Int J Legal Med (2008) 122:67–71
DOI 10.1007/s00414-006-0149-z

TECHNICAL NOTE

Complex variability of intron 40 of the von Willebrand factor (vWF) gene

Sandra Hering · Christa Augustin · Jeanett Edelmann · Micaela Heidel · Kathrin Chamaon · Jan Dressler · Reinhard Szibor

Received: 9 May 2006 / Accepted: 13 December 2006 / Published online: 2 February 2007

Abstract Intron 40 of the von Willebrand factor (vWF) gene exhibits a highly variable region of about 0.65 kb, which contains 5 juxtaposed STRs. We sequenced 0.65 kb amplicons from 68 chromosomes and found 2 frequent indel polymorphisms and 5 SNPs. The 68 chromosomes investigated here presented a total of 47 different haplotypes. Regarding the SNP allele distribution in our sample, we arranged our results of the vWF intron 40 into a system of 3 haplotypes, i.e. haplotypes a, b and c. Our review may be valuable in further optimising vWF typing in forensic applications and in avoiding pitfalls. Further attempts to develop sophisticated techniques may soon enable haplotyping using autosomale STR clusters.

Keywords vWF · Intron 40 · Short tandem repeats · SNPs · Indel polymorphisms · Kinship testing

Electronic supplementary material The online version of this article (doi:10.1007/s00414-006-0149-z) contains supplementary material, which is available to authorized users.

S. Hering · M. Heidel · J. Dressler
Institut für Rechtsmedizin,
Technische Universität Dresden,
Fetscherstrasse 74,
01307 Dresden, Germany

C. Augustin
Institut für Rechtsmedizin,
Universitätsklinikum Hamburg-Eppendorf,
Butenfeld 34,
22529 Hamburg, Germany

J. Edelmann
Institut für Rechtsmedizin, Universität Leipzig,
Johannisallee 28,
04103 Leipzig, Germany

K. Chamaon
Institut für Neuropathologie,
Otto-von-Guericke-Universität Magdeburg,
Leipziger Strasse 44,
39120 Magdeburg, Germany

R. Szibor (✉)
Institut für Rechtsmedizin,
Otto-von-Guericke-Universität Magdeburg,
Leipziger Strasse 44,
39120 Magdeburg, Germany
e-mail: reinhard.szibor@medizin.uni-magdeburg.de

Introduction

Intron 40 of the von Willebrand factor (vWF) gene [12] located at 12p13 exhibits a highly variable region of about 0.65 kb, which contains 3 juxtaposed STRs [10, 12, 14, 17, 22] and several diallelic polymorphisms. In our paper, we name the three microsatellites orientated in the downstream direction Pol K (vWF–Kimpton), Pol F and Pol P (according to Dewa [4]). Pol K is known as the CODIS marker vWA (synonym VWF31A, vWF) and has gained great significance in forensic science. vWF (Pol K) displays considerable structural variability within the repeat region [16], which has sporadically been used to obtain additional information for kinship testing [20]. Furthermore, null alleles and allele imbalance were observed, which require attention [1, 8, 11]. Attempts were also made to involve the juxtaposed microsatellites in identity and kinship testing [4, 23]. Pena et al. [18] designated Pol F and Pol P vWF2 and vWF1 and analysed 338 haplotypes. Linkage disequilibrium (LD) between alleles was found to be highly significant. Dewa [4] examined the potential usefulness of haplotype analysis in forensic investigations using the STRs Pol K and Pol P and 29 haplotypes were identified in 116 unrelated individuals with the heterozygosity and polymorphism information content being 0.948 and 0.921, respectively. Because the separate amplification of the juxtaposed

STRs provides no haplotypes, both groups [4, 18] inferred haplotype pairs from un-phased genotype data via segregation analyses of family trios. Haplotypes can accidentally be obtained if one of the two loci Pol K and Pol P is homozygous. Therefore, as was shown in a few cases of kinship testing, family constellations may incidentally be suitable for vWF haplotyping [4]. However, the benefit of using closely linked markers for forensic purpose normally requires that these markers are jointly analysed as haplotypes. Such conditions are generally given when Y chromosomal (ChrY) markers are investigated. ChrX typing of hemizygous male DNA also directly reveals haplotypes, and therefore the complete haplotype constellation can also be recognised in females when (confirmed) father–daughter pairs are investigated. The same can partially apply to mothers when very closely linked markers are investigated in father–mother–daughter trios and mother–son duos. Hence, ChrY [5, 9] and ChrX [19, 21, 22] haplotyping were successfully applied in solving complex kinship cases. The inference of haplotype pairs directly from dizygote single individuals requires more sophisticated approaches. Our attempts to create appropriate methods for haplotyping the vWF intron 40 polymorphism require in-depth knowledge of the molecular structure. It is surprising to note that during our basic research in this matter, we found variability to be higher than expected. This is caused not only by the STRs but also by some SNPs and indel polymorphisms (indels). A comprehensive review of human indels was published by Weber et al. [25].

The section of the intron 40 sequence dealt with in this paper spans the region nt941–nt1740 of the vWF intron sequence, which was retrieved from Ensembl v34, Human Exon VIEW http://www.ensembl.org/Homo_sapiens). In this study, we present vWF intron 40 sequencing results on 68 chromosomes and an appropriate allele STR nomenclature, which is in compliance with the ISFG recommendations [3]. Our survey regarding the intron 40 SNPs may assist primer selection for special purposes such as the creation of short amplicon PCR kits [13, 27]. Furthermore, SNP studies can provide inferences of the population and ancestry to which an individual belong [28].

Materials and methods

DNA samples (blood and buccal swabs) were taken from paternity cases in routine kinship testing. German specimens for sequencing analysis were selected, choosing different alleles in the known vWF Pol K polymorphism [13–19]. In addition, some people from Asia and Africa and cell line DNA were included. DNA extraction was carried out using the commercially available QIAamp DNA blood kit (Qiagen, Hilden, Germany). Primers were designed using the Primer3 software http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). The amplification was carried out in a 25 µl PCR reaction volume containing approximately 0.1–1 ng DNA, 200 µM each dNTP, 2 mM $MgCl_2$, 0.5 µM of each primer, 1 U Taq polymerase (Applied Biosystems, Foster City, CA) and 1X PCR buffer. The cycle conditions in a PTC-200 cycler (MJ Research, Watertown, MA) were as follows: 95°C for 3 min soaking, 94°C for 30 s, 56°C for 1 min.

- Primer 1: 5′-TGTGAAAGCCCTAGTGGATG-3′
- Primer 2: 5′-CCTGTGAGTGGGATGCTACA-3′

Before sequencing, we separated the PCR products of heterozygous DNA donors using native horizontal polyacrylamide gel electrophoresis (T 5%, C 3.3%; gel thickness 500 µm, gel length 25 cm; 600 V, 5–8 W, 3–4 h) on the Multiphor II electrophoresis equipment (Pharmacia, Upsala, Sweden). Bands of interest were excised, eluted and re-amplified. Both strands of the PCR products were sequenced with the direct Taq cycle sequencing method using the BigDye-Terminator kit v1.1 (Applied Biosystems) and the PCR primers 1 and 2, respectively. A final volume of 12 µl sequencing mixture contained 2.5 µl reaction mix, 1.5 µl primer (1 pmol/µl) and 10–20 ng PCR product. For amplification, 25 cycles were carried out as follows: 96°C for 15 s, 50°C for 15 s, 60°C for 4 min. Cycle sequencing products were separated in a 47 cm capillary (module *Seq POP4RapidE*), 32 min run time, 30 s injection time. Data were analysed by means of the PE/ABD software Sequencing Analysis 3.7 (Applied Biosystems).

Results

To address primers and polymorphisms, the sequence of intron 40 part nt941–nt1740 is depicted in Fig. 1. Results of sequence analysis are shown in Tables 1 and S1, which is supplied as ESM, demonstrating the three polymorphic regions K, F and P with five variable tetranucleotide blocks, several SNPs and two indel polymorphisms.

Processing in the 5′–3′ direction the first microsatellite (Pol K) in the region of interest is known as CODIS marker vWA and can be described as a tetranucleotide repeat containing a variable number of TCTA and TCTG repeats. The total of the two motifs summarised normally lies in the range 13–22.

To avoid confusion, we continue the conventional vWF nomenclature beginning with a TCTA repeat, which is not strictly in compliance with the ISFG recommendations [3]. The next TCTA polymorphism, named here Pol F, begins 159 bp downstream and is variable in the range of 5 to 12 repeats. A further very complex TCTA marker begins

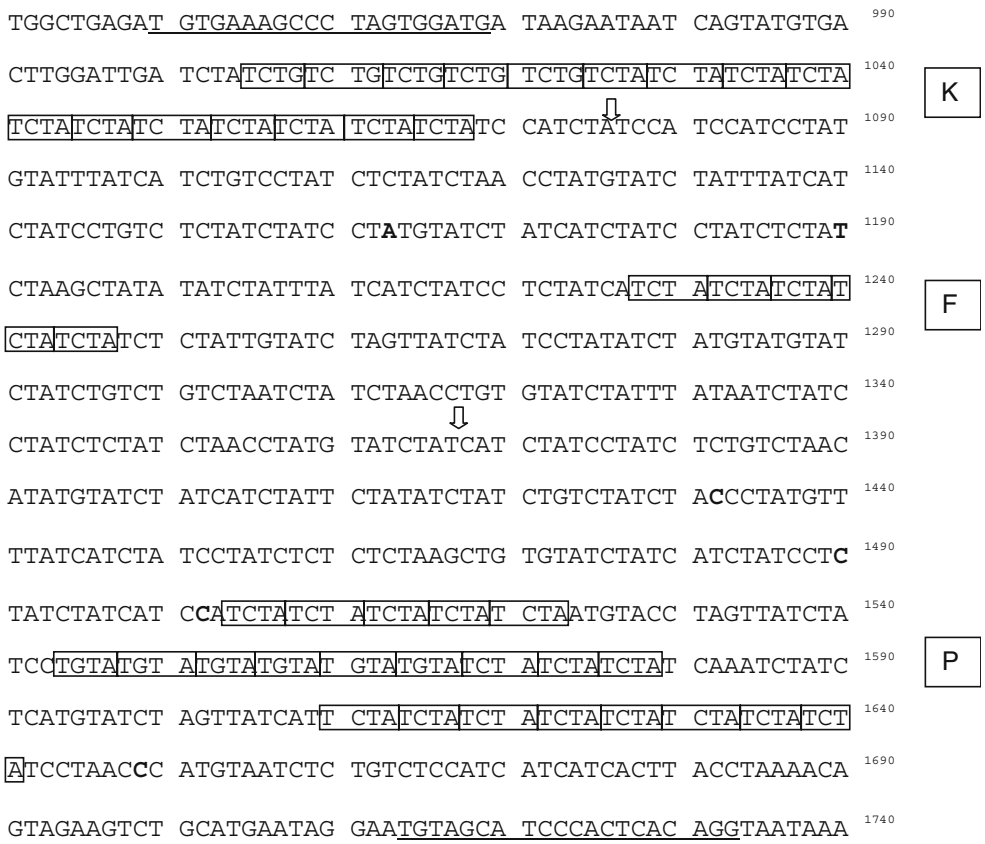

Fig. 1 Section of the vWF intron 40 sequence: the numerals refer to the intron position. Tetranucleotide repeats are boxed. SNP positions are shown as bold letters. Arrows mark the indel positions. Primer binding regions are underlined. STR blocks are named K, F and P

256 bp downstream and is named Pol P. This region shows four variable repeat blocks and can be described with the following common formula in which *m* was found to be 5, 6, 10 or 11, *n* was 6 or 7, *o* was 2–5 and *p* was 8–14:

$$(\mathrm{TCTA})_m 20\,\mathrm{bp} (\mathrm{TGTA})_n (\mathrm{TCTA})_o 30\,\mathrm{bp} (\mathrm{TCTA})_p.$$

Because at least two sections of Pol P can easily be investigated separately, it is justified to treat these polymorphisms as two separate STR systems called Pol P1 and Pol P2.

Table 1 Frequent SNPs and indel polymorphisms and their assignment to the three haplogroups a, b and c

Type of polymorphism	Intron 40 position	Alleles	Haplogroups		
			a	b	c
SNP^{TV}	1,163	A/T	A	T	A
SNP	1,432	C/T	C	C	T
SNP	1,490	C/T	C	C	T
SNP	1,502	C/T	C	C	T
SNP	1,649	C/T	C	C	T
Indel	1,076/1,077	–/(TCTA)$_2$ (TCCA)$_2$	–	–	Ins
Indel	1,367/1,368	–/TTAT	–	–	Ins

Haplogroup a corresponds to the GenBank sequence.
TV: transversion, *Ins*: insertion.

Furthermore, the intron 40 section investigated here contains some SNPs and inserted sequences (Table 1). A 16 bp indel (TCTA)$_2$ (TCCA)$_2$ was found 8 bp downstream of the Pol K. In our sample, only chromosomes exhibiting the common vWF allele 14 were affected by the insertion. In such cases, we found the atypical Pol K repeat structure (TCTA) (TCTG) (TCTA) (TCTG)$_4$ (TCTG)$_3$, which, together with the insertion, gave a total of 14 repeats. A further 4 bp indel (TTAT) occurred 120 bp downstream of the Pol F repeat. This insertion allele also seems to be strictly associated with the common vWF allele 14.

In intron 40, 4 C to T transition SNPs and 1 A to T transversion SNP were detected and are printed in bold in Fig. 1. Owing to the SNP results, we would postulate three types of chromosomes for the German population sample exhibiting haplogroups a, b and c. With regard to the Ensembl sequence, haplogroup a shows no deviation at the SNP loci, haplogroup b is solely characterised by the A–T transversion, while haplogroup c exhibits all four C–T transitions, but no transversion. The 13 alleles of Asian origin and 1 allele of African origin can also be arranged in this haplogroup system. In addition, we found one

chromosome with a T to C transition at intron position 1,190 exhibiting all the characteristics of haplogroup c. Table S1 orders all sequenced chromosomes according to the SNP haplogroup. A total of 47 different haplotypes was found in 68 sequenced amplicons.

Discussion

STR polymorphisms

Sequence variations within the "classical" vWF polymorphism were described earlier [16]. Pol F was first described by Peake et al. [17] who identified 8 alleles ranging from 6 to 14 ATCT repeats. Rare alleles 9 and 15 were reported by Haddad and Sparrow [6]. This is in accordance with our finding of 5 to 12 TCTA repeats. van Amstel and Reitsma [2], Pena et al. [18] and Dewa [4] found 6 to 8 alleles in the polymorphic regions called WBII, vWF2 or vWF-p, respectively. The sub-structure of vWF2 was examined by sequencing and called vWF2-a and vWF2-b [7]. This corresponds to Pol P2 in our paper. A sequencing study, which included some non-Caucasians and cell line DNA revealed 7 alleles with summarised 17 to 23 repeats, but 14 different repeat structures indicating a high degree of genetic variation. Between the Pol F and the Pol P2 region, we found a further variable TCTA repeat block, which we named vWF P1. Therefore, the intron 40 of the vWF gene contains 5 juxtaposed STRs: K, F, P1, P2-a and P2-b.

SNPs and indel polymorphisms

SNP and indel localisation in STR flanking regions is a substantial issue in forensic genetics. During sequencing, several SNPs were detected. The A to T transversion in position 1,163 and the C to T transition in position 1,649 are known in Ensembl SNP VIEW as rs216870 and rs2192205, respectively. An additional G/T SNP is listed as rs28455627, which corresponds to intron position 1,490 where we found a C to T transition. Walsh [24] and Alves et al. [1] described an A to T transversion in position 949 responsible for some concordant typing results using the Perkin Elmer Profiler Plus kit and the Promega PowerPlex kit. Three further SNPs upstream of the K polymorphism were reported by Hendrickson et al. [8] as (−72 T–A), (−77 G–A) and (−90 A–G), registered in Ensembl as rs11063969 to rs11063971 (intron positions 929, 924 and 911). An additional SNP downstream of the K polymorphism (+6 C–T) was localised at position 1,078, adjacent to the (+7 C–T) SNP reported by Lazaruk et al. [11] at position 1,079. Further analysis strategies should consider the high SNP prevalence in this region.

Haplogroups

DNA samples for sequencing were chosen according to their different Pol K alleles to obtain an overview of the polymorphisms in intron 40. This selection and the small sample size (only 5 chromosomes of 51 Germans in haplogroup c) made association tests for the SNPs and the two insertions impracticable. But the existence of this haplogroup structure is obvious. Regarding the relative stability for mutation of SNPs, a strong association of the 5 vWF polymorphisms in the 650 bp section of vWF intron 40 has to be assumed. The existence of a strong LD was also reported by Pena et al. [18] for Pol F and Pol P2.

The atypical structure of Pol K allele 14 with the $(TCTA)_2$ $(TCCA)_2$ insertion was described by Möller et al. [16]. Wiegand et al. [26] compared the relatively conserved repeat structure with the sequence found in hominoid primates and assumed that allele 14 could be an ancestral allele. Minaguchi and Takenaka [15] and Lazaruk et al. [11] listed the repeat structure for the K polymorphism (alleles 11 to 22 and some primates) and found the same variation.

Study of the literature and our own sequencing results revealed considerable diversity in the vWF intron 40. Our review may be valuable in further optimising vWF typing in forensic applications and in avoiding pitfalls. Furthermore, the vWF intron 40 carries the potential for the application of a haplotyping approach aimed at solving very complex kinship cases. Progress in using Y-chromosomal [9] and X-chromosomal [22] haplotyping has revealed the potential of such operations. However, gonosomale haplotyping is not suitable for all cases. Further attempts to develop sophisticated techniques may soon enable haplotyping using autosomal STR clusters.

SNPs, indel positions and the P1 polymorphism found in this study are submitted to NCBI (accession numbers ss65824066–65824073).

References

1. Alves C, Amorim A, Gusmao L, Pereira L (2001) VWA STR genotyping: further inconsistencies between Perkin-Elmer and Promega kits. Int J Legal Med 115:97–99
2. van Amstel HK, Reitsma PH (1990) Tetranucleotide repeat polymorphism in the vWF gene. Nucleic Acids Res 18:4957
3. Bär W, Brinkmann B, Budowle B et al (1997) DNA Commission of the ISFH DNA recommendations—further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 110:175–176
4. Dewa K (1996) Haplotype analysis using two tetrameric loci within the vWF gene: its frequency in Japanese and application in forensic medicine. Nippon Hoigaku Zasshi 50:349–356
5. Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R, Zerjal T, Tyler-Smith C (1998) Jefferson fathered slave's last child. Nature 396:27–28

6. Haddad AP, Sparrow RL (1997) Instability in the ATCT variable number tandem repeat locus VWF.VNTR I in intron 40 of von Willebrand factor gene. Br J Haematol 98:662–664
7. Haddad AP, Sparrow RL (2001) The short tandem repeat locus VWF2 in intron 40 of the von Willebrand factor gene consists of two polymorphic sub-loci. Forensic Sci Int 119:299–304
8. Hendrickson BC, Leclair B, Forrest S et al (2004) Accurate STR allele designations at the FGA and MA loci despite primer site polymorphisms. J Forensic Sci 49:250–254
9. Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124
10. Kimpton C, Walton A, Gill P (1992) A further tetranucleotide repeat polymorphism in the vWF gene. Hum Mol Genet 1:287
11. Lazaruk K, Wallin J, Holt C, Nguyen T, Walsh PS (2001) Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing. Forensic Sci Int 119:1–10
12. Mancuso DJ, Tuley EA, Westfield LA, Worrall NK, Shelton-Inloes BB, Sorace JM, Alevy YG, Sadler JE (1989) Structure of the gene for human von Willebrand factor. J Biol Chem 264:19514–19527
13. Meissner C, Bruse P, Mueller E, Oehmichen M (2006) A new sensitive short pentaplex (ShoP) PCR for typing of degraded DNA. Forensic Sci Int DOI 10.1016/j.forsciint.2006.04.014
14. Mercier B, Gaucher C, Mazurier C (1991) Characterization of 98 alleles in 105 unrelated individuals in the F8VWF gene. Nucleic Acids Res 19:4800
15. Minaguchi K, Takenaka O (2000) Structural variations of the VWA locus in humans and comparison with non-human primates. Forensic Sci Int 113:9–16
16. Möller A, Meyer E, Brinkmann B (1994) Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. Int J Legal Med 106:319–323
17. Peake IR, Bowen D, Bignell P, Liddell MB, Sadler JE, Standen G, Bloom AL (1990) Family studies and prenatal diagnosis in severe von Willebrand disease by polymerase chain reaction amplification of a variable number tandem repeat region of the von Willebrand factor gene. Blood 76:555–561
18. Pena SD, de Souza KT, de Andrade M, Chakraborty R (1994) Allelic associations of two polymorphic microsatellites in intron 40 of the human von Willebrand factor gene. Proc Natl Acad Sci USA 91:723–727
19. Schmidtke J, Kuhnau W, Wand D, Edelmann J, Szibor R, Krawczak M (2004) Prenatal exclusion without involving the putative fathers of an incestuous father–daughter parenthood. Prenat Diagn 24:662–664
20. Szibor R, Plate I, Krause D (1997) Heteroduplex analysis is a rapid method for detection of suballeles caused by mixed length and sequence variability in STR systems. Adv Forensic Haemog 6:346–348
21. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74
22. Szibor R, Hering S, Kuhlisch E, Plate I, Demberger S, Krawczak M, Edelmann J (2005) Haplotyping of STR cluster DXS6801–DXS6809–DXS6789 on Xq21 provides a powerful tool for kinship testing. Int J Legal Med 119:363–369
23. Tamura A, Tsuji H, Nishio H, Suzuki K (1999) Haplotype analysis of a de novo allele at a vWF STR locus using flanking STR loci. Leg Med (Tokyo) 1:188–192
24. Walsh S (1998) Commentary on Kline MC, Jenkins B, Rogers S. Non-amplification of a vWA allele. J Forensic Sci 43:1103–1104
25. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G (2002) Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71:854–862
26. Wiegand P, Meyer E, Brinkmann B (2000) Microsatellite structures in the context of human evolution. Electrophoresis 21:889–895
27. Wiegand P, Klein R, Braunschweiger G, Hohoff C, Brinkmann B (2006) Short amplicon STR multiplex for stain typing. Int J Legal Med 120:160–164
28. Yuasa I, Umetsu K, Watanabe G, Nakamura H, Endoh M, Irizawa Y (2004) MATP polymorphisms in Germans and Japanese: the L374F mutation as a population marker for caucasoids. Int J Legal Med 118:364–366