

Identification of Four Novel Factor VIII Gene Mutations and Protein Structure Analysis using Molecular Dynamic Simulation

Faisal A Al-Allaf^{1,3*}, Tarek MA Owaidah⁴, Zainularifeen Abduljaleel^{1,2*}, Mohiuddin M Taher^{1,2}, Mohammad Athar^{1,2}, Halah Abalkhail⁴, Wajahatullah Khan⁵ 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

⁵Department of Basic Sciences, College of Science and Health Professions, King Saud Bin Abdulaziz University for Health Sciences, PO Box 3660, Riyadh 11426, Saudi Arabia

Abstract

Hemophilia A is an X-linked recessive hemorrhagic disorder caused by mutations in the factor VIII gene. To find out known and novel causative mutations in Hemophilia A, we carried out genetic analysis among Saudi patients. Twenty six Patients who were negative for inv-1/inv-22 were selected for analysis with Multiplex Ligation-dependent Probe Amplification (MLPA) and Sanger sequencing. Furthermore the functional and structural effects of the novel mutations were predicted using Molecular dynamic simulation. The results showed three known large deletions in 6 samples (delE8,9,10,11,12,13,14; delE7,8,9,10; and delE4) and twelve mutations in 18 samples; four of them were novel. The novel mutations detected were two substitutions in exon 8 at position c.1021G>C, p.(Ala341Pro) and position c.1183A>C, p.(Lys395Gln), one substitution in exon13 at position c.1930T>A, p.(Leu644Met), and one substitution in exon22 at position c.6322G>C, p.(Ala2108Pro). Clinical data of Patients with novel mutations showed <1% Factor VIII levels (severe hemophilia) with episodic bleeding and were on a routine treatment of plasma derived clotting factor concentrate. This data is in line with MD simulation showed significant changes of the different mutations on the protein structure compared to native protein. These results should enrich the spectrum of mutations and enlarge the factor VIII protein's database in Saudi Arabian population; furthermore it showed that the *in silico* MD simulation for functional and structural studies could be a reasonable approach for investigating the advance genotype-phenotype correlation.

Keywords: Hemophilia A; Novel mutation; Saudi Arabia; Blood coagulation; Molecular dynamic (MD) simulation

Introduction

Hemophilia A is a recessive bleeding disease caused due to the deficiency in factor VIII (FVIII; hemophilia A) [1]. The protein is encoded by F8 gene and considered among one of the largest genes located on the long arm of chromosome X (Xq28 position, comprising of 25 introns and 26 exons organized in five domains, A1, A2 and A3 and C1-C2). The region for F8 gene is characterized with a high GC content and within the 9.1 kb coding region there are about 70 CpG dinucleotides. This resulted into hyper-mutation and approximately 30% of variants are usually novel [2]. A total of 2320 mutations so far have been reported in F8 gene in the Human Gene Mutation Database (HGMD) [3]. The most frequent mutation has been found in Intron 22 inversion with a frequency rate of 42% among individuals with severe haemophilia-A [4]. The second most frequent mutation has been observed in intron 1 as an inversion with a prevalence of about 1–5% among haemophilia-A patients [5,6]. A total of 1388 missense/nonsense, 377 small deletions, 156 splicing, 117 small insertions, 27 small indels, 213 gross deletions, 26 gross insertions, 12 complex rearrangements, and four regulatory mutations have so far been detected [3,7]. Several studies have been conducted describing the mutations in the Western populations [8–12], however, the spectrum and nature of common mutations carrying haemophilia A in Arab population is still lacking investigation. Specifically the data available for frequent mutations is insufficient among Saudi population. Hence, the present study was designed to explore the known and novel mutations among Saudi population through genetic analysis of hemophilia A patients and examination of the mutation spectrum.

Materials and Methods

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 factor VIII 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. The hemophilia A patients were selected based on the criteria and guidelines as indicated by the British Committee for

***Corresponding author:** Faisal A Al-Allaf, Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P.O. Saudi Arabia ox 715, Makkah 21955, Saudi Arabia, Tel: +966509897560; E-mail: faallaf@uqu.edu.sa

Abdellatif Bouazzaoui, Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, Saudi Arabia P.O. Box 715, Makkah 21955, Saudi Arabia, Tel: +966571297636; E-mail: alazzauoi@uqu.edu.sa; ab1971@hotmail.de

Zainularifeen Abduljaleel, Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P.O. Box 715, Makkah 21955, Saudi Arabia, Tel: +966570552361; E-mail: zaabdujaleel@uqu.edu.sa; zainulbio@gmail.com

Received February 09, 2017; **Accepted** February 20, 2017; **Published** February 27, 2017

Citation: Al-Allaf FA, Owaidah TMA, Abduljaleel Z, Taher MM, Athar M, et al. (2017) Identification of Four Novel Factor VIII Gene Mutations and Protein Structure Analysis using Molecular Dynamic Simulation. J Genet Syndr Gene Ther 8: 317. doi: 10.4172/2157-7412.1000317

Copyright: © 2017 Al-Allaf FA, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Standards in Hematology. The genomic DNA used to be 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.

Multiplex ligation probe-dependent amplification (MLPA)

Multiplex ligation probe-dependent amplification (MLPA) of exons in factor VIII gene was performed using SALSA MLPA Kit, P178-F8 B2 (MRC Holland, Amsterdam, the Netherlands) on the DNA of samples following the manufacturer's instruction. The DNA from four male unrelated individuals without family history of hemophilia A was included as the control. The data obtained from the ABI Prism 3500 Genetic Analyser (Applied Biosystems, Foster City, CA, USA) were analyzed using Coffalyser software (MRC Holland, Amsterdam, the Netherlands).

Polymerase chain reaction (PCR) amplification and Capillary Sequencing

Polymerase chain reaction (PCR) amplification of exons for factor VIII hemophilia gene was executed 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 10x PCR buffer, 2 μ l 25 mM $MgCl_2$, 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 [13,14]. 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 30 s, extension at 68°C for 1 min and final extension at 68°C for 5 min. The amplified product was separated on 1.5% Agarose gel to ensure the size and quality of the band. The PCR products were purified by magnetic beads method using Agencourt AMPure 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 EDTA/Ethanol method and 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 Applied Biosystems of Applied Biosystems Corporation subsidiaries in the US.

Protein structure modeling and identification of coding SNPs

The single amino acid polymorphism database (SAAP) [15] and dbSNPs were utilized to recognize the protein and for the crystal structure analysis of coagulation Factor VIII (PDB ID: 2R7E) encoded by F8 gene and identification of single point mutation residue location. The particular mutation residue damaging effects were confirmed with PolyPhen2 [16] and SIFT programs [17]. The energy minimization of the mutant protein structure was executed by way of ANOLEA (Atomic Non-Local Environment Assessment), a server that carried out energy calculations of a protein sequence and investigated the "Non-Local Environment" (NLE) of every heavy atom within the molecule [18]. The energy of each and every pairwise interaction in this non-local environment used to be get from a distance-dependent knowledge-based on the mean force abilities as a result of a database that conducts energy calculations at the atomic level for protein structures. The calculations had been executed on the non-local interactions for all the heavy atoms of the tolerable amino acids actual in the molecule and present an energy value for every amino acid of the protein as output energy profile. High-power zones (HEZs) within the profile have been associated with errors or with potential interacting zones of proteins. The application yields the structure in three dimensions, making a choice on the high-energy amino acids in the protein.

In silico solvent accessibility of amino acid residues

Accessible surface area (ASA) used to exhibit the solvent accessibility of amino acid residues in protein structure. The protein structure was retrieved from DSSP (database of secondary structure assignments) [14-18], this program that calculates DSSP entries from PDB (protein data bank), the entry implementation of ASA view for all protein or individual chain. The criterion used to add chain or PDB query in the input file, for generating ASA View plot for co-ordinate or PDB file. The amino acid residues under the solvent conditions confirmed three cons of solvent accessibility, i.e., buried, partially buried and exposed, indicating, low, moderate and excessive accessibility relatively [19]. The protein's secondary structure is principal for studying the association between amino acid and protein structure. The DSSP application recognizes secondary structure, geometrical appearances and solvent exposure of proteins in Protein Data Bank layout.

Functional effect and stability analysis

For the bulk of mutant variants (single amino acid changes or nsSNPs) in humans, it's have an effect on protein function is unknown. This procedure provides binary classifications (impact/neutral) attended through extra detailed score. In addition, we gained insights about the protein's stability using the program Schrodinger, USA, considering the mutant stability predictions on a protein of unknown structure. In case of the mutant variant, the region of the mutated residue was designated, apart from the wild type amino acids. Several known disease causing nsSNPs with recognized 3D-dimensional protein structures have structural influences on key residues and sites, which are associated with protein function. Moreover, the predictions of the functional effects dimension by way of the Screening for Non-suited Polymorphisms (SNAP) were compared [17]. SNAP scores ranged from -100 (strongly predicted as neutral) to 100 (strongly estimated to alter perform); the distance was directly correlated with the binary determination boundary (0), which measures the impact reliability [18] to illustrate the disease-associated mutations which could potentially affect the protein interactions as well [19]. Protein function will typically be associated with evolutionarily conserved residues [20]. A dangerous signal corresponds to a mutation as a way to be expected as being stabilizing. Changes in the folding free energy upon mutation ($\Delta\Delta G$) will aid the support that a mutation will impact in a disease specifically because it damages a main area of the protein. SNAP scores will relate to further realistic effects [21].

Molecular dynamics (MD) simulation and energy minimization

The molecular dynamics (MD) simulations have been applied by making use of Biomolecular simulation application CHARMM, an academic research application most likely used for macromolecular mechanics and dynamics with resourceful analysis and manipulation tools of atomic coordinates and dynamics trajectories. CHARMM-GUI (<http://www.Charmm-gui.Org>), offers an online-established graphical user interface to analyze a range of enter input data and molecular programs to standardize and make use of common as well as evolved simulation techniques in CHARMM. Even though on this simulation, the solvated system was neither minimized nor equilibrated, nevertheless 0.15 M ions can also be brought in the simulation field through specifying ions (KCl) and concentration (C). The numbers of ions automatically established by the ion-obtainable volume (V), the whole charge of the system (Qsys) and via the positive ion (z+) valency, to neutralize the complete procedure charge, (z+N+-N=-Qsys). The ion-accessible volume, V, used to be anticipated by using subtracting

molecular volume from the complete system volume, identifying ion-placing system of Monte Carlo. The preliminary configuration of ions was confirmed utilizing short Monte Carlo (MC) simulations with a primitive model, for instance van der Waals interactions. The solvation free energy was expressed as non-polar and electrostatic contributions; however the nonpolar contribution was partitioned as repulsive and dispersive contributions utilizing the weeks. To scale down the computational attempt, the free energy simulations were carried out with few specific solvent water molecules in close vicinity to the solute, even as the effect of the rest of the solvent mass was once proven implicitly as an effective solvent boundary potential (SSBP). KCl used to be integrated to neutralize the total negative charge of the structures. Molecular dynamics simulations had been carried out with a 2 fs time step at a consistent temperature of 300 K and a constant pressure of 1 atm below periodic solvent boundary stipulations. The Particle Mesh Ewald process [22] used to be utilized for electrostatics and a 12-Å cutoff was used for van der Waals interactions.

Results

MLPA analyses of large deletion/insertion

To find out the disease-causing gene variants, DNA samples from patients who were negative for inv-1 and inv-22 were used for MLPA method to screen for large deletion or insertion in factor VIII gene. The results showed that 6 patients exhibited large known deletions in various exons. Figure 1 showed large deletions of E8,9,10,11,12,13,14 (delE8-14). Other patients also showed large deletions in E7,8,9 and 10 (delE7-10) whereas one individual had a large deletion in E4 (delE4). The clinical data for patients with delE8-14 and delE7-10 showed <1% of factor VIII (severe hemophilia) along with episodic bleeding and patients were on clotting factor concentrate extracted from plasma or having more than one treatment. Furthermore, one patient (Pt ID#2) with delE8-14 showed presence of chronic joint disability on the left ankle and right elbow (Tables 1A and 1B).

Mutations analysis using capillary sequencing

After the MLPA analysis, 21 samples of Saudi Arabian patients were sequenced for selected exons of factor VIII. Out of the 21 DNA

samples screened, we found a total of twelve mutations in 18 samples, out of which four were novel mutations whereas eight of them were known (Table 2). The electrophoregram in Figure 2 showed the novel mutation c.1021G>C, p.(Ala341Pro) which we founded in the exon 8 of one 30 years old male patient (Pt ID #AB24). This G>C mutation occurred in GCT of alanine to become CCT (proline) and this is present in A1 domain of FactorVIII, both amino acid are neutral amino acid however this mutation is associated with severe form of haemophilia. In another 41 years old mal patient (Pt ID #AB31) we founded also in exon 8 a substitution at c.1183A>C, p.(Lys395Gln) corresponding to the A2 domain of Factor VIII. This mutation was in the genetic code AAG coding for lysine which is a positively charged (basic) amino acid and altered it into CAG coding for glutamine which is neutral amide amino acid which conducted to structural modification resulted in the loss of function. The Third novel mutation was found in exon 13 of one 20 years old patient (Pt ID #AB34) at c.1930T>A position the substitution occurred in TTG to become ATG and it is present in A2 domain of Factor VIII. This mutation altered leucine into methionine p.(Leu644Met) and both amino acids are non polar aliphatic amino acid. The last novel mutation at c.6322G>C, p.(Ala2108Pro) was found in the exon 22 of one 31 years old male patient (Pt ID #AB41). This mutation is present in C1 domain of FactorVIII and occurred in GCC of alanine to become CCC (Proline). Both amino acid are neutral amino acid but the function of HF8 protein was missed as the clinical data showed (Table 2A). Except one patient (Pt ID# 43) how showed mild hemophilia, all patiente with novel or reported mutations present less than 1% of Factor VIII levels (severe hemophilia) with episodic bleeding and they were on plasma derived clotting factor VIII concentrat or on treatment with recombinant factor VIII. Furthermore, some of them showed presence of chronic joint disability (Table 2A).

Causative mutations predicted by SIFT and PolyPhen2

In this study, the sequence homology-based tool SIFT was used to determine the level of conservation for a particular amino acid position in the protein. The result was shown to be inversely proportional to the tolerance and observed mutations were predicted as 'DAMAGING' with a score ≤ 0.05 (Table 2B). The SIFT program distinguish the disease associated SNPs from the neutral ones with only 20-25% false positive error rate and were comparable to homologs in protein sequence databases. PolyPhen2 program predicts the fate of the structural and functional aspect of a protein resulting from amino acid alterations in the sequence based on the precise empirical rules. The empirical rules applied on the sequences show 72% accuracy with 8% false positive prediction. Based on the results of PolyPhen2, it is concluded that all novel variants p.(Lys395Gln), p.(A2108P), p.(A341P), and p.(L644M) was a possible damaging type.

Patient ID	Gender	Age (years)	Status	Exon	T. o mutation
AB1	Male	21	Reported	E8,9,10,11,12,13,14	del
AB2	Male	16	Reported	E8,9,10,11,12,13,14	del
AB3	Male	11	Reported	E8,9,10,11,12,13,14	del
AB4	Male	11	Reported	E8,9,10,11,12,13,14	del
AB5	Male	9	Reported	E7,8,9,10	del
AB6	Male	nd	Reported	E4	del

Table 1A: Large deletion variants identified in hemophilia-A patients.

Patient ID	Hgb	MCV	Platelet	PTT	F8	F8 inh	BFA	TOC	TPU	PChJDi	SPECIFY
AB1	148	79.2	246	83.4	<0.01	11.8	Severe Hemophilia (<1%)	Episodic	More than one	No	
AB2	121	77.2	418	102	<0.01	341.3	Severe Hemophilia (<1%)	Prophylaxis	Plasma Derived Clotting Factor Concentrate	Yes	Lt. Ankle & Rt. Elbow
AB3	128	79.6	490	142	<0.01	39	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	No	
AB4	101	85.7	482	115	<0.01	151.6	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	No	
AB5	13.3	75.4	379	108.5	<0.01	0.38	Severe Hemophilia (<1%)	Episodic	Recombinant Clotting Factor Concentrate	nd	
AB6	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	

Abbreviations: T. o mutation: Type of Mutation; N. P: Nucleotide Position; A. a: Amino Acid; DOB: Date of Birth; BFA: Baseline Factor Activity; TOC: Type of Care; TPU: Treatment Products Used; Hgb: Hemoglobin; MCV: Mean Cell Volume; PTT: Partial Thromboplastin Time; F 8: Factor VIII; F 8 inh: Factor VIII Inhibitor, PDCFC: Plasma Derived Clotting Factor Concentrate; PChJDi: Presence of Chronic Joint Disability

Table 1B: Clinical characteristics of the studies patients with large deletion.

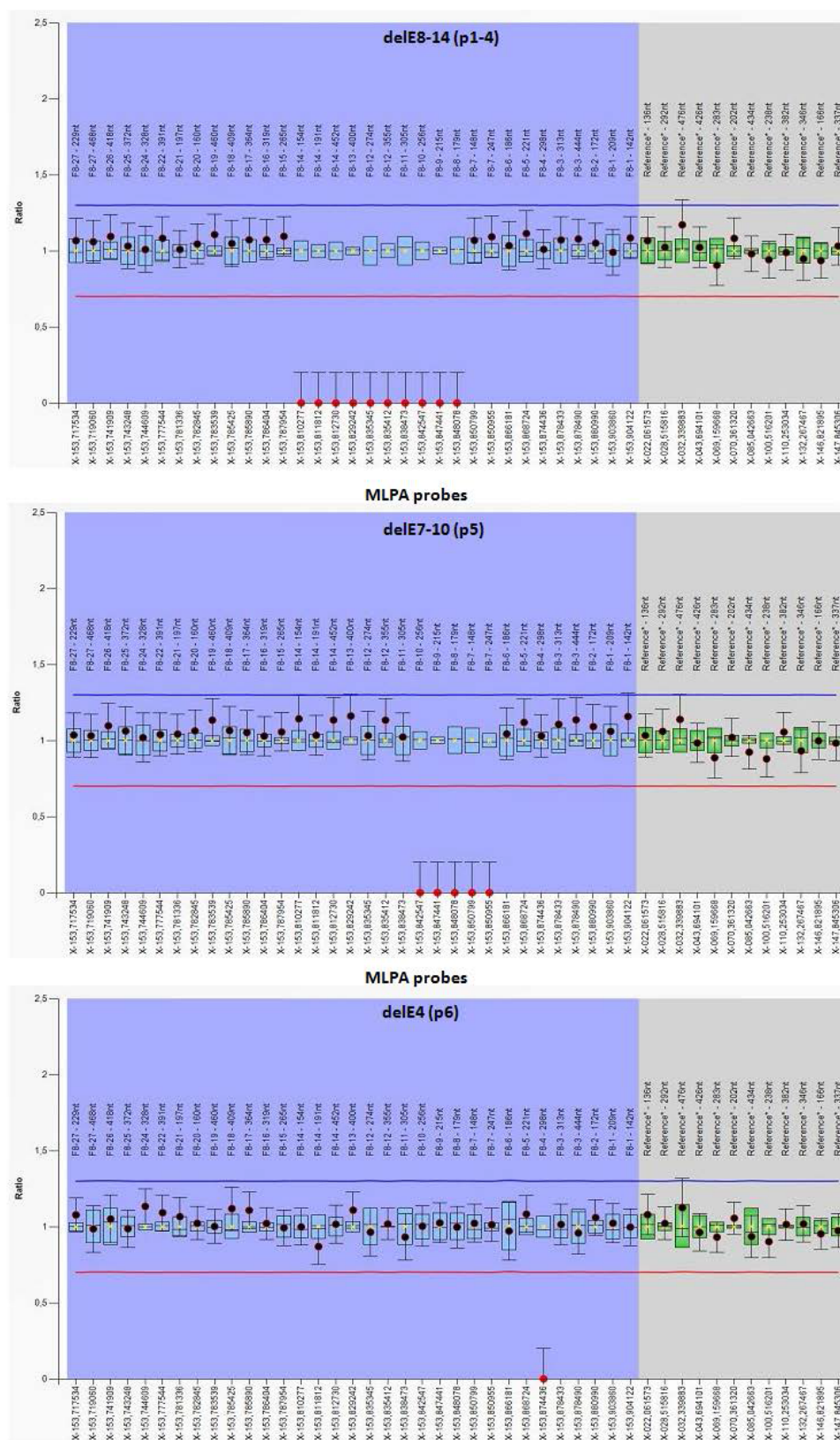


Figure 1: Results of MLPA analysis in factor VIII gene.

Plots of dosage ratio of multiplex ligation probe-dependent amplification probes in exons of Factor VIII gene for family I (p1-4), family II (p5) and family III (p6). The dosage ratios of exons E8,9,10,11,12,13,14; exons E7,8,9,10 and exon E4 probe are 0 for FI (p1-4), FII (p5) and FIII (p6), respectively

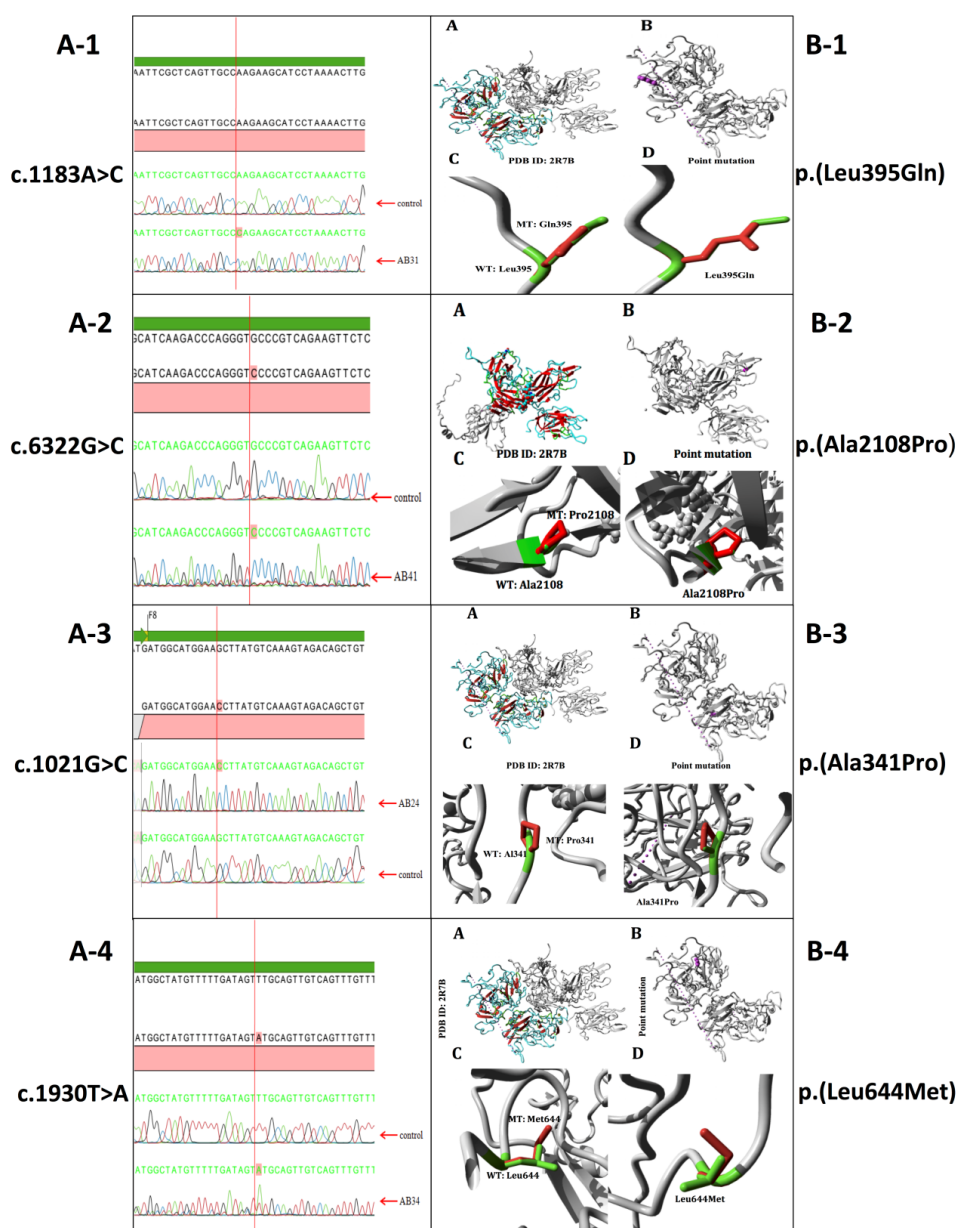


Figure 2: Identified mutations on protein structure modeling.

A1-A4) all novel variants [c.1183A>C, p.(K395Q)], [c.6322G>C, p.(A2108P)], [c.1021G>C, p.(A341P)] and [c.1930T>A, p.(L644M)] were founded by capillary. B1-B4) Protein was illustrated common for all mutations colored by element; alpha helix-blue, beta stand-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 and mutant type residues are shown and colored green and red respectively. D) Seen from a slightly different angle of both wild and mutant residues shown green and red

MD simulation and solvent accessibility

The factor novel mutations that cause substitute in amino acid may significantly moderate the stability of the protein structure; for this purpose protein structural data furnish useful comprehensive understanding of its functionality. The factor VIII gene novel mutations, i.e., p.(Lys395Gln), p.(Ala2108Pro), p.(Ala341Pro), and p.(Leu644Met) is not available in the single nucleotide polymorphism database (dbSNP) (<http://www.ncbi.nlm.nih.gov/SNP>). Additional confirmation for this mutation was obtained from the Human Genome

Variation database (HGVBASE) (<http://www.hgvs.org/>) along with the evidence-indicating gene coding regions and the location of all novel variants. The factor VIII crystal protein structure was available for the first novel mutation Lys395Gln was located at position 395. The amino acid position was present near the (αHelix2-K395Q) where “Gln” was the mutant residue instead of “Lys”. The second novel mutation Ala2108Pro was located at 2108. The amino acid was near the (αHelix5-A2108P) where “Ala” was mutated instead of “Pro”. The third and fourth novel mutations p.(Ala341Pro); p.(Leu644Met) were located at loop region. The hemophilia risk associated protein structure

Patient ID	Hgb	MCV	Platelet	PTT	F8	F8 Inh	BFA	TOC	TPU	PChJDi	SPECIFY
AB7	98	57.4	428	67	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Rt. Ankle
AB8	102	54.7	361	86.6	<0.01	0	Severe Hemophilia (<1%)	Episodic	More than one	n.d	
AB11	168	79.7	240	104	<0.01	263.5	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Both Knee
AB15	127	81.5	237	57.7	<0.01	0	Severe Hemophilia (<1%)	Prophylaxis	Plasma Derived Clotting Factor Concentrate	Yes	Lt. Knee, Rt and Lt. Elbow
AB16	139	80.1	352	111	<0.01	0	Severe Hemophilia (<1%)	Prophylaxis	More than one	Yes	Rt. and Lt. Ankle
AB24	153	86.6	182	113	<0.01	0	Severe Hemophilia (<1%)	Prophylaxis	recombinant factor VIII	No	
AB26	167	84.2	207	105	<0.01	2.4	Severe Hemophilia (<1%)	Episodic	recombinant factor VIII	Yes	Rt. and Lt. Ankles
AB31	154	87	412	71.3	N.A	N.A	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	No	
AB34	115	73.4	392	101	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Rt. Knee
AB36	159	85.3	161	103	<0.01	0	Severe Hemophilia (<1%)	Episodic	Fresh Frozen Plasma	Yes	Both elbows and Rt. Knee
AB39	151	90.1	273	95.7	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Rt. Knee, Lt. elbow, Lt. Ankle
AB41	155	77.8	252	68.1	<0.01	35.8	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Rt. Elbow
AB43	166	82.4	231	52.4	0.16	0.23	Mild Hemophilia (5-40%)	Episodic	Plasma Derived Clotting Factor Concentrate	No	
AB46	132	85.3	310	64.9	<0.01	0	Severe Hemophilia (<1%)	Prophylaxis	Plasma Derived Clotting Factor Concentrate	Yes	Rt. Knee
AB48	128	76.8	355	65.9	<0.01	0	Severe Hemophilia (<1%)	Prophylaxis	Recombinant Clotting Factor Concentrate	Yes	Lt. Ankle
AB49	131	79.2	432	107.3	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	No	
AB50	149	87.4	264	89.5	<0.01	0	Severe Hemophilia (<1%)	Episodic	Plasma Derived Clotting Factor Concentrate	Yes	Rt. and Lt. Ankles
AB76	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d

Table 2A: Clinical characteristics of the studies patients with single mutations.

Patient ID	Gender	Age(year)	Exon	Status	N. P	A. a changes	SIFT/PolyPhen2
AB7	Male	16	E25	reported	c.6769A>G	p.(Met2257Val)	DAMAGING
AB8	Male	19	E25	reported	c.6769A>G	p.(Met2257Val)	DAMAGING
AB11	Male	32	E14	reported	c.5143C>T	p.(Arg715')	DAMAGING
AB15	Male	24	E2	novel	c.209delTTGT	p.(Phe70')	DAMAGING
AB16	Male	18	E6	reported	c.760A>G	p.(Asn254Asp)	DAMAGING
AB24	Male	30	E8	novel	c.1021G>C	p.(Ala341Pro)	DAMAGING
AB26	Male	29	E4	reported	c.409A>C	p.(Thr137Pro)	DAMAGING
AB31	Male	41	E8	novel	c.1183A>C	p.(Leu395Gln)	DAMAGING
AB34	Male	20	E13	novel	c.1930T>A	p.(Leu644Met)	DAMAGING
AB36	Male	34	E2	novel	c.209delTTGT	p.(Phe70')	DAMAGING
AB39	Male	23	E12	reported	c.1835G>C	p.(Arg612Pro)	DAMAGING
AB41	Male	31	E22	novel	c.6322G>C	p.(Ala2108Pro)	DAMAGING
AB43	Male	26	E7	reported	c.923C>T	p.(Ser308Leu)	TOLERATED
AB46	Male	15	E23	reported	c.6545G>A	p.(Arg2182His)	DAMAGING
AB48	Male	18	E23	reported	c.6545G>A	p.(Arg2182His)	DAMAGING
AB49	Male	13	E23	reported	c.6545G>A	p.(Arg2182His)	DAMAGING
AB50	Male	32	E23	reported	c.6545G>A	p.(Arg2182His)	DAMAGING
AB76	n.d	n.d	E23	reported	c.6545G>A	p.(Arg2182His)	DAMAGING

Abbreviations: T. o mutation: Type of Mutation; N. P: Nucleotide Position; A. a: Amino Acid; DOB: Date of Birth; BFA: Baseline Factor Activity; TOC: Type of Care; TPU: Treatment Products Used; Hgb: Hemoglobin; MCV: Mean Cell Volume; PTT: Partial Thromboplastin Time; F 8: Factor VIII; F 8 inh: Factor VIII Inhibitor; PDCFC: Plasma Derived Clotting Factor Concentrate; PChJDi: Presence of Chronic Joint Disability

Table 2B: Single mutations identified in hemophilia-A patients.

before all novel mutations was obtained from the SWISS-PORT and using the program CCP4 (QtMG). The protein structure after the novel mutations showed higher energy value compared to the native protein structure. The KoBaMIN program was used to minimize the energy of the mutated structure (-146.850 kJ/mol; score 0.80) compared to the native structure energy (36.584 kJ/mol; score: -3.33). After energy minimization enhancements the structure using the potential of mean force showed an RMSD with 0.2461 Å. Moreover, CHARMM and Gromacs were used for MD simulations to observe the implications of the mutation by comparing the native and mutated structures under solvation. The solvate was acquired successfully around the factor VIII

protein structure with water and exhibited a fully solvated molecule with edge distance 10.0. The results indicate $\Delta\Delta G$ -1.73kcal/mol energy before mutation, which was stabilizing, however after the mutation with all mutations p.(Lys395Gln), p.(Ala2108Pro), p.(Ala341Pro), and p.(Lys644Met) the energy was higher ($\Delta\Delta G$ 2.03 kcal/mol) and was possibly destabilizing. The native and mutant structures when superimposed showed an RMSD with 1.04 Å.

Discussion

Hemophilia A is an X-linked recessive bleeding disease caused due to a deficiency in factor VIII protein. Numerous publications have

described various mutations in factor VIII gene, however the spectrum and nature of common mutations causing hemophilia-A in the Arab population particularly among Saudis is very limited [23-25]. The present study was hence conducted to explore the known and novel mutations among Saudi hemophilia patients through genetic analysis and analyze the mutation spectrum. In this study, after the analysis of Intron 22 and Intron 1 inversion [26], which carries the most mutation in Hemophilia factor VIII gene [4], we analyzed samples, which were negative for these mutations using MLPA and Sanger methods. From a total of 26 samples analyzed, we found three large deletions in 6 samples, which were present at a frequency of about 11.54% compared to the 9.18% in HGMD (213 large deletions in 2320 total cases) and 18 single mutations (69.23%) versus 59.8% present in the HGMD. The results with regard to the large deletion frequency and the single mutations observed in our study were comparable to HGMD; furthermore we found four novel mutations from the 18 mutations. This present a frequency about 22.22% which is comparable to early publication [2]. The clinical data of the samples with single mutation in Table 2B showed that the patient present <1% of factor VIII (severe hemophilia) with episodic bleeding and the patients were on one or more than one treatment regimens. In the next step we investigated the relationship between predicted consequences of nsSNPs using *in silico* approaches based on the evolutionary biology, protein structure, and identification of real phenotypes variants confirmation by Sanger sequencing method. The 3D structure of factor VIII by X-ray crystallography PDB ID: 2R7E [27] and modeling studies made it possible to predict the effects of nsSNPs at the structural level by mapping them on corresponding structures. The functional consequences of most factor VIII gene SNPs are still unknown, although our all one novel missense variants p.(Lys395Gln), p.(Ala2108Pro), p.(Ala341Pro) and p.(Leu644Met) have so far been associated with X-linked inherited bleeding disorder. In vivo and in vitro studies on the function of nsSNPs have found that the genetic mutations in factor VIII gene were responsible for Haemophilia-A. Several reports are available to validate the importance of single amino acid substitutions in Haemophilia-A at activation cleavage sites [27], affecting factor VIII binding to von Willebrand factor [28]. It is seemly clear that implementation of the molecular evolutionary proceeding may be strong tools for prioritizing SNPs to be genotyped for future molecular epidemiological studies. Furthermore, from an evolutionary perspective, SNPs enhancing a conserved amino acid site are usually served to have functional importance. *In silico* instruments like SIFT and PolyPhen2 are in a position to predict about 90-95% of damaging SNPs. PolyPhen2 and SIFT application have shown to efficiently predict the effect of over 80% of amino acid substitutions [15]. The solvent accessibility prediction has previously been used to study the predicted tertiary structure of proteins [29]. Our result showed that the solvent accessibility (Acc) modeling altered the protein structure, indicating that the Acc for the mutant structures with novel variants was 39.01 (31%) compared to the 30.82 (30%) for the wild structure. The structural weakness was considerably higher than the normal, indicating that the mutant may result in the alteration of the critical functioning of a protein when it is un-stabilized. Other reports also supported similar results observed during experimental data analyses [30,31]. Disease causing mutations typically occur the protein (buried) and at hydrogen bonding residues [32]. In protein kinases, comparisons show up to cluster within the functionally fundamental catalytic core [17], adopted by using residue scanning to fix polar and neutral residues for analysis of the protein stability and solvent accessibility. To observe the impact of the novel variants on the protein structure and its consequences on the function of the protein, we analyzed the native and mutant structures at different time scales. The results clearly showed a progress loss of helix of

the native structure compared to the mutant structure till the end of the simulation as a result of energy minimization. The results indicate that all novel causative mutations caused an increase in the α -helical content in the protein structure along with total absence of β -sheet amino acid residues. The mutant structure energy was -146.850 kJ/mol (score 0.80) compared to the native structure energy 36.584 kJ/mol (score: -3.33). Furthermore, following the energy minimization improvements the structure based on the potential mean force resulted in RMSD score 0.2461 Å but the mutations due to p.(Lys395), p.(Ala2108), p.(Ala341) and p.(Leu644) resulted in higher energy ($\Delta\Delta G$ 2.03 kcal/mol) and was destabilizing. The RMSD score of 1.04 Å was observed when the two structures (native and mutant) were superimposed. Earlier studies in Detailed structural information of factor VIII is limited which is based homology model for factor VIII in this mutagenesis and functional study has proved a useful resource in that mutation at approximately half of the target residues predicted to participate in hydrogen bonding affected protein stability. During the preparation of this article, a 3.75-Å crystal structure of the B-domain less, human factor VIII procofactor was reported [27]. Furthermore, this approach applied to other bonding interactions should yield insights into additional structural changes that occur during conversion of the procofactor to cofactor and may identify other gain-of-function variants that could have application for therapeutic use.

Conclusion

The results obtained from the present study should enrich the factor VIII mutation database with regard to Saudi Arabian population however more samples should be investigated to calculate the frequency of various mutations. 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 genetic screen for potential haemophilia-A patients.

Acknowledgement

We are indebted to the subjects under this study and family members for their cooperation. The authors would like to thank the Science and Technology Unit at Umm Al-Qura University for the 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, MA 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 interests.

References

1. Au PK, Lin CS, Lau ET, Tang MH (2015) Characterization of a novel mutation in F8 gene causing severe haemophilia A by deletion mapping with STS markers. *Haemophilia* 21: e136-139.
2. 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.
3. Human Gene Mutation Database (2015).
4. Antonarakis SE, Rossiter JP, Young M, Horst J, de Moerloose P et al. (1995) Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood* 86: 2206-2212.
5. Salvati 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.

6. 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.
7. Peyvandi F, Kunicki T, Lillicrap D (2013) Genetic sequence analysis of inherited bleeding diseases. *Blood* 122: 3423-3431.
8. Boekhorst J, Verbruggen B, Lavergne JM, Costa JM, Schoormans SC, et al. (2005) Thirteen novel mutations in the factor VIII gene in the Nijmegen haemophilia A patient population. *Br J Haematol* 131: 109-117.
9. Repesse Y, Slaoui M, Ferrandiz D, Gautier P, Costa C, et al. (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.
10. Graw J, Brackmann HH, Oldenburg J, Schneppenheim R, Spannagl M, et al. (2005) Haemophilia A: from mutation analysis to new therapies. *Nat Rev Genet* 6: 488-501.
11. 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.
12. Citron M, Godmilow L, Ganguly T, Ganguly T (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.
13. 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.
14. Cavallo A, Martin AC (2005) Mapping SNPs to protein sequence and structure data. *Bioinformatics* 21: 1443-1450.
15. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7: 20.
16. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4: 1073-1081.
17. Bromberg Y, Rost B (2007) SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res* 35: 3823-3835.
18. Bromberg Y, Ovberton J, Vaisse C, Leibel RL, Rost B (2009) In silico mutagenesis: A case study of the melanocortin 4 receptor. *Faseb J* 23: 3059-3069.
19. Torkamani A, Verkhivker G, Schork NJ (2009) Cancer driver mutations in protein kinase genes. *Cancer Lett* 281: 117-127.
20. Wang Z, Moulton J (2001) SNPs, protein structure, and disease. *Hum Mutat* 17: 263-270.
21. Bromberg Y, Rost B (2008) Comprehensive *in silico* mutagenesis highlights functionally important residues in proteins. *Bioinformatics* 24: i207-212.
22. Hockney RW (1981) Computer simulation using particles. New York: MacGraw-Hill.
23. Fang H, Wang L, Wang H (2007) The protein structure and effect of factor VIII. *Thromb Res* 119: 1-13.
24. Thompson AR (2003) Structure and function of the factor VIII gene and protein. *Semin Thromb Hemost* 29: 11-22.
25. Al-Allaf FA, Taher MM, Abduljaleel Z, Athar M, Ba-hammam FA, et al. (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.
26. Shen BW, Spiegel PC, Chang CH, Huh JW, Lee JS, et al. (2008) The tertiary structure and domain organization of coagulation factor VIII. *Blood* 111: 1240-1247.
27. Hamaguchi M, Matsushita T, Tanimoto M, Takahashi I, Yamamoto K, et al. (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.
28. Higuchi M, Wong C, Kochchan L, Olek K, Aronis S, et al. (1990) Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA. *Genomics* 6: 65-71.
29. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577-2637.
30. Zhang KH, Li Z, Lei J, Pang T, Xu B, et al. (2011) Oculocutaneous albinism type 3 (OCA3): Analysis of two novel mutations in TYRP1 gene in two Chinese patients. *Cell Biochem Biophys* 61: 523-529.
31. Simeonov DR, Wang X, Wang C, Sergeev Y, Dolinska M, et al. (2013) DNA variations in oculocutaneous albinism: An updated mutation list and current outstanding issues in molecular diagnostics. *Hum Mutat* 34: 827-835.
32. Sunyaev S, Ramensky V, Bork P (2000) Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet* 16: 198-200.