



Original Article

Identification of Diagnostic SNPs in Exon 28 of the Von Willebrand Factor Gene for Diagnosis of VWD Type 2B in Patients Referred to IBTO Coagulation Lab: In-silico Approach

Roshanak Shamriz ¹ M.Sc., Kamran Atarodi ² Ph.D., Minoo Ahmadinejad ² M.D, AP/CP., Nasrin Ghasemi ^{3*} M.D., Ph.D.

¹ International Campus, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

² Blood Transfusion Research Center, High Institute for Research and Education in Transfusion Medicine, Tehran, Iran

³ Abortion Research Center, Reproductive Sciences Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

ABSTRACT

Article history

Received: 11 Mar 2023

Accepted: 31 May 2023

Available online: 15 Jun 2023

Keywords

In-silico

Single nucleotide polymorphism

Von Willebrand factor

Background and Aims: von Willebrand disease (VWD) is brought on by hereditary impairments in the von Willebrand factor (VWF). It has been determined that most of the molecular abnormalities are present in exon 28 of the VWF gene qualitative variations. In this regard, investigating mutations in exon 28 of the VWF coding gene can help identify the VWD type and can be used to manage patients by using appropriate strategies.

Materials and Methods: 10 suspected VWD type 2B patients were investigated for the genotype of exon 28. Routine coagulation tests were performed for the patients. Molecular sequencing was also used for the evaluation of mutations in exon 28. Determining the tertiary structure of VWF can aid in understanding its functional residues and interactions. The nsSNVs' pathogenicity was examined utilizing potent bioinformatics methods. Step-by-step testing of the high-risk mutations was done using SIFT, PolyPhen-2, I-Mutant 2.0, PHD-SNPg, and SNPs&GO. Then, the secondary structure, amino acid conservation, and feature of amino acids were investigated via Protparam, Cofactor, Interprosurf, ConSurf, NetSurfP-2.0, and HOPE.

Results: N1231T, V1229G, V1316M, and P1266Q missense mutations in VWF were detected. 3D structure of VWF predicted and evaluated. Putative conserved domains were identified. P1266Q and V1316M amino acid substitutions are predicted as "Not tolerated substitution" damaging and Disease, while V1229G and N1231T amino acid substitutions are predicted as "tolerated substitution" benign and Neutral.

Conclusion: V1316M and P1266Q amino acid substitutions were determined as high-risk mutations using powerful bioinformatics tools in VWD patients.

*Corresponding Author: Abortion Research Center, Reproductive Sciences Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran. Email: nghasemi479@gmail.com

Introduction

Von Willebrand disease (VWD), one of the most common inherited bleeding disorders, affects one in 10,000 persons [1]. von Willebrand factor (VWF) structure or inherited function deficiencies lead to VWD [2]. VWF is a great glycoprotein produced by megakaryocytes and endothelial cells and released into circulation through a constitutive pathway and upon stimulation [3]. Erik von Willebrand first identified VWF characteristic properties in a Scandinavian family [4]. The highly conserved protein VWF is produced by the *VWF* gene on the short arm of chromosome 12 [5].

VWD is divided into three major groups. A partial quantitative deficiency is kind 1, a qualitative defect is type 2, and a nearly whole deficiency of VWF is a type 3 defect [6]. There are four secondary classifications for VWD type 2. Variants of Type 2A have reduced platelet adhesion due to a selective lack of high-molecular-weight VWF multimers [6-9]. Type 2B is a rare disease in VWD2B, mutations located in exon 28 of the VWF A1domain. Variants of type 2B have a stronger affinity for platelet glycoprotein [10]. Despite a somewhat normal size distribution of VWF multimers, Type 2M includes variants with clearly impaired platelet adherence. Variants of type 2N have a significantly reduced affinity for factor VIII [11]. These six VWD classifications are correlated with critical clinical characteristics and therapy needs [1].

Given the complexity of the condition, the capacity to correctly and reliably diagnose

people with VWD remains a crucial and hotly debated area of interest [12]. *VWF* polymorphisms are important regulators of the *VWF* gene's expression. Exon, 28 of the *VWF* gene has been revealed to include most molecular abnormalities [5]. In this regard, investigating mutations in exon 28 of the *VWF* coding gene can help identify the VWD type and can be used to manage patients using appropriate strategies.

Due to a large number of patients with bleeding disorders who are referred to the Coagulation Reference Laboratory of the Iranian Blood Transfusion Organization (IBTO) for confirmation and differentiation of VWD 2B by molecular analysis, a study was carried out to provide a more thorough diagnosis for these patients. Four significant single nucleotide polymorphisms (SNPs) that were strongly linked to bleeding were found. We explain the pathogenicity potential of these missense mutations using several bioinformatics prediction tools. Additionally, analyze how this mutation affects the structure and function of proteins.

Materials and Methods

Patients selection and experimental investigations

In this cross-sectional study, ten bleeding disorder patients who were sent to the IBTO coagulation laboratory between 2018 and 2021 were chosen, and the preliminary and main VWD assays based on ISTH were classified as VWD type 2 B/platelet type VWD. The

inclusion criteria were the lack of a prior history of factor VIII and platelet-VWD abnormalities and a declaration of informed consent to participate in the trial. Receiving blood products or factor coagulants was an exclusion condition. These patients did not receive blood products two weeks before the sampling, and it did not affect the phenotypic tests of the patients, and they were included in the study. Using the STA Compact coagulometer (STAGO, France), routine coagulation tests such as the activated partial thromboplastin time (APTT) and prothrombin time (PT) were carried out. Using a one-stage procedure, factor VIII coagulant activity (FVIII:C) was measured (normal range: 50-150 IU/dl). An enzyme-linked immunosorbent test (ELISA) was also used to quantify VWF antigen (VWF:Ag; normal range: 50-150 IU/dl). A polyclonal antibody against VWF was utilized for this purpose. Utilising the aggregometry method, VWF Ristocetin cofactor activity (VWF: RCo) was measured. Formalin-fixed platelets generated from healthy donors were employed in this technique. The platelet-rich plasma sample was combined with various amounts of ristocetin to perform ristocetin-induced platelet aggregation (RIPA). The manufacturer's instructions for the QIAamp DNA Mini kit (QIAGEN Hilden, Germany) were followed in order to extract DNA. DNA was taken and used in a polymerase chain reaction. The amplified DNA was analyzed using the Sanger sequencing technique (bioSystems 3130XL - ABI USA) to verify the alterations. This study is approved by the Ethics Committee of the

abortion research center, reproductive sciences institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran (Ethics code: IR.SSU.MEDICINE.REC.1399.211).

Sequence and structure availability and homology search

The FASTA formatted VWF protein sequence was saved for additional analysis. The sequence was used as a query for protein BLAST against a non-redundant protein database at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. In this regard, the query protein's conserved domains were detected [13]

Template search

To find homologous structures, the query protein sequence was utilized as input data for PSI-BLAST against the protein data bank (PDB) [14] at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

Sequence examination

For estimation and determination of various physical and chemical parameters such as molecular weight, theoretical pI, amino acid composition, the total number of negatively and positively charged residues, instability index, aliphatic index, and grand average of hydropathicity (GRAVY), Protparam at <http://expasy.org/tools/protparam.html> was used [15].

Structural forecast

The I-TASSER [16] server at <http://zhanglab.ccmb.med.umich.edu/I-TASSER/> forecasts the structures and activities of the VWF protein by using ab initio and multiple-threading alignments to build 3D models. Phyre2 [17] at <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index> substantially increases alignment

accuracy and detection rate by using HHsearch to align hidden Markov models by building a 3D model of VWF protein. To model protein areas lacking homology to experimentally determined structures, Phyre2 also integrates a new ab initio folding simulation. In addition, LOMETS (Local Meta-Threading-Server), a web service for predicting protein structure [18] at <http://zhanglab.ccmb.med.umich.edu/LOMETS/>, was used to build a 3D model of VWF protein. It produces 3D models by compiling ten locally-installed threading programs with top-scoring target-to-template alignments (FUGUE, HHsearch, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKS, SP3, FFAS, and PRC).

Model assessments

The VWF protein's three-dimensional models were qualitatively assessed using the Qmean server [19] at <http://swissmodel.expasy.org/qmean/cgi/index.cgi?page=help> QMEAN. Qmean is a composite scoring function that, using just one model, can yield both global (i.e., for the entire structure) and local (i.e., per residue) error estimates.

Model improvement

ModRefiner [20] at <http://zhanglab.ccmb.med.umich.edu/ModRefiner/> was used to improve the final structure of VWF protein at the atomic level with high resolution. It can start from the C-alpha trace, the main-chain model, or the entire atomic model of the best-selected VWF protein model. Both side-chain and backbone atoms are completely flexible during structure refinement simulations, where the conformational search is guided by a composite of physics- and knowledge-based force field.

Prediction of ligand binding sites

COFACTOR [21], available at <https://zhanggroup.org/COFACTOR>, was used for annotating the biological function of the VWF protein. To find VWF protein functional sites, COFACTOR analyses the 3D structure and then threads the query through the BioLiP protein function database using local and global structure matches. The best functional homology templates, such as ligand-binding sites, will extract functional insights.

VWF important residues selection

The ConSurf Server [22] for identification of Functional Regions at <https://consurf.tau.ac.il/> used the on Willebrand factor structure as an input file. ConSurf can identify conserved functional and structural amino acids based on the evolutionary relationships between the protein and its homologs. ConSurf depicts the VWF protein's three-dimensional structure on a scale of 1 to 9, where 1 (9) corresponds to a hypervariable (highly conserved) amino acid. InterProSurf [23] at <http://curie.utmb.edu/pattest9.html> was used to predict functional locations and interface residues in VWF protein surfaces.

Effects of amino acid substitution

The effects of amino acid substitution on VWF protein function can be predicted using the Sorting Intolerant From Tolerant (SIFT) [24] analyzer at <https://sift.bii.a-star.edu.sg/>. If the score determined by SIFT is less than 0.05, the mutation is considered to affect VWF protein function. Sequences that contain one of the fundamental amino acids are referred to as "Seq Rep." Low fraction means there is not enough data for this position. Low confidence

prediction may be brought on by substantially gapped or unaligned data in this position.

Based on the multiple sequence alignment of the VWF protein 3D structure, PolyPhen-2 [25], at <http://genetics.bwh.harvard.edu/pph2/>, forecasts the impact of an amino acid substitution on the structure and function of the protein. This tool can determine the position-specific independent count (PSIC) score, which ranges from 0 to 1. According to PolyPhen-2, the outcome is benign (scoring 0-0.15), potentially detrimental (score 0.15-0.85), and very certainly damaging (score 0.85-1). For forecasts with confidence, this tool offers the two values “sensitivity” and “specificity”. SNPs&GO [26] at <https://snps.biofold.org/snps-and-go//pages/help.html> Use GO phrases to predict alterations related to diseases. To determine if a particular single nucleotide polymorphism in VWF protein is disease-related or unrelated, SNPs&GO have been optimized.

The machine learning technique PHD-SNPg [27] at <https://snps.biofold.org/phd-snp/> (Predicting Human Deleterious SNPs in the Human Genome) relies on sequence-based characteristics. This program considers both the coding and non-coding effects of SNVs in VWF protein. It is determined whether the SNV is a harmful or benign mutation. A probabilistic score ranges from 0 to 1, and variants are considered harmful mutations if the value is more than 0.5.

I-Mutant2.0 [28] (<http://folding.biofold.org/i-mutant/i-mutant2.0.html>) is a support vector machine (SVM) that forecasts changes in VWF protein stability caused by single point

mutations based on protein structure or sequence. I-Mutant2.0 determines the VWF protein's free energy changes using empirical thermodynamic data (DDG). In light of this, VWF protein stability changes depending on whether DDG is lower or higher than 0.

Consideration of the structure

As a sequence-based tool, NetSurfP 2.0 [29] at <http://www.cbs.dtu.dk/services/NetSurfP/> predicts the secondary structure, surface accessibility, structural disorder, and backbone dihedral angles (Phi and Psi angles) for each residue in the VWF protein sequence.

With a threshold of 25%, this method forecasts 2-class relative solvent accessibility (RSA) for an amino acid that is either buried or exposed. Additionally, the output of absolute solvent accessibility (ASA) is determined by multiplying RSA by ASAm_{max}. HOPE was used to examine the structural implications of a point mutation in the VWF protein sequence. HOPE [30] at <https://www3.cmbi.umcn.nl/hope/combines> the information from several web services and databases.

Results

Patients selection and experimental investigations

This study examined ten patients who were thought to have VWF 2B. No mutation was observed in exon 28 in two patients. More than one mutation was observed in three patients. In most patients, ecchymosis was observed as a bleeding manifestation. In addition, many patients had a history of transfusion of blood and blood products. Some patients had undergone surgery due to the underlying

disease. Regarding the condition of the family members of the patients, information was not available in most cases; however, in four patients, the family members of the patient had a history of bleeding. Four missense mutations, including N1231T, V1229G, V1316M, and P1266Q amino acid substitutions, were detected in the VWF protein.

Homology search and availability of sequence and structure

The VWF preproprotein [Homo sapiens] NCBI Reference Sequence, Accession number NP 000543.3 and GI: 1813372009, has 2813 residues and was made available by the National Centre for Biotechnology Information (NCBI). The VWF preproprotein sequence was used as a query in the BLAST method, and a set of sequences were introduced as the highest comparable sequences. The BLAST results revealed several hits to the preproprotein of the VWF.

BLAST calculates a pairwise alignment between a query and the database sequence iterates. It does not execute multiple alignments or compute an explicit alignment between the database sequences. Based on the alignment of those (database) sequences to the query, an implicit alignment between the database sequences is created for the sake of this sequence tree display. Two database sequences may frequently align to different portions of the query, causing them to barely or entirely overlap. If that is the case, it is impossible to determine how far apart these two sequences are, so only the sequence with the higher score is represented in the tree. Putative conserved domains were detected

within this sequence. The sequence residues from 1198-1276 belong to VWF type A (VWA) N-terminal; This domain is found in VWF proteins, where it is found in the N-terminus of the first VWA domain (pfam00092).

NCBI released the 2813-residue VWF preproprotein [Homo sapiens] NCBI Reference Sequence, Accession number NP 000543.3 and GI: 1813372009, public. A group of sequences was introduced as the highest comparable sequences using the VWF preproprotein sequence as the query in the BLAST algorithm. Numerous hits to the preproprotein of the VWF were found in the BLAST results. A pairwise alignment between a query and the database sequences is calculated iteratively by BLAST. It does not compute an explicit alignment between the various database sequences or perform multiple alignments. The VWA domain, first identified in the blood coagulation protein VWF (vWF), comprises the residues from 1276 to 1384. The VWA domain typically consists of 200 amino acid residues folded into the traditional a/b para-Rossmann fold. Since its discovery, the VWA domain has attracted much attention due to its frequent occurrence and involvement in various crucial cellular processes. These include the development of the basal membrane, cell migration, differentiation, adhesion, hemostasis, signaling, chromosomal stability, malignant transformation, and immune defenses, where these domains form heterodimers in integrins while multimers in vWF. The numerous molecules this domain combines with reveal

the diverse interaction surfaces of this domain. Most of the time, a metal ion-dependent adhesion site known as the MIDAS motif mediates ligand binding, which is a defining trait of most, if not all, domains. Figure 1 and Table 1 display putative conserved domains found in the sequence. P1266Q, V1229G, and N1231T, found in the VWA N-terminal, and V1316M, located in the VWA domain, are three SNPs. The sequences of these two potential domains are displayed in Figure 2.

Template selection

PDB was searched for a template. The 8D3C PDB record was detected as the template for the Human VWF homology modeling. The Structure of Chain A, VWF [Homo sapiens] (Accession: 4DCH -A, Max score: 2862, Total Score: 3291, Query coverage: 77%, Max ident: 99.59%, E value: 0.0 and Accession Length: 1469) was selected as a template from Protein Data Bank and saved in PDB format.

Sequence scrutiny

The physicochemical properties of mutated sequences were compared to their parent protein. Molecular weight, PI, Estimated half-life, instability index, aliphatic index, and GRAVY were calculated for wild-type and mutated sequences. All results are summarized in Table 1.

3D modeling

Five models were created by I-TASSER and rated according to their C-scores. The best model among those provided, with the highest C-score, was chosen for validation studies. A confidence score, or C-score, is used by I-TASSER to gauge the caliber of predicted models. It is determined using the importance of threading template alignments and the

convergence parameters from simulations of the structure construction. A model with a high confidence level has a higher C-score, and vice versa. The C-score typically falls between [-5, 2]. Phyre2 predicted one three-dimensional model for the protein while Lomets meta server predicted ten. The models were also selected for additional analyses.

Model evaluation and improvement

The QMEAN server qualitatively estimated the 3D models. The I-TASSER model was chosen carefully as the best-predicted structure. The selected model was refined with a moderefiner server. The QMEAN validation Scores of the initial and the final models after refinement are shown in Figure 1. QMEAN is a single model approach that linearly combines statistical potentials and agreement terms. A linear combination of four statistically possible words is known as QMEAN4. It has been trained to forecast IDDT scores globally between 0 and 1. It has been converted into a z-score to compare the value shown here with what might be anticipated from high resolution X-ray structures.

Prediction of the ligand binding location

The initial I-TASSER model is matched to every structure in the PDB database using the TM-align structural alignment algorithm after the structure assembly simulation. The top 10 proteins from the PDB with the highest structural similarity, or highest TM-score, to the projected I-TASSER model are listed in this section. These proteins frequently perform similarly to the target because of their similar structural makeup. Ligand binding sites were identified utilizing the COFACTOR software

that involved conserved residues, particularly 606, 608, 325, 329, 342, 348, 213, 250, 255, 265, 275, 304, 312, 342 and 348 with 0.04 C-score. The prediction's confidence score is known as the C-score. C-scores are between 0 and 1, and a higher value denotes a more trustworthy forecast.

Selection of significant residues in VWF preproprotein

An amino acid's level of evolutionary conservation in a protein represents a compromise between the necessity to maintain the structural integrity and function of the macromolecule and the amino acid's inherent propensity to change. The VWF preproprotein's evolutionary pattern of amino acids is analyzed via the ConSurf web service to identify key regions for structure and function. The mapping of the conservation grades onto the 3D model is shown in Figure 2. Interprosurf results reveal functional protein surface sites. The anticipated functional residues at the protein structure surface are shown in Figure 3 by Interprosurf.

Amino acid substitution effects

Using SIFT, P1266Q and V1316M amino acid substitutions are projected as a "Not tolerated substitution"; in contrast, V1229G and N1231T amino acid substitutions are predicted as a "tolerated substitution." The SIFT findings are shown in Table 2.

Then PolyPhen-2 examined the amino acid replacements. P1266Q has a computed score of 0.999 and was projected to be PROBABLY DAMAGING. V1316M was given a score of 1.000 and was classified as PROBABLY DAMAGING. The N1231T score was

estimated at 0.002 and is BENIGN. The V1229G score was determined to be 0.000 and was projected to be BENIGN. 75 amino acid residues around the variant's location in the query sequence are aligned several times by PolyPhen-2 (Fig. 4). In order to find protein structures, PolyPhen-2 BLASTs the query sequence against the PDB. If an amino acid in a hit differs from an amino acid in the input sequence at the corresponding location, the hit is by default rejected. The positions of all retained hits are then mapped from the substitution's position to the relevant places. According to the sequence identity or E-value of the sequence alignment with the query protein, hits are arranged in ascending order (Fig. 4).

The SNPs&GO results are shown as a table with the number of the modified site in the protein sequence, the wild-type residue, the new residue, and a prediction of the disease-relatedness of the associated mutation (Disease or neutral polymorphism Neutral). Reliability Index (RI) value evaluation is based on support vector machine output. More details about the output of the server are reported in Table 3.

The I-Mutant 2.0 tool was used to study the stability of mutant protein sequences in the following step. The DDG value of an amino acid substitution was computed using I-Mutant2.0 (DG (NewProtein)-DG (WildType) in Kcal/mol), and this tool also indicated a decrease in the stability of mutations (see Table 3). PHD-SNPg also considered the beneficial or unfavorable consequences of the amino acid substitution. PhD-SNPg prediction

of harmful SNPs in the human genome. A binary classifier that can identify both coding and non-coding harmful variations. N1231T and V1229G are considered Neural Polymorphism, whereas V1316M and P1266Q are predicted to be Disease-related Polymorphism. The PHD-SNPg results are displayed in Table 3.

Structural consideration

Using the NetSurfP-2.0 online service, the secondary structure, RSA, ASA, Phi, and Psi, as well as class assignment of amino acid replacements, were investigated. It is

important to note that when solvent accessibility increases, mutant protein stability decreases. With a threshold of 25%, this tool forecasts 2-class relative solvent accessibility for an amino acid that is either buried or exposed. Additionally, by multiplying RSA and ASAm_{max}, absolute solvent accessibility output is determined. The class assignment, RSA, ASA, and secondary structure of the native and mutant amino acids were taken into account based on the NetSurfP-2.0 results, as shown in Table 4.

Table 1. Physicochemical properties of mutated sequence vs. their parent protein

	Wild type	V1229G	P1266Q	V1316M	N1231T
Molecular weight	309264.51	309222.43	309295.53	309296.57	309251.51
Theoretical pI	5.29	5.29	5.29	5.29	5.29
Estimated half-life	30 hours	30 hours	30 hours	30 hours	30 hours
Instability index	51.64	51.64	51.64	51.69	51.61
Aliphatic index	73.42	73.31	73.42	73.31	73.42
GRAVY	-0.208	-0.209	-0.209	-0.209	-0.207

Table 2. Sorting intolerant from tolerant results

Position	Seq Rep	Predict Tolerated
	w 1229V 0.28 c f m y h i l P V n R q t d G s k E A	
	w m i f c v y 1231N 0.28 r p h q L t e k a s d G N	
y w v t s r q n m l k i h g f e d c a 1266P 0.14 P		
y w v t s r q p n m k i h g f e d c a 1316V 0.14 V		

The threshold for intolerance is 0.05. Amino acid color code: nonpolar, **uncharged** polar, **basic**, **acidic**. Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction. 'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information.

Table 3. SNPs&GO, I-Mutant2.0, and PhD-SNPg predictions

Mutation	SNPs&GO Predictions	SNPs&GO Probability	I-Mutant Stability	PhD-SNPg predictions	DDG
V1229G	Neutral	0.102	Decrease	Neutral	-0.97
N1231T	Neutral	0.115	Decrease	Neutral	-0.87
P1266Q	Disease	0.758	Decrease	Disease	-2.33
V1316M	Disease	0.569	Decrease	Disease	-1.77

SNPs & GO Prediction= Neutral (Neutral Variation) and Disease (Disease associated variation) SNPs & GO Probability: Disease probability (if > 0.5 mutations is predicted Disease) I-Mutant Stability: Stability of mutations
PhD-SNPg prediction: Neutral (Neutral SNP) and Disease (Disease-related SNP)
PhD-SNPg DDG: DDG < 0: Decrease Stability and DDG > 0: Increase stability

Table 4. NetSurfP-2.0 results. RSA, ASA, Phi, Psi and the secondary structure of the native and mutant amino acids were considered.

	Wild				Mutant			
	RSA	ASA	Phi	Psi	RSA	ASA	Phi	Psi
1229	55%	84Å	-31°	27°	62%	49Å	59°	-20°
1266	40%	56Å	-69°	133°	68%	154Å	-90°	124°
1231	52%	76Å	-113°	119°	43%	59Å	-117°	135°
1316	1%	2Å	-122°	134°	0%	1Å	-127°	138°

RSA= Relative solvent accessibility; ASA= Absolute solvent accessibility

Table 5. Mutation table of studied patients

Patients	Mutation (EX28)	Mutation type	Genotype
A	NOT FOUND	-	WT
B	c.3797C>A p.P1266Q,c.3686T>G p.V1229G,c.3692A>Cp.N1231T	Missense	Hetero
C	c.3946G>A p.V1316M	Missense	Hetero
D	c.3692A>C p.N1231T	Missense	Hetero
E	c.3946G>A p.V1316M	Missense	Hetero
F	c.3797C>A p.P1266Q	Missense	Hetero
G	c.3946G>A p.V1316M	Missense	Hetero
H	c.3946G>A p.V1316M	Missense	Hetero
I	c.3797C>A p.P1266Q, c.3686T>G p.V1229G,c.3692A>C .N1231T	Missense	Hetero
J	NOT FOUND	-	WT

The structural alterations brought on by the amino acid substitutions V1316M, P1266Q, N1231T, and V1229G were taken into account using the HOPE web service. Each amino acid is unique in size, charge, and hydrophobicity. These characteristics frequently vary between the original wild-type residue and the newly introduced mutant residue. Figure 5 depicts the schematic structures of the original (left) and mutant (right) amino acids. The backbone is the same for each amino acid and is colored red. The side chain, unique for each amino acid, is colored black. HOPE results were summarized below:

Mutation of a Valine into a Glycine at position 1229

The sizes of the mutant and wild-type amino acids are different. The interaction may be lost

because the mutant residue is smaller than the wild-type residue. Given that the mutant residue (glycine) is less hydrophobic than the wild-type residue (alanine), any hydrophobic interactions in the protein's surface or core will no longer exist. Glycine has a high degree of flexibility, which may interfere with the protein's need for rigidity in this location. At this location, the wild-type residue (Valine) is not preserved. The amino acid glycine was more frequently seen in other comparable sequences at this location. This indicates that proteins containing the mutant residue are more prevalent than proteins with the wild-type residue (Valine). It is probably unlikely that this mutation will harm the protein (Valine to Glycine).

Mutation of a Proline into a Glutamine at position 1266

The sizes of the mutant and wild-type amino acids are different. Because the mutant residue is larger, bumps could result. The wild-type and mutant residues have different hydrophobicities. More hydrophobic than the mutant residue is the wild-type residue. Hydrophobic interactions will be lost in the protein's inside or on its surface. Proline is the residue in the wild-type. Prolines are known to be extremely rigid, and as a result, they cause a unique backbone conformation that may be necessary for this position. This unique conformation may be disturbed by the mutation. The wild-type residue appears frequently at this point in the sequence, but other residues have also been seen. In other analogous sequences, neither the mutant residue nor another type of residue with related characteristics was found at this location. The protein could suffer from the mutation. It is well known that prolines have an extremely stiff structure that occasionally forces the backbone into a certain shape. The mutation may disrupt the local structure by changing a proline with such a function into another residue.

Mutation of an Asparagine into a Threonine at position 1231

The sizes of the mutant and wild-type amino acids are different. Because the mutant residue is smaller, interactions may be lost. The wild-type and mutant residues have different hydrophobicities. More hydrophobic than the wild-type residue is the mutant residue. At this location, the mutation adds a more

hydrophobic residue. This may lead to the breakup of hydrogen bonds and/or interfere with proper folding. The wild-type residue appears frequently at this point in the sequence, but other residues have also been seen. The mutant residue is one of the other forms of residue seen in homologous sequences at this location. This indicates that this mutation is potentially harmless to the protein and can happen at this location.

Mutation of a Valine into a Methionine at position 1316

The sizes of the mutant and wild-type amino acids are different. Because the mutant residue is larger, bumps could result. The mutation is found in a domain UniProt designated as VWFA 1, platelet glycoprotein Ib binding site. The mutation introduces an amino acid with distinct characteristics, disrupting this domain's functionality. A few additional residue types have been seen at this location, in addition to the wild-type residue, which is highly conserved. None of the other residue types seen at this location in other homologous proteins had the mutant residue. However, it was discovered that other residues shared some of the altered residue's characteristics. The mutant residue is close to a location that is highly conserved. A flowchart of the methodology is shown in Figure 6, in which all the software and servers are mentioned, and their coherence and consistency in expressing the results are illustrated.

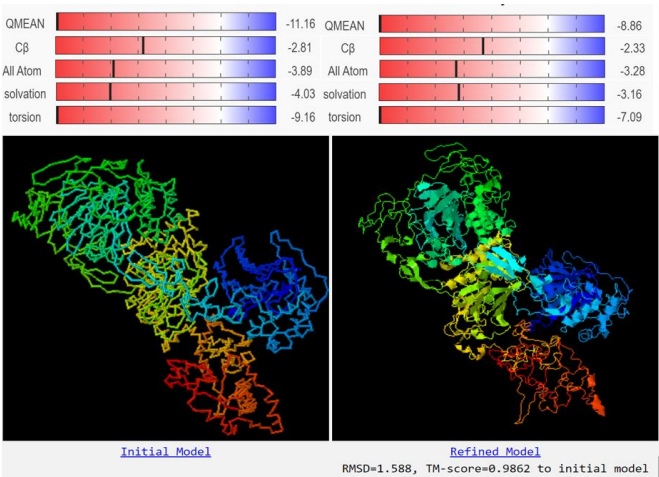


Fig. 1. Top: Validation Scores of initial and the final models after refinement. Below: 3D Visualization of refinement of the I-TASSER model with a mode-refiner server.

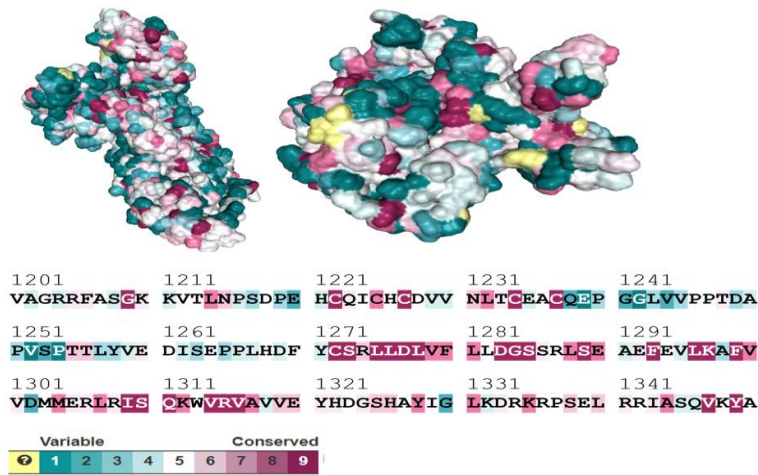
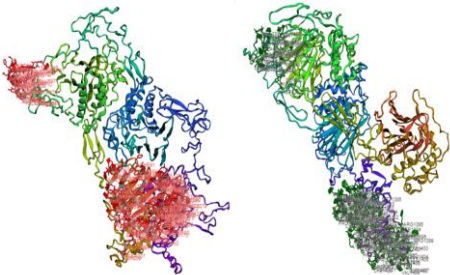


Fig. 2. Consurf results. Structure of the von Willebrand factor preproprotein colored by conservation using the color-code bar.



Cluster number	Functional residues at the protein structure surface predicted by Interprosulf
1 (Red color)	27,30,279,322,324,1241,1242,1262,1265,1266,1267,1268,1269,1270,1271,1272,1273,1274,1275,1276,1277,1278,1279,1280,1292,1293,1294,1295,1296,1297,1298,1299,1300,1301,1302,1303,1304,138640, 54, 56, 76, 77, 78, 80,81,144,184,185,192,199, 200, 201, 202, 203, 204, 205, 206, 207, 208
2 (Green color)	1316,1317,1318,1319,1320,1321,1322,1323,1324,1325,1326,1327,1328,1329,1330,1331,1333,1336,1337,1338,1340,1341,1342,1343,1344,1345,1346,1347,1348,1349,1263,1305,1306,1307,1308,1309, 1443,1444,1445,1446,1448,1449,1450,1451,1452,1455,1457,1458,1459,1462,1467,1468,1470,1472,

Fig. 3. Interprosulf predicts functional residues at the protein structure surface. Overview of the protein in ribbon presentation. Functional residues at the protein structure surface predicted by interprosulf are labeled and shown in stick representation.

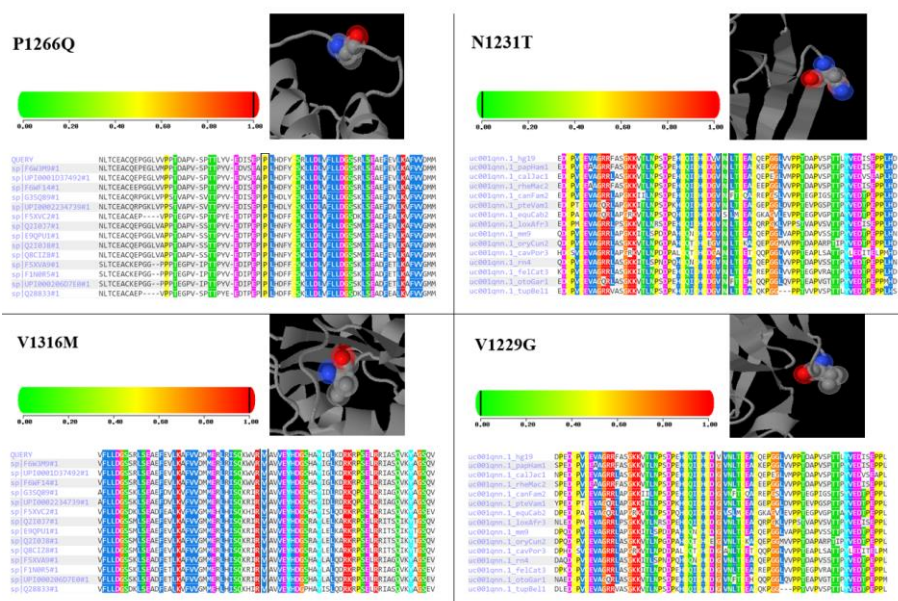


Fig. 4. PolyPhen-2 result. According to the PolyPhen-2 result, V1316M and P1266Q are predicted to be probably damaging, and N1231T and V1229G are predicted to be benign 3D Visualization of mutations in von Willebrand factor structure and 75 amino acids surrounding the mutation position (marked with a black box) are shown.

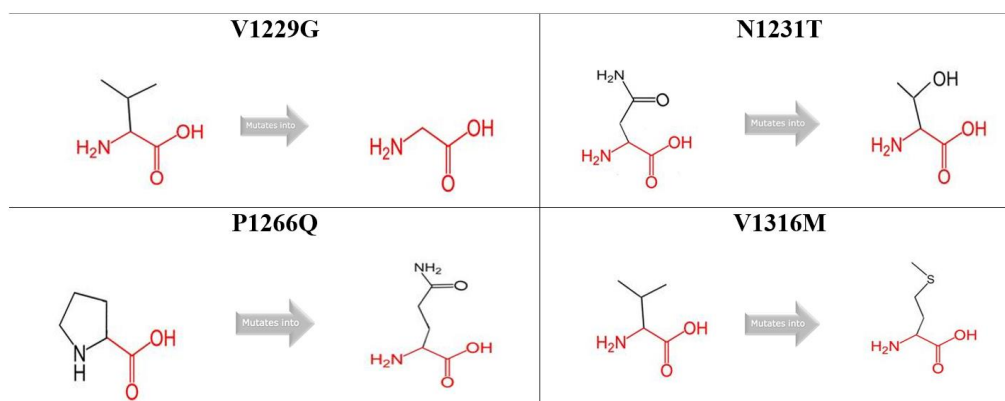


Fig. 5. The schematic structures of the original (left) and the mutant (right) amino acid.

Discussion

Most inherited bleeding disorders are thought to be caused by VWD [31]. Following a clinical and physical examination, a family history of (particularly mucocutaneous) bleeding, and laboratory tests, VWD is diagnosed [4]. Methods for identifying VWD have been established; however, they lack the sensitivity and specificity to distinguish between different kinds of VWD [32]. According to earlier research, detecting exon 28 mutations in the VWF coding gene can be useful in identifying

high-risk individuals who are exposed to VWD 2B [33]. Therefore, we investigated the mutations occurring in exon 28 in this study. As regards, the investigation of mutations in exon 28 of the VWF coding gene can help to identify the VWD type and can be used to manage patients by using appropriate strategies [34], V1316M, P1266Q, N1231T, and V1229G amino acid substitutions was identified and scrutinized in the present study (Table 5).

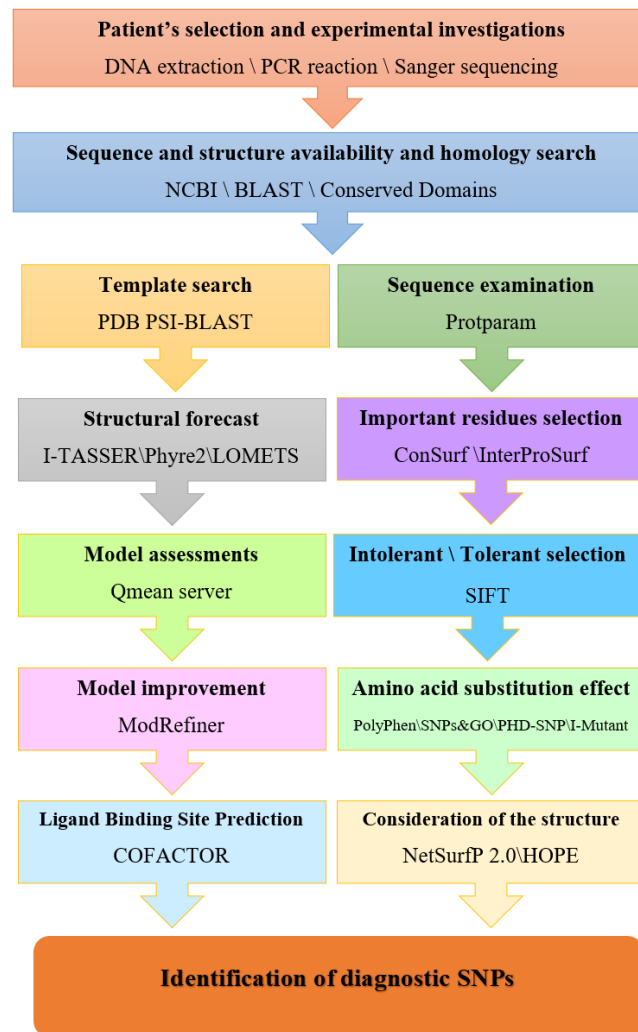


Fig. 6. Flowchart of the methodology, all the software and servers are mentioned, and their coherence and consistency in expressing the results are illustrated.

To achieve the goal, it was necessary to predict the structure of VWF. The architecture of proteins typically takes on biochemical roles. In their natural habitats, amino acid linear chains take on a distinctive three-dimensional form. Experimental and theoretical methods could both be used to evaluate protein structures. Due to difficulties encountered in the experimental determination of 3D protein structures, bioinformatic techniques are significant [35]. To determine the best template, the VWF sequence was used as a query for a BLAST search against the PDB [14]. Lower

E-value, greater query coverage, and maximum identity are suitable selection criteria [36]. Consequently, the one with the greatest overall score can be the most trustworthy template. VWF 3D models were created using various in silico techniques, including ab initio, threading, homology modeling, and combinations. The models created by different servers had to be compared to get the best model. Choosing among the predicted 3D structures required the implementation of scoring programs that reflected conformational energy. The best model

was determined based on the quality estimation and was created by the I-TASSER server.

The quality of the model might be raised through refinement. One goal of ModRefiner is to bring the basic starting models' hydrogen bonds, backbone structure, and side-chain placement closer to those of their natural form. Additionally, it produces a noticeable improvement in the physical condition of nearby structures [20]. Qmean Global Quality Score of the refined model was improved. Putative conserved domains were detected within this protein. The sequence residues from 1198-1276 belong to VWA N-terminal, and residues from 1276-1384 belong to the VWA domain.

Changes in protein structure and, consequently, function result from the presence of SNVs. Additionally, About 90% of human variability is caused by SNVs. A unique characteristic of amino acids is their size, charge, and hydrophobicity rating. These characteristics frequently vary between the mutant and original wild-type residues [37]. In light of this, protein function, structure, stability, and ultimately protein-protein interaction may all be impacted by the amino acid change.

Although expensive and time-consuming, experimental procedures are the most trustworthy approach. Additionally, in some circumstances, it is hard to recover mutant proteins using *in vitro* and *in vivo* mutagenesis methods [38]. Therefore, researchers can forecast the impact of mutations using bioinformatics tools. There is not just one tool for this; various instruments can be utilized to get trustworthy findings. Strong bioinformatics

methods can be used to assess protein structure and function changes. These methods allow for the prediction of atom distances, the pathogenicity of mutations, modifications to protein structure, and polar interactions.

Additionally, a measurement of the total energy of proteins can be used to examine the stability of mutant proteins. The most significant biological challenge is figuring out protein structure in the presence of mutations. The pathogenicity potential of a missense mutation is thoroughly investigated in this study using various bioinformatics prediction tools. Additionally, analyzes were performed to show how this mutation affects the structure and function of proteins. The mutant amino acid's alterations in structure and hydrophobicity are compared with the natural residue to corroborate the results. Also taken into account is the preservation of the original residue. This computational strategy can be utilized as a preliminary step in designing a focused molecular procedure to support the findings of the bioinformatics study. Due to the high cost and time requirements involved in discovering and investigating SNPs, computational tools can be beneficial in planning focused molecular methods [38].

Additionally, it is occasionally impossible to perform molecular techniques like protein extraction or mutagenesis. Studies demonstrated that the combined bioinformatics methods can be used to identify high-risk SNPs. Regardless of the particular condition, our study thoroughly analyzes the pathogen nsSNVs in VWF protein utilizing various bioinformatics tools and diverse methodologies. These high-risk mutations'

conservation, hydrophobicity changes, and structural modifications were also considered. ProtParam, Cofactor, InterProSurf, ConSurf, SIFT, PolyPhen-2, I-Mutant2.0, SNPs&GO, PHD-SNPg, NetSurfP-2.0, and HOPE are the tools employed in our investigation.

P1266Q and V1316M amino acid substitutions are predicted as a “Not tolerated substitution” damaging and Disease, while V1229G and N1231T amino acid substitutions are predicted as a “tolerated substitution” benign and neutral. According to the results of bioinformatics techniques, the V1316M and P1266Q amino acid substitutions have been identified as the high-risk mutations in the current consideration through N1231T, V1229G, V1316M, and P1266Q amino acid replacements.

Conclusions

Among N1231T, V1229G, V1316M, and P1266Q amino acid substitutions in VWF protein, V1316M and P1266Q amino acid substitutions were determined as the high-risk mutations using powerful bioinformatics tools in VWD patients.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors thank Shahid Sadoughi University of Medical Sciences and Blood Transfusion Research Center, High Institute for Research and Education in Transfusion Medicine for support to conduct this work.

References

- [1]. Sadler JE, Budde U, Eikenboom J, Favaloro E, Hill F, Holmberg L, et al. Update on the pathophysiology and classification of von Willebrand disease: a report of the subcommittee on von Willebrand Factor. *Journal of Thrombosis and Haemostasis* 2006; 4(10): 2103-114.
- [2]. Rassoulzadegan M, Ala F, Jazebi M, Enayat MS, Tabibian S, Shams M, et al. Molecular and clinical profile of type 2 von Willebrand disease in Iran: a thirteen-year experience. *International Journal of Hematology* 2020; 111(4): 535-43.
- [3]. Goodeve AC. The genetic basis of von Willebrand disease. *Blood Reviews* 2010; 24(3): 123-34.
- [4]. Nichols WC, Ginsburg D. von Willebrand disease. *Medicine* 1997; 76(1): 1-20.
- [5]. Porter CA, Goodman M, Stanhope MJ. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. *Molecular Phylogenetics and Evolution* 1996; 5(1): 89-101.
- [6]. Hamilton A, Ozelo M, Leggo J, Notley C, Brown H, Frontroth JP, et al. Frequency of platelet type versus type 2B von Willebrand disease. *Thrombosis and Haemostasis* 2011; 105(3): 501-508.
- [7]. Bowman M, Tuttle A, Notley C, Brown C, Tinlin S, Deforest M, et al. The genetics of Canadian type 3 von Willebrand disease: further evidence for co-dominant inheritance of mutant alleles. *Journal of Thrombosis and Haemostasis* 2013; 11(3): 512-20.
- [8]. Casaña P, Martínez F, Haya S, Tavares A, Aznar JA. New mutations in exon 28 of the von Willebrand factor gene detected in patients with different types of von Willebrand's disease. *Haematologica* 2001; 86(4): 414-19.
- [9]. Gok V, Isik E, Yilmaz E, Aydin F, Ozcan A, Unal E, et al. Type 2B Von Willebrand disease mimicking autoimmune thrombocytopenia in the neonatal period. *Erciyes Medical Journal* 2021; 43(2): 201-204.
- [10]. Yee A, Kretz CA. Von Willebrand factor: form for function. *Seminars in thrombosis and hemostasis*: Thieme Medical Publishers, 2014.
- [11]. Bellissimo DB, Christopherson PA, Flood VH, Gill JC, Friedman KD, Haberichter SL, et al. VWF mutations and new sequence variations identified in healthy controls are more frequent in the African-American population. *Blood* 2012; 119(9): 2135-140.

- [12]. Ng C, Motto DG, Di Paola J. Diagnostic approach to von Willebrand disease. *Blood* 2015; 125(13): 2029-2037.
- [13]. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* 2007; 35(S1): 61-5.
- [14]. Mount DW. Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols*. 2007;2007(7):pdb.top17.
- [15]. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. 1999; 112: 531-52.
- [16]. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008; 9(1): 1-8.
- [17]. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 2015; 10(6): 845-58.
- [18]. Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Research* 2007; 35(10): 3375-382.
- [19]. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Research* 2009; 37(S2): 510-54.
- [20]. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal* 2011; 101(10): 2525-534.
- [21]. Zhang C, Freddolino PL, Zhang Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Research* 2017; 45(1): 291-99.
- [22]. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, et al. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 2003; 19(1): 163-64.
- [23]. Negi SS, Schein CH, Oezguen N, Power TD, Braun W. InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics* 2007; 23(24): 3397-399.
- [24]. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research* 2012; 40(1): 452-57.
- [25]. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics* 2013; 76(1): 1-7.
- [26]. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC genomics* 2013; 14(3): 1-7.
- [27]. Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Research* 2017; 45(1): 247-52.
- [28]. Calabrese R, Capriotti E, Fariselli P, Martelli P, Casadio R. Protein Folding, Misfolding and Diseases: The I-Mutant Suite. *BITS'09 Sixth Annual Meeting of the Bioinformatics Italian Society*; 2009.
- [29]. Yaseen A, Li Y. Context-based features enhance protein secondary structure prediction accuracy. *Journal of Chemical Information and Modeling* 2014; 54(3): 992-1002.
- [30]. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 2010; 11(1): 1-10.
- [31]. Kasatkar P, Shetty S, Ghosh K. Genetic heterogeneity in a large cohort of Indian type 3 von Willebrand disease patients. *PLoS One* 2014; 9(3): 92575.
- [32]. Mancuso D, Tuley E, Westfield L, Worrall N, Shelton-Inloes B, Sorace J, et al. Structure of the gene for human von Willebrand factor. *Journal of Biological Chemistry* 1989; 264(33): 19514-9527.
- [33]. Lenting P, Casari C, Christophe O, Denis C. von Willebrand factor: the old, the new and the unknown. *Journal of Thrombosis and Haemostasis* 2012; 10(12): 2428-437.
- [34]. Keightley AM, Lam YM, Brady JN, Cameron CL, Lillicrap D. Variation at the von Willebrand factor (vWF) gene locus is associated with plasma vWF: Ag levels: identification of three novel single nucleotide polymorphisms in the vWF gene promoter. *Blood* 1999; 93(12): 4277-283.
- [35]. Dorn M, e Silva MB, Buriol LS, Lamb LC. Three-dimensional protein structure prediction: Methods and computational strategies. *Computational Biology and Chemistry* 2014; 53(2): 251-76.
- [36]. Sefid F, Rasooli I, Jahangiri A. In silico determination and validation of baumannii acinetobactin utilization a structure and ligand binding site. *BioMed Research International* 2013; 172784: 1-14.
- [37]. Wang Z, Moulton J. SNPs, protein structure, and disease. *Human Mutation* 2001; 17(4): 263-70.
- [38]. Kumar A, Rajendran V, Sethumadhavan R, Shukla P, Tiwari S, Purohit R. Computational SNP analysis: current approaches and future prospects. *Cell Biochemistry and Biophysics* 2014; 68(2): 233-39.