# Phenotypic classification of the mutations in coagulation factor IX segregate to different locations in its protein structure

Pavithra M. Rallapalli<sup>1</sup>, Edward G. Tuddenham<sup>2</sup>, Keith Gomez<sup>2</sup> and Stephen J. Perkins<sup>1</sup>

<sup>1</sup>Research Department of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT <sup>2</sup> The Katharine Dormandy Haemophilia Centre and Thrombosis Unit, Royal Free Hampstead NHS Trust, London NW3 2QG, UK;

#### Introduction

Coagulation factor IX (FIX), a zymogen synthesised in the liver, is a single-chain, vitamin Kdependent plasma glycoprotein. It is an important component of the intrinsic system of the coagulation cascade. Defects in the F9 gene leads to Haemophilia B, a gender-linked recessive incidence coagulation disorder with an approximately 1 in 50,000, and occur almost exclusively in males. Haemophilia B has attracted much recent interest because of the development of ground-breaking gene therapy methods for this at UCL by Nathwani et al.

#### Objective

To develop an interactive database in which the F9 mutations are presented in searchable formats, and viewed in conjunction with the FIX protein structure and sequence along with the clinical phenotypes for Haemophilia B.

#### Methods

- Literature searches to collate all previously known and novel FIX mutations.
- Homology modelling and energy minimisation to create the human FIX structure starting from the crystal structure of porcine FIX.
- Use of PHP, MySQL, XML, Perl and Jmol tools for database and web-interface, and for statistical, structural and sequence analyses.

#### **Novel Mutations**

Of the 1094 unique mutations, 121 unique mutations are novel and not published since the last update of the Haemophilia B mutation database in 2004. There are also 42 (3%) multiple mutations within F9 comprising of 40 double, 1 triple and 1 quadruple mutants, an additional total of 87 individual mutations.

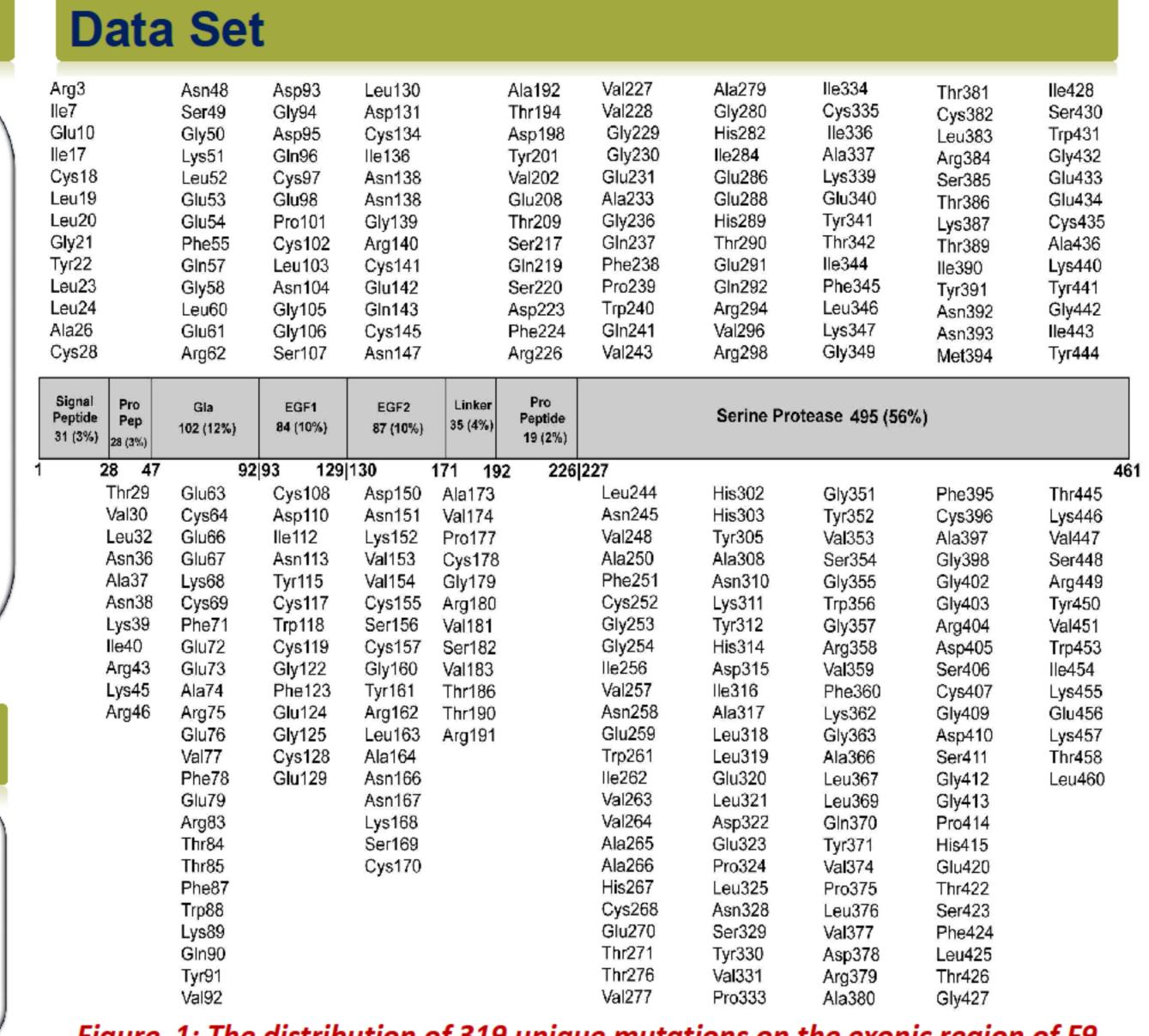


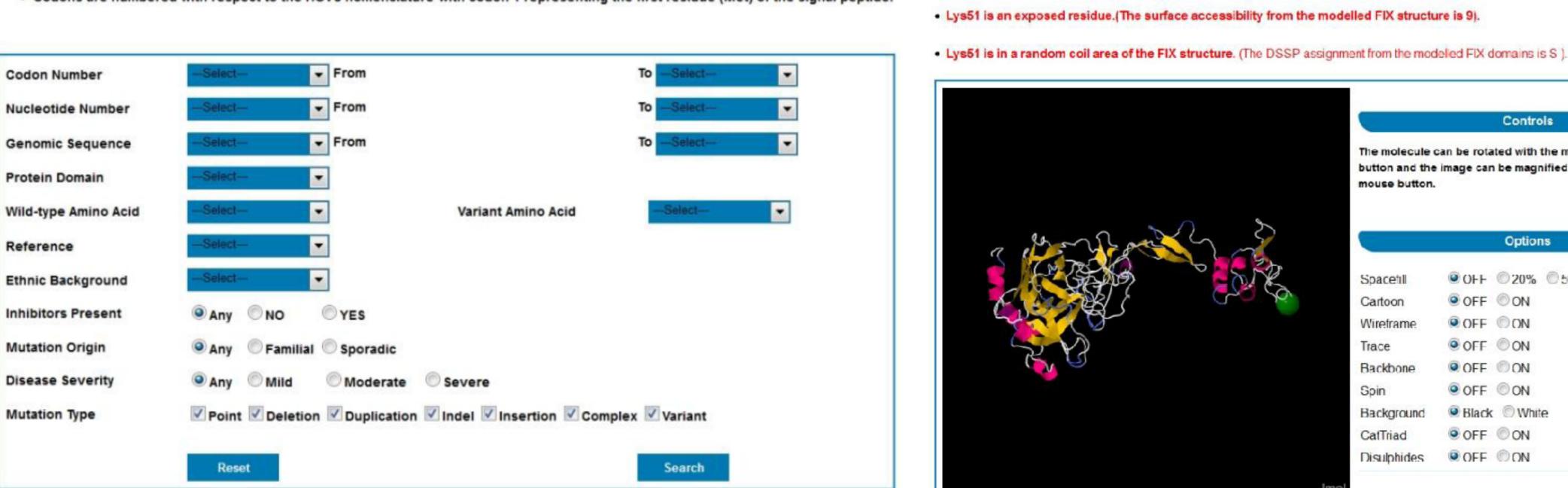
Figure 1: The distribution of 319 unique mutations on the exonic region of F9 associated with Haemophilia B

#### **Statistics** Types of 1094 unique mutations Location of 1094 unique mutations Complex, 0.5% Polymorphism, 5.0% ---Duplication, 0.2% Insertion, 3.5% Indel, 1.5% Effect of 1094 unique mutations Distribution of 1094 unique mutations on protein sequence 29 (2.7%) 395 (36.8%) Silent, 1.7% 38 (3.5%) 19 (1.8%) **15** (1.4%) 22 (2.1%) 27 (2.5%) 250 300 350 400

Figure 2: Statistics of the FIX mutations available in the newly-updated Factor IX mutation database.

#### www.factorix.org In Depth Mutation Analysis: c. 151 A>G (p.Lys51Glu) Factor IX Mutation Database . 151 A>G p.Lys51Glu (5) Structures Mutation Type:Point Mutation Effect: Missense Domain:Gla Location:exor(2) Simple Nucleotide Search Comments: Inhibitors:NO CpG:N Reference: Rowley G (UP) Nucleotide ---Select--- ▼ Search Haemophilia B is caused by mutations in the gene that codes for blood protein factor IX. There are currently 1094 unique disease causing mutations in the F9 Patient Information Hide gene compiled within this database corresponding to 3443 Haemophilia B PATIENT 1 Simple Amino Acid Search Antigen (%): 121 Clotting (%): 4 Genotype: -Phenotype: Moderate 2012 Release- Version 1.1 (May 2012) Amino Acid ---Select--- ▼ Search Ethnicity: United Kingdom Reference: Rowley et al (UP) Type: | Inhibitors: NO Rallapalli PM, Tuddenham EG, Gomez K, Perkins SJ (2012)- Manuscript in Comments: Mutation type Search PATIENT 2 Clotting (%): 7 Antigen (%): Type ---Select--- ▼ Search Phenotype: Moderate Have you or someone you know been diagnosed with haemophilia B? Ethnicity: United Kingdom Reference: Koeberl et al (1990) Inhibitors: NO The information contained on this web site is provided for scientific research Exon and Intron based search purposes only. We do not give medical advice or recommend any particular treatment for specific individuals. Click <u>here</u> for patient information. Exon -- ▼ G0 Intron -- ▼ GO Residue Information Charge Hydrophobicity dvanced Search hydrophilic postive hydrophilic Records from mutations can be retrieved using any combination of the following search criteria. Structural Implications Lys51 is an exposed residue.(The surface accessibility from the modelled FIX structure is 9).

Codons are numbered with respect to the HGVS nomenclature-with codon 1 representing the first residue (Met) of the signal peptide.



Options ● OFF ◎ 20% ◎ 50% ◎ 100% OFF ON Cartoon ● OFF ● ON OFF ON OFF ON OFF ON Black White

Figure 5: Screenshots of the various mutational analysis options available in the newly-developed Factor IX mutation database.

## Acknowledgement

We thank Pfizer and the Special Trustees of the Royal Free Hospital and the Katharine Dormandy Trust for Haemophilia and Related Disorders for their support.

Structure record Domain References organisation Mutations [F9] **Protein Data** Patient Gene data Sequence AA information information Nucleotide Codon table Admin Login Download Advanced Contact Search Haemophilia E Structures Sequences Mutations Statistics User Manual Sequence Mutation Map Structure World Map

Database work flow

Figure 3: Organisation of the Haemophilia B mutation database. (A) General architecture of the backend database. (B) Network of the webpages inside the Factor IX mutation database.

### Phenotype and location of mutations

Haemophilia B-associated mutations have been reported for 319 out of the 461 residues in FIX, with activity and antigen levels reported for 164 of these. Of the 150 unique mutations in the four domains with known phenotypes, 18% (26) are quantitative type I mutations, 68% (102) qualitative type II mutations, and 14% (22) are unclassified. Note that the SP domain is formed as two subdomains, between which lies the catalytic active site cleft. In this, 46 of the 64 type II mutations (blue) are located in subdomain 2, and the majority of these lead to severe haemophilia B. In contrast, 13 of the 21 type I mutations (yellow) are in the subdomain 1 of the SP domain.

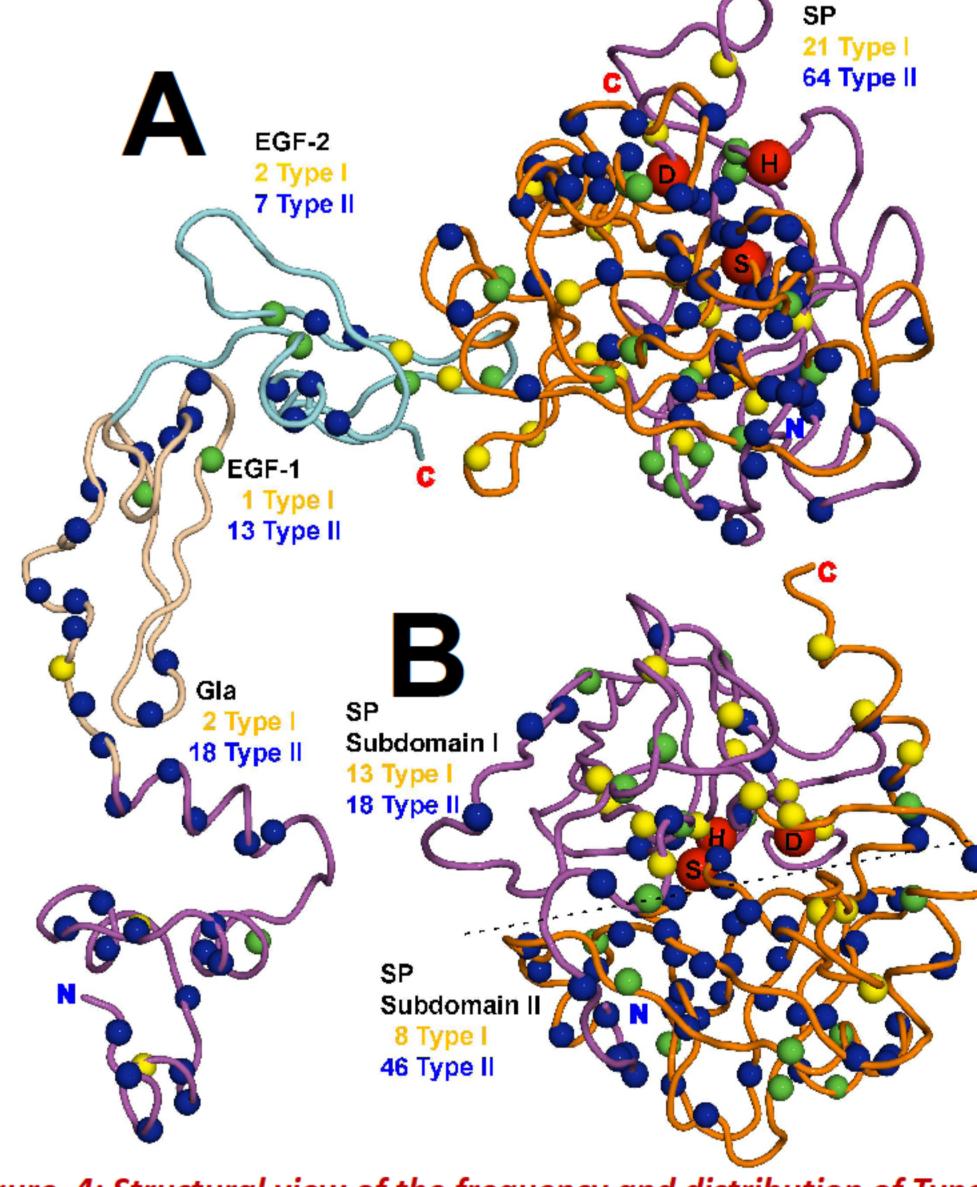


Figure 4: Structural view of the frequency and distribution of Type I and Type II phenotypic mutations within the four domain regions of human FIX protein.

#### Conclusion

By mapping the mutations with a 3D structural model for FIX, we can now correlate the phenotype and severity of the FIX mutations with their location in the FIX structure. The type II phenotype is dominant in the small Gla, EGF-1 and EGF-2 domains, and in the subdomain II of the catalytic SP domain. Curiously the type I phenotype is dominant in the subdomain I of the SP domain. This finding implies that most of the FIX structure is involved with functional interactions, while subdomain I is prone to protein misfolding or degradation.



Controls

OFF ON

OFF ON

Disulphides



