Characterizing the electrophoretic mobility of Huntington Disease Alleles 13-77: Are the deviations from the theoretical values intrinsic or internal sizing artifacts?

Email: margaret.kline@nist.gov
Phone: 301-975-3134

Margaret C. Kline, David L. Duewer, and John M. Butler

National Institute of Standards and Technology (NIST), 100 Bureau Drive MS 6312, Gaithersburg, MD 20899-6312

Poster # G09 :
Association for Molecular Pathology 2010 Annual Meeting.
San Jose, CA, November 17-20, 2010

Disclosure of Potential Competing Interests: *All authors are employed by the National Institute of Standards and Technology which produces and sells Standard Reference Materials*

Introduction: Huntington's disease (HD) is an inherited neurodegenerative disease with clinical manifestations associated with the expansion of a polymorphic CAG repeat. The Applied Genetics Group at the National Institute of Standards and Technology (NIST) developed certified reference material SRM 2393 CAG Repeat Length Mutation in Huntington's Disease, to support accurate clinical determination of HD allele repeat number. The certified values delivered by SRM 2393 are based on Sanger sequencing and were verified by analysis of polymerase chain reaction (PCR) products. Systematic differences were observed between the base pair (bp) size of the alleles determined by sequence and bp size implied by relative electrophoretic mobility. The repeat number of the electrophoretic peak of greatest amplitude within a given envelope of product peaks was observed to decrease with increasing PCR stutter. These phenomena were investigated to assure accurate allele assignment of the HD alleles.

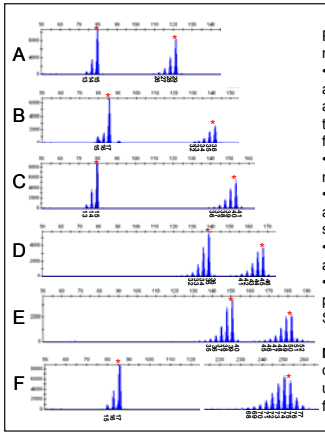
Methods: DNA samples were obtained from Coriell Cell Repositories. Samples were PCR amplified with a 6-FAM labeled forward primer and unlabeled reverse primer using the Roche GC-Rich PCR System (Roche) for genotyping purposes with the primers and amplification conditions listed in Table 1. PCR products were separately analyzed by capillary electrophoresis (3130xl Genetic Analyzer, POP 4, 36 cm array) using two internal size standards (ISSs).

Table 1. PCR parameters for Genotyping the HD locus (low stutter)

Primers [1]:	Forward–6FAM_GCCTTCGAGTCCCTCAAGTCCTTC Reverse–GCGGCGGTGGCGGCTGTTG
PCR amplifications:	10 µL reaction volumes
Master mix:	0.4 unit of GC-RICH enzyme mix, 80 mM dNTPs 1 M GC-RICH reaction buffer 0.5 M GC-RICH resolution solution 0.4 µM of each forward and reverse primer.
Thermocycler:	GeneAmp 9700 Applied Biosystems in 9600-emulation mode (ramp speeds of 1 °C/s)
Thermal cycling:	95 °C for 3 min 10 cycles of {95 °C for 30 s, 68 °C for 30 s, 72 °C for 45 s}, 20 cycles as above but with an additional 0.5 s at 72 °C per cycle
Holds:	72 °C for 7 min 4 °C hold until removed from the thermal cycler

Figure 1 displays genotyping profiles obtained using the above low stutter amplification parameters. Peak areas were determined for the primary and {-3, -6, -9, …, +3 , +6} bp amplification products. These stutter peaks are identified as {n-1, n-2, n-3, …, n+1, n+2} respectively. The profiles from low stutter genotyping of the SRM 2393 materials are in complete agreement with the sequencing data. However, even with this low-stutter system, a primary allele of more than ≈65 repeats is unlikely to be the largest peak in the envelope of amplification products.

Figure 1. Low stutter electropherograms



Primary alleles in the electropherograms are noted with * directly above the peak apex.
•As the number of CAG repeats of the primary allele increases there is an increase in the peak area of the associated stutter products, first in the reverse (n-1, n-2, …) positions, then in the forward (n+1, n+2).
•The primary alleles 15 (A,C) and 17 (B,F) have n-1 and n-2 stutter products.
•n-4 stutter is first noticeable with the primary 29 allele (A); n-5 stutter with the 40 allele (C); n-7 stutter with the 75 allele (F).
•n+1 stutter is noticeable with the primary 35 allele (D); n+2 stutter with the 50 allele (E).
•the primary alleles with associated stutter products are identifiable in all components of SRM 2393.

Note: the size scale (x-axis) is interrupted for component F; the RFU (relative fluorescent units, Y-axis) region beyond 215 bp is expanded for clarity.

Results: Since ISSs are used to infer the size of electrophorectically analyzed sample alleles, the apparent bp sizes are dependent on the mobility of both the ISS components and the sample alleles. The sequence-assigned CAG repeat numbers for the SRM 2393 materials enabled assigning repeat numbers to the apparent size of stutter products and alleles from other materials. The relationships between the number of CAG repeats and the apparent size for the two ISSs are quite smooth, with an average bp difference between consecutive alleles of 2.9 bp ± 0.1 bp. However, the differences between the sequence-assigned and apparent bp sizes systematically increase with ISS-1 while they fluctuate with ISS-2 (Figure 2).

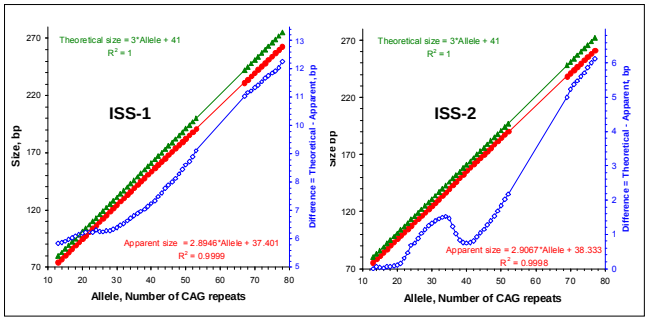


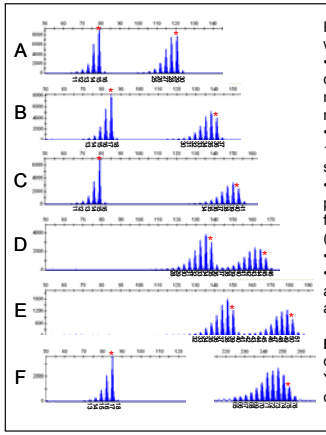
Table 2. ISS assigned bp sizes

Allele	ISS-1 bp	ISS-2 bp	Difference
15	80.09 ± 0.08	81.36 ± 0.05	1.27 bp
17	85.95 ± 0.06	87.33 ± 0.06	1.38 bp
29	121.65 ± 0.05	122.14 ± 0.04	0.49 bp
35	139.18 ± 0.08	139.92 ± 0.06	0.74 bp
36	142.11 ± 0.09	143.16 ± 0.09	1.05 bp
39	150.85 ± 0.07	152.63 ± 0.07	1.78 bp
40	153.75 ± 0.10	155.64 ± 0.05	1.89 bp
45	168.11 ± 0.09	170.24 ± 0.06	2.13 bp
50	182.39 ± 0.11	184.54 ± 0.08	2.15 bp
75	254.14 ± 0.11	255.52 ± 0.07	1.38 bp

The difference between the apparent bp sizes as determined by the two ISSs for the HD alleles delivered by the six SRM 2393 components range from 0.5 bp to 2.15 bp (Table 2). Use of different ISS thus could lead to ambiguity in the assessment of the number of CAG repeats. *This potential ambiguity can be addressed by calibrating genotyping systems with the SRM 2393 materials when assigning allele calls.*

PCR stutter is expected to increase with increasing allele repeat number. When the components of SRM 2393 were amplified using the thermal cycling conditions in Table 1 but using a different manufacturer's "GC rich" reagents, high stutter was observed (Figure 3). For the alleles of more than of 17 CAG repeats delivered by SRM 2393, the primary allele is no longer the largest peak in the amplification product envelope.

Figure 3. High stutter electropherograms

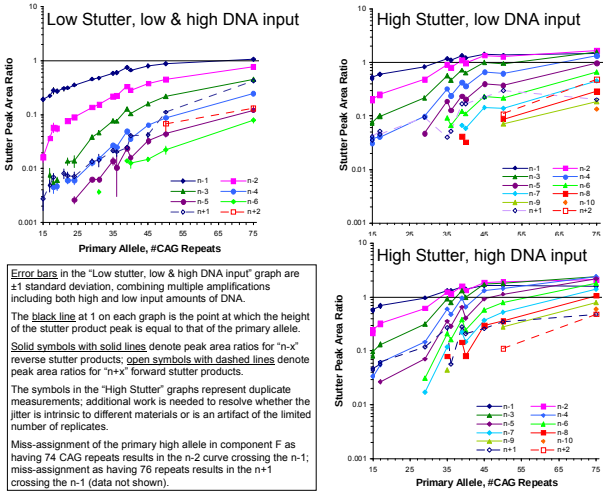


Primary alleles in the electropherograms are noted with * directly above the peak apex.
•As in Fig. 1, there is an increase in the peak area of the associated stutter products with increasing number of CAG repeats, first in the reverse (n-1, n-2, …) positions, then in the forward (n+1, n+2).
•The primary alleles 15 (A,C) and 17 (B,F) have n-1 through n-4 stutter products as well as n+1 stutter.
•n-1 stutter has a larger peak height than the primary alleles beginning at 35 (D); n-6 stutter is first noticeable at 35 (D); n-10 stutter at 75 allele (F).
•n+2 stutter is first noticeable with the 75 allele (F).
•the primary alleles of components D and E are 10 and 11 alleles apart and have a continuum of associated stutter products between.

Note: the size scale (x-axis) is interrupted for component F; the RFU (relative fluorescent units, Y-axis) region beyond 215 bp is expanded for clarity.

To better understand the effect of stutter on the shape of the envelope, peak area ratios of the PCR products were studied in both low and high stutter amplification systems and with both high (35 ng) and low (14 ng) input DNA (Figure 4).

Figure 4. Stutter Peak Areas Ratioed to their Primary Allele



Error bars in the "Low stutter, low & high DNA input" graph are ±1 standard deviation, combining multiple amplifications including both high and low input amounts of DNA.

The black line at 1 on each graph is the point at which the height of the stutter product peak is equal to that of the primary allele.

Solid symbols with solid lines denote peak area ratios for "n-x" reverse stutter products; open symbols with dashed lines denote peak area ratios for "n+x" forward stutter products.

The symbols in the "High Stutter" graphs represent duplicate measurements; additional work is needed to resolve whether the jitter is intrinsic to different materials or is an artifact of the limited number of replicates.

Miss-assignment of the primary high allele in component F as having 74 CAG repeats results in the n-2 curve crossing the n-1; miss-assignment as having 76 repeats results in the n+1 crossing the n-1 (data not shown).

When compared to the peak area of the primary allele, the area ratios of the n-1 to n-4 stutter products increase fairly steadily under all conditions. As previously noted, under low stutter conditions the peak area of the n-1 product exceeds that of the primary ≈ 65 repeats. However, with high stutter low input DNA conditions the n-1 product is expected to be as large or larger than the primary with as few as 35 repeats (see Fig 3). With high stutter high input DNA conditions the n-1 product is expected to be as large or larger than the primary with as few as 29 repeats, this can also be observed in Fig 3.

Genotyping of HD samples must take stutter into consideration. Under high stutter conditions, if the primary allele is chosen as the amplification product with the largest peak height then the critical 39 and 40 alleles will be identified as 38 and 39 alleles (Table 3). *SRM 2393 can be used to quantify the stutter characteristics of a PCR amplification system.*

Table 3. Technical Standards and Guidelines for HD [2]

HD allele definitions

Allele	# CAG repeats
Normal	≤ 26
Mutable normal	27 to 35
HD reduced penetrance	36 to 39
HD	≥ 40

Clinical sizing accuracy

Allele	Repeats
≤ 43	± 1
44 to 50	± 2
51 to 75	± 3
> 75	± 4

The European HD Network recently conducted an interlaboratory study involving 1326 paired results from 121 laboratories representing 15 counties. The results were presented at the European HD Network Plenary Meeting held in September 2010 (Prague). SRM 2393 pre-release material was analyzed by 2 of the laboratories and found concordant and useful [3].

Conclusions: Apparent differences between sequencing and electrophoretic allele sizes are fully explained as ISS artifacts and stutter phenomena. The use of a certified reference material such as SRM 2393 that has been well characterized as to the primary alleles present and stutter behavior are extremely valuable for method calibration and characterization.

References:

[1] Kalman, L. et al. Development of genomic reference materials for Huntington disease genetic testing. Genet Med 2007:9 910 0;719-723.

[2] Potter, N.T. et al., American College of Medical Genetic: Standards and Guidelines for Clinical Genetics Laboratories 2006 Edition; Technical Standards and Guidelines for Huntington Disease.

[3] REGISTRY study http://www.euro-hd.net/html/registry; accessed October 20, 2010.