Am. J. Hum. Genet. 72:931–939, 2003

Rapid Direct Sequence Analysis of the Dystrophin Gene

Kevin M. Flanigan,[1,2,3,4] Andrew von Niederhausern,[2] Diane M. Dunn,[2] Jonathan Alder,[2] Jerry R. Mendell,[5] and Robert B. Weiss[2]

Departments of [1]Neurology, [2]Human Genetics, [3]Pathology, and [4]Pediatrics, University of Utah, Salt Lake City, and [5]Department of Neurology, Ohio State University, Columbus

Mutations in the dystrophin gene result in both Duchenne and Becker muscular dystrophy (DMD and BMD), as well as X-linked dilated cardiomyopathy. Mutational analysis is complicated by the large size of the gene, which consists of 79 exons and 8 promoters spread over 2.2 million base pairs of genomic DNA. Deletions of one or more exons account for 55%–65% of cases of DMD and BMD, and a multiplex polymerase chain reaction method—currently the most widely available method of mutational analysis—detects ~98% of deletions. Detection of point mutations and small subexonic rearrangements has remained challenging. We report the development of a method that allows direct sequence analysis of the dystrophin gene in a rapid, accurate, and economical fashion. This same method, termed "SCAIP" (single condition amplification/internal primer) sequencing, is applicable to other genes and should allow the development of widely available assays for any number of large, multiexon genes.

Introduction

The dystrophinopathies—Duchenne muscular dystrophy (DMD [MIM 310200]) and Becker muscular dystrophy (BMD [MIM 300376])—are the most common inherited disorders of muscle. Although reliable prevalence data are lacking, the prevalence of DMD is generally estimated at 1:3,500 live male births (Emery 1991). Both DMD and BMD are due to mutations in the dystrophin gene (MIM 300377), located at Xp21, which comprises 79 exons and 8 tissue-specific promoters distributed across ~2.2 Mb of genomic sequence—making dystrophin the largest gene yet described. Dystrophin gene deletions are found in ~55% of patients with BMD and 65% of patients DMD; point mutations account for ~30% of mutations, and duplications account for the remainder (Miller and Hoffman 1994). Genetic testing for deletions relies on a multiplex PCR technique, with amplification of fragments containing 18–25 of the gene's 79 exons (Beggs et al. 1990; Chamberlain et al. 1990) and with deletions detected as absent or size-shifted bands on agarose gel analysis. Because deletions tend to occur in "hotspots" within the dystrophin gene, analysis of this limited number of exons can detect 98% of dystrophin deletions.

Testing for dystrophin point mutations has been available only on a research basis, from specialized laboratories. Such analysis requires sequencing of all 79 exons and 8 promoters; there are no particularly common point mutations or point mutation hotspots, and each affected family may carry a unique mutation in this enormous gene (termed "private mutations" because they are exclusive to individual families). Instead of direct sequence analysis, some research laboratories perform point mutation analysis on cDNA derived by RT-PCR from muscle mRNA. As an alternative, other laboratories have utilized the protein truncation test (PTT), which may be performed using peripheral blood lymphocyte DNA (Roest et al. 1993) but often uses mRNA derived from muscle biopsy (Tuffery-Giraud et al. 1999). There is an obvious drawback to approaches that require muscle biopsy, an invasive procedure with a generally accepted risk of complications (e.g., bleeding, infection, and hematoma formation) of ~1%, and one that may often be associated with psychological distress for children.

Because direct sequence analysis of the dystrophin gene has been considered too labor intensive, expensive, and time consuming (Bennett et al. 2001), several groups have recently developed strategies to detect exonic sequence variations by use of screening methods followed by direct sequence analysis of variant fragments only. One of these strategies is based on SSCP analysis (Mendell et al. 2001). This strategy, termed "DOVAM-S" (detection of virtually all mutations–SSCP) relies on multiplexing as many as 23 amplicons per lane, with SSCP in as many as five conditions. The authors report that as many as 75% of nondeletion mutations may be detected by this method, but there are several drawbacks. One is that all band variations

Received November 22, 2002; accepted for publication January 9, 2003; electronically published March 11, 2003.

Address for correspondence and reprints: Dr. Kevin M. Flanigan, Eccles Institute of Human Genetics, Room 7110B, 15 N. 2030 E. Street, Salt Lake City, UT 84112. E-mail: kevin.flanigan@genetics.utah.edu

0002-9297/2003/7204-00014$15.00

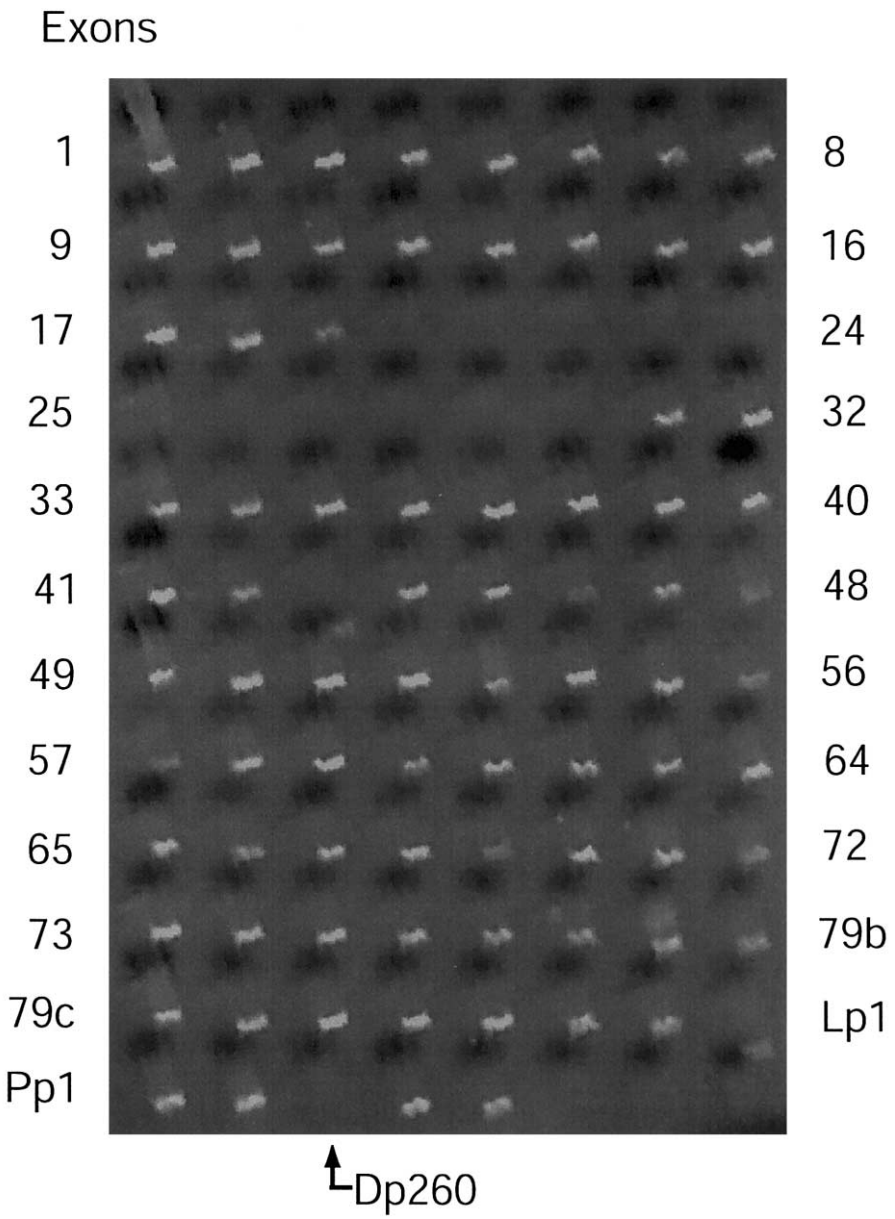

Figure 1 Agarose gel analysis of primary PCR products. An aliquot of each well from the 96-well PCR amplification plate is loaded in a 96-well format onto an agarose gel. Electrophoretic separation distance for each band is ~1.8 cm, because the wells are angled slightly in relation to the migration path. The products are from a patient with a multiexon deletion who is missing exons 20–30 and the DMD260 promoter. Products corresponding to exons 1–78 are located in sequential wells, starting left to right and top to bottom, followed by the multiple exon 79 and alternate promoter products. Note the absence of products in wells corresponding to exons 20–30 and to Dp260.

detected by SSCP techniques still need to be sequenced to determine whether they represent pathogenic mutations; the dystrophin gene, because of its size, has many reported polymorphisms. Another drawback is that, for economies of scale in the use of reagents and technician time, individual samples may need to be saved until multiple samples are available for simultaneous analysis of band variation.

A second strategy relies on denaturing high-performance liquid chromatography (DHPLC) as a screening method (Bennett et al. 2001). This strategy screens for DNA variations by separating heteroduplex and homoduplex DNA fragments by reverse-phase liquid chromatography, again followed by direct sequence analysis of variant amplicons. Using this method, Bennett et al. (2001) detected point mutations in six of eight DNA samples from patients without deletions, and those authors argued for its use, on the basis of both economic

and scientific concerns. Another alternative includes double gradient, denaturing gradient gel electrophoresis (DGGE) (Cremonesi et al. 1997). A drawback to all such approaches is that they are screening methods, with limitations in sensitivity inherent in each underlying technique. Each can detect both mutations and non–disease-associated polymorphisms, and a further sequencing step is required to distinguish between these possibilities.

We have developed a method that allows for the rapid, accurate, and economical analysis of any large multiexon gene. This method, single condition amplification/internal primer (SCAIP) sequencing, relies on amplification of a large number of exons at a single set of PCR temperatures. Sequencing specificity is gained by uniform use of a second, internal set of sequencing primers. Sufficient sequencing specificity is obtained without optimization of individual amplicon conditions. In the case of the dystrophin gene, our method results in complete double-stranded sequencing coverage of all known coding regions and in seven of the eight tissue-specific promoters. This sequencing can be routinely performed within three working days (following DNA purification) at a reasonable cost (including both reagents and personnel costs). Although the dystrophin muscle mRNA consists of 14 kb, our method analyzes an average of nearly 110 kb of sequence, allowing detection of polymorphisms in flanking intronic regions, as well as the 3′ and 5′ UTRs. In addition, our method allows detection of the ~2% of patients with exonic deletions not detected by the widely available multiplex PCR technique.

Patients and Methods

Patients

We ascertained patients from the University of Utah's Muscular Dystrophy Association clinic. The diagnosis of a dystrophinopathy was determined by the presence of clinical features consistent with DMD or BMD, along with either (1) absent or altered dystrophin expression, as determined by immunohistochemical, immunofluorescent, or immunoblot analysis or (2) a clear X-linked family history. Some patients had previously had confirmation of dystrophin deletions by clinical testing. Probands from 42 families were enrolled; 41 were male individuals with dystrophinopathy, according to the above criteria, and the remaining proband was a woman who was an obligate carrier (and the mother of two deceased patients with DMD). She had adult-onset limb-girdle weakness, which led to wheelchair dependence in her sixth decade. Nine additional DNA samples were obtained from self- or physician-referred patients nationwide who had been shown to be deletion negative on standard screening.

Patients were catalogued according to whether they harbored large-scale dystrophin deletions detectable by standard clinical multiplex PCR analysis. Blood samples for DNA analysis were obtained, according to a protocol approved by the institutional review boards of participating institutions, from patients who either had no clinical record of testing for dystrophin deletion (unknown deletion status) or who had no detectable deletion by commercial testing. DNA was obtained from each blood sample, using a salting-out method (PureGene, Gentra Systems).

Direct sequence analysis was also performed on 66 DNA samples from one clinical center (Ohio State University [OSU]). Of these, 64 had been previously evaluated by the DOVAM-S technique. The clinical phenotype of this set of patients was confirmed by one investigator (J.M.).

SCAIP Sequencing

The genomic organization of the dystrophin gene was assembled from contigs downloaded from the UCSC Genome Bioinformatics Web page (Kent et al. 2002) (see also the International Human Genome Sequencing Consortium 2001 [Lander et al. 2001]). Assembly and exon-intron annotation was performed using task-specific Perl scripts. Our assembly reveals that the DMD region is currently contiguous and gap free for the dystrophin Dp427m muscle isoform (accession number NM_0040006.1) spanning 2.09 Mb, and the dystrophin Dp427c brain isoform (accession number NM_000109) spanning 2.22 Mb of chromosome Xp21.2. Primer systems for PCR were designed to amplify DNA fragments, which span each exon and seven of the eight promoters (Dp427m, Dp427p, Dp427c, Dp427l, Dp260, Dp140, and Dp116). Each amplicon was designed for an optimal size range of 1.2–1.4 kb, with the exon, including unique promoters, centered within the amplicon (with the exception of exon 79, which was broken into seven fragments) to maintain uniform conditions. These were designed to produce 93 amplicons with a nearly universal size; this uniformity allows us to predict likely amplification conditions, using a single set of PCR temperatures. Aliquots (15 pmol) of each primer are placed into individual wells of a 96-well tray, evaporated to dryness in a speed vacuum system, and stored in a freezer at −20°C until use. For PCR amplification, 10 μg of patient template DNA is aliquoted into a master PCR mixture, and, subsequently, 25 μl of the mixture is aliquoted into the 96-well dish with dry primers. PCR amplifications were performed using the Platinum *Taq* DNA Polymerase High Fidelity System (Invitrogen), and each 25 μl of aliquoted reaction

contained 100 ng of genomic DNA, 200 μM dNTPs, 15 pmol of each PCR primer, 1 × reaction buffer, 2 mM Mg_2SO_4, and 0.5 U *Taq* DNA polymerase. PCR is performed in a thermocycler for 25 cycles under the following conditions: denaturation at 94°C for 20 s, annealing at 55°C for 30 s, and extension for 68°C for 4 min, followed by a final extension at 68°C for 7 min.

To validate PCR amplification and detect any deletions, 3 μl of the PCR reaction is run on a .75% agarose gel/ethidium bromide gel. The resulting gel is photographed and analyzed for absence of one or more bands. Because the absence of a single band may result from a primer site polymorphism, PCR is repeated, in such cases, using each of the following: the same primers, the internal sequencing primers, and combinations of original and internal primers. The absence of more than one adjacent exon is interpreted as being consistent with a multiexon deletion. The PCR products are then transferred and bound to a 96-well filter plate (MAFBN0B10; Millipore), with a 1.0-μm glass fiber type B filter, in the presence of a 5 M guanidine HCl/potassium acetate solution. Four 80% ethanol washes remove unincorporated primers, nucleotides, and excess salt and are followed by an elution of the fragments with warm nanopure H_2O.

Internal sequencing primers were designed to anneal to unique intronic flanking sequences, with attention to the specific 3′ sequence for each primer. As with the PCR, the primers are stored in 384-well plates so that both PCR set-up and sequence reaction set-up can be done with multichannel pipettors and pipetting robots. The sequence reactions are assembled by transfer of 6 μl of PCR product, along with 10 pmol of sequencing primer, to 2 wells of a 384-well cycle-sequencing plate, and the samples with primers are evaporated to dryness in a speed vacuum system. The fragments are then rehydrated with a 5-μl mixture of 0.5 μl ABI Prism BigDye terminators (version 3.0), 0.75 μl 5 × buffer mix, and 3.75 μl nanopure H_2O; the plates are heat-sealed with foil and are placed on thermocycling blocks for cycle sequencing. Cycling conditions included an initial denaturation at 96°C for 30 s, followed by 46 cycles at 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. Upon completion of cycle sequencing, 20 μl of 62.5% ethanol/1M potassium acetate (pH 4.5) was added to each reaction, and the sequence plates were centrifuged at 4,000 rpm at 4°C for 45 min. The samples were resuspended in 15 μl of HiDi formamide (ABI) and were electrophoresed on an ABI 3700 DNA analyzer prepared with POP-5 capillary gel matrix. All plates within the system have bar code labels with plain sample identifiers on them. These bar codes are captured at multiple steps of the process, using a Web-based system for plate tracking.

Sequence Analysis

After initial data processing on the ABI 3700 instruments, sequence-trace files are transferred onto a Linux disk server. The base calls are reanalyzed with the Phred program (Ewing et al. 1998), which adds a quantitative base-quality value. This base-quality value provides a probabilistic estimate of the correctness of the base call. The quality values are the log of the probability that the base call is correct, such that a Phred value of 20 corresponds to a 99% probability that the base call is accurate, and a Phred value of 30 corresponds to a 99.9% probability that the base call is accurate. The sequence is assembled with dystrophin consensus sequence, using the Phrap program (Ewing and Green 1998), and potential mutations are identified using the Consed program (Gordon et al. 1998). The read assembly is done on a per-PCR-fragment basis, and a single PCR Phrap assembly will consist of the consensus genomic sequence and all sequence reads relating to the PCR. The read sequence and Phred quality values are compared with the assembled consensus sequence using Cross_match (Ewing and Green 1998), and all discrepancies are tagged and ranked according to the Phred quality of the base (cutoff of 15). All PCR assemblies (reads plus consensus sequence and tagged discrepancies) are then compiled into one Consed project for review. Potential base discrepancies are catalogued using Perl scripts, and the discrepancies undergo human review of original trace files. This final list of reviewed discrepancies is loaded into an Oracle database, where they can then be further reviewed in a Web browser.

Nucleotide sequence position is based on the annotated mRNA sequence, found in GenBank (accession number NM_0040006.1), that encodes the dystrophin Dp427m isoform. The oligonucleotide sequences of primers used in amplification and sequencing of the dystrophin gene are available in supplementary tables A and B (online only).

Results

Large-Scale (≥1 Exon) Deletions

Deletion status was determined by reviewing clinic records or obtaining results of clinical (multiplex PCR) testing in 42 Utah probands. Large-scale deletions were found in 25/42 (59.5%) patient samples. (As discussed below, a single Utah sample had a nonhotspot single-exon deletion, bringing the total found in the Utah cohort to 26/42 [62%] probands.) SCAIP sequence analysis was not performed in individuals with these large-scale deletions.

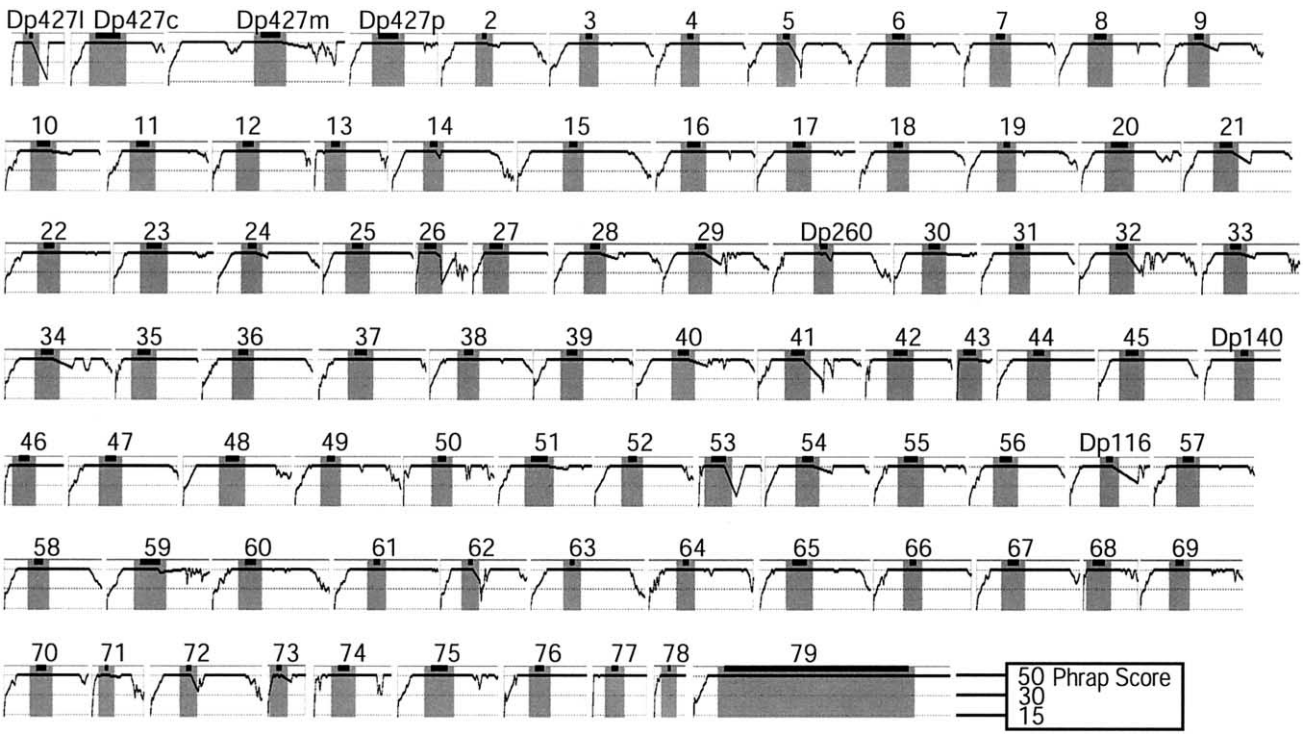

Figure 2 Average Phrap score coverage of DMD exons and promoter regions. Each block represents the length of the individual PCR products, with the exonic sequence indicated by the thick line on the top horizontal axis. The average Phrap score observed in the present study is plotted along its horizontal position, with the vertical axis ranging from Phrap score 15 to 50. Phrap scores >50 are not shown, and the portions of the plot corresponding to the exons ±100 nt are indicated in gray.

Direct Sequence Analysis by SCAIP Sequencing

Amplification efficiency and deletion detection.—In anticipation of direct sequence analysis, PCR amplification was performed on 94 samples. These included the remaining 17 Utah probands without multiplex deletions and the 9 referral samples (total number of unique families is 26), 2 relatives of Utah probands (1 asymptomatic carrier mother and 1 affected sibling), and 66 samples from OSU (64 DOVAM-screened and 2 unscreened). Thus, PCR amplification was performed on a total of 94 specimens.

Analysis of PCR products by visualization on agarose gels resulted in the identification of three individuals with deletions ⩾1 exon. In one proband studied at OSU, multiple amplification products from adjacent exons (the DMD260 promoter and exons 20–30 [fig. 1]) were missing; review of records (which were unblinded only after the entire sample set was analyzed) showed that this had been detected by DOVAM analysis. In two patients, single amplification products were not present in exons not screened in commonly used multiplex screening sets; in each case, PCR was repeated with internal primers, to exclude the presence of polymorphisms at the primer sites, and the absence of a product on the second round of amplification was interpreted as representing single exon deletions. One Utah patient had a deletion of exon 18. One OSU patient had a deletion of exon 21; unblinded postamplification review of the DOVAM results showed that a possible deletion had been suspected but that a primer site polymorphism could not be excluded. A total of 8,742 PCR analyses were performed (94 individuals × 93 PCRs each); this number of analyses excluded exons determined to be deleted in these three patients (−14 deleted exons, leaving 8,728 potential products), the efficiency of primary PCR recovery (defined as the presence of a band on first pass, using single-plate amplification) was 8,716/8,728 (99.86%).

Sequencing efficiency and quality.—Sequencing was not performed on samples from the three patients in whom deletions were detected; therefore, direct sequence analysis was performed on 91 individual patient samples (91 × 93 exons = 8,463 exons sequenced). The overall quality of sequence recovery is shown in figure 2. The Phrap score over coding regions of the gene is generally >60. The efficiency of primary sequencing recovery (defined as high quality sequence on the first sequencing reaction) was 8,396/8,463 (99.21%).

Mutation and polymorphism detection.—Among the

samples from the 16 Utah probands and 9 referral samples, mutations were detected by SCAIP sequence analysis in 16; 5 additional samples harbored duplications (see below), resulting in an overall detection efficiency of 80% in this group (16/20 patients without duplications). The mutations are summarized in tables 1 and 2 and include 10 stop codon mutations, 1 single-base-pair insertion, and 1 single-base-pair deletion. The single-base-pair insertions and deletions were easily detectable as mixed base calls in the two females tested.

In two referral samples, we detected sequence variations that may cause disease by altering intronic splice signals. One is a highly likely cause of disease, because it occurs in the highly conserved +1 position in intron 25 (changing a G to a C). The other is less definitively causative, because it occurs in the less conserved −9 position in intron 11. Both are unique in our series ($n = 91$) and are previously unreported, according to the Leiden database of dystrophin mutations. Definitive assignment of a causative status to these two mutations will require analysis of dystrophin transcripts; at present, muscle samples are unavailable to us, although further studies are planned.

Of particular interest are two substitutions that result in nonsynonymous changes in amino acid sequence in highly conserved functional domains of the dystrophin protein. One of these, in a boy with a DMD phenotype (loss of ambulation at age 10 years) substitutes a phenylalanine for a cysteine in the dystroglycan-binding domain, in a residue conserved in the dystrophin protein through *Caenorhabditis elegans*. The second, in a boy with a BMD phenotype (still ambulant at age 16 years), substitutes a valine for an asparagine at a similarly conserved residue in the actin-binding domain.

After direct sequence analysis was performed, dystrophin duplication analysis was performed in 13 samples, including the 9/25 Utah or referral samples without detectable mutations, and the 4 with presumed mutations discussed above (2 intronic and 2 missense). Duplication analysis was performed using the multiplex amplifiable probe hybridization (MAPH) technique (White et al. 2002). No duplications were detected in the samples with the four presumed mutations. Of the remaining nine samples, duplications were found in five (data not shown). Of the four remaining patients without detected mutations, one individual (patient 42965) was reported to have dystrophin of an *increased* molecular weight when analyzed with the use of commercially obtained immunoblots, raising the possibility that a duplication remains undetected by the MAPH technique. Our strategy is not currently optimized for the detection of duplications in agarose gels, because (1) detection by template dosage is qualitative, not quantitative, after 25 cycles of PCR amplification, reaching a nonexponential phase of PCR; and (2) detection by altered band size would require duplications to occur within the 1.1–1.3 kb of sequence contained within the primary amplification product, which would be expected to be a rare event.

Comparison of sensitivity to DOVAM.—Using SCAIP, we studied 66 samples from a second center, in a blinded fashion; 64 had previously been studied by DOVAM, which detected subexonic mutations in 44 samples and possible exonic deletions in 2 (discussed above). SCAIP analysis detected all 44 mutations, as well as a previously undetected stop codon mutation (Glu2035X in exon 42, GAG:2035:TAG) in 1 of the 20 other nondeleted samples. This position is 2 nucelotides (nt) 5′ of a common variant, GAT:2035:GAG (Asp:Glu), that may have interfered with the SSCP analysis used in the DOVAM test.

Discussion

We describe the initial application of the SCAIP sequencing method to mutational analysis of the dystrophin gene. This method has a sensitivity at least equal to that of DOVAM screening, and it detected at least one mutation not detected by the DOVAM method. It gives highly reproducible and accurate results and can be performed economically on single samples. Overall, SCAIP detected dystrophin mutations in 70% of patient samples that did not have deletions of ⩾1 exon. When five patients with duplications from the Utah/referral set were excluded, the detection increased to 73% (61/84). This is probably an underestimate of the actual rate of detection in the general nonduplication sample population, because duplication testing was not performed on the DOVAM-negative/SCAIP-negative set ($n = 17$).

Attempts to correlate these numbers with the general dystrophinopathy population are unhelpful, because our patient set was not a random sample; it likely represented a population enriched in duplications, stop codons, and subexonic rearrangements (to judge by the referral samples). The absence of detectable mutations in the remaining patients is not yet explained; however, the situation is unlike the one produced when DOVAM or DHPLC screening is performed, in that we can state unequivocally that the known coding regions of the dystrophin gene do not contain disease-causing subexonic mutations.

The primary determinant of sequence specificity and base-call quality is the uniform use of internal sequencing primers. The assay design is robust, in that it can tolerate secondary, nonspecific PCR amplification products, as opposed to assays that use a single set of primers or use secondary primers to universal sequences on the 5′ end of the PCR primers. An original design criterion was to optimize a single 96-well–plate assay, in which all coding regions and promoters of the dystrophin gene

Table 1

Mutations Detected in Utah and Referral Patients

Sample (No. of Probands), Mutation Status (No. of Probands), and Patient ID (Number)	Disease Type	Age at		Mutation				
		Presentation (years)	Loss of Ambulation (Time of Writing) (years)	Exon	Type (Size)	nt[a]	Amino Acid	Novel[b]
Utah nondeletion, nonduplication (12):								
With mutation (9):								
42172 (1)	DMD	1.25	9	47	Stop	6868A→T	Lys2290X	+
42588 (2)	BMD	3	NA (10)	31	Stop	4250T→A	Leu1417X	−
42719 (3)	BMD	13	NA (19)	31	Stop	4240C→T	Gln1414X	+
42953 (4)	DMD	6	9	64	Stop	9337C→T	Arg3113X	−
42970 (5)	BMD	20	NA (58)	1	Stop	9G→A	Trp3X	+
42390 (6)	DMD	3	NA (4)	30	Deletion (1 bp)	4103delG	Frameshift	+
42389 (6a)[c]	NA	NA	NA	30	Deletion (1 bp)	4103delG	Frameshift	+
42359 (7)	Carrier[d]	30	NA (58)	8	Insertion (1 bp)	783_784insT	Frameshift	+
42458 (8)	DMD	5	11	68	Missense	9938G→T	Cys3313Phe	+
42515 (9)	BMD	6	NA (16)	6	Missense	494A→T	Asp165Val	+
Without mutation (3):								
40818 (10)	DMD	7	10		ND		None	
42273 (11)	BMD	8	NA (18)		ND		None	
42965 (12)	BMD	13	NA (21)		ND		None	
Referral (8):								
With mutation (7):								
42962 (13)	DMD	4	NA (5)	53	Stop	7720C→T	Gln2574X	+
42964 (14)	DMD	4	NA (7)	34	Stop	4693C→T	Gln1565X	+
42968 (15)[e]	IMD	2.5	NA (13)	58	Stop	8608C→T	Arg2870X	+
42969 (16)	BMD	3	NA (11)	5	Stop	355C→T	Gln119X	+
42971 (17)	BMD	5	NA (21)		Splice		IVS25+1G→C	+
42974 (18)	DMD	4	12		Splice		IVS11−9G→A	+
42986 (19)	DMD	2.5	10	34	Stop	4693C→T	Gln1565X	+
Without mutation (1):								
42963 (20)	BMD	5	NA (11)		ND			

Note.—NA = not applicable because patient is still ambulant. ND = none detected.

[a] Nucleotide sequence position is based on the annotated mRNA sequence found in GenBank (accession number NM_004006) which encodes the dystrophin Dp427m isoform (see text for discussion of mutation nomenclature).

[b] Novel refers to whether a mutation was (−) or was not (+) found in the Leiden DMD mutation database.

[c] Individual 42389 is the asymptomatic mother of individual 42390.

[d] Individual 42359 is a manifesting female carrier.

[e] Individual 42968 is listed as an intermediate phenotype [IMD]; he is ambulant at age 13 years, after receiving steroids for several years, but family history suggested a DMD phenotype with loss of ambulation before age 12 years in other members.

Table 2

Summary of Mutation Detection in Nondeleted, Nonduplicated Probands

Sample	No. of Mutations	No. of Samples (% with Mutation)
Utah samples/referrals	16	20 (80)
DOVAM positive	44	44 (100)
DOVAM negative	1	18 (5)
Unscreened by DOVAM	0	2 (0)
Total	61	84

could be amplified on a single PCR plate. The PCR products are then to be purified in plate format, using multichannel pipetting robots, and two cycle sequencing plates are prepared and processed. The one patient–one plate assay is designed to meet the requirements of both a rapid turnaround time for the assay and the ability to scale the assay in response to increasing demand.

Phenotype/Genotype Correlations

The rapid and economical detection of stop codons and small rearrangements will facilitate the study of sequence context effects on disease expression. However, in the present study, only limited correlations between phenotype and genotype are to be drawn, although our results raise several interesting examples. One patient with BMD—the most mildly affected patient in the Utah group, who is still walking at age 58 years—has a mutation resulting in a premature stop signal in the third amino acid of the muscle isoform; the next methionine is at position 124. Another intriguing result is the presence, in our relatively small sample, of two stop codon mutations in exon 31, both resulting in the BMD phenotype. Although stop codon mutations are expected to be essentially randomly distributed across the gene (unlike the hotspots found for exonic deletions) (Roberts et al. 1994), the presence of two exon 31 stop codon mutations raises the possibility that stop codons in certain exons may predispose to a milder phenotype, perhaps because of the influence of such mutations in promoting exon skipping (as seen in the *mdx* mouse [Wilton et al. 1997; Lu et al. 2000]). Further studies are planned to determine the mRNA and protein sequences in these and other patients.

Two patients had a previously undescribed Gln1565X mutation. These patients are not known to be related, and analysis of SNPs reveals different haplotypes over at least a portion of the dystrophin gene, supporting the idea that they are unrelated (although distant relatedness with intragenic recombination cannot be excluded). This example illustrates one of the additional advantages of SCAIP analysis. That is, SNPs are found throughout the gene; some are quite common, others less so. With the use of SCAIP analysis—unlike the use of screening strategies, such as SSCP or DHPLC—a sequence variation can be immediately identified with certainty, and the frequency of such variations can be quickly established by comparison with our large and growing database of specific polymorphisms. By cataloging these SNPs throughout the dystrophin gene's coding and control regions, and establishing a rigorous and standardized phenotyping process, we expect to generate testable hypotheses regarding the role of such SNPs in the presentation and/or progression of disease. For example, polymorphisms in the primary cardiac or brain isoform promoters could conceivably alter the clinical expression of cardiomyopathy or cognitive dysfunction. Studies to address these possibilities are under way.

Implications for Clinical Use, Including Genetic Counseling

Application of this method to the study and clinical care of dystrophin-related diseases will obviate the need for muscle biopsy in a large number of patients. It will routinely allow rapid detection, in an economical fashion, of the following gene variations in patients with dystrophinopathy: (1) all deletions of >1 exon; (2) rearrangements <1 exon (deletions and insertions); (3) premature stop codon mutations; (4) splice signal site mutations; and (5) missense mutations. Total direct costs of the SCAIP assay, including labor and consumable supplies, are estimated to be <$1,000. The actual cost of such a clinical test will depend on the indirect and overhead cost structures of a laboratory offering this service. Reports of nonsynonymous polymorphisms as disease-causing missense mutations in the dystrophinopathies are rare. We expect that analysis of data generated by our method will allow identification of variants at highly conserved amino acids in patients without any other sequence variation, leading to identification of greater numbers of missense mutations.

The availability of rapid direct sequence analysis will have an immediate impact on genetic counseling in the dystrophinopathies. Because approximately one-third of all patients with dystrophinopathy harbor de novo mutations, X-linked family histories are often absent, and testing of both known and presumptive carriers can, at present, be performed with high reliability only if a proband's specific mutation is known. In the absence of large-scale deletions, carrier testing relies (when possible) on haplotype analysis. The high-quality sequence-acquisition method we describe here allows ready identification of point mutations (or small-scale rearrangements) in the heterozygous state and will lead to improved genetic counseling related to dystrophinopathies, as well as to other diseases to which it is applied. Further studies to address the sensitivity of this technique

in autosomal disorders (with only heterozygous mutations) are under way.

Summary

Direct sequence analysis of large multiexon genes from individual patient samples has heretofore been considered too labor intensive, reagent intensive, and time consuming. As a result, many disease-related genes are not sequenced in clinical practice. The SCAIP method is applicable to the sequencing of any gene, allowing rapid, economical, and efficient sequence analysis of large multiexon genes in single patient samples.

Acknowledgments

This study was supported by Parent Project Muscular Dystrophy, the Muscular Dystrophy Association, the Primary Children's Research Foundation, and National Institutes of Health grants R01 NS43264-01 (to K.F.) and U01 HG02138-04 (to R.W.). The authors wish to acknowledge M. Bromberg, B. Wong, and T. Mozaffar, for referral of patients; M. Howard, for helpful discussions; R. Kuhn, G. O'Neill, A. Aoyagi, R. Mao, K. Ward, J. Bernaducci, and C. Schilling, for assistance; and J. den Dunnen and S. White of Leiden University, for duplication analysis.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for sequences of the dystrophin isoforms Dp427m [accession number NM_0040006.1] and Dp427c [accession number NM_000109])

Leiden Database, http://www.dmd.nl/

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for DMD [MIM 310200], BMD [MIM 300376], and the dystrophin gene [MIM 300377])

UCSC Genome Bioinformatics, http://genome.cse.ucsc.edu/ (for contigs)

References

Beggs AH, Koenig M, Boyce FM, Kunkel LM (1990) Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. Hum Genet 86:45–48

Bennett RR, Dunnen J, O'Brien KF, Darras BT, Kunkel LM (2001) Detection of mutations in the dystrophin gene via automated DHPLC screening and direct sequencing. BMC Genet 2:17

Chamberlain JS, Gibbs RA, Ranier JE, Caskey CT (1990) Multiplex PCR for the diagnosis of Duchenne muscular dystrophy. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Francisco, pp 272–281

Cremonesi L, Firpo S, Ferrari M, Righetti PG, Gelfi C (1997) Double-gradient DGGE for optimized detection of DNA point mutations. Biotechniques 22:326–330

Emery AE (1991) Population frequencies of inherited neuromuscular diseases: a world survey. Neuromuscul Disord 1: 19–29

Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res 8: 186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:175–185

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8:195–202

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res 12:996–1006

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Lu QL, Morris GE, Wilton SD, Ly T, Artem'yeva OV, Strong P, Partridge TA (2000) Massive idiosyncratic exon skipping corrects the nonsense mutation in dystrophic mouse muscle and produces functional revertant fibers by clonal expansion. J Cell Biol 148:985–996

Mendell JR, Buzin CH, Feng J, Yan J, Serrano C, Sangani DS, Prior TW, Sommer SS (2001) Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. Neurology 57:645–650

Miller RG, Hoffman EP (1994) Molecular diagnosis and modern management of Duchenne muscular dystrophy. Neurol Clin 12:699–725

Roberts RG, Gardner RJ, Bobrow M (1994) Searching for the 1 in 2,400,000: a review of dystrophin gene point mutations. Hum Mutat 4:1–11

Roest PA, Roberts RG, van der Tuijn AC, Heikoop JC, van Ommen GJ, den Dunnen JT (1993) Protein truncation test (PTT) to rapidly screen the DMD gene for translation terminating mutations. Neuromuscul Disord 3:391–394

Tuffery-Giraud S, Chambert S, Demaille J, Claustres M (1999) Point mutations in the dystrophin gene: evidence for frequent use of cryptic splice sites as a result of splicing defects. Hum Mutat 14:359–368

White S, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van Ommen GJ, Breuning MH, den Dunnen JT (2002) Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. Am J Hum Genet 71: 365–374

Wilton SD, Dye DE, Laing NG (1997) Dystrophin gene transcripts skipping the mdx mutation. Muscle Nerve 20:728–734