Verification of somatic CAG repeat expansion by pre-PCR fractionation

Jesse M. Hunter\textsuperscript{a,b,c}, Andrew B. Crouse\textsuperscript{b}, Mathieu Lesort\textsuperscript{b,c}, Gail V.W. Johnson\textsuperscript{b,c}, Peter J. Detloff\textsuperscript{b,∗}

\textsuperscript{a} Department of Cell Biology, University of Alabama at Birmingham, Birmingham, AL 35294, USA
\textsuperscript{b} Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, 540 Kaul Building, 720 20th Street South, Birmingham, AL 35294, USA
\textsuperscript{c} Department of Psychiatry, University of Alabama at Birmingham, 1061 Sparks Center, 1720 7th Avenue South, Birmingham, AL 35294-0017, USA

Received 12 August 2004; received in revised form 7 October 2004; accepted 8 October 2004

Abstract

The inheritance of a long CAG repeat causes several late onset neurological disorders including Huntington’s disease (HD). Longer CAG repeats correlate with earlier onset of HD suggesting an increased toxicity for the products of long repeat alleles. PCR based data has been used to show that HD CAG repeat expansion beyond the inherited length occurs in affected tissues indicating a possible role for somatic instability in the disease process. PCR, however, is prone to artifacts resulting from expansion of repeat sequences during amplification. We describe a method to distinguish between CAG repeat expansions that exist in vivo and those that potentially occur during PCR. The method involves size fractionation of genomic restriction fragments containing the expanded repeats followed by PCR amplification. The application of this method confirms the presence of somatic expansions in the brains of a knock-in mouse model of HD.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Repeat instability; Trinucleotide; Polyglutamine; PCR; Shadow band

1. Introduction

Huntington’s disease (HD) is fatal, late onset and neurodegenerative. The clinical manifestations of HD include psychiatric disturbances, chorea, cognitive decline, and dementia (Kremer, 2002). The genetic cause of HD is the inheritance of an expanded CAG repeat in exon 1 of the Huntingtin gene (Huntington’s Disease Collaborative Research Group, 1993). Alleles vary in repeat number from 6 to 35 CAGs for non-pathogenic alleles to 36 or greater for disease alleles. The length of pathogenic CAG repeats is inversely correlated to the age of onset, with longer repeats tending to cause an earlier onset (Gusella et al., 1996).

The length of the CAG repeat of the \textit{HD} gene can change from generation to generation and large expansions responsible for most juvenile onset cases are typically inherited from a father carrying an expanded HD allele (MacDonald et al., 1993). Such germline expansions are thought to be responsible for the tendency of offspring to have ages of onset earlier than preceding generations. Although somatic changes in repeat length occur much less frequently than in the germline, knock-in mice with expanded HD repeats exhibit late-onset expansions of long repeat HD alleles in the striatum, a brain region that is severely affected by the disease process (Kennedy and Shelbourne, 2000; Kennedy et al., 2003). Somatic instability in the brains of other HD mouse models as well as HD patient material has also been reported (Kennedy et al., 2003; Telenius et al., 1994; Wheeler et al., 1999). These data highlight the possibility that the selective vulnerability of the striatum might be caused by tissue specific expansions and suggest that inhibiting somatic expansion might be therapeutically beneficial (Kennedy and Shelbourne, 2000; Wheeler et al., 2003).
The molecular details of repeat instability have yet to be elucidated. Several models have been proposed to explain changes in repeat size seen in many model systems (Parniewski and Staczek, 2002). Such mechanisms involve a single strand of a repeat base pairing out of register with its complementary strand during recombination or replication, including DNA repair related replication. Evidence that single strands of some repeats can form stable secondary structures in vitro and in vivo from S. cerevisiae suggest a means by which structures base paired out of register can be stabilized and highlight an intrinsic quality of such repeats thought to contribute to their length instability (McMurray, 1999; Moore et al., 1999).

Direct methods of determining instability are PCR based and thus rely on multiple rounds of DNA replication to amplify the repeat containing regions. The typical product of a long repeat PCR reaction contains several additional bands of lesser quantity approximately the same size as the most abundant product. Instability is sometimes determined by a skewing of the distribution of such products (De Rooij et al., 1995; Gourdon et al., 1997; Ishiguro et al., 2001; Kennedy and Shelbourne, 2000; Mangiarini et al., 1997; Monckton et al., 1997; Telenius et al., 1994; Wheeler et al., 1999). A more convincing method involves PCR using a small number of template molecules (~10) which allows the amplification of very long expanded templates without the great excess of competing short repeat template (Kennedy and Shelbourne, 2000; Monckton et al., 1995). The caveat of these methods is the possibility that alterations in repeat length might take place during the early rounds of replication in the PCR reaction. The finding that PCR products with repeat expansions (beyond the inherited length) are only found in selected tissues of knock-in HD mice reveals important tissue to tissue differences in the HD template DNA prior to PCR. This difference might reflect an expansion that has occurred prior to PCR (Kennedy and Shelbourne, 2000). Alternatively, the difference might reflect a tissue specific alteration to the DNA that enhances expansion during PCR. One of many possible mechanisms of expansion during PCR is shown in Fig. 1b. Tissue specific double strand breaks and other DNA modifications have been described in HD mouse models and patient material (Bogdanov et al., 2001; Butterworth et al., 1998; Draganov et al., 1995; Giuliano et al., 2003; Portera-Cailliau et al., 1995; Thomas et al., 1995). Furthermore, it would be expected that the amount of any substrate involving a break or nick would either be enhanced or reduced by mutations in DNA repair systems, which might explain alterations in the amount of expanded repeat PCR products from such mutants (Gomes-Pereira et al., 2004; Spiro and McMurray, 2003; Wheeler et al., 2003). Finally, in vitro expansion might also explain the propensity for products of PCR from tissues to favor increases in repeat length over decreases. Some specialized methods of PCR that favor expansions of repeat sequences have been used as a means of creating in vitro expansions (e.g. Ordway and Detloff, 1996). Thus, it is not yet clear whether lengthened PCR products reflect in vivo expansions.

We describe here a method to determine whether expansions occur in vivo or during PCR. This method, pre-PCR fractionation, involves the sizing and isolation of restrictive DNA template molecules (µl) which allows the amplification of expanded repeat templates. Single strands of repeat DNA are depicted (shaded boxes) flanked by PCR primer-binding sites (open boxes) within the context of the chromosome (lines with breaks). These modifications might enhance the stability of an intramolecular structure formed during PCR on either the new (6) or template (7) strands or provide a break that could be extended by polymerase (8). An example of a mechanism where expansion during PCR would be favored over contraction (1) Tissue specific, age-dependent nicks or double strand breaks might occur at the end of a CAG repeat. Sites internal to the repeat might be protected by proteins known to bind tightly to CAG repeats or the CAG sequence might not be recognized by endogenous endonucleases. (2) During early cycles of PCR, a DNA strand broken at the end of a CAG repeat could anneal out of register and since it is already full length, it could only be expanded beyond the inherited length. (3) The polymerase could then add the second primer-binding site resulting in an expanded CAG repeat template. (4) The expanded CAG repeat formed during early PCR cycles could then be amplified during later PCR cycles.
tion fragments containing trinucleotide repeats prior to PCR. Such methods have been previously applied to examine the instability of human minisatellite sequences (Jeffreys and Neumann, 1997; Jeffreys et al., 1990). We have used pre-PCR fractionation of DNA from a knock-in mouse line carrying a long CAG repeat allele of the mouse HD gene to confirm the presence of somatic repeat expansions prior to amplification by PCR.

2. Materials and methods

2.1. Mice, brain dissection, and DNA isolation

The knock-in HD mouse model used in this study, CHL2, was described previously (Lin et al., 2001). Briefly, this model was made by expanding the 21 base pair endogenous repeat CAGCAGCAGCAGCAGCAG in exon 1 of the mouse HD homolog, Hdh, to 150 CAGs in ES cells. Standard blastocyst injections of these cells were used to produce mice. Three wild type, seven heterozygous and six homozygous Hdh(CAG)150 mice were used in this study. All experiments were conducted according to IACUC guidelines. Tail biopsies from all mice were taken at three weeks of age. DNA isolated from these tail samples was used for genotyping mice and as a control of the CAG repeat size initially inherited. Mice at 1 year of age were euthanized by CO2 inhalation and decapitated. The brain was rapidly extracted from the skull, and the striatum was isolated by dissection then flash-frozen on dry ice. DNA was isolated from the striatum by standard proteinase K digestion followed by phenol/chloroform extraction (Strauss, 1998).

2.2. DNA digestion and Pre-PCR fractionation

An outline of the general procedure is shown in Fig. 2. Striatal CHL2 mouse DNA (1–3 μg) was digested with 10 Units of the restriction enzymes PstI and AatII (NewEngland Biolabs) at 37°C for 3 h. This digest yields a 996 bp DNA fragment corresponding to the wild type allele and a 1425 bp mutant DNA fragment corresponding to an unexpanded Hdh(CAG)150 allele. Digested striatal DNA was mixed with 10× DNA loading buffer without dye and loaded onto a 25 cm long 2% TAE-agarose gel. DNA Ladder was run in a lane that was not adjacent to the samples to avoid contamination. Gels were run at 150 V for approximately 6 h. Running buffer was changed after 3 h to ensure that the buffer capacity was not lost, and to prevent gels from overheating. After electrophoresis, the lane containing the DNA size markers was cut from the gel and stained with ethidium bromide (EtBr) for visualization under UV light. The position of the DNA ladder fragments in the gel was recorded. Lanes containing samples were not stained, since EtBr might interfere with PCR. Gel slices (1–3 mm, approximately 100 μl volume each) were taken from 800–1200 bp for wild type DNA and 1200–2500 bp for the Hdh(CAG)150 allele. A clean razor blade was used for each fraction, to avoid cross contamination between fractions. Gel slices were used as template in PCR amplification. Care was taken to ensure that buffers, gels, electrophoresis equipment, and tools used to isolate gel slices were free of DNA contamination.

Fig. 2. Diagram of the method of Pre-PCR fractionation of CAG repeats. (1) Genomic DNA is cut at specific restriction sites outside of the CAG repeat. If expansion or contraction of the CAG repeat are present in vivo different sized fragments will result. (2) DNA is size fractionated by agarose gel electrophoresis before PCR amplification. (3) DNA fragments of various sizes are excised from the gel. Each fraction is then amplified separately by PCR. (4) The PCR products from each fraction are separated again according to size by agarose gel electrophoresis. Possible outcomes: (a) if no CAG repeat instability occurs in vivo or during PCR, amplification products will appear as single bands of the same size in all fractions containing template in the second electrophoresis; (b) if CAG repeat instability only occurs in vivo, there will be a correspondence between the restriction fragment template size and the product of PCR, resulting in a staircase pattern; (c) if CAG repeat instability only occurs during PCR amplification, then one or a few fractions containing long repeat template should show multiple bands or a large smear after PCR; (d) if CAG repeat instability occurs in vivo and during PCR several bands would be produced in each lane in addition to a staircase pattern.
2.3. DNA amplification by PCR

The PCR protocol used in this study replicated, as closely as possible, the conditions described by Kennedy and Shelbourne (2000), the only differences being an increase of 10 PCR cycles at 60 °C to increase product yield and the use of Taq DNA polymerase from Fisher Scientific. Primers were MHD16 (5′-GCC ATT CAT TGC CTT GCT GCT AAC-3′) and MHD18 (5′-GAC TCA CCG TCG TGT CAG CCG TTC C-3′) from Kennedy and Shelbourne (2000). For conventional genotyping PCR the final reaction mix contained 10% DMSO, 1× Fisher buffer B (10 mM Tris–HCl pH 9.0 at room temperature, 50 mM KCl), 1.0 mM MgCl2, 200 μM dNTPs, 1 μM MHD16 and 1 μM MHD18, 1 U of Taq DNA polymerase per 10 μl of reaction mix. The thermocycler program was as follows: (1) 94 °C for 4 min, (2) 94 °C for 30 s, (3) 71 °C for 30 s, decreasing 0.5 °C cycle, (4) 72 °C for 30 s, (5) GOTO step 2 for 19 more cycles, (6) 94 °C for 30 s, (7) 60 °C for 30 s, (8) 72 °C for 30 s, (9) GOTO step 6 for 19 more cycles, (10) 72 °C for 7 min, (11) hold at 4 °C. PCR of gel slice templates was carried out with the same protocol except it was necessary to increase the MgCl2 concentration from 1.0 to 1.5 mM to counteract the Mg2+ chelating affect of EDTA present in TAE buffered gels. In order to add DNA from a template in solidified agarose, the gel slice templates from the first electrophoresis were melted in a heat block at 95 °C. Ten microlitres of molten agarose gel template (approximately one-tenth of each gel slice) was used in a 50 μl PCR reaction. Each lane from the fractionation gel contained portions of polymerase; less than 1 U of polymerase per 10 μl of reaction mix. The molten PCR products were then loaded into dry wells of the gel in order of fragment size isolated. The molten PCR products solidified within the well. After loading all of the lanes, the gel was submerged in TAE running buffer. Electrophoresis was carried out as described for the first electrophoresis (above). The gel was then stained in EB and destained in water prior to visualization under UV light.

2.5. Second electrophoresis

Amplification products from PCR were loaded on another 25 cm long 2% TAE-agarose gel. Since agarose was present in the template added to the PCR, cooled PCR products solidified. In order to load the PCR products on a gel they were heated to 95 °C and 10× loading buffer with no dye (20% Ficoll 400, 0.1 M EDTA, 1% SDS) was added to the PCR products. The PCR products were then cooled slowly to 60 °C. The molten PCR products were then loaded into dry wells of the gel in order of fragment size isolated. The molten PCR products solidified within the well. After loading all of the lanes, the gel was submerged in TAE running buffer. Electrophoresis was carried out as described for the first electrophoresis (above). The gel was then stained in EB and destained in water prior to visualization under UV light.

3. Results

The overall strategy as outlined in Fig. 2 involved digesting sample DNA with restriction enzymes that recognize sites flanking the CAG repeat. The resulting fragments were separated by agarose gel electrophoresis and isolated by cutting portions of the gel corresponding to different potential sizes of restriction fragments. The resulting size fractionated template DNAs were amplified by PCR and then analyzed by standard agarose gel electrophoresis. In vivo alterations in repeat size were expected to result in several positive fractions and a correspondence in fraction size with the size of the final PCR product. In contrast, the lack of in vivo size differences was expected to result in PCR product from only one (or two adjacent) template fractions representing the inherited allele. Furthermore, the lack of a correspondence between template fraction size and PCR product size would have suggested that alternation in repeat length had occurred during the PCR reaction. Fig. 2 shows some potential outcomes of the procedure to highlight differences expected, if expansions had occurred in vivo versus in vitro.

Pre-PCR fractionation was applied to striatal DNA from Hdh(CAG)150 knock-in mice to determine if in vivo repeat expansions occur in wild type, heterozygous, and homozygous Hdh(CAG)150 mice. Previous work has shown longer repeats to be much less stable than shorter ones (Parniewski and Staczek, 2002). Consistent with these previous findings we found no increases for the short repeat in wild type mice (Fig. 3a). Gel fractions isolated from striatum of mice with long repeats, however, yield PCR products expected from in vivo expansions. Restriction fragment fractions of greater length than the inherited CAG repeat length yielded PCR products of greater length than the inherited allele. Further-
more, there was a good correspondence in length of the restriction fragment template (as sized by the first gel electrophoresis) and the length of the PCR product (as shown by the second electrophoresis). This correlation of size resulted in the staircase pattern expected for in vivo repeat expansion seen in Fig. 3b and c. The sizes of the bands shown in these expanded fractions (lanes 1–8 of Fig. 3b and c) had similar lengths to smears in PCR from unfractinated striatum (lane S of Fig. 3b and c), suggesting that these smears represent a population of in vivo expansions. Comparisons from mouse to mouse, showed differences in the degree of repeat expansion (data not shown), but no mice in this study revealed a banding pattern consistent with somatic contractions. In general, as the repeat number increased the amplified product became less visible, suggesting either less template was present in fractions representing very long expansions or a lesser efficiency of PCR for longer repeat lengths as previously reported (Mutter and Boynton, 1995).

An unusual extra band slightly smaller and of lesser intensity than the large repeat band was observed in some lanes representing long repeats in the pre-PCR fractionation experiments (Fig. 3c). This unusual band increased in size with increasing size of the long CAG repeat fragment, shadowing the larger band. Southern analysis showed this ‘shadow’ band hybridized to a radiolabeled CAG repeat probe indicating it contained homology with the repeat region (data not shown). A similar band was also observed in controls that were not fractionated prior to PCR (lanes T and S of Fig. 3c). We performed further analyses to determine whether this band might represent an in vivo contraction of the long CAG repeat. Direct PCR (template not fractionated) from tail biopsy DNA from heterozygotes (n > 50) always had a similar extra band migrating slightly faster than the long repeat band (Fig. 4a). This band was not seen in homozygotes or in wild type mice, which had only one band of the expected size (see Fig. 4, n > 50 for each genotype). The band also appeared when DNA from a homozygous Hdh(CAG)150 mouse was mixed with DNA from a wild type mouse then subject to PCR amplification. The presence of the band did not rely on multiple PCR cycles as shown by mixing wild type and homozygous PCR products then subjecting the mix to a single cycle of PCR (Fig. 4a, lanes 4–5) or simply heat denaturing and rapidly cooling the mixture (data not shown). The presence of the shadow band and several other bands that might be mistaken for expansions or contractions of the CAG repeat
also depend on the ratios of this mixed DNA (compare lanes 4 and 5 of Fig. 4a). A shadow band appeared after PCR of a mixture of wild type DNA and sized fractions of striatal DNA from a homozygous Hdh(CAG)150 mouse (Fig. 4b). A shadow band was less prominent after pre-PCR fractionation of heterozygotes than it was in direct PCR, suggesting the gel fractionation removed most but not all of the wild type length allele from the longer repeat fractions. Incomplete digestion of the wild type allele might have provided minute amounts of wild type fragments with lengths similar to expanded alleles.

4. Discussion

We have used size fractionation of template DNA prior to PCR to confirm the presence of CAG repeat expansions in the striatum of mice that had inherited the Hdh(CAG)150 allele. The need for such a method is based on the theoretical possibility that CAG repeat expansions might occur during PCR and on results from experiments where PCR has been used as a means of lengthening repeats. Prior to the development of this method, there was ambiguity as to whether expanded PCR products represented in vivo or in vitro expansions. The correspondence of fragment size before and after amplification provides evidence that the expansions exist in vivo.

We also observed additional bands when PCR was performed on samples that included both wild type Hdh and the long repeat Hdh(CAG)150 alleles. These bands could easily be mistaken for expansions or contractions of the CAG repeat. Our experiments, however, suggest they represent unusual structures formed by denaturation and reannealing of DNA. For example, the shadow band seen in Fig. 4 is most likely a combination of a single strand of DNA from a long CAG repeat allele base paired to a single strand of DNA from the short repeat allele. The presence of extra bands in the mixing experiment highlights the problems with the assumption that a new band on a gel from a PCR product represents an in vivo change in repeat size and demonstrates the necessity of pre-PCR fractionation to verify any such claim.

The limitations and problems of this technique are common to other PCR based techniques. First, conclusions should not be based on the absence of amplification products. PCR can be inhibited by a number of conditions and the presence or absence of a band in non-optimal conditions might rely on stochastic events occurring during the first few PCR cycles. Thus, the failure to detect certain expansions should not be used to draw the conclusion that those expansions do not exist. Pre-PCR fractionation is best applied to tissues where expansions are known to be present. Second, despite the simplicity of this method, which allows its use in any lab that can perform PCR, the extra manipulations of the template increase the likelihood of contamination by amplifiable sequence. Contamination during fractionation is particularly troublesome since the amount of expanded template is probably very low in some fractions. Kennedy and Shelbourne (2000) have estimated one large expansion per 250 cells. Electrophoresis buffers and portions of the apparatus that contact the gel used to obtain size fractions should be free of even trace quantities of the products of the PCR. Thus it may be necessary to acid wash or UV treat any surface that the gel or buffer will contact prior to the first electrophoresis and use a separate apparatus for the second electrophoresis. The presence of a wild type band (and a shadow band) in fractions that should have only represented expanded repeats (Fig. 3c) emphasizes the difficulty of eliminating such contamination. Third, pre-PCR fractionation is not the most straightforward way to determine the absolute size of repeat expansions. Although, one could devise a method of taking a fraction for each possible repeat size, the number of fractions involved would be prohibitive. Thus, pre-PCR fractionation is best used as a qualitative means of determining whether expansions occur in vivo and as a complement to direct PCR based techniques more suited for a determination of exact repeat size (De Rooij et al., 1995; Ishiguro et al., 2003; Kennedy and Shelbourne, 2000; Mangiarini et al., 1997; Telenius et al., 1994; Wheeler et al., 2003). Fourth, these results should only be considered valid for a given set of PCR conditions, since the alteration of such conditions might produce in vitro changes in repeat length or additional non-specific PCR products. We chose conditions similar to Kennedy and Shelbourne (2000), given their elegant use of small pool PCR. As such our results most strongly support their arguments that somatic CAG repeat expansions exist in vivo.

The existence of these expansions, now verified by pre-PCR fractionation, leaves us with age-old question of consequence versus causation. The expansions might be a downstream event resulting from polyglutamine induced DNA damage followed by inaccurate DNA repair. Alternatively, the expansions may contribute to the disease process, leaving us with the hopeful hypothesis that inhibition of such expansions might provide therapy for CAG repeat disorders.

Acknowledgements

Appreciation is expressed to the members of the Detloff, Johnson and Lesort laboratories for expert technical assistance and helpful discussions. This research was funded by a NIH Predoctoral Neuroscience Training Grant, NIH grants NS41744 (GWW), NS41552 (ML), and the HDF Cure HD Initiative (PD).

References
