

Identification of six novel factor VIII gene variants using next generation sequencing and molecular dynamics simulation

Faisal A. Al-Allaf^{1,2,3§}, Zainularifeen Abduljaleel^{1,2§}, Neda M. Bogari¹, Tarek M. A. Owaidah⁴, Mohiuddin M. Taher^{1,2}, Mohammad Athar^{1,2}, Ahmed Elsendiony⁵, Halah Abalkhail⁴, Ahmed Abdellatif⁶, Wafa Elbjeirami³ and Abdellatif Bouazzaoui^{1,2§}

¹Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P.O. Box 715, Makkah 21955, Saudi Arabia; ²Science and Technology Unit, Umm Al Qura University, P.O. Box 715, Makkah 21955, Saudi Arabia; ³Molecular Diagnostics Unit, Department of Laboratory and Blood Bank, King Abdullah Medical City, Makkah 21955, Saudi Arabia; ⁴Department of Pathology and Laboratory Medicine, King Faisal Specialist Hospital and Research Center, Riyadh, Kingdom of Saudi Arabia; ⁵Makkah Regional Laboratory, Makkah, Saudi Arabia; ⁶Faculty of Pharmacy, Qassim University, Qassim, Saudi Arabia.

Hemophilia A is an X-linked recessive hemorrhagic disorder caused by variants in the *F8* gene. To identify known and novel causative variants in hemophilia A, we have carried out genetic analysis among Saudi patients. Twenty-one patients, who were negative for *inv-1/inv-22*, were selected for analysis by next generation sequencing, thereafter confirmed by Sanger sequencing. In addition, the functionality and structural changes in the variant proteins were assessed using Molecular dynamics (MD) simulation and compared with wild-type and native proteins. In the samples we analyzed, we found 10 variants in 12 individuals; among them, five were novel and five were previously reported. The novel variants were located at positions: c.6130_6131insC, c.5815G>C, c.5493C>G, c.3734_3740delinsATTCT and c.3744A>T. With the exception of one variant which was silent, the MD simulation revealed that the observed variants were causing severe structural changes when compared to the native protein and resulted in a loss of the protein's function. The MD analysis is in line with clinical data of patients who had <1% Factor VIII levels (severe hemophilia) with episodic bleeding, and were on more than one treatment. Moreover, some patients presented with chronic joint disability. These results will enrich the spectrum of variants and enlarge the factor VIII protein's database in the Saudi Arabian population.

Key words: next generation sequencing, hemophilia A; novel variant; Saudi Arabia; blood coagulation; molecular dynamics (MD) simulation

Received: 10 November, 2017; revised: 20 November, 2018; accepted: 22 January, 2019; available on-line: 22 February, 2019

✉ e-mail: faallaf@uqu.edu.sa (FAAI-A); alazzauoi@uqu.edu.sa, ab1971@hotmail.de (AB); zaabduljaleel@uqu.edu.sa, zainulbio@gmail.com (ZA)

§Authors Contributed Equally to the Manuscript

Acknowledgment of Financial Support: This work was supported by the National Science, Technology, and Innovation Plan (MAARI-FAH) of the Kingdom of Saudi Arabia to Dr. Faisal A. Al-Allaf, (Grant Code: 09-BIO920-10).

Abbreviations: Pat ID, patient ID; N. P, nucleotide position; A, a, amino acid; AiY, age in year; BFA, baseline factor activity; TOC, type of care; TPU, treatment products used; Hgb, hemoglobin; MCV, mean cell volume; PTT, partial thromboplastin time; FVIII, factor VIII; FVIII inh, factor VIII inhibitor; BU, Bethesda unit; PDCFC, plasma derived clotting factor concentrate; PChJDi, presence of chronic joint disability

INTRODUCTION

Hemophilia A (OMIM 306700) and hemophilia B (OMIM 306900) are sex-linked genetic disorders resulting in deficiency of plasma coagulant activities of FVIII and FIX, respectively (Hoyer, 1994; Lillicrap, 1998), and are transmitted in an X-linked recessive mode of inheritance. There, the heterozygous females transmit the affected gene and the disease phenotype presents almost exclusively in males. However, some females can present two defective copies, which also results in an hemophilia disorder, but such case rarely occurs (Lenting *et al.*, 1998). Hemophilia A and B can affect people from different races and ethnic groups (Mannucci & Tuddenham, 2001; Bolton-Maggs & Pasi, 2003). The prevalence of hemophilia A affects 1 in 5000 males, and hemophilia B – 1 in 30000 males (Lillicrap, 1998, Bolton-Maggs & Pasi, 2003). Protein for FVIII is encoded by the *F8* gene and is considered as one of the largest genes located on the long arm of chromosome X (Xq28 position). The *F8* gene is composed of 26 exons and 25 introns, and encodes a mature protein of 2,332 amino acids consisting of six domains: A1-A2-B-A3-C1-C2, which is activated by thrombin after proteolytic cleavage of the B domain. The *F8* gene region is characterized by high GC content; within the 9.1 kb-coding region there are about 70 CpG dinucleotides. High GC content is hyper-mutable, and about 30% of variants found in such a region are usually novel (Nair *et al.*, 2014). A total of 2320 *F8* gene variants have been reported so far in the Human Gene Mutation Database (HGMD) (Human Gene Mutation Database (accessed on 29th Nov 2015)). Among individuals with severe hemophilia A, the most frequent variant has been found in the intron 22 inversion, with a frequency rate of 42% (Antonarakis *et al.*, 1995). In the second most frequent variant, we found an inversion in intron 1, with a prevalence of about 15% among hemophilia A patients (Salviato *et al.*, 2004; Schroder *et al.*, 2006). A total of 1388 missense/nonsense, 156 splicing, 377 small deletions, 117 small insertions, 213 gross deletions, 27 small indels, 26 gross insertions, 4 regulatory variants and 12 complex rearrangements have been detected so far (Peyvandi *et al.*, 2013; Human Gene Mutation Database (accessed on 29th Nov 2015)).

In normal cases, the FVIII range varied between 0.5 and 2 IU/mL, while hemophilia is classified as severe at <0.01 IU/mL (<1%), moderate at 0.01–0.05

IU/mL (1–5%), and mild at 0.05–0.4 IU/mL (5–30%) (White *et al.*, 2001; Bolton-Maggs & Pasi, 2003). Several studies have been conducted describing different variants in the Western populations (Citron *et al.*, 2002; Boekhorst *et al.*, 2005; Graw *et al.*, 2005; Repesse *et al.*, 2007; Rossetti *et al.*, 2008), and the hemophilia care in these countries is well-organized; however, such organized care is not available in the Arab countries. Based on the prevalence of hemophilia A (1 in 5000 males) we estimated that at least 2200 Saudis are most likely affected by hemophilia A in Saudi Arabia, but the spectrum and nature of common variants causing hemophilia A in the Arab populations is still lacking investigation, and specifically the data available for frequent variants is insufficient. Taken together, the study presented here was planned to explore the known and novel variants through genetic analysis of hemophilia A patients among the Saudi population. Furthermore, examination of the variant spectrum using *in silico* functional studies based on the scores of SIFT, and PolyPhen2, was done in order to identify the deleterious variants that are likely to affect the FVIII protein structure. We have identified potential pathologically significant variants, proposed a model structure for the variant proteins, and compared them with the wild-type protein. Moreover, analysis of native and variant modeled protein for stability and solvent accessibility were determined by an MD simulation.

MATERIALS AND METHODS

Clinical evaluation and blood testing. All patients were assessed to determine their age at the time of diagnosis, type of factor replacement, bleeding history, and joint or organ with recurrent bleeding. The laboratory testing included a complete blood count and partial thromboplastin time (PTT). For the blood testing, a 10-mL sample was collected into 3.2% sodium citrate solution and then centrifuged. The separated plasma and blood samples were transferred to the central laboratory (in Riyadh, Saudi Arabia) to determine the FVIII levels and confirm diagnosis. FVIII inhibitors were measured using the modified Nijmegen–Bethesda method (Miller *et al.*, 2012) and the FVIII activities were carried out with an Behring Coagulation System (BCS; Siemens, Marburg GmbH, Germany) using a chromogenic assay (Al-Allaf & Bouazzaoui, 2016).

Subjects and DNA isolation. DNA samples were obtained from Saudi Arabian patients undergoing treatment at the King Faisal Specialist Hospital and Research Centre (KFSH&RC), Riyadh, Saudi Arabia. The blood samples were tested for FVIII coagulant activity using Behring Coagulation System (BCS; Siemens, Marburg, Germany). An informed consent was obtained from the patients and an ethical approval was obtained from the KFSH&RC-IRB. Detailed medical history was also obtained to confirm the pattern of inheritance. Hemophilia A patients were selected based on the criteria and guidelines indicated by the British Committee for Standards in Hematology (Guidelines). Genomic DNA was isolated from EDTA entire blood samples with the MagNA pure compact nucleic acid isolation kit-I (Roche, Mannheim, Germany) according to the product's guidelines.

Library enrichment and sequencing. Library preparations have been performed using an AmpliSeq library

preparation kit and primer pool from the inherited disease panel (Life Technologies, CA, USA). Next, the library concentration was quantified using real time PCR with ion Universal Library Quantitation Kit (Life Technologies, CA, USA) and the size distribution was evaluated with an Agilent Bioanalyzer and a high sensitivity DNA kit (Agilent, Santa Clara, CA). Thereafter, emulsion PCR and Ion Sphere Particles (ISP) enrichment was done using the Ion Xpress Template kit (Life Technologies, CA, USA) with Ion one touch template kit v2.0 (Life Technologies, CA, USA), according to the manufacturer's instructions. Finally, sequencing was performed with PGM using Ion PGM 200 sequencing kit (Life Technologies, CA, USA) and Ion 318 Chip.

Identifying coding variants and functional analysis. After next generation sequencing, the SOAP aligner, BWA and GATK programs were used to identify variants and insertions or deletions (InDels). Variants and InDels were annotated by using the exome-assistant program (<http://122.228.158.106/exomeassistant>). There are three algorithms were used to evaluate the non-synonymous variants, SIFT (Kumar *et al.*, 2009), PolyPhen2 (Adzhubei *et al.*, 2013) (<http://genetics.bwh.harvard.edu/pph2/>) and HOPE (<http://www.cmbi.ru.nl/hope/>). First two algorithms can be tested only on variants (pathogenic or benign). “SIFT” it examines the evolutionary conservation of amino acids, which highly conserved sites tend to be intolerant to substitutions, whereas sites with a low degree of conservation confer a greater degree of tolerance. “Polyphen2” it calculates a position-specific independent counts profile for each candidate mutation and predicts whether it will be benign, possibly damaging or probably damaging, according to the probability intervals (0, 0.2), (0.2, 0.85) and (0.85, 1). The third algorithmsproject-HOPE was used to determine the native conformation of a protein using energy minimization. Multiple sequence alignments were performed using ESPript3.0 (<http://espript.ibcp.fr/ESPrpt/cgi-bin/ESPrpt.cgi>). Furthermore, we used the COSMIC3 (version 64), and CDC Hemophilia-A Mutation Project (CHAMP) databases (<http://www.cdc.gov/ncbddd/hemophilia/champs.html>). Additionally, the known and novel variants were confirmed by Sanger's sequencing.

PCR amplification and capillary sequencing. Amplification of the *F8 gene* exons was done as previously described (Al-Allaf *et al.*, 2017a; Al-Allaf, 2017b). In summary, the gene was sequenced with 50 ng of genomic DNA as template in a 20 μ l reaction mixture using 0.4 μ l HotStarTaq plus DNA Polymerase (Qiagen, Hilden, Germany), 2 μ l 10 \times PCR buffer, 2 μ l 25 mM MgCl₂, 0.4 μ l 10 mM dNTPs, 2 μ l of 10 μ M forward and reverse primers; descriptions of the primers used for PCR amplification and sequencing have been reported previously (Vidal *et al.*, 2001). The PCR program consisted of Taq polymerase activation at 95°C for 5 min, followed by 40 cycles of denaturing at 95°C for 30 s, annealing at 60°C for 3s, extension at 68°C for 1 min, and final extension at 68°C for 5 min. The amplified product was separated on a 1.5% agarose gel to ensure the size and quality of the band. PCR products were then purified by the magnetic beads method using an Agencourt AM-Pure XP kit (Beckman Coulter, Munich, Germany), and used as templates for direct sequencing with a BigDye Terminator v3.1 cycle sequencing ready reaction kit (Applied Biosystems, CA, USA). The sequencing reaction products were purified using an EDTA/

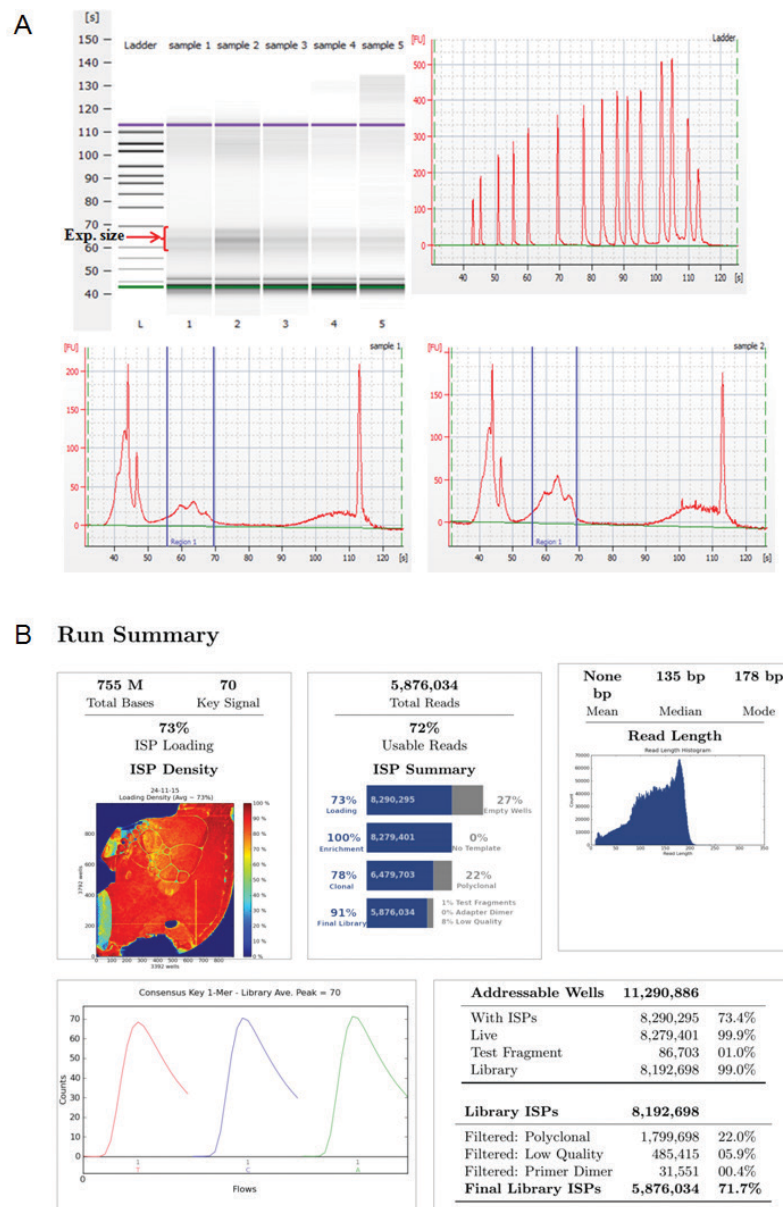


Figure 1. Representative gel for size selection and run

The NGS was performed using an AmpliSeq strategy with inherited diseases panel kit on Ion torrent PGM system. **(A)** Gel and electropherogram for the representative samples. **(B)** Run summary with run quality assessment

Ethanol method and sequenced by capillary electrophoresis on an ABI 3500 Genetic analyzer, and the end analysis was accomplished with Sequence Analysis Software v5.4 Applied ABI PRISM and CLC genomic workbench (Qiagen bioinformatics, Manchester, UK)

Molecular dynamics (MD) simulation. The molecular dynamics (MD) simulation program CHARMM-GUI (<http://www.Charmm-gui.Org>) was used for comparison of the wild-type and variant protein. This simulation did not minimize or equalize the solvated system. However, 0.15 M ions can likewise be added to the simulation procedure through the process specification ion (KCl) and concentration (C). To start with, ions were configured and confirmed using the short Monte Carlo (MC) simulations with a primitive model, for example, the van der Waals associations. In general, the free energy simulation was completed with a couple of particular dissolvable

water molecules while checking the solute regularly. This was despite the fact that the effect of whatever remains of the solvent mass was once demonstrated undoubtedly as an efficient solvent boundary potential (SSBP). Nevertheless, the molecular dynamics simulation was done with a 2 fs time step at a steady temperature of 300 K and a consistent pressure of 1-atm, which is below the stipulations of the periodic solvent boundary.

RESULTS

NGS run and quality

To identify out the disease-causing gene variants, we chose the inherited disease panel kit (IDP kit), which was already available to us. The kit has three primer

Table 1A. Single variants identified in hemophilia A patients

Pat ID	Gender	Age	Exon	Status	N. P	A. a changes	Function	SIFT	SIFT: Score	Polyphen2	Polyphen2: Score
AB07	Male	16	E14	rs1800291	c.3780C>G	p.(Asp1260Glu)	Missense	Tolerated	0.52	Benign	0.000
AB08	Male	19	E14	rs1800291	c.3780C>G	p.(Asp1260Glu)					
AB09	Male	18	E14	rs1800292	c.3864A>C	p.(Ser1288Ser)	Silent	-	-	-	-
AB10	Male	8	E14	rs1800291	c.3780 C>G	p.(Asp1260Glu)	Missense	Tolerated	0.52	Benign	0.000
AB17	Male	nd	E20	Novel	c.6130_6131insC	p.(Leu-2044Profs*9)	Frameshift	-	-	-	-
AB18	Male	28	E20	Novel	c.6130_6131insC	p.(Leu-2044Profs*9)					
AB19	Male	9	E16	Novel	c.5493C>G	p.(Thr1831Thr)	Silent	-	-	-	-
			E17	Novel	c.5815G>C	p.(Ala1939Pro)	Missense/Splicing site	Damaging	0	damaging	1.000
AB27	Male	30	E14	Novel	c.3744A>T	p.(Leu1248Phe)	Missense	Tolerated	0.39	Possibly damaging	0.868
			E14	Novel	c.3734delCCin-sATTCT	p.(Asn-1245Lysfs*9)	Frameshift	-	-	-	-
AB28	Male	42	E14	Novel	c.3744A>T	p.(Leu1248Phe)	Missense	Tolerated	0.39	Possibly damaging	0.868
			E14	rs1800292	c.3864 A>C	p.(Ser1288Ser)	Silent	-	-	-	-
AB47	Male	12	E23	rs137852466	c.6545G>A	p.(Arg2182His)	Missense	Damaging	0	Possibly damaging	1.000
AB52	Male	8	E19	rs1050705	c.6107A>G	p.(Tyr2036Cys)	Missense	Damaging	0.04	Probably damaging	1.000
AB53	Male	12	E8	rs137852368	c.1063C>T	p.(Arg355*)	Nonsense	-	-	-	-

pools with more than 9000 primer pairs, also including the genes for hemophilia. For this reason, for NGS we selected 21 samples that were from patients who were negative for inv-1 and inv-22. Examination of the size distribution of the libraries is shown for a representative gel (Fig. 1A) and the size was about 227 bp. Next, the enriched DNA was sequenced at a single base resolution using Ion-torrent PGM and the read of each probe reached approximately 135 bp in length (Fig. 1B) with about 73% loading. After the run, samples from twenty-one patients were analyzed along with two healthy adult human samples as controls, and the mapping was done with the hg19 reference human genome. About 90% yielded clean reads uniquely corresponding to the target regions, with about 98% of the targeted region covering at least 95% mean depth cov-

erage for each sample. The average depth coverage for exons among the 21 patient samples was 95%, with the highest depth coverage of 98%.

The discovered variants

We have identified a total of 10 variants in 12 individual DNA samples, out of which in 5 individual samples we discovered 5 novel variants, and in 8 individual samples there were 5 previously reported variants (Table 1A). After NGS, all 10 variants were confirmed by capillary method sequencing. The first novel variant was found at position c.6130_6131insC (Fig. 2B1) in the exon 20 of two related patients (Pt ID #AB17 and Pt ID #AB18). The C insertion occurred in the CTG codon for leucine, which is a non polar aliphatic amino acid, and altered it into CCT codon for proline

Table 1B. Clinical characteristics of the studies patients with single variants

Pat ID	Hgb (g/l)	MCV (fm)	Platelet X (1000/ μ l)	PTT (Sec)	FVIII (IU/ml)	FVIII Inh (BU/ml)	BFA	TOC	TPU	PChJ-Di	SPECIFY
AB07	98	57.4	428	67	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Rt. Ankle
AB08	102	54.7	361	86.6	<0.01	0	Severe Hemophilia (<1%)	Episodic	More than one	nd	-
AB09	134	75.1	344	66.7	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Rt. Elbow, Rt. Knee and Rt. Toe
AB10	126	71.5	457	83.8	<0.01	0	Severe Hemophilia (<1%)	Episodic	Recombinant factor VIII	No	-
AB17	111	79.2	381	75.7	<0.01	32	Severe Hemophilia (<1%)	Episodic	More than one	Yes	Rt. & Lt. Knee
AB18	156	87.6	220	51.5	<0.01	151.6	Severe Hemophilia (<1%)	Episodic	More than one	Yes	Lt. Knee
AB19	120	79.2	527	121	<0.01	0	Severe Hemophilia (<1%)	Episodic	Fresh Frozen Plasma	No	-
AB27	157	81.2	299	125.3	<0.01	0	Severe Hemophilia (<1%)	Prophylaxis	Fresh Frozen Plasma	Yes	Rt. Knee
AB28	157	81.8	354	89.6	<0.01	0	Severe Hemophilia (<1%)	Episodic	Fresh Frozen Plasma	Yes	Rt. & Lt. Knee
AB47	111	80	312	67	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	No	-
AB52	13.3	75.4	379	108.5	<0.01	0.38	Severe Hemophilia (<1%)	Episodic	Recombinant Clotting Factor Concentrate	nd	nd
AB53	131	80.2	343	120.8	<0.01	0.44	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	nd	-

which is non polar aromatic amino acid and is present in the C1 domain of FVIII. Furthermore, this variant caused a frameshift leading to an early stop codon p.(Leu2044Profs*9). Based on our prediction study, we assume that the protein synthesized will not be a complete protein and will be eventually degraded. The clinical data is in agreement with the molecular analysis and had shown that both patients present with <1% in the FVIII level (severe hemophilia), with more than one treatment and episodic bleeding. Moreover, both patients presented with chronic joint disability on either both or only left knee, respectively (Table 1B).

The next variant was found at position c.5815G>C in the exon 17 (Fig. 2B2) of one 9 years old male (Pt ID #AB19). The G>C variant occurred in the GCA codon to become CCA p.(Ala1939Pro) and is present in the A3 domain of FVIII protein. Although both amino acids are neutral, the patient has severe hemophilia with <1% of FVIII levels, and was on a fresh frozen plasma treatment. Furthermore, the same patient also displayed another novel variant at c.5493C>G in exon 16; there, ACC was changed into ACG, however the nucleotide change did not change the amino acid p.(Thr1831Thr). In exon 14 of two related male patients (Pt ID #AB27 and Pt ID #AB28) we have found a novel variant at position c.3744A>T (Fig. 2B3). It occurred in the TTA codon for leucine, which is a

non-polar aliphatic amino acid, and altered it into TTT codon for phenylalanine, which is a non-polar aromatic amino acid p.(Leu1248Phe). Based on our prediction study, it is probable that this variant caused a structural change leading to loss of function, which was supported by the clinical data which indicated that the 30 year old males (Table 1B) presented with <1% FVIII levels (severe hemophilia). Furthermore, a 42 year old patient (Pt ID #AB28) presented a delinsion variant at position c.3734_3735delinsATTTCCT; p.(Asn1245Lysfs*9) (Fig. 2B4). Based on our prediction study, these changes would cause a frameshift and therefore we assume that the protein synthesized would not be a complete protein and thus eventually it would be degraded. Clinical data had shown that the 42 year old patient (Pt ID #AB28) presented with <1% FVIII levels (severe hemophilia), with episodic bleeding and he was on fresh frozen plasma treatment for hemophilia. Moreover, the patient presented with chronic joint disability on both knees.

Structure and functional impact of the novel variants

Out of five novel variants, two were frameshift variants c.3734_3735delinsATTTCCT p.(Asn1245Lysfs*9), c.6130_6131insC, p.(Leu2044Profs*9), one was a silent variant c.5493C>G, p.(Thr1831Thr), and two

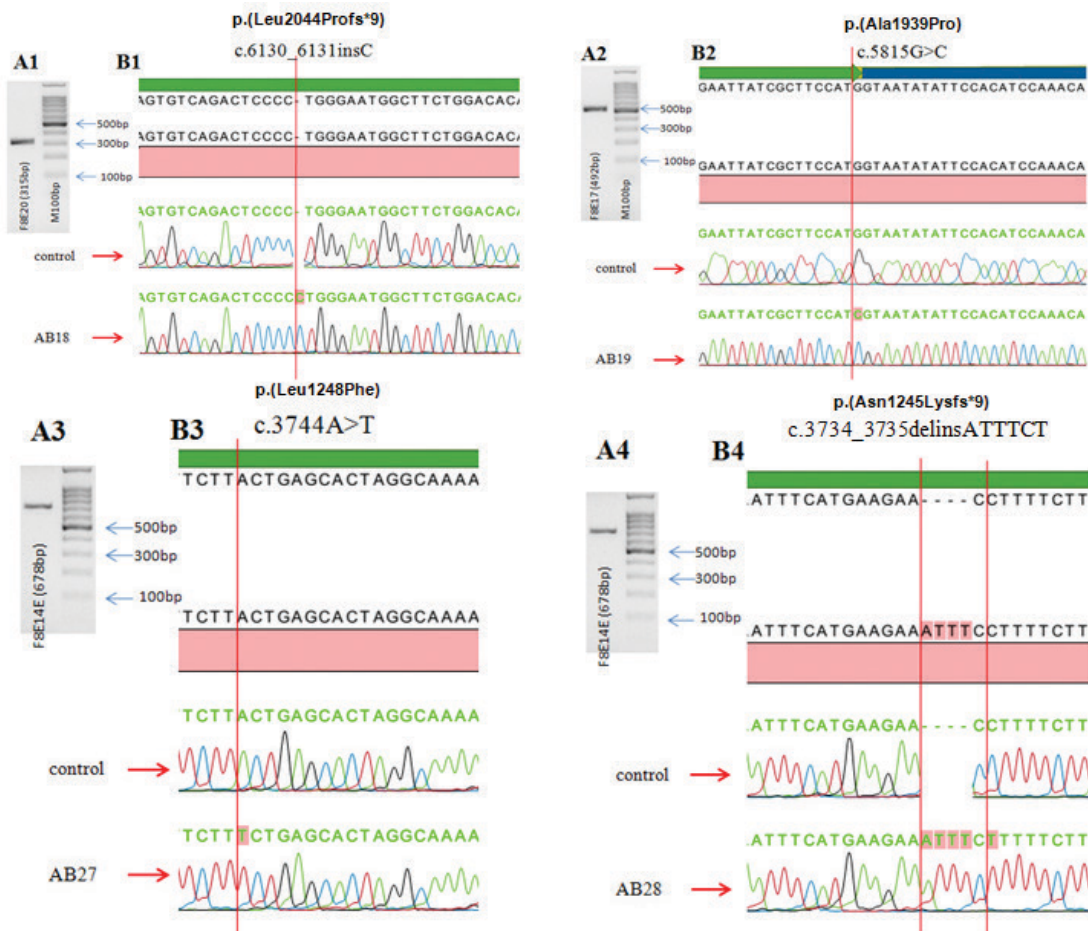


Figure 2. Identified FVIIIa variants

B1-B4) electropherogram for all novel variants [p.(Leu2044Profs*9), p.(Ala1939Pro), p.Leu1248Phe), p.(Asn1245Lys) and p.(Leu1246Phe)] which were identified by NGS and confirmed by a capillary sequencing method.

were missense variants c.5815G>C p.(Ala1939Pro) and c.3744A>T, p.(Leu1248Phe). The two variants c.3734_3735delinsATTCT, p.(Asn1245Lysfs*9) and c.6130_6131insC, p.(Leu2044Profs*9), yielded frameshift variants predicted to produce truncated proteins. For the other two missense variants p.(Ala1939Pro) and p.(Leu1248Phe) at exon14, we carried out an *in-silico* analysis to assess the phenotypic effect of the missense variants, which were predicted as deleterious and highly damaging by the SIFT and Polyphen v2 programs. The protein structural effect analysis of these missense variants was carried out using Project hope. Such analysis allowed to identify that these variants could be critical *F8* gene targets, which can result in hemophilia A and other pathological disorders. In two patients (Pt ID #AB27 and Pt ID #AB28) with the missense variant p.(Leu1248Phe), the discovered mutation introduces a charge at this position which could cause repulsion between the variant residue and neighboring residues. The differences in amino acid properties can disturb this region and thus disturb its function. This residue is part of an interpro domain named Coagulation Factor 5/8 (IPR024715). The wild type and variant amino acids differ in size. This variant amino acid is bigger and thus might lead to protein misfolding. In the second variant, c.5815G>C p.(Ala1939Pro), the 1939Pro residue is bigger than the wild-type residue Ala1939. This variant is located within a domain annotated in UniProtKB as: F5/8 type A3, and Plastocyanin-like 6. The sequence

variant introduces an amino acid with different properties, which can disturb this domain and abolish its function. The effect of this variant was annotated as Hemophilia-A [HEMA (MIM: 306700)]. The effect of this variant is annotated in HEMA as having well-known pathological significance.

Molecular dynamics simulation of native and variant proteins

The resulting domains' structure is a regular non-covalent heterodimer consisting of a heavy chain (A1-A2-B domains) and a light chain (A3-C1-C2 domains). The wild type residues of the two missense variants (leu1248, and Ala1939) (Fig. 3: A1 and B1) in FVIII, as well as in the other thermostable site in variant residues (1248Phe and 1939pro) were focused as critical residues and were substituted utilizing Schrodinger (BioLuminate). Later, minimization of energy of FVIII receding variants were performed at 300 400.516 kJ/mol p.(Leu1248Phe), and 300 485.173 kJ/mol p.(Ala1939pro). Later lowered to -510 112.111 kJ/mol p.(Leu1248Phe) and -489 034.074 kJ/mol p.(Ala1939pro), after completing variants as calculated by BioLuminate. Various polar and non-polar residues were used to screen and base the energy minimization procedures that enabled fixing a refinement cut of 0.00 Å with a reasonable solvent reduction. The consequence of diminishing the limited energy estimation of the β helix variants in wild type (Leu1248 and

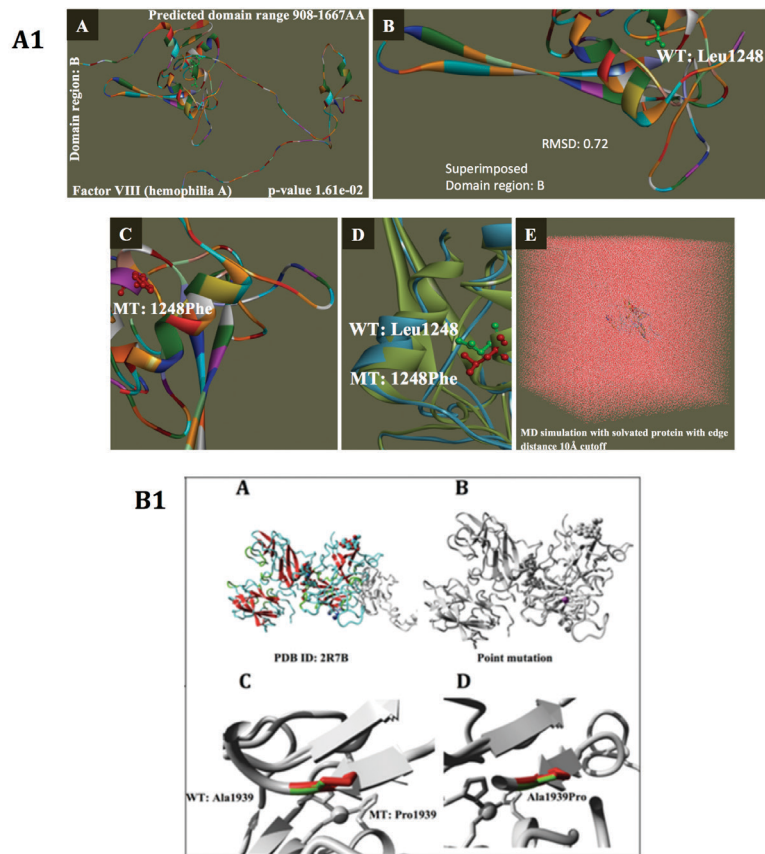


Figure 3. The MD simulation showing octahedron boundary explicit water solvated

(A1) Homology model structure domain region B of the human FVIII protein with structural alterations in the regions due to the variants (Leu1248Phe). (B) Wild type structure of FVIII domain region B with a point variant α Helix 1248 (green), in a ball and stick representation of the helix region. (C) Variant type structure of FVIII domain region B with a variant α Helix 1248Phe (red), a stick representation of the helix region. (D) Wild and Variant type structures superimposed at the FVIII domain region B have wild type residues α Helix Leu1248 (green) and variant residue α Helix1246Phe and 1248Phe (red), made by using CCP4/QTMG. (E) The molecular dynamics simulation used in water box surrounds the entire protein in the middle. The visual inspection also allows identifying the side chain of the histidine residue involved in the hydrogen bonding with surrounding molecules and in that case the δ nitrogen of the histidine (HSB) was a protonated residue. (B1) Structural and functional effect of the novel missense variant c.5815G>C, p.(Ala1939Pro). (A) The protein is colored coded by structural elements: alpha helix-blue, beta strand-red, turn-green, 3/10 helix-yellow and random coil-cyan. Other molecules in the complex are colored grey. (B) The protein is colored grey; the side chain of the mutated residue is colored magenta and shown as small balls. (C) The side chains of both, the wild-type and variant type residues are shown and colored green and red, respectively. (D) Seen from a slightly different angle, the wild-type and variant residues are shown in green and red.

Ala1939) demonstrated that the substitution of (Leu1248 and Ala1939) to (1248Phe and 1939Pro) may add to enhance conservativeness and may increase protein folding. Additionally, two β helix variants depended on the reasonable solvent, as indicated by MD simulations. Amid these reenactments, proteins ceaselessly fold and unfold and give clear understanding for this process. When contrasted with the outcomes at a temperature of 300 K, the deviation of the FVIII protein stretched from 2.0 Å to 10 ns. Interestingly, the deviation of these two variants p.(Leu1248Phe), and p.(Ala1939pro) was kept up at 1.2 Å until the simulation was completed ($t=10$ ns). This demonstrated that it had achieved its folded state while the little peak at 1.2 ns showed that mutation settled the protein structure. The resulting evidence demonstrated that the variant structure was steady and could keep up its adaptation at 300 K, at a pressure bar of 1.00040. Furthermore, with a surface tension of 5,000.0 Å in an aggregate simulation time (ns) of 1.2/slipped were by 0.0, and recording intervals (ps) energy of 1.2. In any case, there were inconspicuous changes that were seen between the wild type and variant FVIII protein. This can be seen when the superimposed structures with

an RMSD estimation of 1.023 Å were resolved by use of MOE. Additionally, the p.(Leu1248Phe), and p.(Ala1939pro) variants of the FVIII protein demonstrated a higher dissolvable and accessible surface area than the 1248Phe and 1939pro variants within the protein structure. The structural discharge of Leu1248 and 1939pro brought about all the same less hydrophobic properties within the protein structure, while addition of these variants could expand the minimization and clearing of the residues in the protein structure.

DISCUSSION

Hemophilia A is an X-linked recessive bleeding disease caused by a deficiency in the FVIII protein. Numerous publications have described various variants in the *F8* gene, however the spectrum and nature of common variants causing hemophilia A in the Arab population, and particularly among the Saudis, is very limited. The study presented here was hence conducted to explore the known and novel variants among the Saudi hemophilia A patients through genetic analysis and to analyze the variants' spectrum. In this study, after anal-

ysis of Intron 22 and Intron 1 inversion (Al-Allaf & Bouazzaoui, 2016), which carries the most variants in the *F8 gene* (Antonarakis *et al.*, 1995), we analyzed samples which were negative for these variants, by using the next generation sequencing method (NGS). From a total of 21 samples analyzed, we found a total of 10 variants in 12 individual DNA samples, out of which 5 are novel variants in 5 individual samples and 5 were previously reported variants in 8 individual samples (Table 1A). This presents a frequency of about 50% in single variants, and is comparable to single variant frequency in HGMD (about 59.8%). Furthermore, clinical data for patients associated with the samples (Table 1B) showed that the patients presented with <1% of FVIII (severe hemophilia) with episodic bleeding, and those patients were on one or more than one treatment regimens.

The functional significance of the *F8 gene* variants is still unknown, although our all-novel variants have so far been associated with the X-linked inherited bleeding disorder. The targeted FVIII protein structure predicted for the region which spans 908-1667AA provides additional detail for these analyses, particularly for variants located within interfaces on the B-region and their nearest residue neighbors in the full-length molecule. In the five novel variants that yield severe disease symptoms and phenotypes of less than 1% of normal circulating FVIII activity, the variations are found on the surface of the B region at positions not known to be involved in binding interactions with membranes. These residues are located in the different domains, especially for the two novel missense variants, p.(Leu1248Phe), and p.(Ala1939pro), that were located on the same B region. These two missense variants located at the A chain, which is highly similar to the PDB: 5IP2 structure, based on the homology modeling. The results of this study are based on MD simulation. These two novel missense variants that introduce differences in amino acid properties can disturb this region and disturb its function. Both of these variants were part of an interpro domain Coagulation Factor 5/8 (IPR024715). The wild type and variant amino acids differ in size. This variant residue is bigger and thus might lead to protein misfolding and changes in hydrophobicity.

Furthermore, several reports are available to substantiate the importance of single amino acid substitutions in hemophilia A at the initiation cleavage sites (Hamaguchi *et al.*, 1991), affecting FVIII binding to the von Willebrand factor (Higuchi *et al.*, 1990). Moreover, from an evolutionary standpoint, variants enhancing a preserved amino acid site are usually thought to have functional significance. Functional analyses based on *in silico* studies by the SIFT and PolyPhen2 programs are in a position to predict about 90–95% of damaging variants. PolyPhen2 and SIFT have shown to predict the effect of over 80% of amino acid substitutions (Adzhubei *et al.*, 2013). The prioritizing changes that are likely to cause a loss of protein function and low specificity predictions with caution and additional confirmation to support the pathogenicity should be sought before reporting novel missense changes.

CONCLUSION

The results obtained from the study presented here should enrich the FVIII variant database with regard to the Saudi Arabian population, however, more samples should be investigated to calculate the frequency of various variants. The results suggest that the protein functional and structural studies using MD simulation

approach could be a reasonable strategy for investigating the structural and functional effect. Furthermore, the finding of this study could be useful in developing a genetic screen for potential hemophilia A patients.

Acknowledgement

We are indebted to the subjects under this study and family members for their cooperation. The authors would like to thank Dr Abdulrahman Ghaleb Al-Ahdal and all the staff of Science and Technology Unit (STU) and Deanship of Scientific Research (DSR) at Umm Al-Qura University for their continuous support. This work was supported by the National Science, Technology, and Innovation Plan (MAARIFAH) of the Kingdom of Saudi Arabia to Dr. Faisal A. Al-Allaf, (Grant Code: 09-BIO920-10).

Author contributions

Conceived and designed the study: FA and AB. Performed the experiments: FA, AB and MT. Analyzed the data: ZA, NB, TO, MT, MA, AE, HA, AA, WE and AB. Wrote the paper: ZA, AB. Principle Investigator: FA. All of the authors read and approved the final manuscript.

Conflict of Interest Disclosures

The authors declare no conflict of interest.

REFERENCES

- Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7: Unit 7.20.
- Al-Allaf FA, Taher MM, Abduljaleel Z, Athar M, Ba-hammam FA, Abdulla M, Bouazzaoui A, Abalkhail H, Owaidah TMA (2016) Mutation screening of the factor VIII gene in hemophilia A in Saudi Arabia: two novel mutations and genotype-phenotype correlation. *J Mol Genet Med* **10**: 2. <https://doi.org/10.4172/1747-0862.1000211>
- Al-Allaf FA, Taher MM, Abduljaleel Z, Bouazzaoui A, Athar M, Bogari NM, Abalkhail HA, Owaidah TM (2017a) Molecular analysis of factor VIII and factor IX genes in hemophilia patients: identification of novel mutations and molecular dynamics studies. *J Clin Med Res* **9**: 317–331. <https://doi.org/10.14740/jocmr2876w>
- Al-Allaf FA, Owaidah TMA, Abduljaleel Z, Mohiuddin M, Taher MM, Athar M, Abalkhail H, Khan W, Bouazzaoui A (2017b) Identification of four novel factor VIII gene mutations and protein structure analysis using molecular dynamic simulation. *J Genet Syndr Gene Ther* **8**: 317. <https://doi.org/10.4172/2157-7412.1000317>
- Antonarakis SE, Rossiter JP, Young M, Horst J, de Moerloose P, Sommer SS, Ketterling RP, Kazazian HH Jr, Négrier C, Vinciguerra C, Gitschier J, Goossens M, Girodon E, Ghanem N, Plassa F, Lavergne JM, Vidaud M, Costa JM, Laurian Y, Lin SW, Lin SR, Shen MC, Lillicrap D, Taylor SA, Windsor S, Valleix SV, Nafa K, Sultan Y, Delpech M, Vnencak-Jones CL, Phillips JA 3rd, Ljung RC, Koumbarelis E, Gialeraki A, Mandalaki T, Jenkins PV, Collins PW, Pasi KJ, Goodeve A, Peake I, Preston FE, Schwartz M, Scheibel E, Ingerslev J, Cooper DN, Millar DS, Kakkur VV, Giannelli F, Naylor JA, Tizzano EF, Baiget M, Domenech M, Altisent C, Tusell J, Beneyto M, Lorenzo JI, Gaucher C, Mazurier C, Peerlinck K, Matthijs G, Cassiman JJ, Vermeylen J, Mori PG, Acquila M, Caprino D, Inaba H (1995) Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood* **86**: 2206–2212. PMID: 7662970
- Boekhorst J, Verbruggen B, Lavergne JM, Costa JM, Schoormans SCM, Brons PPT, van Kraaij MGJ, Novakova IR, van Heerde LW (2005) Thirteen novel mutations in the factor VIII gene in the Nijmegen haemophilia A patient population. *Br J Haematol* **131**: 109–117. <https://doi.org/10.1111/j.1365-2141.2005.05737.x>
- Bolton-Maggs PH, Pasi KJ (2003) Haemophilias A and B. *Lancet* **361**: 1801–1809. [https://doi.org/10.1016/s0140-6736\(03\)13405-8](https://doi.org/10.1016/s0140-6736(03)13405-8)
- Citron M, Godmilow L, Ganguly T, Ganguly A (2002) High throughput mutation screening of the factor VIII gene (F8C) in hemophilia A: 37 novel mutations and genotype-phenotype correlation. *Hum Mutat* **20**: 267–274. <https://doi.org/10.1002/humu.10119>
- Graw J, Brackmann HH, Oldenburg J, Schneppenheim R, Spannagl M, Schwaab R (2005) Haemophilia A: from mutation analysis to

- new therapies. *Nat Rev Genet* **6**: 488–501. <https://doi.org/10.1038/nrg1617>
- Guidelines BC&SH, www.bcsghguidelines.org
- Hamaguchi M, Matsushita T, Tanimoto M, Takahashi I, Yamamoto K, Sugiura I, Takamatsu J, Ogata K, Kamiya T, Saito H (1991) Three distinct point mutations in the factor IX gene of three Japanese CRM+ hemophilia B patients (factor IX BMNagoya 2 factor IX Nagoya 3 and 4). *Thromb Haemost* **65**: 514–520. PMID: 1871712
- Higuchi M, Wong C, Kochhan L, Olek K, Aronis S, Kasper CK, Kazazian HH Jr, Antonarakis SE (1990) Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA. *Genomics* **6**: 65–71. PMID: 2105906
- Hoyer LW (1994) Hemophilia A. *N Engl J Med* **330**: 38–47. <https://doi.org/10.1056/nejm199401063300108>
- Human Gene Mutation Database (accessed on 29th Nov 2015) <http://www.hgmd.org/>
- Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protoc* **4**: 1073–1081. <https://doi.org/10.1038/nprot.2009.86>
- Lenting PJ, Mourik AJ, Mertens K (1998) The life cycle of coagulation factor VIII in view of its structure and function. *Blood* **92**: 3983–3996. PMID: 9834200
- Lillicrap D (1998) The molecular basis of haemophilia B. *Haemophilia* **4**: 350–357. PMID: 9873754
- Mannucci PM, Tuddenham EG (2001) The hemophilias – from royal genes to gene therapy. *N Engl J Med* **344**: 1773–1779. <https://doi.org/10.1056/nejm200106073442307>
- Miller CH, Platt SJ, Rice AS, Kelly F, Soucie JM (2012) Validation of Nijmegen-Bethesda assay modifications to allow inhibitor measurement during replacement therapy and facilitate inhibitor surveillance. *J Thromb Haemost* **10**: 1055–1061. <https://doi.org/10.1111/j.1538-7836.2012.04705.x>
- Nair PS, Shetty SD, Chandrakala S, Ghosh K (2014) Mutations in intron 1 and intron 22 inversion negative haemophilia A patients from Western India. *PLoS One* **9**: e97337. <https://doi.org/10.1371/journal.pone.0097337>
- Peyvandi F, Kunicki T, Lillicrap D (2013) Genetic sequence analysis of inherited bleeding diseases. *Blood* **122**: 3423–3431. <https://doi.org/10.1182/blood-2013-05-505511>
- Repesse Y, Slaoui M, Ferrandiz D, Gautier P, Costa C, Costa JM, Lavergne JM, Borel-Derlon A (2007) Factor VIII (FVIII) gene mutations in 120 patients with hemophilia A: detection of 26 novel mutations and correlation with FVIII inhibitor development. *J Thromb Haemost* **5**: 1469–1476. <https://doi.org/10.1111/j.1538-7836.2007.02591.x>
- Rossetti LC, Radic CP, Larripa IB, De Brasi CD (2008) Developing a new generation of tests for genotyping hemophilia-causative rearrangements involving int22h and int1h hotspots in the factor VIII gene. *J Thromb Haemost* **6**: 830–836. <https://doi.org/10.1111/j.1538-7836.2008.02926.x>
- Salviato R, Belvini D, Radossi P, Tagariello G (2004) Factor VIII gene intron 1 inversion: lower than expected prevalence in Italian haemophilic severe patients. *Haemophilia* **10**: 194–196. PMID: 14962212
- Schroder J, El-Maarri O, Schwaab R, Muller CR, Oldenburg J (2006) Factor VIII intron-1 inversion: frequency and inhibitor prevalence. *J Thromb Haemost* **4**: 1141–1143. <https://doi.org/10.1111/j.1538-7836.2006.01884.x>
- Vidal F, Farssac E, Altisent C, Puig L, Gallardo D (2001) Rapid hemophilia A molecular diagnosis by a simple DNA sequencing procedure: identification of 14 novel mutations. *Thromb Haemost* **85**: 580–583. PMID: 11341489
- White G, Rosendaal F, Aledort L, Lusher J, Rothschild C, Ingerslev J (2001) Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. *Thromb Haemost* **85**: 560. PMID: 11307831