

IN SILICO ANALYSIS THROUGH A NEWLY GENERATED FACTOR VIII MODEL SUPPORTS TOLERANCE OF B DOMAIN TO MULTIPLE AMINO ACID SUBSTITUTIONS LIMITING THE LOCAL IMPACT OF MISSENSE VARIANTS

A. Branchini¹, M.F. Testa¹, M. Pinotti¹, R. De Cristofaro^{2,3}

¹Department of Life Sciences and Biotechnology, University of Ferrara, Italy; ²Department of Translational Medicine and Surgery, Catholic University of the Sacred Heart, Roma, Italy; ³Center for Hemorrhagic and Thrombotic Diseases, Foundation University Hospital "A. Gemelli" IRCCS, Roma, Italy

Background and Aims. Missense variants, a relevant haemophilia A (HA) cause, are scarcely reported in the large factor VIII (FVIII) B domain (3%) that, among FVIII domains, remains the least-characterized and structurally unresolved. Here, to provide insights into the impact of multiple amino acid substitutions in B domain at a single-residue scale, we performed systematic *in-silico* analyses through a newly generated FVIII model (FVIII_{IT}) with assembled B domain.

Methods. Generation of the FVIII_{IT} model by the I-TASSER tool and integrated *in-silico* analysis through DDMut (protein stability) and REVEL (variant pathogenicity) tools.

Results. HA missense variants are mainly reported in FVIII A1-A2-A3-C1-C2 (A1-A3, 77%; C1-C2, 20%; hereafter A1-C2), particularly those associated with severe HA as highlighted by frequency normalized per domain size (A1-C2, 0.36-0.49; B, 0.02). The newly generated FVIII_{IT} model was first validated for resolved A1-C2 domains towards reference structures by an alanine scanning approach, showing minor impact when extended to B domain. Noticeably, a mutational scanning at a single-residue scale mimicking all possible single-nucleotide changes resulting in missense variants (total >15,000; >6,000 in B domain) showed significant differences ($p < 0.0001$) between B and A1-C2 domains, with lower predicted impact for a high proportion (~90%) of

B-domain variants. Single-nucleotide changes are also associated with clinically relevant nonsense mutations, causing premature termination codons (PTCs), whose suppression by the so-called translational readthrough (TRT) may result in full-length proteins bearing subsets of amino acid substitutions. A three-level and progressively refined analysis, validated in relation to experimentally obtained data, revealed significantly higher values for B-domain TRT-deriving missense variants arising from i) either possible ($n=2439$) or HA ($n=1094$) PTCs, corresponding to 2439/1094 (B, $n=1007/396$) predicted missense variants, ii) possible ($n=457$) or HA ($n=190$) PTCs, corresponding to 1371/570 (B, $n=651/228$) predicted missense variants, at which TRT-mediated re-insertion of the original amino acid is not predicted, and iii) PTCs at tryptophan positions (possible/HA, $n=35/31$) at which missense changes ($n=114/105$; B, $n=27$) are expected to be detrimental, as in A1-C2 domains, while overlapping missense changes are not reported in B domain.

Conclusions. Altogether, these original insights, obtained by rational *in-silico* analyses through the newly generated FVIII-IT model, may contribute evidence toward the hypothesis that B domain is tolerant to multiple amino acid changes, supporting potential underestimation of HA-associated B-domain missense variants due to limited local impact of amino acid substitutions.