A Polymerase Chain Reaction, Nuclease Digestion, and Mass Spectrometry Based Assay for the Trinucleotide Repeat Status of the *Fragile X Mental Retardation 1 Gene*

Eric D. Dodds[a], **Flora Tassone**[b], **Paul J. Hagerman**[b,*], and **Carlito B. Lebrilla**[a,b,*]

[a]Department of Chemistry University of California Davis One Shields Avenue, Davis, CA, 95616 USA

[b]School of Medicine, Department of Biochemistry and Molecular Medicine University of California Davis One Shields Avenue, Davis, CA, 95616 USA

Abstract

CGG repeat expansions in the 5′ non-coding region of the fragile X mental retardation 1 gene (*FMR1*) give rise to both neurodevelopmental and neurodegenerative human diseases depending on the length of the expansion. Expansions beyond 200 repeats (full mutation) generally result in gene silencing and fragile X syndrome (FXS), the leading heritable form of cognitive impairment and autism. Smaller expansions (55-200 CGG repeats; "premutation") give rise to the neurodegenerative disorder fragile X-associated tremor/ataxia syndrome (FXTAS) through an entirely distinct, toxic mRNA gain-of-function mechanism. A rapid means for both high-risk and newborn screening for allele size would provide greater opportunity for early intervention and family counseling, as well as furnish critical data on repeat size distribution and expanded allele frequencies. In the current work, we propose a novel mass spectrometry (MS) based method for the rapid identification of expanded CGG repeats to complement a recently described polymerase chain reaction (PCR) method for large population screening. In this combined approach, the optimized PCR method is used to amplify the relevant region of *FMR1*, followed by extensive non-specific nuclease digestion. The resulting oligonucleotides are analyzed by MS in a manner that provides the relative proportion of triplet repeat oligonucleotides in seconds per sample. This assay enables swift and reproducible detection of expanded CGG alleles using a single blood spot, and in principle is suitable for large scale studies and newborn screening. Moreover, this analytical scheme establishes a unique new intersection of MS with molecular biology, with potential for significant interdisciplinary impact.

Introduction

The fragile X mental retardation 1 gene (*FMR1*; OMIM*309550) is subject to expansions of a CGG repeat in the 5′ non-coding region of the gene. Expansions in excess of 200 CGG repeats (normal range < 45 repeats) generally lead to transcriptional silencing of the gene and, due to the absence of the *FMR1* protein (FMRP), the neurodevelopmental disorder fragile X syndrome (FXS). FXS is the leading heritable form of cognitive impairment and the leading known single gene associated with autism.[1] Smaller expansions (55-200 CGG repeats) often give rise to primary ovarian insufficiency (POI)[2, 3] and the neurodegenerative disorder fragile X-associated tremor/ataxia syndrome (FXTAS).[4-6] For premutation expansions, disease formation is believed to arise through an entirely distinct pathogenic mechanism, one involving

*To whom correspondence should be addressed P.J.H.: Telephone 1-530-754-7266; Fax 1-530-754-7269; pjhagerman@ucdavis.edu
C.B.L.: Telephone 1-530-752-6364; Fax 1-530-754-5609; cblebrilla@ucdavis.edu.

the direct toxic gain-of-function of the *FMR1* mRNA,[6, 7] which is produced at elevated levels in the premutation range.[8-10]

The implementation of routine neonatal screening for *FMR1* mutations has been a focus of substantial interest and debate for a number of years. The potential benefits of newborn testing include: the possibility to provide early intervention for learning delays associated with both full mutations and premutations; the ability to prevent a protracted and arduous diagnostic interval for children with FXS; the opportunity to offer reproductive counseling and other services to families affected by *FMR1* expansions; and a means to obtain more accurate estimates of the distribution and frequency of expanded alleles in the general population. [11-13] As effective interventions and treatments continue to develop, *FMR1* screening may well become a candidate for inclusion in newborn screening panels. Current estimates for allele frequencies of full mutation alleles range from ~1/2500 to 1/4000 in males and from ~1/2500 to 1/8000 in females, whereas the premutation is estimated to occur in ~1/120 to 1/260 females and in 1/250 to 1/800 males.[14-20] Thus, expanded *FMR1* alleles and their associated disorders are sufficiently common to warrant further population studies, with the longer term possibility of routine neonatal testing. However, screening efforts on such scale will require the development of an appropriately rapid assay.

Substantial progress has been made toward high-throughput screening through the development of polymerase chain reaction (PCR) techniques for efficiently amplifying the CGG repeat region from blood spots.[21, 22] This methodology has recently been applied to an anonymous screen of approximately 5,000 newborn blood spots samples from an unselected population in Spain.[20] These PCR approaches are potentially amenable to automation and are readily scalable to larger sample numbers with only the need for additional thermal cyclers. Unfortunately, analysis of the PCR products using manual gel electrophoresis (requiring on the order of 1 h per gel) or even automated capillary electrophoresis (requiring on the order of 10 min per sample and considerable additional expense) represents a major bottleneck.

Clearly, the demands of routine large scale screening call for automated methods of much greater throughput. Mass spectrometry (MS) is an analytical platform with many of the attributes necessary to meet those demands. Analysis of nucleic acids by MS is a uniquely capable alternative to more traditional means of characterization, and is well established in terms of analytical development[23-25] and applications in genomics.[26-28] Of particular importance in the present context are the compatibility of MS with automated, large sample volume workflows, and the capability of MS methods to complete a simple sample analysis in seconds.

This report describes a novel technique for rapidly sizing the CGG repeat region of the *FMR1* gene. The assay is comprised of three key steps. First, the relevant region of the gene is selectively amplified using an optimized PCR method.[13] Second, the PCR product is enzymatically hydrolyzed to small oligonucleotides (ONTs) using benzonase, a non-specific endonuclease. Third, the resulting pool of small ONTs is analyzed by MS such that the relative contribution of CGG and non-CGG ONTs can be determined. This approach allowed the repeat numbers of various amplicons to be accurately ranked while circumventing more costly, time consuming, and labor intensive methods. The unique combination of PCR, non-specific endonuclease digestion, and MS analysis furnishes a rapid, cost-effective, and readily automated platform for screening the CGG repeat status of the *FMR1* gene. In addition, this integration of technologies serves to illustrate a new means by which MS can contribute to the fields of nucleic acid analysis and genetic screening.

Materials and Methods

Samples

Following informed consent, blood samples were obtained from individuals seen at the UC Davis M.I.N.D. Institute according to an Institutional Review Board approved protocol. Genomic DNA was isolated from peripheral blood leukocytes (3-5 mL of whole blood) using standard methods and was then amplified by PCR using primers c and f[29] and incorporating the osmolite betaine into the PCR buffer.[13, 21] Alternatively, dried blood spots (FTA cards, 2 $\times$ 1.2 mm disks; Whatman, Piscataway, NJ, USA) were placed in 500 μL PCR tubes and washed with FTA purification solution according to manufacturer guidelines. The supernatants from the final wash was removed, including any excess liquid adhering to the disk. The disk was left to dry for at least one hour and was then combined with PCR master mix (FastStart PCR Kit; Roche Diagnostics, Indianapolis, IN, USA) containing betaine plus primers c and f. PCR conditions were as previously described.[20]

All amplicons were sized using the Qiaxcel capillary electrophoresis system (Qiagen, Valencia, CA, USA) as detailed elsewhere.[20] Prior to nuclease digestion, PCR products were purified using MinElute spin columns (Qiagen) with elution in 30 μL of water. DNA concentration was then measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Typical DNA yields were 20-80 ng/μL.

Nuclease digestion

Purified PCR products were dried in a vacuum centrifuge (approximately 20 min at 35°C) and reconstituted in 20 μL of digestion solution containing 0.25 U/μL benzonase endonuclease (Sigma-Aldrich, St. Louis, MO, USA), 1 mM $MgCl_2$, and 50 mM NH_4HCO_3, pH 8.0. Digestions were carried out for 8 h with incubation at 37°C to produce a mixture of small ONTs. All digests were stored at -20°C until the time of analysis.

Mass spectrometry

Matrix-assisted laser desorption/ionization (MALDI) was accomplished using 2,5-dihydroxybenzoic acid (DHB) as the matrix (obtained from Sigma at the highest available purity). A solution of 50 μg/μL DHB in 50% aqueous CH_3CN was freshly prepared immediately prior to use. Each sample digest was co-spotted with the DHB matrix solution on a stainless steel MALDI target (1 μL digest solution plus 1 μL matrix solution). The spots were allowed to dry completely in a fume hood before the sample probe was introduced to the ion source. All MS analyses were conducted in negative ion mode using a Fourier transform ion cyclotron resonance (FTICR) instrument (Varian IonSpec ProMALDI, Lake Forest, CA, USA) equipped with a 7.0 T superconducting magnet and an external MALDI source capable of hexapole accumulation with vibrational cooling. Ions produced by a variable number of MALDI events were accumulated for mass analysis such that the total ion intensity of all spectra was approximately equivalent. Typically, this involved 5-25 laser pulses of the frequency-tripled Nd:YAG laser (5 ns pulse width, 2.5 Hz repetition rate, 355 nm wavelength). Externally accumulated ions were injected to the ICR cell using a broad band, RF-only quadrupole ion guide in conjunction with gated ion trapping. The quadrupole RF amplitude and ion gating pulse were optimized for ions of mass to charge ratio (*m/z*) of 900, which allowed coverage of *m/z* 500-2000. Trapped ions were accelerated to coherence using a stored waveform inverse Fourier transform (SWIFT) excitation pulse. Time domain ion signal was acquired at an ADC rate of 1 MHz, with a total of 1024 k transient data points being acquired for each spectrum.

Spectrum processing

In order to obtain the most representative relative ion abundance data, the 1024 k data point transients were shortened to include only the first 256 k points prior to further processing. In FTICR-MS, transient signal damping over time is *m/z* dependent and can bias intensity information in favor of higher *m/z* values. This effect is mitigated by using only the earliest time domain data points (*i.e.*, before any significant damping has occurred), thus providing more accurate relative ion intensities.[30-32] The 256 k datasets were padded with one zero fill, Blackman apodized, fast Fourier transformed to the frequency domain, and calibrated from frequency to *m/z* using standard FTICR-MS calibration equations.[33, 34] ONTs of three known compositions served as internal mass calibrants in order to provide optimum mass accuracy: C_1G_1, *m/z* 635.1022; C_1G_2, *m/z* 964.1547; and C_2G_2, *m/z* 1253.2011 (monoisotopic *m/z* corresponding to [M-H]⁻ ions). The noise level of each spectrum was assessed and used to set intensity thresholds (typically, 1-2% relative abundance) for peak detection according to the default settings of the instrument data analysis software (Varian IonSpec Omega, version 8). Peak lists obtained from each mass spectrum were exported in text file format for further processing.

Software and calculations

To facilitate further data reduction and calculation of CGG content metrics, a software package was written and implemented using the IGOR Pro environment (version 6, WaveMetrics, Lake Oswego, OR, USA). The major functions of the CGG Oligonucleotide Recomposition Tool (CORT) were to load MS peak lists, search the lists for small ONT *m/z* values, and compile lists of the matching *m/z* ratios with corresponding relative abundances. MS signals were only assigned to a specific ONT composition if the experimental *m/z* value was within 10 parts per million (ppm) of the theoretical *m/z* value. Once tabulated, the *m/z* and intensity values of matching ONTs were used to calculate the total intensity of ONTs derived from the CGG repeat region (compositions and *m/z* values listed in Table 1) relative to the total intensity of a normalizing set of non-repeat ONTs (compositions and *m/z* values listed in Table 2). Members of the normalizing signal set were chosen due to their relatively high intensity (compared to other non-CGG ONTs) and appropriate distribution across the *m/z* range of interest. The repeat intensity ratio, *R*, was then calculated according to:

$$R = \frac{\Sigma w_i r_i}{\Sigma w_k s_k}$$

[Equation 1]

where r_i represents the relative abundance of the *i*-th ONT mass arising from the CGG repeat region, s_k represents the relative abundance of the *k*-th ONT from the selected subset of non-CGG ONTs, and w_i, w_k are stoichiometric weighting factors accounting for the length of the *i*-th and *k*-th ONTs, respectively. A more detailed description of the CORT algorithm is provided in the Supporting Information.

Results and Discussion

Principle of assay

A schematic overview of the PCR and MS based *FMR1* CGG repeat assay is provided in Figure 1. The *FMR1* region of interest was first amplified using the c and f PCR primers. The amplicons were then purified of primers, residual deoxynucleotide triphosphates, betaine, and buffer salts. Isolated PCR products were then digested using benzonase, an aggressive and non-specific endonuclease capable of hydrolyzing essentially any form of nucleic acid to a pool of small ONTs. Finally, the resulting pool of small ONTs was analyzed by MS, and the *m/z* and relative abundance data were used to infer the CGG repeat length.

Relating the MS data to CGG repeat count was accomplished by first regarding the *FMR1* PCR products as having two portions: a constant tract of sequence flanking the trinucleotide repeat region, and the intervening repeat region itself. Because amplicons corresponding to different *FMR1* genotypes share the constant sequence and differ only in trinucleotide repeat length, it was further reasoned that the intensity contribution of CGG-derived ONTs relative to flanking region ONTs should scale in proportion to the CGG repeat count. Conveniently, each possible small ONT composition corresponded to a unique mass (Supplementary Tables S1-S4 in Supporting Information); therefore, the *m/z* and relative abundance measurements provided by MS analysis allowed relative quantities to be assigned to specific ONT compositions without ambiguity. In this way, the intensity contributions from the CGG repeat region and the constant flanking regions could readily be determined. Because all samples were in effect internally standardized by the flanking sequence of the PCR product, the CGG repeat intensity ratio (calculated according to Equation 1) could be applied to compare CGG repeat lengths of amplicons derived from different sources (for example, the three repeat lengths shown in Figure 1). This normalization and relative quantitation approach allowed the assay to be conducted without the need for a rigorous calibration that would be necessary for absolute quantitation.

Interpretation of mass spectra

As reported previously, benzonase digestion of DNA has been shown to produce a mixture of small ONTs predominantly consisting of dinucleotides, trinucleotides, and tetranucleotides. [35-37] As illustrated in Figure 2, benzonase digests of an *FMR1* amplicon with 30 CGG repeats conformed to this general rule. As such, the size distribution and ionization efficiencies of the resulting ONTs (particularly, in negative ion mode) rendered them highly suitable for routine MS analysis. Typically, trinucleotides were found to be the most abundant components of the ONT mixtures. Dinucleotides and tetranucleotides were also well represented, while pentanucleotide masses made a negligible, often undetectable contribution to the total ion signal. This latter observation also provided an independent measure of the digestion efficiency. While the most abundant signals for each ONT length corresponded to compositions attributable to the CGG repeat region, a number of ONT compositions derived from flanking sequence were also clearly detected.

The same overall features were also found to hold true for PCR products having up to 124 CGG repeats. In Figure 3, all ONT signals that could originate from CGG repeat are labeled with red diamonds (compositions and *m/z* values listed in Table 1), while members of a prime subset of non-CGG ONTs are labeled with green circles (compositions and *m/z* values listed in Table 2). The eight *m/z* ratios labeled with red diamonds represent all dinucleotide, trinucleotide, and tetranucleotide compositions that could be produced by non-specific digestion of (CGG)$_n$ and the complementary strand; however, those compositions were not necessarily unique to the repeat region. Indeed, the regions of flanking sequence were rather GC rich, and thus some contribution to these signals would be expected from the non-CGG portion of the amplicons. Conversely, the eight *m/z* ratios labeled with green circles were selected due to their exclusive origination from non-CGG sequence (assuming a negligible contribution from occasional AGG interruptions within the trinucleotide repeat region). Although pentanucleotide compositions have been listed in **Tables** 1-2 for completeness, they were not found to be significant components of the ONT mixtures.

Another notable point illustrated in Figure 3 is the applicability of the assay not only to blood DNA isolates, but also to direct analysis of dried blood spots - the type of material routinely used for newborn screening. The lower three traces of Figure 3 each resulted from the assay of a single blood spot which had been collected in a manner consistent with current newborn screening protocols. Using dried blood spots as the starting material, sufficient DNA could be amplified to allow benzonase digestion and MS analysis using only about 5% of the digest.

Typically, 1 μL of each 20 μL digest was spotted on a MALDI target for analysis, corresponding to 30-120 ng digested DNA. Thus, the methodology is applicable to a very small amount of sample. The resulting spectra were of essentially the same quality as those from PCR products obtained by amplification from DNA isolated from whole blood.

Determination of repeat intensity ratios

While all of the spectra displayed in Figure 3 were found to be very similar in a qualitative sense, significant quantitative differences became apparent when the signal intensities of repeat region and constant region ONTs were evaluated. Figure 4 highlights the comparison of relative abundances for several select ONT compositions at CGG repeat counts of 30 (normal range) and 66 (premutation range). In all cases, the intensities of repeat region ONTs were found to increase relative to the intensities of flanking region ONTs as the CGG repeat count increased. As a specific example, the intensity difference between the trinucleotide compositions C_1G_2 (repeat-derived) and A_1G_2 (flank-derived) was 72.1% relative abundance in the case of 30 CGG repeats; however, in the case of 66 CGG repeats the difference in intensities was expanded to 82.8% relative abundance. The intensities of C_1G_1 relative to T_1G_1, C_2G relative to $C_1A_1G_1$, and C_2G_2 relative to $C_1T_1G_2$ were similarly increased as a function of repeat count.

In concert, these differences in relative abundance provided a readout indicating the relative trinucleotide repeat status among amplicons of various lengths. When the abundances of the dinucleotide, trinucleotide, and tetranucleotide compositions given in **Tables** 1 and 2 were used to calculate the repeat intensity ratios according to Equation 1, the CGG repeat intensity ratios (R) shown in Figure 5 were obtained. The intensity ratios were found to scale in proportion to CGG repeat length (as independently determined by capillary electrophoresis), with R taking on values ranging from 1.16 ± 0.11 to 5.14 ± 0.15 for *FMR1* genotypes ranging from 30 to 124 CGG repeats. The R values were found to be very reproducible upon replicate analysis (four replicate measurements of each of the seven samples, relative standard deviation < 8% on average).

These results support the initial prediction that benzonase digestion of the *FMR1* amplicons followed by MS analysis of the resulting small ONTs could be used to determine the CGG repeat count with good accuracy and precision. While the potential to size the amplicons directly using this assay is attractive and would appear to be well within reach, the exact size of the CGG repeat need not be determined for the purposes of high-throughput primary screening. Instead, as an initial test, the CGG repeat status could be coarsely assigned as either normal or potentially expanded. The fraction of samples with repeat intensity ratios above some empirically determined threshold would then be flagged for further analysis using a secondary screening method (*e.g.*, gel electrophoresis, capillary electrophoresis, or Southern blotting). Thus, the performance of the PCR and MS based assay surpassed the basic requirement for a primary screening tool.

Conclusions

The previously developed *FMR1* PCR method was effectively interfaced to MS analysis through non-specific endonuclease digestion of the amplicons. Benzonase treatment of the PCR products produced mainly dinucleotides, trinucleotides, and tetranucleotides, each having a molar mass that uniquely identified the corresponding ONT composition. MS analysis of the resulting small ONT mixtures provided *m/z* values and relative abundances for individual components, thus furnishing the basis for a metric related to CGG repeat status. Using this strategy, the relative lengths of trinucleotide repeat regions could be accurately ranked for amplicons having from 30-124 CGG repeats. The assays were accomplished with high reproducibility using blood DNA isolates or individual dried blood spots as the starting material. In principle, the assay should be applicable to the full range of CGG repeat lengths

that can be amplified by the PCR step.[21] There is also the potential to apply this assay in conjunction with a CGG targeted priming approach,[22] thus allowing the presence or absence of a full mutation to be established while concurrently providing the approximate repeat count for normal and premutation alleles.

The coupling of an MS-based readout to the *FMR1* PCR method supplies the potential for a relatively inexpensive and high-throughput primary screen for *FMR1* trinucleotide repeat expansions. MS analysis for a single sample of this type can be conducted in a matter of seconds, and a large number of these analyses can be accomplished in an automated manner. The amenability of MS to high-throughput workflows complements the throughput potential of the PCR method; moreover, the divide between the amplification and analysis steps can be bridged by existing technologies for automated sample handling. For example, PCR reactions carried out in 96 or 384 well plate format could then be processed using a robotic platform for purification and enzymatic digestion. MS analysis would them be performed using an autosampler capable of accepting multi well plates. This approach would be markedly more rapid and efficient than the present readout based on gel electrophoresis or capillary electrophoresis, particularly where large numbers of samples are involved.

In addition to throughput potential, another advantage of this MS approach is the relatively small per sample expense beyond the already low cost of amplification. The only significant expenditures required for the MS approach are for PCR purification cartridges and endonuclease, which for a 96 well format may account for only about fifty cents. These additional expenses would be offset by the elimination of electrophoretic determination at the primary screening stage. Although a significant initial investment would be necessary to obtain the required instrumentation, laboratories already engaged in high-throughput MS based screening will already be equipped with the necessary infrastructure. Indeed, MS is already in routine use for numerous newborn screening tests.[38-40] Finally, while the MS data reported here was obtained using FTICR-MS, it should be noted that such a high performance instrument is not necessary for these analyses. Because ultrahigh mass accuracy and ultrahigh resolution are not essential for assignment of small ONT masses, these assays are well within the capabilities of less expensive mass analyzers.

As presently implemented, this assay does not allow specific genotyping of both alleles in females. Because the two amplicons would be digested and assayed in combination, the information on repeat length would become decoupled from the individual alleles. In this case, the analytical outcome might be expected to reflect an average of the two alleles; however, CGG repeat regions of significantly differing length do not amplify with equal efficiency. Thus, the apparent repeat intensity ratio would be a convolution of the two repeat lengths and their relative PCR yields. Although this does not preclude use of the assay to flag expanded female genotypes for further analysis, female and male samples should be subject to independently determined threshold values for secondary screening.

The possibility to eliminate the nuclease digestion step in favor of directly sizing intact PCR products by MS merits mention. Measurement of *m/z* values for large oligonucleotides is within reach of modern MS technologies, and could be considered less circuitous than the presently described assay. Nevertheless, the analysis of oligonucleotides with molar masses in the hundreds of kilodaltons (kDa) range (*e.g.*, using primers c and f, approximately 96 kDa per strand assuming 30 CGG repeats; 115 kDa per strand for 50 CGG repeats; 254 kDa per strand in the case of 200 CGG repeats) would not be readily portable into existing workflows for MS based newborn screening, which are primarily aimed at analysis of small molecules.

Overall, this novel combination of PCR, nuclease digestion, and MS analysis provides a simple and rapid means of *FMR1* genotyping. This approach serves as a step towards development of

a large scale *FMR1* screening technology with the potential to expedite population studies and neonatal screening, following appropriate clinical validation. Of particular note is that these measurements have been achieved using only a single blood spot as the starting material — an important criterion in the context of newborn screening. Moreover, essentially all stages of the assay are compatible with automation, thus rendering the method suitable for high-throughput screening of large numbers of samples. With appropriate preparative and analytical instrumentation, the analysis of hundreds to thousands of samples per day should be readily attainable at permissive per sample cost. Furthermore, this general strategy establishes a new niche for MS in genetic screening, and should be equally advantageous for any application in which the gross composition of a PCR amplicon or DNA isolate could be informative (for example, in the genotyping of other expanded repeat disorders).[41]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

(1). Hagerman RJ, Rivera SM, Hagerman PJ. Current Pediatric Reviews 2008;4:40–52.

(2). Allen EG, Sullivan AK, Marcus M, Small C, Dominguez C, Epstein MP, Charen K, He W, Taylor KC, Sherman SL. Human Reproduction 2007;22:2142–2152. [PubMed: 17588953]

(3). Nelson LM. New England Journal of Medicine 2009;360:606–614. [PubMed: 19196677]

(4). Jacquemont S, Hagerman RJ, Hagerman PJ, Leehey MA. Lancet Neurology 2007;6:45–55. [PubMed: 17166801]

(5). Amiri K, Hagerman RJ, Hagerman PJ. Archives of Neurology 2008;65:19–25. [PubMed: 18195136]

(6). Brouwer JR, Willemsen R, Oostra BA. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2009in press, DOI 10.1002/ajmg.b.30910

(7). Hagerman PJ, Hagerman RJ. American Journal of Human Genetics 2004;74:805–816. [PubMed: 15052536]

(8). Tassone F, Hagerman RJ, Taylor AK, Gane LW, Godfrey TE, Hagerman PJ. American Journal of Human Genetics 2000;66:6–15. [PubMed: 10631132]

(9). Kenneson A, Zhang F, Hagedorn CH, Warren ST. Human Molecular Genetics 2001;10:1449–1454. [PubMed: 11448936]

(10). Allen EG, He W, Yadav-Shah M, Sherman SL. Human Genetics 2004;114:439–447. [PubMed: 14758538]

(11). Bailey DB Jr. Mental Retardation and Developmental Disabilities Research Reviews 2004;10:3–10. [PubMed: 14994282]

(12). Hagerman RJ, Hagerman PJ. Journal of the American Medical Association 2008;300:2419–2421. [PubMed: 19033593]

(13). Tassone F, Hagerman PJ, Hagerman RJ. Journal of Intellectual Disability Research 2008;52:814.

(14). Rousseau F, Rouillard P, Morel ML, Khandjian EW, Morgan K. American Journal of Human Genetics 1995;57:1006–1018. [PubMed: 7485149]

(15). Toledano-Alhadef H, Basel-Vanagaite L, Magal N, Davidov B, Ehrlich S, Drasinover V, Taub E, Halpern GJ, Ginott N, Shohat M. American Journal of Human Genetics 2001;69:351–360. [PubMed: 11443541]

(16). Dombrowski C, Levesque S, Morel ML, Rouillard P, Morgan K, Rousseau F. Human Molecular Genetics 2002;11:371–378. [PubMed: 11854169]

(17). Song FJ, Barton P, Sleightholme V, Yao GL, Fry-Smith A. Health Technology Assessment 2003;7:1–106.

(18). Berkenstadt M, Ries-Levavi L, Cuckle H, Peleg L, Barkai G. Prenatal Diagnosis 2007;27:991–994. [PubMed: 17705235]

(19). Hagerman PJ. Journal of Medical Genetics 2008;45:498–499. [PubMed: 18413371]

(20). Fernandez-Carvajal I, Walichiewicz P, Xiaosen X, Pan R, Hagerman PJ, Tassone F. Journal of Molecular Diagnostics. 2009in press, DOI 10.2353/jmoldx.2009.080173

(21). Saluto A, Brussino A, Tassone F, Arduino C, Cagnoli C, Pappi P, Hagerman P, Migone N, Brusco A. Journal of Molecular Diagnostics 2005;7:605–612. [PubMed: 16258159]

(22). Tassone F, Pan R, Amiri K, Taylor AK, Hagerman PJ. Journal of Molecular Diagnostics 2008;10:43–49. [PubMed: 18165273]

(23). Nordhoff E, Kirpekar F, Roepstorff P. Mass Spectrometry Reviews 1996;15:67–138.

(24). Banoub JH, Newton RP, Esmans E, Ewing DF, Mackenzie G. Chemical Reviews 2005;105:1869–1915. [PubMed: 15884792]

(25). Hofstadler SA, Sannes-Lowery KA, Hannis JC. Mass Spectrometry Reviews 2005;24:265–285. [PubMed: 15389854]

(26). Boom, D. v. d.; Beaulieu, M.; Oeth, P.; Roth, R.; Honisch, C.; Nelson, MR.; Jurinke, C.; Cantor, C. International Journal of Mass Spectrometry 2004;238:173–188.

(27). Meng Z, Simmons-Willis TA, Limbach PA. Biomolecular Engineering 2004;21:1–13. [PubMed: 14715314]

(28). Tost J, Gut IG. Journal of Mass Spectrometry 2006;41:981–995. [PubMed: 16921576]

(29). Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, Holden JJ, Fenwick RG Jr. Warren ST, Oostra BA, Nelson DL, Caskey CT. Cell 1991;67:1047–1058. [PubMed: 1760838]

(30). Farrar TC, Elling JW, Krahling MD. Analytical Chemistry 1992;64:2770–2774.

(31). Goodner KL, Milgram KE, Williams KR, Watson CH, Eyler JR. Journal of the American Society for Mass Spectrometry 1998;9:1204–1212.

(32). Gordon EF, Muddiman DC. Journal of Mass Spectrometry 2001;36:195–203. [PubMed: 11288202]

(33). Marshall AG, Hendrickson CL, Jackson GS. Mass Spectrometry Reviews 1998;17:1–35. [PubMed: 9768511]

(34). Zhang LK, Rempel D, Pramanik BN, Gross ML. Mass Spectrometry Reviews 2005;24:286–309.

(35). Janning P, Schrader W, Linscheid M. Rapid Communications in Mass Spectrometry 1994;8:1035–1040. [PubMed: 7696699]

(36). Liao Q, Chiu NH, Shen C, Chen Y, Vouros P. Analytical Chemistry 2007;79:1907–1917. [PubMed: 17261027]

(37). Quinlivan EP, Gregory JF 3rd. Analytical Biochemistry 2008;373:383–385. [PubMed: 18028864]

(38). Chace DH. Chemical Reviews 2001;101:445–478. [PubMed: 11712254]

(39). Carpenter KH, Wiley V. Clinica Chimica Acta 2002;322:1–10.

(40). Chace DH. Journal of Mass Spectrometry 2009;44:163–170. [PubMed: 19034889]

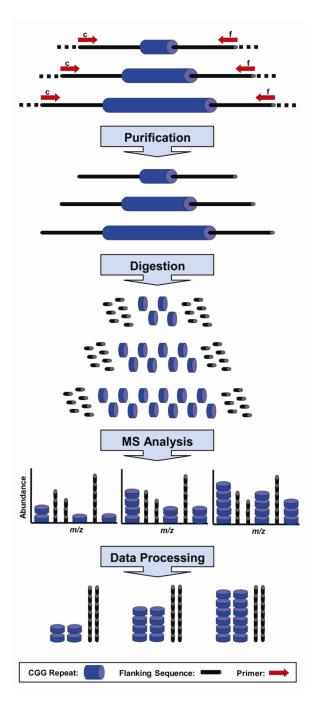(41). Richards RI. Human Molecular Genetics 2001;10:2187–2194. [PubMed: 11673400]

Figure 1.
Schematic presentation of the PCR and MS based *FMR1* CGG repeat assay, illustrated for three different CGG repeat lengths.
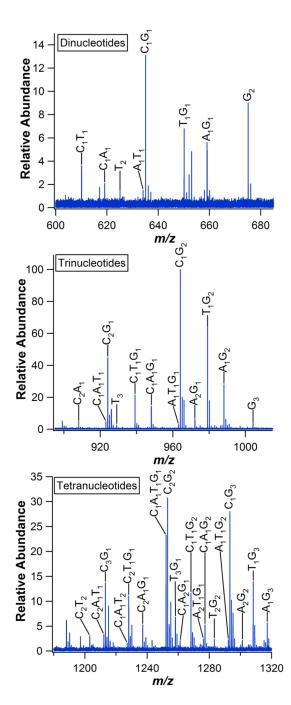
Figure 2.
Dinucleotide, trinucleotide, and tetranucleotide regions from the negative ion MALDI-FTICR
mass spectrum of a benzonase digested *FMR1* PCR product. The amplicon was derived from
a male individual with 30 CGG repeats. Peaks matching ONT masses to within 10 ppm are
labeled with the corresponding compositions. All ONT signals were observed as singly
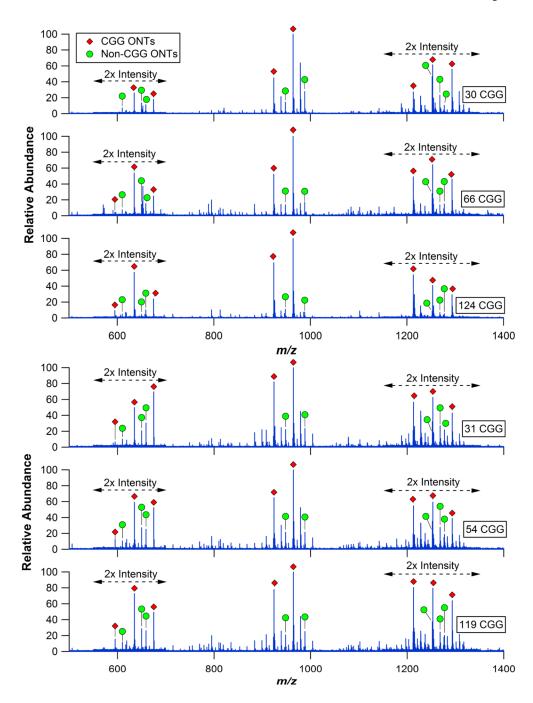deprotonated ions.

Figure 3.
Negative mode MALDI-FTICR-MS analysis of benzonase digested *FMR1* PCR products derived from male individuals having 30, 66, and 124 CGG repeats (from blood DNA isolates, upper three traces) and male individuals having 31, 54, and 119 CGG repeats (directly from dried blood spots, lower three traces). The abundance scales have been expanded two fold in the indicated regions.
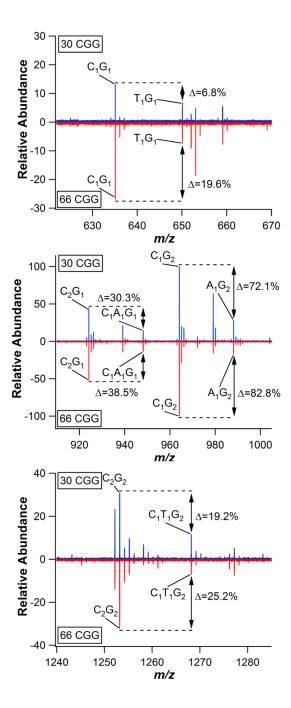
Figure 4.
Comparison of CGG and non-CGG small oligonucleotide intensities for *FMR1* PCR amplicons containing 30 CGG repeats (normal range; upright traces in blue) and 66 CGG repeats (premutation range; inverse traces in red).
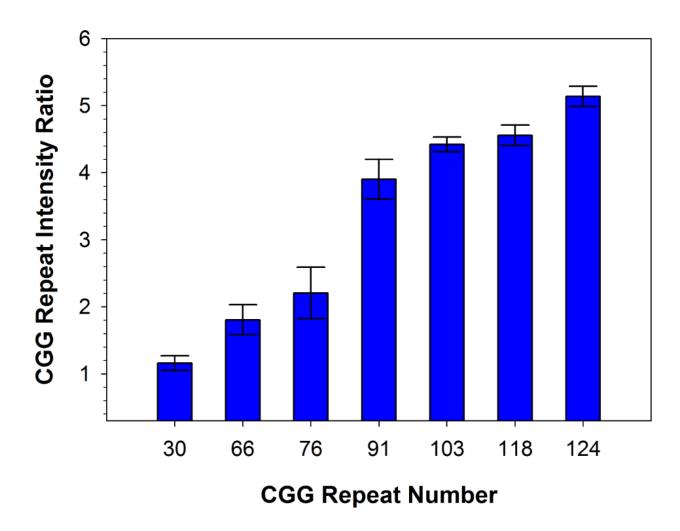
Figure 5.
CGG repeat intensity ratios corresponding to CGG repeat counts ranging from 30 to 124. Each value corresponds to the mean of quadruplicate analyses; error bars represent the standard deviation. The CGG repeat intensity ratios were calculated according to Equation 1.

Table 1

Compositions of all dinucleotides, trinucleotides, tetranucleotides, and pentanucleotides that can be produced by benzonase digestion of the *FMR1* PCR amplicon CGG repeat region. Compositions arising from both complementary strands of DNA are included. For each composition, the corresponding monoisotopic and average mass to charge ratios are provided for the singly deprotonated ions.

Oligonucleotide Composition	Monoisotopic [M-H]⁻ (*m/z*)	Average [M-H]⁻ (*m/z*)
C_2	595.09603	595.377
C_1G_1	635.10218	635.402
G_2	675.10832	675.427
C_2G_1	924.14855	924.587
C_1G_2	964.15470	964.612
C_3G_1	1213.19492	1213.772
C_2G_2	1253.20107	1253.797
C_1G_3	1293.20722	1293.822
C_4G_1	1502.24129	1502.957
C_3G_2	1542.24744	1542.982
C_2G_3	1582.25359	1583.007
C_1G_4	1622.25973	1623.032

Table 2

Compositions of selected dinucleotides, trinucleotides, tetranucleotides, and pentanucleotides produced by benzonase digestion of the *FMR1* PCR amplicon constant region. These signals were chosen to serve as normalizers for the CGG oligonucleotide signals when calculating CGG repeat intensity ratios according to Equation 1. For each composition, the corresponding monoisotopic and average mass to charge ratios are provided for the singly deprotonated ions.

Oligonucleotide Composition	Monoisotopic [M-H]⁻ (*m/z*)	Average [M-H]⁻ (*m/z*)
C_1T_1	610.09569	610.389
T_1G_1	650.10184	650.414
A_1G_1	659.11341	659.427
$C_1A_1G_1$	948.15978	948.612
A_1G_2	988.16593	988.637
$C_1A_1T_1G_1$	1252.20582	1252.809
$C_1T_1G_2$	1268.20073	1268.809
$C_1A_1G_2$	1277.21230	1277.822
$C_2A_1T_2$	1516.24571	1516.981
$C_3T_1G_1$	1517.24096	1517.969
$C_1T_3G_1$	1547.24029	1547.993
$C_1A_1T_1G_2$	1581.25834	1582.019