Int J Legal Med (1996) 109:62–65


ORIGINAL ARTICLE

M. D. Barber · B. H. Parkin

# Sequence analysis and allelic designation of the two short tandem repeat loci D18S51 and D8S1179

Received: 7 May 1995 / Received in revised form: 23 may 1996

**Abstract** This paper reports the sequences of eleven D8S1179 and twenty one D18S51 alleles. The D8S1179 alleles ranged in size from 162 bp to 202 bp and increased in size by regular 4 bp increments. They were shown to possess a compound repeat region composed of the tetranucleotides TCTA and TCTG. Alleles at the D18S51 locus ranged in size from 271 bp to 343 bp and possessed a simple repeat region composed of the tetranucleotide AGAA. The majority of alleles increased in size by 4 bp increments corresponding to the addition of one tetranucleotide repeat unit. However, three alleles differed in size by 2 bp from the 4 bp increment as a result of a dinucleotide insertion within the 3′ flanking region. These alleles also exhibited an altered 3′ flanking sequence in the first four nucleotides following the repeat region. The allelic designations proposed for these loci on the basis of this sequence data are currently being used in a multiplex PCR profiling system employed in a National DNA database in the United Kingdom.

**Key words** Polymerase chain reaction · Short tandem repeat loci · D18S51 · D8S1179 · Tetranucleotide repeat · Allelic designation

## Introduction

The polymorphic nature of many short tandem repeat (STR) loci, and their accessiblity to amplification using the polymerase chain reaction (PCR), has seen their increasing use in forensic identitiy testing. A number of DNA profiling systems based on the singleplex (Alonso et al. 1993; Wiegand et al. 1993; Möller and Brinkmann 1994; Möller et al. 1995) or multiplex amplification (Hochmeister et al. 1995; Pestoni et al. 1995) of STR loci have been evaluated for use with forensic samples. Two DNA profiling systems which are based on the muliplex amplification of STR loci, combined with an automated fluorescent detection system, are currently employed in this laboratory and those of the Forensic Science Service (Kimpton et al. 1994; Oldroyd et al. 1995; Urquhart et al. 1995). As part of the validation of these two profiling systems the alleles of each STR locus employed have been sequenced in order to establish the repeat unit sequence and assign allelic designations. This paper reports the results of these sequencing studies on alleles identified at the D18S51 (Straub et al. 1993) and D8S1179 (selected from the Co-operative Human Linkage Centre database (CHLC), Oldroyd et al. 1995) loci. Data from the repeat unit and flanking regions are presented and allelic designations are proposed.

It has recently emerged that the STR locus D8S1179 had previously been incorrectly designated as D6S502. This was due to a labelling error at a laboratory contributing to the CHLC database (A. Urquhart. Forensic Science Service, personal communication). It should be noted that the oligonucleotide primer sequences and the CHLC designation of GATA7G07 have remained unchanged. Furthermore, there are no other loci on chromosome 8, used in the multiplex system employing D8S1179, to which it could be linked.

M. D. Barber (✉) · B. H. Parkin
The Metropolitan Police Forensic Science Laboratory,
109 Lambeth Road, London SE1 7LP, UK

## Materials and methods

### Sample preparation

Sequence analysis of individual D8S1179 and D18S51 alleles, isolated from donated blood samples, was carried out as described previously (Barber et al. 1995, 1996). However, the following modifications were made: the 1st and 2nd round PCR amplification conditions were 95° C for 60 s, 60° C for 60 s, 72° C for 60 s and a final extension at 72° C for 10 min. D8S1179 and D18S51 alleles were amplified separately using the primer sequences:

D8-1 $^{5'}$TTTTTGTATTTCATGTGTACATTCG$_{3'}$
D8-2 $^{5'}$CGTAGCTATAATTAGTTCATTTTC$_{3'}$ (Oldroyd et al. 1995)

D18-1 $^{5'}$CAAACCCGACTACCAGCAAC$_{3'}$
D18-2 $^{5'}$GAGCCATGTTCATGCCACTG$_{3'}$ (Straub et al. 1993)

In the 2nd round amplification step two separate reactions were carried out using primers with a 5′ biotin label on either the sense or anti-sense strands in order to obtain sequence data from both DNA strands.

### Solid phase sequencing reactions

Solid phase sequencing of the sense and anti-sense strands of D8S1179 alleles and the antisense (TCTT) strand of D18S51 alleles was carried out using a Prism T7 Sequenase dye terminator single stranded sequencing kit (Applied Biosystems) as described previously (Barber et al. 1995, 1996). The sequencing primer was the non-biotinylated primer used in the second round PCR amplification reaction. However, sequence analysis of the sense strand (AGAA) of D18S51 alleles was carried out by Thermo Sequenase (Amersham Life Science) dye terminator cycle sequencing as follows: an aliquot of the second round amplification, containing 0.50 µg of DNA, was added to 300 µg of Dynabeads M-280 Streptavidin (Dynal) in the presence of a final concentration of 2 M lithium chloride and incubated at 48° C for 15 min. All the following wash steps were carried out in a total volume of 500 µl with the aid of an MPC-E magnetic block (Dynal). The immobilised DNA was washed in TT buffer (250 mM Tris-HCl pH 8.0, 0.1% Tween 20) and then in double distilled water prior to being denatured in 100 µl of 1.5 M NaOH for 4 min at room temperature. The eluted nonbiotinylated strand was removed and the DNA/bead complex washed sequentially in TT buffer and TET buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% Twen 20) before being resuspended in an appropriate volume of water.

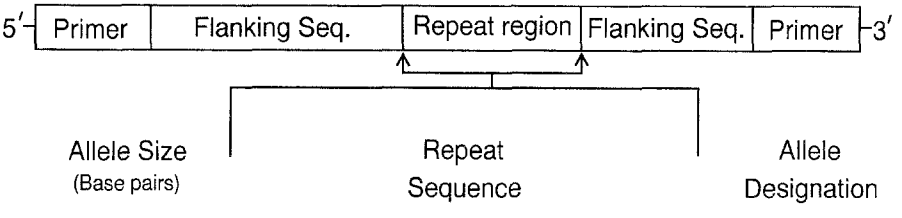
The sequencing reactions were carried out in a Perkin Elmer Cetus 9600 thermal cycling block for 25 cycles of 96° C for 30 s, 54° C for 15 s and 60° C for 240 s using a total reaction volume of 20 µl. The sequencing primer was the non-biotinylated primer used in the second round PCR amplification reaction. The completed sequencing reactions were made to a total volume of 100 µl and the unincorporated labelled terminators removed by extraction with an equal volume of phenol:water:chloroform (68:18:14). The samples were vigorously mixed with a vortex and centrifuged at 13000 rpm for 5 min. The aqueous phase was removed and re-extracted with a 100 µl aliquot of phenol: water:chloroform as before. The aqueous phase was removed and the sequencing products precipitated by the addition of 15µl of 2 M sodium acetate (pH 4.5) and 300 µl of 100% ethanol. The samples were placed on ice for 15 min then centrifuged at 13000 rpm for 15 min at room temperature. The DNA pellet was washed with 500 µl of 70% ethanol then dried under vacuum. The samples were then resuspended in 4 µl of loading buffer (deionised formamide containing 5 mM EDTA, pH 8.0), denatured at 90° C for 4 min and placed on ice.

The sequencing products were loaded onto a 6% polyacrylamide gel containing 8.3 M urea and 1 × TBE buffer (90 mM Tris-HCl, 90 mM boric acid, 0.125 mM EDTA, pH 8.3) and analysed on a model 373A automated DNA sequencer (Applied Biosystems). Electrophoresis was carried out at 2500 V, 40 mA, 30 W and 40° C for 12 h. Sequence data were analysed using 373A version 1.2.1 analysis software (Applied Biosystems). Consensus sequences were created from the sense and antisense strands using SeqEd version 1.0.3 software (Applied Biosystems).

## Results

### D8S1179

A total of eleven D8S1179 alleles have been identified and sequenced in the present study that range in size from 162 bp to 202 bp when amplified with the PCR primers previously described. All the alleles identified increased in size by regular 4 bp increments and the nucleotide sequences of the repeat regions are shown Fig. 1. Alleles at




| Allele Size (Base pairs) | Repeat Sequence | Allele Designation |
|---|---|---|
| 162 | $(TCTA)_8$ | 8 |
| 166 | $(TCTA)_9$ | 9 |
| 170 | $(TCTA)_{10}$ | 10 |
| 174 | $(TCTA)_{11}$ | 11 |
| 178 | $(TCTA)_{12}$ | 12 |
| 182 | TCTA TCTG $(TCTA)_{11}$ | 13 |
| 186 | $(TCTA)_2$ TCTG $(TCTA)_{11}$ | 14 |
| 190 | $(TCTA)_2$ TCTG $(TCTA)_{12}$ | 15 |
| 194 | $(TCTA)_2$ TCTG $(TCTA)_{13}$ | 16 |
| 198 | $(TCTA)_2$ $(TCTG)_2$ $(TCTA)_{13}$ | 17 |
| 202 | $(TCTA)_2$ TCTG $(TCTA)_{15}$ | 18 |

**Fig. 1** Schematic representation of the eleven D8S1179 alleles (TCTR strand) identified in this study, showing the repeat region sequence

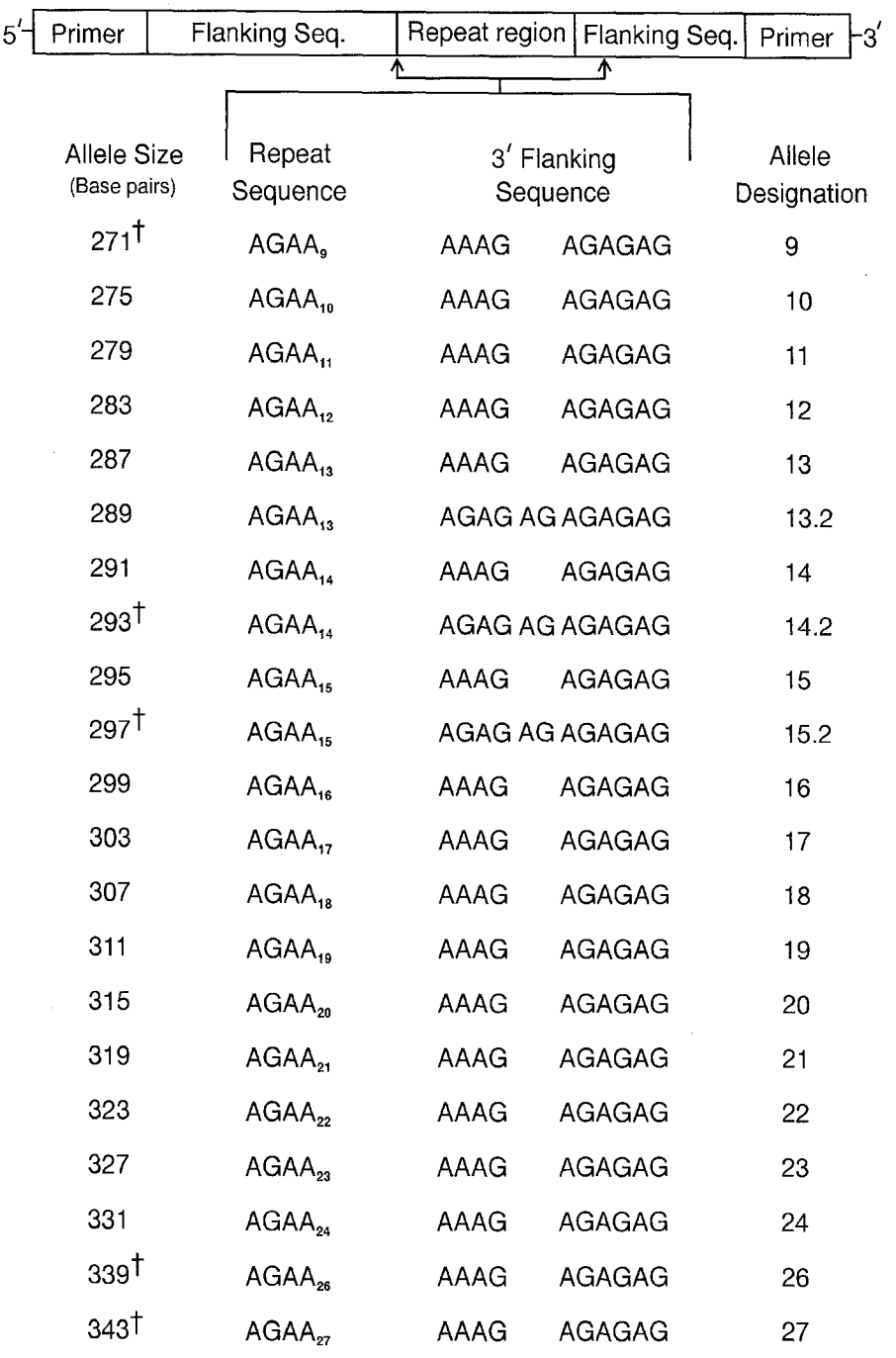



**Fig. 2** Schematic representation of the twenty one D18S51 alleles (AGAA strand) identified in this study, showing the repeat region sequence and the start of the 3′ flanking sequence. † Alleles identified in databasing studies by the Forensic Science Service (Dr. A.Urquhart, personal communication)

the lower end of the observed size range (162–178 bp) possess a simple repeat region based on the tetranucleotide TCTA. However, alleles at the upper end of the observed size range (182–202 bp) possess a compound repeat region composed of the tetranucleotides TCTA and TCTG. No sequence variation was observed in either the 5′ or 3′ flanking regions within the sample of alleles examined in this study. On the basis of this sequence data the repeat unit of D8S1179 alleles is designated as TCTR, where R represents A or G, in accordance with the recommendations of the Nomenclature Committee of the International Union of Biochemistry (1985).

D18S51

A total of twenty one D18S51 alleles have been identified and sequenced in the present study that range in size from 271 bp to 343 bp when amplified with the PCR primers previously described. Sequence analysis has shown that D18S51 alleles possess a simple repeat region composed of the tetranucleotide AGAA (see Fig. 2). Eighteen of these alleles were seen to increase in size by regular 4 bp increments corresponding to the addition of one tetranucleotide (AGAA) repeat unit. It was noted, however, that the difference in size between the 331 and 339 bp alleles was 8 bp so the existance of an allele of 335 bp could be

inferred but has, as yet, not been observed. In addition, three alleles (289, 293 and 297 bp) differed in size by 2 bp from the 4 bp increment. These alleles exhibited an altered 3′ flanking sequence in the first four nucleotides following the repeat region and also appeared to have an AG dinucleotide insertion as shown in Fig. 2. The exact location of this insertion is not known as it occurs within a region containing a number of tandemly repeated AT dinucleotides. This altered 3′ flanking sequence could have arisen as a result of an A to G transition in the first four nucleotides following the repeat region (AAAG to AGAG). However, this would have to be accompanied by a duplication of an AG dinucleotide within this region. Such a mutation may have arisen through slipped mispairing, which has been shown to be a significant mutational mechanism in regions of DNA containing tandemly repeat sequences (Streisinger et al. 1996; Efstratiadis et al. 1980; Mahtani and Willard 1993), although the exact mutational mechanisms is unknown.

## Discussion

The objective of the present study was to sequence representatives of each allele size observed at the D8S1179 and D18S51 STR loci in order to establish their repeat unit structure and assign allelic designations. In accordance with the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics alleles of STR loci are designated according to the number of repeat units each allele contains (ISFH recommendations 1992, 1994). Such a system eliminates problems associated with a nomenclature based on amplified fragment size which requires very accurate sizing of individual alleles. It also simplifies the presentation of STR evidence in court and facilitates direct comparison of STR typing results obtained in different laboratories where alternative primer sequences may have been used. When an allele does not conform to the standard repeat unit motif it is designated with a suffix showing the number of base pairs present in the partial repeat. This system has been used to designate alleles differing by 2 bp from the regular 4 bp increment at the HumFIBRA/FGA locus (Barber et al. 1996). The designations for D8S1179 alleles (Fig. 1) range from 8 for the 162 bp allele to 18 for the 202 bp allele. The designations for D18S51 alleles (Fig. 2) range from 9 for the 271 bp allele to 27 for the 343 bp allele. The three alleles (289, 293 and 297 bp) which differ by 2 bp from the regular 4 bp increment are designated as 13.2, 14.2 and 15.2 respectively. These designations are currently being employed in assigning D8S1179 and D18S51 alleles in a multiplex profiling system employed in a National DNA database in the United Kingdom.

**Acknowledgements** The authors would like to thank Dr. A Urquhart (Forensic Science Service, Birmingham) for many helpful discussions regarding allele nomenclature and for supplying a number of the D18S51 alleles characterised in this study.

## References

Alonso A, Martin P, Albarran C, Sancho M (1993) Amplified fragment length polymorphism analysis of the VNTR locus D1S80 in central spain. Int J Leg Med 105:311–314

Barber MD, Piercy RC, Andersen JF, Parkin BH (1995) Structural variation of novel alleles at the Hum vWA and Hum FES/FPS short tandem repeat loci. Int J Legal Med 108:31–35

Barber MD, McKeown BJ, Parkin BH (1996) Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus. Int J Legal Med 108:180–185

DNA Commission of the International Society for Forensic Haemogenetics (1992) Recommendations of the DNA Comission of the International Society for Forensic Haemogenetics relating to the use of PCR based polymorphisms. Forensic Sci Int 55:1–3

DNA recommendations – 1994 report concerning further recommendations of the DNA commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. Int J Legal Med 107:159–160

Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, DeReil JK, Forget BG, Weissman SM, Slightom JL, Blechl AE, Smithis O, Baralle FE, Shoulders CC, Proudfoot NJ (1980) The structure and evolution of the human β-globin gene family. Cell 21:653–668

Hochmeister MN, Budowle R, Schumm JW, Sprecher CJ, Borer UV, Dirnhofer R (1995) Swiss population data and forensic efficiency values on 3 tetrameric short tandem repeat loci-HUMTHO1, TPOX and CSF1PO – drived using a STR multiplex system. Int J Legal Med 107:246–249

Kimpton CP, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, Gill P (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Legal Med 106:302–311

Mahtani MM, Willard HF (1993) A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci. Hum Mol Genet 2:431–437

Möller A, Brinkmann B (1994) Locus ACTBP2 (SE33): sequencing data reveals considerable polymorhism. Int J Legal Med 106:262–267

Möller A, Schürenkamp M, Brinkmann B (1995) Evaluation of an ACTBP2 ladder composed of 26 sequenced alleles. Int J Legal Med 108:75–78

Nomenclature Committee of the International Union of Biochemistry (1985) Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. Eur J Biochem 150:1–5

Oldroyd NJ, Urquhart AJ, Kimpton CP, Millican ES, Watson SK, Downes T, Gill PD (1995) A highly discriminating octoplex short tandem repeat PCR system suitable for human individual identification. Electrophoresis 16:334–337

Pestoni C, Lareu MV, Rodriguez MS, Muñoz I, Barros F, Carracedo A (1995) The use of the STRs HUMTHO1, HUMVWA31/A, HUMF13A1, HUMFES/FPS, HUMLPL in forensic application: validation studies and population data for Galicia (NW Spain). Int J Legal Med 107:283–290

Straub RE, Speer MC, Luo Y, Rojas K, Overhauser J, Ott J, Gilliam C (1993) A microsatellite genetic linkage map of human chromosome 18. Genomics 15:48–56

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M (1966) Frameshift mutations and the genetic code. Cold Spring Harb Symp Quant Biol 31:77–84

Urquhart A, Oldroyd NJ, Kimpton CP, Gill P (1995) A highly discriminating heptaplex short tandem repeat PCR system for forensic identification. BioTechniques 18:116–121

Wiegand P, Budowle B, Rand S, Brinkmann B (1993) Forensic validation of the STR systems SE33 and TC11. Int J Legal Med 105:315–320