

A standard nomenclature for von Willebrand factor gene mutations and polymorphisms

On behalf of von Willebrand Factor Subcommittee of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis

Goodeve AC^{1*}, Eikenboom JCJ², Ginsburg D³, Hilbert L⁴, Mazurier C⁴, Peake IR¹, Sadler JE⁵, Rodeghiero F⁶.

From ¹Division of Genomic Medicine, Royal Hallamshire Hospital, Sheffield, UK; ²Department of Hematology, Leiden University Medical Center, Leiden, the Netherlands; ³Howard Hughes Medical Institute, University Of Michigan, Ann Arbor MI, USA; ⁴Laboratoire français du Fractionnement et des Biotechnologies, Lille, France; ⁵Washington University School of Medicine, St Louis MO, USA; ⁶Hematology Department, S. Bortolo Hospital, Vicenza, Italy.

Keywords; standard nomenclature, mutation, polymorphism, recommendation, VWF, VWD

Introduction

Mutations in the von Willebrand factor (VWF) gene responsible for von Willebrand disease (VWD) and predominantly those responsible for type 2 VWD have been identified by investigators for several years. The plethora of mutation screening techniques now available and their enhanced sample throughput capability is enabling an increasing number of investigations in patients with types 1 and 3 VWD. An unambiguous VWF nucleotide and amino acid nomenclature is now essential.

Historical perspective

cDNA sequence

The cDNA nucleotide (nt) sequence for VWF was published by Bonthron et al in 1986 (1). Nt were numbered from the A of the initiator methionine ATG codon as the +1 position, as recommended by Sadler and Ginsburg in 1993 (2).

Mutations currently entered onto the VWF electronic database (<http://mmg2.im.med.umich.edu/vWF/>) are numbered from the initiator ATG. The ISTH SSC subcommittee on VWF (3) however recommended previously that numbering should initiate from the transcription start site, 250 nt 5' to the A of the initiator ATG.

In 1993 and 1996, recommendations from the "nomenclature working group" (see below) were the use of the mRNA cap site as the +1 position for all genes (4, 5), whereas more recently (1998), this group also recommended the use of the A of the initiator methionine ATG as the +1 nt position (6) for nucleotides of newly characterised genes.

Gene sequence

The VWF gene is located on the short arm of chromosome 12 at 12p12-pter. A partial genomic sequence for VWF was published by Mancuso et al in 1989 (7) based on 38 separate genomic clones covering the 52 exons plus 5' and 3' untranslated regions of

* Corresponding Author: Dr Anne Goodeve, Division of Genomic Medicine, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF, UK, Tel : 0114 271 2679, Fax: 0114 272 1104
E-mail: a.goodeve@shef.ac.uk

VWF. Nt are numbered by cloned sequence block number and nt number within that block (5/1234 means block 5, nt 1234 of block 5).

The entire VWF genomic DNA sequence has recently become available in first draft form as a result of the human genome project (<http://www.ncbi.nlm.nih.gov/>), but is not currently in a user-friendly format. The current chromosome 12 draft sequence (September 2000) contains the entire VWF gene, with the exception of a small section at the extreme 5' end. However, this sequence was previously reported by a number of groups. Locus identifiers for this 5' sequence are HUMVWFA01 (exons 1-3 and Alu repeats) and HSVWF123 (exons 1-3). The remainder of the gene is encompassed by two overlapping contigs (regions of sequenced DNA) in the database; 111,904 bp of sequence representing exons 4-34 are identified by AC005846, while the 3' region of the gene is located within seven unordered clones in the antisense orientation under accession number AC005904.

Pseudogene sequence

A partial unprocessed VWF pseudogene is located on chromosome 22 at 22p12-pter. The sequence for the majority of the pseudogene, including its 5' boundary, was published by Mancuso et al in 1991 (8). The sequence spans 21-29 kb and is equivalent to exons 23-34 of the VWF gene (including intron sequences). Nt were numbered sequentially from 1-21,034. The pseudogene sequence is available under accession number M60676. Chromosome 22 has now been completely sequenced, and therefore the entire VWF pseudogene plus surrounding sequence is available. The relevant contig is identified by the code Hs22_2569. Knowledge of the pseudogene sequence is essential when designing primers for PCR amplification from genomic DNA representing exons 23-34 of VWF.

Amino acid sequence

Amino acids (aa) were numbered by Bonthron et al (1) and by Mancuso et al (7) from the translation start site, with the initiator methionine ATG set as +1. However, as the mature subunit of VWF has been much more intensively studied than the entire pre-pro-VWF, many investigators have numbered mature VWF, starting from the first amino acid, serine (S764 of pre-pro-VWF) (for examples, see Sadler et al, (9)).

Nomenclature Recommendations

The ISTH SSC factor VIII and IX subcommittee has already suggested a standard nomenclature for FVIII and IX gene mutations (10) and now suggest a nomenclature for VWF.

A series of papers on proposed mutation nomenclature schemes for human genes have been published in Human Mutation produced by the "nomenclature working group", a group of researchers representing interest in mutation detection from different fields of research (4-6). Readers are recommended to refer to these publications, particularly to that of Antonarakis et al (6) for further details of all circumstances where mutation nomenclature is required. As several different numbering schemes have been used for VWF in the past, authors are encouraged to indicate the scheme that is used.

At their first mention in a text, mutations may be described in detail, for example alanine substituted by glycine at codon 123; insertion of a guanine following the adenine nucleotide at position 1234. Thereafter, the nomenclature described below should be used.

cDNA sequence

The recommendation for the VWF cDNA is to continue to number from the A of the initiator ATG site as +1. This numbering scheme results in nucleotide numbers 250 nt

lower than when numbering VWF from the transcription initiation site (3). The sequence is available under accession number NM_000552.

Gene DNA sequence

When the entire VWF genomic sequence becomes readily available, genomic DNA nt numbering should start from the initiator ATG as +1, as for the cDNA. However, when numbering, a lower case g for genomic DNA should be used in front of the nt number (eg g.1234), to demonstrate which sequence is referred to. Numbering should be sequential throughout the gene, from 1-178,000 as for the FIX gene (11). The exon or intron number should also be given for orientation. Until the complete VWF gene sequence is readily available, the partial VWF genomic sequence of Mancuso et al (7) may be used when required.

Nucleotide changes

Nucleotide alterations should be referred to as indicated by examples in Table 1. The original and mutated nucleotides are both given following the nucleotide position number, to avoid confusion with amino acid numbering. Various symbols representing the nt substitution can be utilised, such as G A, G->A and G>A.

For insertion and deletion mutations, where one or two nt are involved, they should be named. Where >2 nt are involved, only the number involved is required. For deletions or insertions in repetitive sequence eg. AAAAAA, the most 3' nt number is arbitrarily assigned.

For intronic changes, where nt close to the intron/exon boundary are numbered, it is simple to use the cDNA numbering with a + or - sign to designate distance into the intron. For alterations deeper into the intron, the full VWF genomic DNA sequence should be used once it is readily available, until then, the sequence of Mancuso et al (7) may be used. When using genomic VWF DNA sequence numbering to describe intronic sequence, the intron number should also be given for orientation.

Changes in the upstream untranslated region are numbered from the A of the initiator ATG (+1) as negative numbers.

Amino acid sequence

Numbering should start from the initiator methionine as the +1 position. Separate numbering schemes for mature VWF should be discontinued, so avoiding confusion with amino acid numbering in the propeptide. The use of single letter aa codes will facilitate discrimination from the previous nomenclature, where the three letter amino acid codes were widely used for aa in the mature VWF.

Amino acid alterations

Amino acid alterations should be referred to as indicated by examples in Table 2. The wild type amino acid is given before the codon number and the mutant amino acid following the number. This placement avoids confusion with the nucleotide numbering scheme, where all nucleotides are given following the nucleotide position number. Termination (stop) codons are designated by an X.

For insertion or deletion mutations, where one or two aa are involved, they should be named. Where >2 aa are involved, only the number involved is required. For deletions or insertions in repetitive sequence eg. LysLysLys, the most carboxyl (3') aa number is arbitrarily assigned.

To facilitate comparison with earlier publications, at the first description in a manuscript of a mutation in mature VWF, the previous nomenclature (for mature VWF, starting from Ser764 of pre-pro VWF) should be given in parenthesis. For this purpose, the three letter aa codes should be utilised.

Polymorphisms

Polymorphic alterations should be referred to as indicated by examples in Table 3. Two alleles which co-exist in the population at an arbitrarily designated polymorphic frequency of >1% (12) and have no known deleterious effect on phenotype can be shown at a specified location using a “/” symbol. The first allele shown should be that in the reference sequence available on the VWF electronic database and the second the polymorphic variant.

To avoid confusion, for polymorphic aa residues, the aa designations are given before the codon position number, whereas for polymorphic nt, the nt designations are given following the nt position number.

REFERENCES

1. Bonthron D, Orr EC, Mitsock LM, Ginsburg D, Handin RI, Orkin SH. Nucleotide-sequence of pre-pro-von Willebrand factor cDNA. *Nucl Acids Res* 1986; 14: 7125-7127.
2. Sadler JE, Ginsburg D. A database of polymorphisms in the von Willebrand-factor gene and pseudogene. *Thromb Haemost* 1993; 69: 185-191.
3. Sadler JE. A revised classification of von Willebrand disease. *Thromb Haemost* 1994; 71:520-525.
4. Beaudet AL, Tsui L-C. A suggested nomenclature for designating mutations. *Human Mutation* 1993; 2: 245-248.
5. Ad Hoc Committee on Mutation Nomenclature. Update on nomenclature for human gene mutations. *Human Mutation* 1996; 8: 197-202.
6. Antonarakis SE, Nomenclature Working Group. Recommendations for a nomenclature system for human gene mutations. *Human Mutation* 1998; 11: 1-3.
7. Mancuso DJ, Tuley EA, Westfield LA, Worrall NK, Shelton-Inloes BB, Sorace JM, Alevy YG, Sadler JE. Structure of the gene for human von Willebrand-factor. *J Biol Chem* 1989; 264: 19514-19527.
8. Mancuso DJ, Tuley EA, Westfield LA, Lester Mancuso TL, Lebeau MM, Sorace JM, Sadler JE. Human von Willebrand-factor gene and pseudogene - structural-analysis and differentiation by polymerase chain-reaction. *Biochemistry* 1991; 30: 253-269.
9. Sadler JE, Matsushita T, Dong Z, Tuley EA, Westfield LA. Molecular mechanisms and classification of von Willebrand disease. *Thromb Haemost* 1995; 74: 161-166.
10. Peake I, Tuddenham E. A standard nomenclature for factor-VIII and factor-IX gene-mutations and associated amino-acid alterations. *Thromb Haemost* 1994; 72: 475-476.
11. Yoshitake S, Schach BG, Foster DC, Davie EW, Kurachi K. Nucleotide-sequence of the gene for human factor-IX (antihemophilic factor-B). *Biochemistry* 1985; 24: 3736-3750.
12. Peake I, Winship P. What is a polymorphism. *Blood* 1991; 78: 251.

Table 1. Recommendations for VWF nucleotide alterations

Nucleotide change	Nomenclature example	Description
Substitution	1234G>A	Guanine substituted by adenine at position 1234 in the VWF cDNA sequence
Substitution	g1234G>A	Guanine substituted by adenine at position 1234 in the complete VWF genomic sequence
Substitution	5/1234G>A	Guanine substituted by adenine at position 1234 in block 5 of the partial VWF genomic sequence of Mancuso et al (7)
Deletion	1234delG	Deletion of guanine from position 1234 of the VWF cDNA sequence
Deletion	1234delN	N nt deleted at position 1234 in the VWF cDNA sequence, N >2 nt
Insertion	1234insGT	Insertion of guanine and thymidine after nt 1234 in the VWF cDNA sequence
Insertion	1234insN	N nt inserted after nt 1234 in the VWF cDNA sequence, N >2nt
Intronic change	1234+1G>T	Substitution of guanine of GT splice donor, 1st position of intron. Nt number 1234 is last nt of preceding exon of the VWF cDNA sequence
Intronic change	1234-2A>G	Substitution of adenine of AG splice acceptor, 2nd last position of intron. Nt 1234 is the first nt of the succeeding exon of the VWF cDNA sequence

Table 2. Recommendations for VWF amino acid alterations

Amino acid change	Nomenclature example	Description
Missense	R123G	Arginine residue substituted by glycine at position 123 in the pre-pro VWF sequence
Nonsense	R123X	Arginine residue substituted by a stop codon at position 123 in the pre-pro VWF sequence
Insertion	R123insS	Insertion of a serine residue following the arginine residue at position 123 in the pre-pro VWF sequence
Deletion	R123del	Deletion of an arginine residue from position 123 in the pre-pro VWF sequence

Table 3. Recommendations for VWF polymorphism nomenclature

Polymorphic change	Nomenclature example	Description
Single nucleotide polymorphism	1234A/G	Nucleotide 1234 in the VWF cDNA sequence is polymorphic. Two alleles known (adenine and guanine), that in reference sequence shown first
Insertion/deletion polymorphism	1234C/delC	Presence/absence of cytosine at position 1234 in the VWF cDNA sequence is polymorphic
Single amino acid polymorphism	A/G1234	Amino acid residue at position 1234 in the pre-pro VWF sequence is polymorphic. Two alleles known (alanine and glycine), that in reference sequence shown first