![The University of Edinburgh logo]

THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

A general method for the detection of large CAG repeat expansions by fluorescent PCR

Link:
[Link to publication record in Edinburgh Research Explorer](#)

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of Medical Genetics

OPEN ACCESS

A general method for the detection of large CAG repeat expansions by fluorescent PCR

Jon P Warner, Lilias H Barron, David Goudie, Kevin Kelly, David Dow, David R Fitzpatrick, David J H Brock

Abstract
The expansion of a tandemly repeated trinucleotide sequence, CAG, is the mutational mechanism for several human genetic diseases. We present a generally applicable PCR amplification method using a fluorescently labelled locus specific primer flanking the CAG repeat together with paired primers amplifying from multiple priming sites within the CAG repeat. Triplet repeat primed PCR (TP PCR) gives a characteristic ladder on the fluorescence trace enabling the rapid identification of large pathogenetic CAG repeats that cannot be amplified using flanking primers. We used our method to test a cohort of 183 people from myotonic dystrophy families including unaffected subjects and spouses. Eighty five clinically affected subjects with expanded alleles on Southern blot analysis were all correctly identified by TP PCR. This method is applicable for any human diseases involving CAG repeat expansions.
(J Med Genet 1996;33:1022–1026)

Key words: fluorescent PCR; CAG repeat; myotonic dystrophy; TP PCR.

**Human Genetics Unit,
Molecular Medicine
Centre, Western
General Hospital,
Edinburgh EH4 2XU,
UK**
J P Warner
L H Barron
D R Fitzpatrick
D J H Brock

**Department of
Pathology,
Ninewells Hospital and
Medical School,
Dundee DD1 9SY, UK**
D Goudie

**Medical Genetics
Laboratory,
University of
Aberdeen Medical
School,
Aberdeen AB9 2ZD,
UK**
K Kelly

**East Anglian Region
Genetics Service,
Addenbrooke's
Hospital, Hills Road,
Cambridge CB2 2QQ,
UK**
D Dow

Correspondence to:
Dr Warner.

The expansion of the trinucleotide sequence CAG is the mutational mechanism for several human genetic diseases including myotonic dystrophy, Huntington's disease, spinocerebellar ataxia I and III, and dentatorubralpalludoluysian atrophy.[1-5] Currently, for these diseases, PCR using flanking primers allows amplification up to approximately 100 CAG repeats but is unreliable above this size. The diagnostic threshold for this series of diseases is at between 40 to 50 repeats with alleles above this size being in the affected range. The myotonic dystrophy CAG repeat can expand to give alleles of greater than 5 kb.[1] PCR amplification of DNA from an affected person carrying such an expansion gives a single normal allele and fails to pick up a larger allele. For people apparently homozygous for small common repeat size, the presence of a large unamplifiable allele can be excluded by performing a Southern blot and probing with a locus specific fragment flanking or containing the CAG repeat. We have developed a simple fluorescent PCR system that can rapidly identify but not size the largest alleles for any disorder with a CAG repeat expansion. This will reduce the number of Southern blot analyses required by allowing the identification of

people homozygous for smaller alleles. Samples with large CAG expansions identified by TP PCR will, however, still require Southern blotting if an accurate estimation of size is needed.

Materials and methods
PCR DESIGN
Traditionally, PCR amplification of these unstable DNA sequences involves the use of a pair of locus specific primers (P1, P2) that flank the variable trinucleotide sequence. Our PCR assay uses a specific flanking fluorescently labelled primer (P1) in concert with a pair of primers (P3, P4) which have a common 5' sequence (tail) (fig 1). The common 5' sequence or tail primer (P3) was selected from an artificially produced random DNA sequence. The selection criteria for the sequence of the tail primer were (1) that it should contain little or no self complementarity, (2) no complementarity with either $(GCA)_5$ or $(TGC)_5$, and (3) no homology to known human sequences. The paired primer (P4) has the sequence $(GCA)_5$ or $(TGC)_5$ at its 3' terminus depending on the strand of the CAG repeat to be amplified.

In the early amplification cycles the repeat specific 3' terminus of P4 primes at multiple sites within CAG repeat alleles giving rise to a mixture of products. Specificity is dictated by the fluorescent locus specific primer (P1). A 10:1 molar ratio of P3 to P4 ensures that primer P4 is exhausted in the early amplification cycles. This reduces priming at $(CAG)_n$ sites internal to the PCR products produced in earlier rounds, which would result in a gradual shortening of the average PCR product size (fig 1). The phenomenon of maintaining the allele size distribution by using a combination of tailed primer pairs was first described by Neil and Jeffreys[2] in the context of satellite DNA repeat amplification. The primer P3 preferentially binds to the end of products from previous amplification rounds owing to the stabilising effect of the 5' tail sequence. A long extension time is used to allow complete extension of the larger sized products within the PCR product mixture.

POLYMERASE CHAIN REACTION
DNA used for PCR amplification was extracted from venous blood leucocytes using standard DNA extraction methods. Primer sequences are shown in table 1. The sequences of HD1 and DRPLA 2 were selected from the HD[3] and DRPLA[5] gene sequences respectively. P1
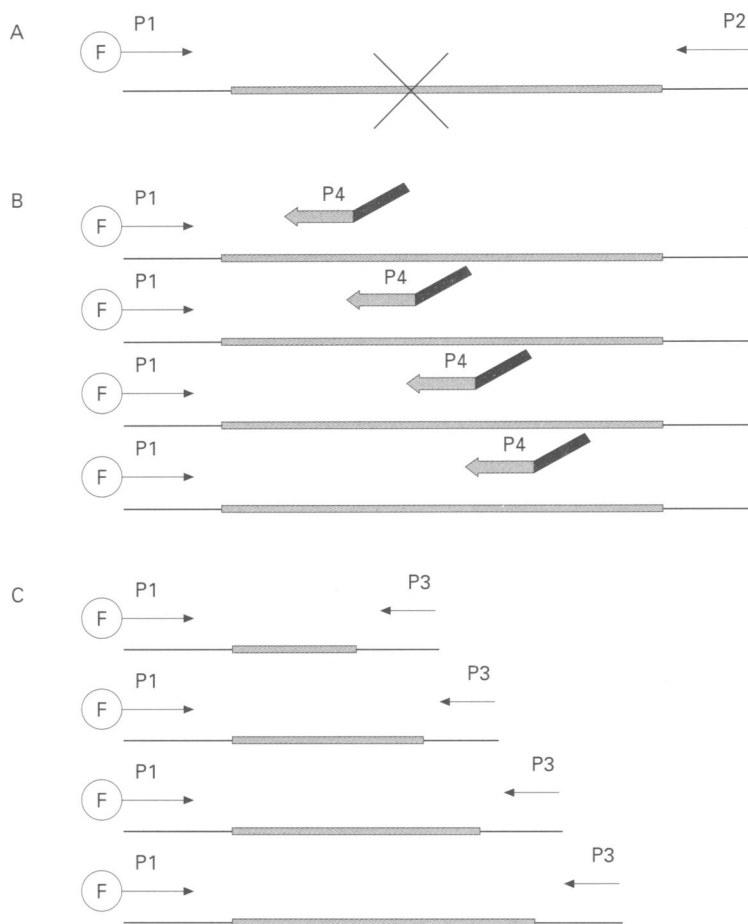
Figure 1   Design details for internally primed repeat specific PCR. Stippled box represents (CAG)ₙ repeat. F shows 5' fluoresceinated primer. (A) For large alleles exceeding 100 CAG repeats the PCR using flanking primers P1 and P2 fails to give a product. (B) In the early amplification cycles primer the repeat specific 3' terminus of P4 binds at multiple sites within CAG repeat alleles giving rise to a mixture of products. Specificity is dictated by the fluorescent locus specific primer (P1). A 10:1 molar ratio of P3 to P4 ensures that primer P4 is exhausted in the early amplification cycles. (C) The primer P3 amplifies from the end of products from previous amplification rounds. A long extension time is used to allow complete extension of the larger sized products within the PCR product mixture and conserve the representation of the longer products.

and P2 were selected from the 3' non-coding region of the myotonin protein kinase gene DNA sequence.[1] The primers flank the myotonin CAG repeat, have matched annealing temperatures, and are also appropriate for amplification in conjunction with the primers P3R, P4CTG, and P4CAG. PCR reactions were performed in 25 µl reaction volumes. The following buffer components were found to give optimal amplification: 1.5 mmol/l MgCl$_2$, 10 mmol/l Tris HCl, pH 8.3, 50 mmol/l KCl, 10% DMSO, 200 µmol/l dATP, 200 µmol/l dCTP, 200 µmol/l dGTP, and 200 µmol/l dTTP. Genomic DNA (200 ng-1 µg) and approximately 2 U of Taq polymerase (Cetus) were used for each reaction. The following primer combinations were used. Mix A: P1 (5'

Table 1   Primer sequences

Primer	Sequence (5'-3')
P1	AGA AAG AAA TGG TTC TGT GAT CCC
P2	GAA CGG GGC TCG AAG GGT CCT TGT AGC CG
P3R	TAC GCA TCC CAG TTT GAG ACG
P4CTG	TAC GCA TCC GAG TTT GAG ACG TGC TGC TGC TGC TGC T
P4CAG	TAC GCA TCC CAG TTT GAG ACG CAG CAG CAG CAG CAG CA
HD1	ATG AAG GCC TTC GAG TCC CTC AAG TCC TTC
DRPLA 2	TCA CCA TCA CCA CCA GCA ACA GCA A

fluoresceinated 1 µmol/l) and P2 (1 µmol/l) for conventional amplification of the myotonic dystrophy CAG repeat, mix B: P1 (5' fluoresceinated 1 µmol/l) with P3R (1 µmol/l) and P4CTG (0.1 µmol/l), and mix C: P2 (5' fluoresceinated 1 µmol/l) with P3 (1 µmol/l) and P4CAG (0.1 µmol/l) for amplifications priming from the CAG repeat allele. Samples were overlaid with mineral oil (Sigma) and incubated for four minutes at 94°C before the addition of the Taq polymerase. The reactions were subjected to 30 cycles of 94°C for one minute, 60°C for one minute, and 72°C for two minutes followed by a 10 minute extension at 72°C in a Perkin Elmer DNA thermal cycler.

Five microlitres of each PCR product were added to 5 µl of distilled water and 10 µl of loading buffer (10 mg/ml blue dextran in deionised formamide) and incubated at 95°C for three minutes. Tubes were cooled on ice before loading on a denaturing 6% polyacrylamide 7 mol/l urea gel. PCR products were resolved by electrophoresis in 0.6 × TBE and visualised using an automatic laser fluorescent sequencer (ALF Pharmacia). Traces were analysed using the Pharmacia fragment manager software.

SUBJECTS
Blood samples from 183 people from 48 myotonic dystrophy families had been collected by the Clinical Genetics Department in Edinburgh. DNA had been extracted and banked by the Molecular Genetics Service Laboratory in Edinburgh. DNA samples had been sent for Southern and conventional PCR analysis by another Scottish Molecular Genetics Consortium member laboratory in Aberdeen. This cohort included samples from unaffected people and spouses. The samples had been banked with no indication of the disease status of individual family members. The DNA samples were tested using PCR mixes A and B and scored blind.

Results and discussion
Typical traces obtained using mix A are shown in fig 2. The PCR amplification across the myotonin CAG repeat gives reliable amplification up to alleles of approximately 100 CAG repeats but fails to amplify above this size. Analysis of severely affected myotonic dystrophy patients by genomic Southern blotting typically shows expansions between 1.5 and 4 kb. PCR amplification of DNA from an affected subject carrying such an expansion gives a single normal allele and fails to pick up a larger allele, and is thus indistinguishable from patients genuinely homozygous for a small allele.

Typical traces obtained using mix B are shown in figs 3 and 4. Mix B gives similar traces to mix C (not shown). Amplification products show a characteristic ladder with a three base pair periodicity when products are analysed on a laser fluorescence DNA sequencer. The ladder peaks at the size of the largest allele present or, for very large repeat alleles, the peak height diminishes gradually
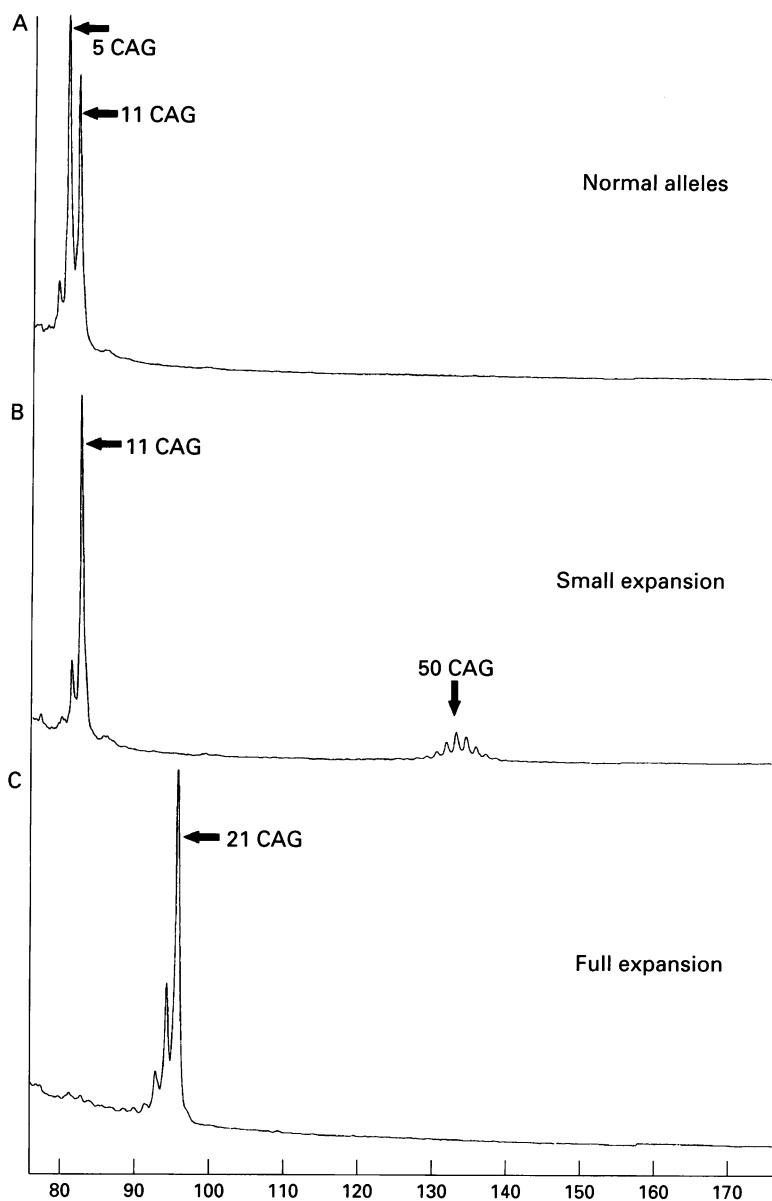
Figure 2   PCR amplification at the myotonin locus using flanking primers. The axis shows migration time in minutes. CAG allele sizes shown with the arrows. (A) Trace obtained from a heterozygous normal subject. (B) Trace obtained from a heterozygous subject with a small expansion. (C) Trace obtained from a patient with myotonic dystrophy and an expanded allele size of >4 kb as determined by Southern blot analysis. The larger allele fails to amplify.

with mix A gave the characteristic ladder which extended beyond 50 CAG repeats with the peak height diminishing with increasing product size (figs 3 and 4, bottom trace). These 76 people were scored as affected with myotonic dystrophy with CAG repeat alleles of greater than 50 CAG repeats. The remaining 107 DNA samples did not give the long characteristic ladder seen for large CAG expansions with mix B. Amplification with mix A showed that 76 people were heterozygous for CAG repeat alleles in the normal range with 31 patients homozygous for alleles in the normal range.

The clinical status of the patients and the results of the PCR and genomic Southern analysis reported by the Aberdeen laboratory were consulted. A total of 74 of the 76 people with CAG repeat alleles in the affected range were symptomatic and classed with an affected disease status. Two patients had alleles just within the affected range and were as yet asymptomatic. Both had either children or close relatives affected with myotonic dystrophy in whom we had also detected expanded CAG repeat alleles. The results of the conventional DNA analysis were that 75 patients had expanded CAG alleles in the affected range, from 50 CAG repeats to expansions of >5 kb. All 75 were in the group of 76 patients for whom our PCR assay had given evidence for the presence of a large CAG repeat allele. The one patient we identified who had not been ascertained by conventional testing was said to be clinically affected. The result of the Southern analysis for this DNA had been ambiguous because of poor DNA quality and a fresh sample will be collected for retesting.

On the basis of this series of tests we are confident that our TP PCR test for myotonic dystrophy is robust and reliable. For this large cohort of patients we obtained no false negative results and the results we obtained showed absolute concordance with the results obtained by conventional means. The only difference we saw suggested that our method was able to identify expanded alleles in poor quality DNA samples where Southern analysis failed to give a reliable result. This method does not measure the size of large expanded CAG repeat alleles but reliably identifies their presence. DNA samples can be screened using this method and only samples with large expansions would need to be analysed further using genomic Southern blotting in order to establish the exact size of large expanded alleles. The recently described XL PCR based method of Cheng *et al*[6] could be used for samples where insufficient DNA for Southern analysis is available. This method is, however, technically demanding and labour intensive as PCR products are resolved by agarose gel electrophoresis, the gel being then blotted and probed with a locus specific probe. Our TP PCR method works if the fluorescent primer P1 is replaced by ^{32}P end labelled P1. These radiolabelled products are, however, more difficult to size accurately in the 30 to 60 CAG repeat size range (data not shown). Obviously silver staining is inappropriate for TP PCR as many non-specific inter(CAG)$_n$ products are generated.
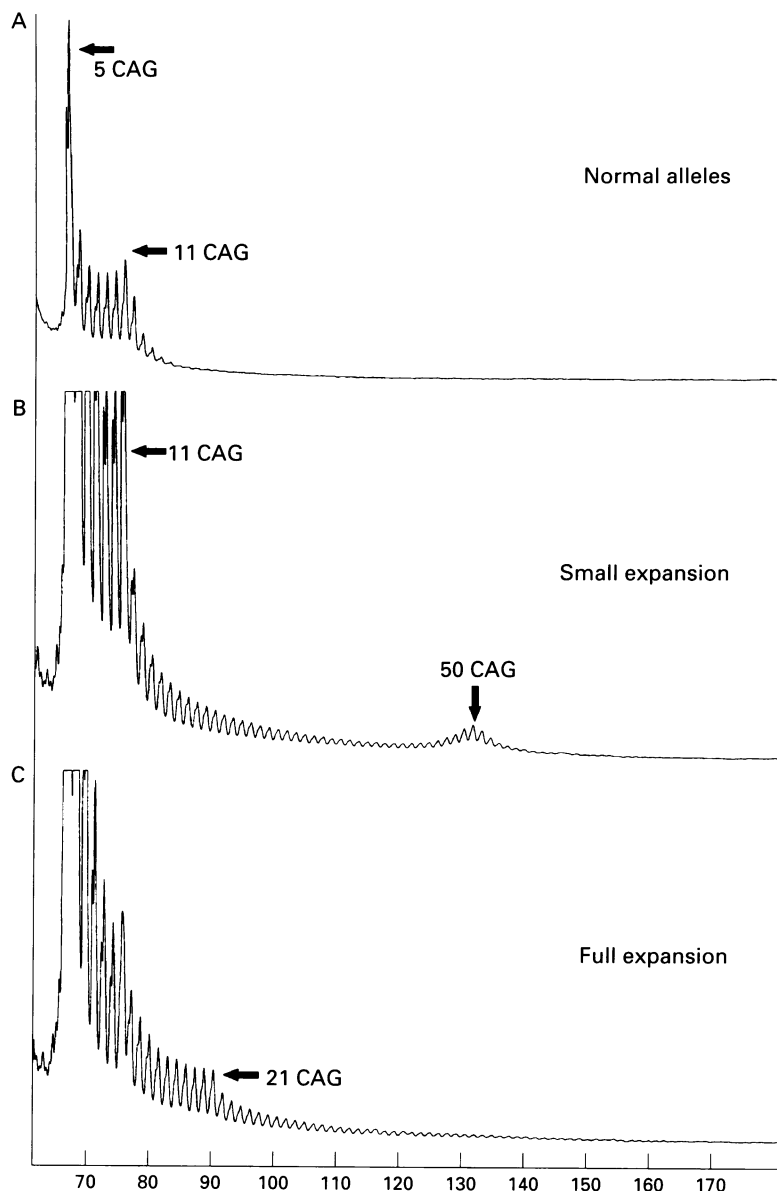
with increasing product size (fig 4). Allele sizes were determined by comparing with amplified PCR product from a person whose myotonin CAG alleles had been cloned and sequenced. Once a reference point is established, the three base pair periodicity of the ladder for the mix B traces gives an excellent size ladder for use with both systems, as the internally primed PCR products are smaller than the products obtained using the flanking primers.

Of the 183 samples that were tested with PCR mix A, four were found to have CAG repeat allele between 50 and 89 CAG repeats in the small affected allele size range. When these four samples were tested with mix B they gave the characteristic ladder with a 3 bp periodicity which extended to and peaked at the expected allele size for the larger allele (figs 3 and 4, central trace). A further 72 samples that had given one normal CAG repeat allele

We have carried out some pilot TP PCR experiments using specific flanking primers for other diseases where expanded CAG repeats are the mutational mechanism. For Huntington's disease the same PCR conditions as described above for mix B replacing the primer P1 with HD1[7] give reliable results (not shown). It has been suggested that for people with juvenile onset of Huntington's disease, where an expanded allele appears to be absent and the patient is homozygous for a smaller allele, a Southern analysis should be performed to exclude the presence of an unamplifiable large allele (Clinical Molecular Genetics Society of Great Britain, personal communication, 1995). Despite never having failed to amplify both alleles for juvenile Huntington's disease referrals up to 96 CAG repeats using flanking primers in Edinburgh, we would recommend the use of TP PCR to show the presence of these large, more difficult to amplify alleles. Similarly, we have shown that reliable results are obtained for large alleles for DRPLA when the primer DRPLA replaces P2 in mix C.

In conclusion we have developed a novel PCR assay that can be used for screening for expanded CAG repeat alleles including full mutation alleles of >5 kb in myotonic dystrophy. TP PCR is robust and reliable and can in theory also be used in conjunction with locus specific flanking primers for any disease with a similar mutational mechanism.

Figure 3   PCR amplification at the myotonin locus using internal primers. The axis shows migration time in minutes. CAG allele sizes shown with the arrows. The same people typed in fig 2 were retyped using mix B. Note the characteristic ladder with a 3 bp periodicity. (A) Both alleles give peaks and all the intermediate priming sites give peaks. There is a slight continuation of the pattern beyond the maximum allele size owing to mispriming at the end of the repeat. This usually extends for less than 30 bp. (B) Both alleles give peaks as in (A). (C) The ladder shows the presence of a large CAG allele undetectable using flanking primers.
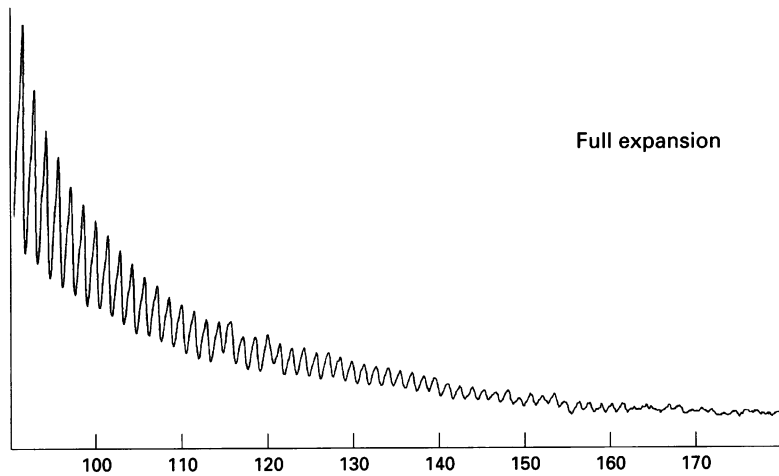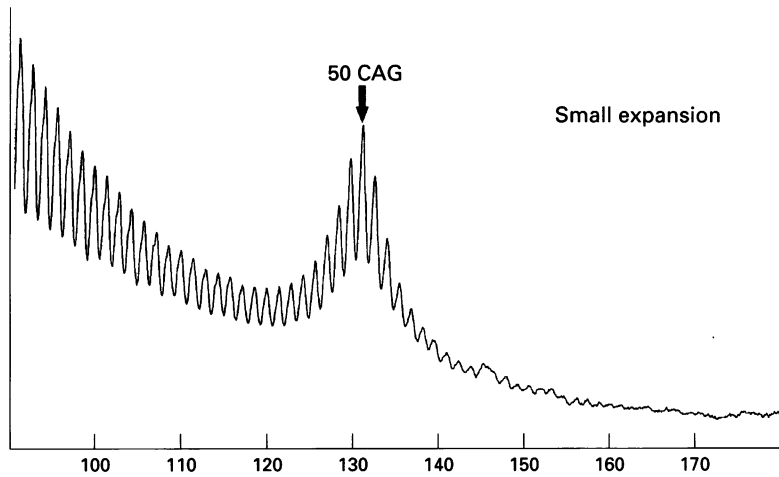
Figure 4 *Close up of traces shown in fig 3. The axis shows migration time in minutes. CAG allele sizes shown with the arrows. (A) Detail from trace shown in fig 3B. (B) Detail from trace shown in fig 3C.*

1 Fu YH, Pizzuti A, Fenwick RG Jr, *et al.* An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 1992;225:1256-8.
2 Neil DL, Jeffreys AJ. Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. *Hum Mol Genet* 1993;2:1129-35.
3 The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971-83.
4 Ashley CT, Warren ST. Trinucleotide repeat expansion and human disease. *Annu Rev Genet* 1995;29:703-28.
5 Nagafuchi S, Yanagisawa H, Sato K, *et al.* Dentatorubral and pallidoluysian atrophy: expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat Genet* 1994;6:14-18.
6 Cheng S, Barcelo JM, Korneluk RG. Characterisation of large CTG repeat expansions in myotonic dystrophy alleles using PCR. *Hum Mutat* 1996;7:304-10.
7 Warner JP, Barron LH, Brock DJH. A new polymerase chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded on Huntington's disease chromosomes. *Mol Cell Probes* 1993;7:235-9.