | 2 | | | 18 |
| Asn | | | | 0 | | | | | 0 | 0 | | 0 | | | | 1 | 0 | | 2 | | 3 |
| Asp | 2 | | 1 | | 0 | 0 | 0 | 3 | 0 | | | | | | | | | | 1 | 0 | 7 |
| Cys | | 5 | | | | | | 1 | | | | | | 0 | | 0 | | 1 | 2 | | 9 |
| Gln | | 0 | | | | | 0 | | 0 | | 0 | 0 | | | 0 | | | | | | 0 |
| Glu | 0 | | | 0 | | 0 | | 3 | | | | 1 | | | | | | | | 0 | 4 |
| Gly | 0 | 0 | | 2 | 1 | | 0 | | | | | | | | | 1 | | | 0 | 0 | 4 |
| His | | 1 | 0 | 2 | | 1 | | | | | | | | | 0 | | | | 0 | | 4 |
| Ile | | 1 | 0 | | | | | | | | 0 | 0 | 0 | 0 | | 1 | 0 | | | 0 | 2 |
| Leu | | 1 | | | | 0 | | | 0 | 0 | | | 0 | 1 | 3 | 1 | | 0 | | 0 | 6 |
| Lys | | 0 | 0 | | | 0 | 1 | | | 0 | | | 1 | | | | 0 | | | | 2 |
| Met | | 0 | | | | | | | | 0 | 0 | 0 | | | | | 1 | | | 1 | 2 |
| Phe | | | | | 0 | | | | | 0 | 2 | | | | | 0 | | | 0 | 0 | 2 |
| Pro | 0 | 1 | | | | 1 | | | 0 | | 3 | | | | | | 0 | | | | 6 |
| Ser | 0 | 1 | 1 | | 1 | | | 1 | | 0 | 2 | | | 2 | 2 | | | 0 | 0 | 1 | 11 |
| Thr | 1 | 0 | 0 | | | | | | | 1 | | 0 | 1 | | 1 | 0 | | | | | 4 |
| Trp | | 1 | | | 0 | | | 0 | | | 0 | | | | | 1 | | | | | 2 |
| Tyr | | | 0 | 2 | 3 | | | | 0 | | | | | 1 | | | | | | | 6 |
| Val | 3 | | | 0 | | | 0 | 3 | | 0 | 0 | | 0 | 0 | | | | | | | 6 |
| SUM | 6 | 11 | 2 | 6 | 6 | 3 | 1 | 15 | 2 | 1 | 9 | 1 | 3 | 4 | 11 | 6 | 3 | 3 | 6 | 1 | 100 |

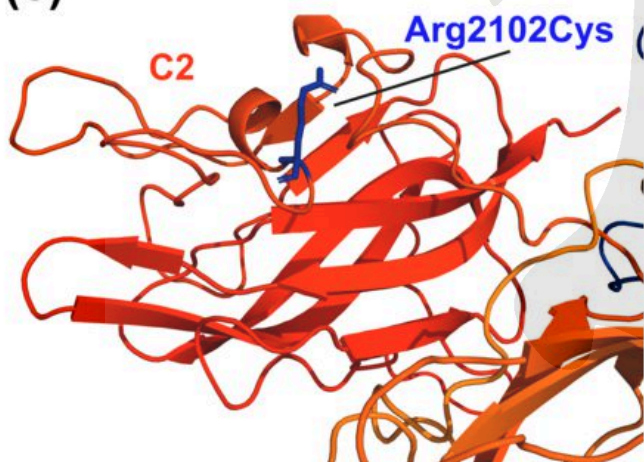
(a)



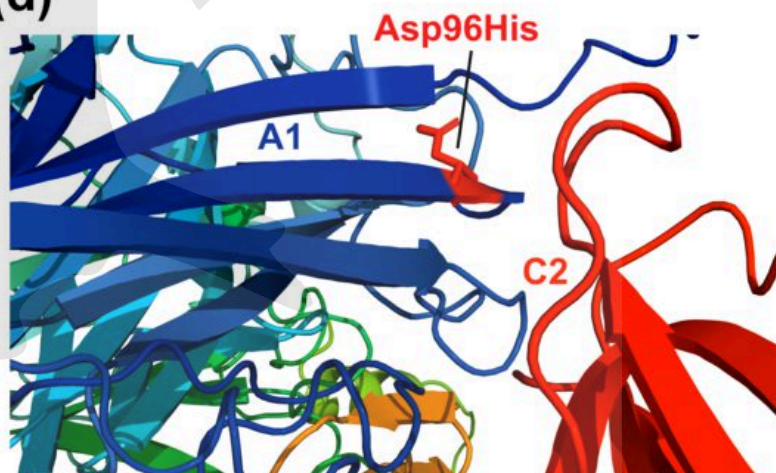
(b)



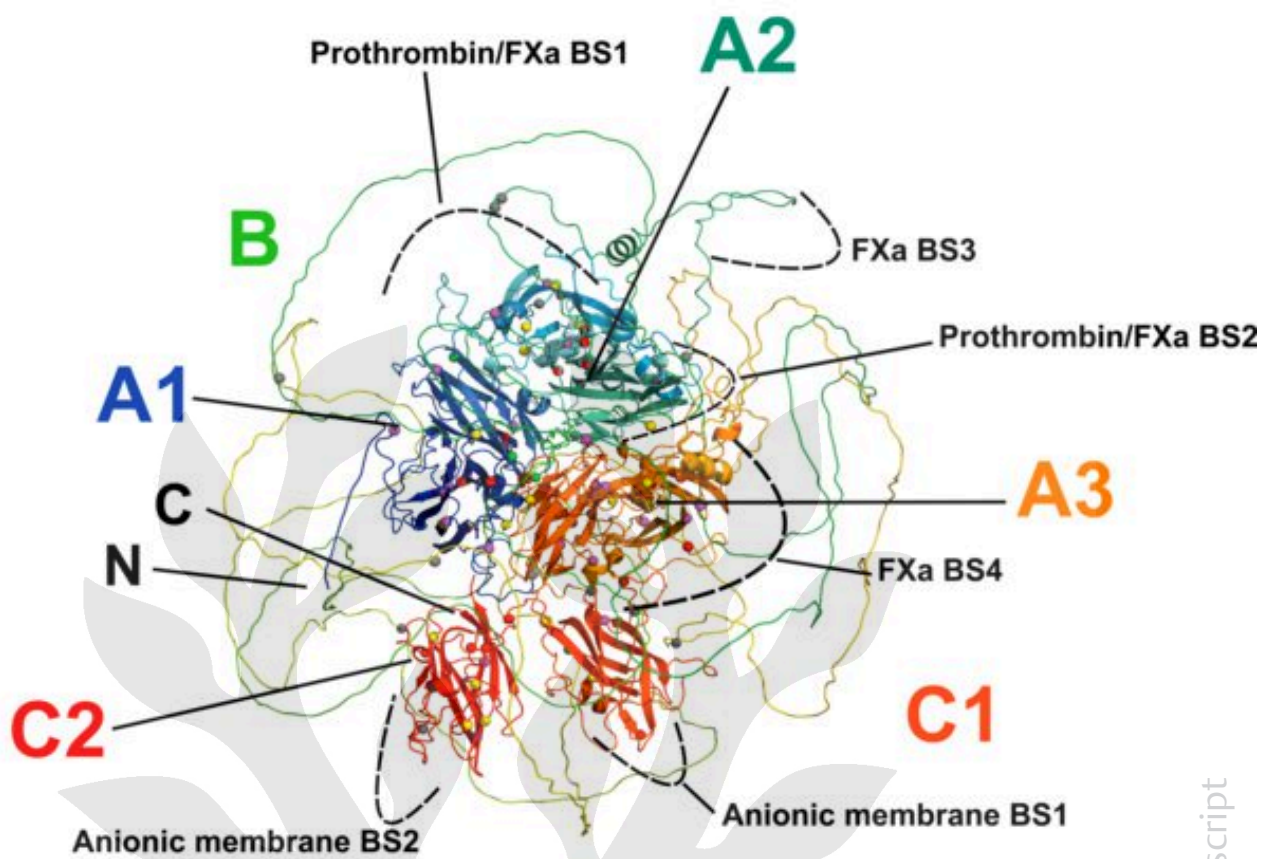
(c)



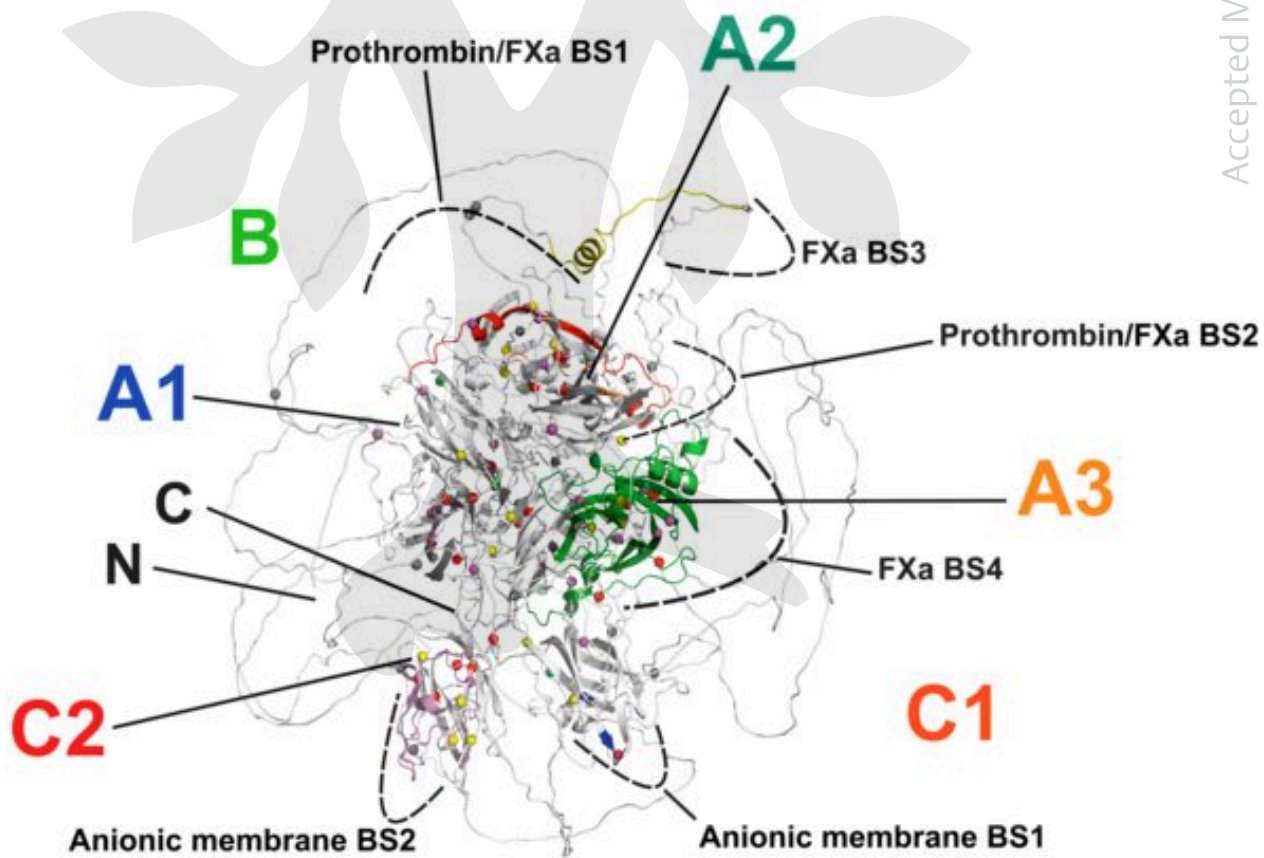
(d)



(a)

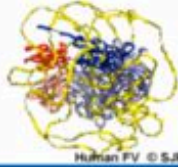


(b)



- Unreported
- Asymptomatic
- Mild
- Moderate
- Severe

F5



Factor V Gene (F5)

Variant Database



UCL

[Home](#)[About FV](#)[Advanced Search](#)[Variants](#)[Structures](#)[AA Alignments](#)[Resources](#)[Support](#)**c.524A>G****p.His175Arg (Legacy AA No. 147)****Variant Type:** Point**Domain:** d5A1**Location:** Exon 4**Variant Effect:** Missense**Codon Change:** 524A>G**No. of Patients Reported:** 8**Phenotype:** Mild**Allele Count *:** 94**Allele Number *:** 282760**Allele Frequency *:** 0.000332

References and Comments:

Liu et al 2014A

prolonged and excessive bleeding after a tooth extraction

Patient Information: Show

Structural Interpretation:

Please click [HERE](#) for in-depth variant analysis.

| Patient_ID | Age (Yrs) | Gender | Race | FV:C(%) | FV:C (IU/dl) | FV:Ag(%) | FV:Ag (IU/dl) | Inheritance | Other Variants | Severity | Comments | Reference |
|------------|-----------|--------|------|---------|--------------|----------|---------------|--------------|--|--------------|---|---------------------------------|
| 22 | NA | F | NA | 2 | NA | -2 | NA | Heterozygous | c.286G>C (p.Asp96His);c.6304C>T (p.Arg2102Cys) | Mild | prolonged and excessive bleeding after a tooth extraction | Liu et al 2014A |
| 23 | NA | F | NA | 2 | NA | -2 | NA | Heterozygous | c.286G>C (p.Asp96His);c.6304C>T (p.Arg2102Cys) | Asymptomatic | sister of case 330 | Liu et al 2014A |
| 58 | NA | F | NA | 2 | NA | -2 | NA | Heterozygous | c.286G>C (p.Asp96His);c.6304C>T (p.Arg2102Cys) | Mild | prolonged and excessive bleeding after a tooth extraction | Liu et al 2014A |
| 59 | NA | F | NA | 2 | NA | -2 | NA | Heterozygous | c.286G>C (p.Asp96His);c.6304C>T (p.Arg2102Cys) | Asymptomatic | sister of case 330 | Liu et al 2014A |
| 60 | NA | F | NA | 60 | NA | 75 | NA | Heterozygous | c.6304C>T (p.Arg2102Cys) | Asymptomatic | mother of cases 330 and 331 | Liu et al 2014A |
| 744 | NA | F | NA | 60 | NA | 75 | NA | Heterozygous | c.6304C>T (p.Arg2102Cys) | Asymptomatic | mother of cases 330 and 331 | Liu et al 2014A |
| 745 | NA | F | NA | 2 | NA | -2 | NA | Heterozygous | c.286G>C (p.Asp96His);c.6304C>T (p.Arg2102Cys) | Mild | prolonged and excessive bleeding after a tooth extraction | Liu et al 2014A |
| 746 | NA | F | NA | 2 | NA | -2 | NA | Heterozygous | c.286G>C (p.Asp96His);c.6304C>T (p.Arg2102Cys) | Asymptomatic | sister of case 330 | Liu et al 2014A |

Residue Information:

| Name | Type | Cyclic | Size | Position | Hydrophobicity | Charge | |
|------------------|------|----------------|---------|----------|----------------|-------------|----------|
| Wild Type | His | aromatic/basic | cyclic | large | surface | hydrophilic | positive |
| Mutated | Arg | basic | acyclic | large | surface | hydrophilic | positive |

Substitution Analysis:

- Grantham Score : 29
- PolyPhen-2 Prediction : **benign (SCORE: 0.026)**
- SIFT Prediction : **Probably Damaging (SCORE: 0.01)**
- PROVEAN Prediction : **Deleterious (SCORE: -5.549)**

Structural Implications:

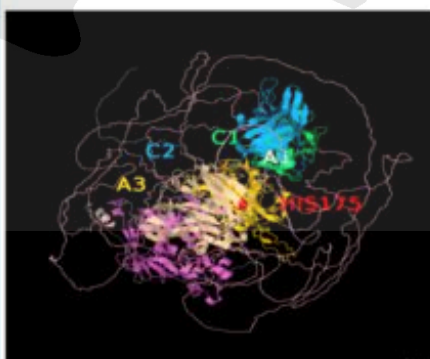
FV **His175** is a buried residue (surface accessibility value = 0).FVa **His175** is an exposed residue (surface accessibility value = 2).**His175** is in a region of secondary structure within the FV domains.The DSSP assignment for this residue is **NA**.

Controls

The molecule can be rotated with the mouse holding the left button and the image can be magnified using the middle mouse button. Note 7KVE is the default structure.

To choose the structure you wish to inspect, please select the Structure at the bottom of the Options menu. If the amino acid is missing from the structure, it will not be labeled and you should select a different structure.

Right Click on the Jmol screen for more options.



Options

Spacefill OFF 20% 50% 100%

Cartoon OFF ON

Wireframe OFF ON

Trace OFF 0.4 0.8

Backbone OFF ON

Spin OFF ON

Background Black White

Disulphides OFF ON

Domains All A1 A2 B A3 C1 C2

Alternative Colouring AA SS None

Labels All Domains Variant None

Structure AlphaFold 7KCY(FVa) 7KVE(FV)

Right Click on the molecule's screen for more options.

This article is protected by copyright. All rights reserved.

Accepted Manuscript

