Rapid capillary zone electrophoresis in isoelectric histidine buffer: high resolution of the poly-T tract allelic variants in intron 8 of the \textit{CFTR} gene

Cecilia Gelfi,\textsuperscript{1} Marilena Perego,\textsuperscript{2} Pier Giorgio Righetti,\textsuperscript{2} Silvia Cainarca,\textsuperscript{3} Stefania Firpo,\textsuperscript{4} Maurizio Ferrari,\textsuperscript{4} and Laura Cremonesi\textsuperscript{4*}

The poly-T tract in intron 8 of the cystic fibrosis conductance transmembrane regulator (\textit{CFTR}) gene exists in three variants, 5T, 7T, and 9T. The 7T and 9T variants generate a predominantly normal \textit{CFTR} transcript, whereas the 5T variant engenders an anomalous product. The analysis of the poly-T tract is assuming increasing relevance, both to assess the implication of the \textit{CFTR} gene in congenital bilateral absence of the vas deferens and to evaluate genotype-phenotype correlation in cystic fibrosis. Mapping of the poly-T tract has been performed by cumbersome and time-consuming methodologies. Capillary zone electrophoresis, combined with laser-induced fluorescence detection, was introduced for a rapid separation of the poly-T tract amplified products. As separation buffer, we adopted 200 mmol/L histidine (pH $\text{pI} = 7.6$), and the capillary was filled with 10\% polyacrylamide, allowing separations in less than 10 min. Capillary zone electrophoresis results were in perfect agreement with dot-blot analysis.

The poly-T tract located in the acceptor splicing site of intron 8 of the cystic fibrosis transmembrane conductance regulator (\textit{CFTR})\textsuperscript{5} gene exists in three allelic variants, 5T, 7T, and 9T (1). The 7T and 9T variants generate a predominantly normal \textit{CFTR} transcript, whereas the 5T variant affects splicing efficiency and generates a high proportion of exon 9-skipped \textit{CFTR} transcripts that engender an anomalous protein product (1). The analysis of the poly-T tract is assuming increasing diagnostic relevance, both to assess the implication of the \textit{CFTR} gene in congenital bilateral absence of the vas deferens (CBAVD) and to evaluate genotype-phenotype correlation in cystic fibrosis (CF). CBAVD is an autosomal recessive disorder accounting for about 1\text{--}2\% of male infertility (2). Although CBAVD is present in the majority of CF males, male patients presenting with isolated CBAVD do not exhibit any of the clinical symptoms characteristic of CF (3). The CF phenotype is characterized by lung disease and/or pancreatic exocrine insufficiency, and almost all male and female patients show failure of reproductive function. The relationship between male sterility in CBAVD and CF patients has been extensively studied at the molecular level to determine the molecular spectrum of mutations involved in this syndrome. A large proportion of isolated CBAVD patients carry either a CF mutation and/or the 5T variant of the poly-T tract in intron 8 in at least one of their \textit{CFTR} genes (4\text{--}15), supporting the hypothesis that CBAVD represents a mild, primary genital form of CF (4). More recently, a correlation between the number of unaffected transcripts and the severity of lung disease in CF patients was found, because it was shown that the amount of exon 9 skipping in respiratory epithelial cells correlates with the severity of the pulmonary disease. Moreover, it was shown that the number of unaffected \textit{CFTR} transcripts transcribed from the 5T allele may be variable among different individuals and different organs of the same individuals (16). These data indicate that the amount of unaffected \textit{CFTR} splicing transcribed from the 5T allele is tissue-specific and that the 5T allele may be associated with variable clinical presentation–from healthy fertile males or CBAVD, to typical CF patients–resulting from variable amounts of exon 9 skip-
ping. Also, our study on a sample of CBAVD patients showed that the frequency of the 5T allele among these subjects is markedly higher than that found in a population of unaffected and CF chromosomes, confirming the implication of the CFTR gene in the pathogenesis of CBAVD.

Until now, mapping of the poly-T tract has been performed with various classical methods, including denaturing gradient gel electrophoresis and direct sequencing (9), dot-blot hybridization with radiolabeled allele-specific oligonucleotides (13), silver-stained, nondenaturing polyacrylamide sequencing format gels (16), single-strand conformation polymorphism analysis and direct sequencing (17), heteroduplex analysis (17), and restriction digestion of the PCR product followed by polyacrylamide gel electrophoresis (8), which are cumbersome and/or time-consuming methodologies. The increasing relevance that the poly-T tract in the CFTR gene is assuming in genotype-phenotype correlation in CF and other CF-related syndromes strongly demands the setting up of a new fast and automatable method.

Capillary zone electrophoresis (CZE) is one of the most promising techniques for performing rapid and reproducible separations of DNA. Compared with conventional slab gel electrophoresis, CZE presents the advantage of automation, gives qualitative and quantitative results, offers high resolution, rapid analysis, low reagent consumption, and minute sample requirement. Several applications have been described using CZE for nucleic acid analysis in mutation detection (18–22), disease diagnosis (23–27), forensic investigations (28–33), and PCR-amplified product quantitation (34–38). Separation by CZE has been achieved by using entangled sieving matrices such as hydroxypropylmethylcellulose, hydroxyethylcellulose, or non-cross-linked polyacrylamide (39–41). In our opinion, viscous solutions of linear polyacrylamide are to be preferred because they offer a vast range of chain lengths, are able to exert sieving over a large interval of DNA fragments lengths, and can either be polymerized in situ from nonviscous solutions of monomers to produce extremely viscous entangled solutions or dissolved directly in separation buffer after prepolymerization and purification (40).

The difficulty in resolving polymorphic DNA regions containing small repeated sequences depends on the size and base composition of the sequence motif. Generally, the longer and more heterogeneous the base composition of the repeated motif, the easier the resolution in a given separation medium. In the past, we set up simplified methods for analyzing by CZE DNA regions containing tetra- (42) and trinucleotide (43) polymorphic repeats, where analysis of double-stranded DNA PCR products was carried out in TBE buffer. Finally, we drew our attention to the resolution of dinucleotide motifs, such as the poly-T tract variants in intron 8 of the CFTR gene, where we could not use the TBE system because it is not the most suitable for resolving double-stranded DNA fragments differing by the same base pair doublet (TT) from each other.

The aim of this study was to develop a fast, reproducible, and efficient protocol for analyzing poly-T tract variants. Recently, analysis of a polyadenine tract of the transforming growth factor-β type II receptor gene in colorectal cancers has been reported, where one or two base deletions of the (A)10 repeat could be detected on single-stranded DNA amplified by an asymmetric PCR and resolved by non-gel-sieving capillary electrophoresis (44). This system allowed the best peak separation at 500 V/cm within 30 min, under ultraviolet detection.

In a continuing effort to improve resolution by exploring new matrices and buffering systems, we recently focused on single-stranded oligonucleotide separations by adopting a new isoelectric buffering system, histidine (His) free base. In this system, we could apply as much as 800-1000 V/cm, thus favoring high resolution, allowing much reduced transit times and improved robustness of the separation column. Because of the quasi-neutral pH of His buffer, these columns can be reutilized hundreds of times while still maintaining a high reproducibility (45, 46).

Taking advantage of our last experience on separations in isoelectric buffers, we developed a method for a rapid typing of short double-stranded DNA molecules, differing by as little as the same 2-bp unit. As a model of a difficult separation, we applied this system to analyze the poly-T tract allelic variants located in intron 8 of the CFTR gene, where the use of such buffering system has been coupled with fluorescently labeled primers to obtain an extraordinary increment of sensitivity by laser-induced fluorescence (LIF) detection (47–49).

Materials and Methods

Materials and Methods

REAGENTS

Acrylamide, ammonium peroxidesulfate, and N,N,N’,N’-tetramethylethlen-diamine were obtained from Bio-Rad. His was from Sigma Chemical. Fused silica capillaries (100 μm i.d., 375 μm o.d.) were from Polymicro Technologies. Marker V was from Boehringer Mannheim, the 10-bp marker was from Life Technologies, Centricon 10 membranes from Millipore, recombinant Dynazyme II polymerase from Finnzyme Oy, and dNTPs from Pharmacia.

DOT-BLOT HYBRIDIZATION WITH RADIOLABELLED ALLELE-SPECIFIC OLIGONUCLEOTIDES

For the first PCR, PCR amplification of the poly-T tract in intron 8 of the CFTR gene used primer 9i-5’, 5’-TAATG-GATCATGGGCCATGT-3’, and primer 9i-3’, 5’-ACAGT-GTGGAATGTGGTGCA-3’ (50). The PCR conditions were as follows: 1 min at 94 °C, 1 min at 56 °C, and 1 min at 72 °C. The product length was 560 bp.

Primers for the second semi-nested PCR were primer 9i-5’ and 9e-3’ (5’-AGAAGAGGCTGTCATCACA-3’).
The same conditions as for first PCR were used, with 1 μL of the first PCR reaction as template. The PCR product length was 207 bp.

The reaction mixture for all amplifications contained 20 pmol of primers, 200 μmol/L dNTPs, 10 mmol/L Tris-HCl, 50 mmol/L KCl, 1.5 mmol/L MgCl₂, 1 g/L Triton X-100, and 2 U of Taq polymerase for 100 μL final volume.

For the dot blot, 2 μL of the second PCR reaction was denatured, blotted onto a nylon membrane, and hybridized with 5′-end labeled allele-specific oligonucleotides complementary to the sequence containing five thymidines (5′-TGTGTGTGTTTTTTAACAG-3′), seven thymidines (5′-TGTGTGTTTTTTTAAACAG-3′), or nine thymidines (5′-TGTGTTTTTTTTTTAACAG-3′), respectively (13) (Fig. 1).

**CZE in amphoteric buffer of PCR amplified products**

For PCR amplification of the poly-T tract in intron 8 of the CFTR gene, the first PCR used primers 9i-5′ and 9i-3′. The template was 400 ng of genomic DNA. The cycling conditions were as follows: 5 min at 94 °C for 1 cycle; 45 s at 94 °C, 45 s at 62 °C, 45 s at 72 °C for 3 cycles, 45 s at 94 °C, 45 s at 60 °C, 45 s at 72 °C for 3 cycles, 45 s at 94 °C, 45 s at 58 °C, 45 s at 72 °C for 18 cycles, 45 s at 94 °C, 45 s at 58 °C, and 10 min at 72 °C for 1 cycle. The product length was 59 bp for the fragment amplified from alleles carrying the 5T variant, 61 bp from the 7T variant, and 63 bp from the 9T variant.

Reaction mixture for 100 μL final volume was: 20 pmol of primers, 200 μmol/L dNTPs, 10 mmol/L Tris-HCl, 50 mmol/L KCl, 1.5 mmol/L MgCl₂, 1 g/L Triton X-100, and 2 U of Taq polymerase.

CZE was performed in a Bio Focus 3000 LIF 2 instrument equipped with a LIF detector providing 488-nm excitation with 520-nm bandpass filter (Bio Rad). To eliminate electroendosmosis, a Beckman-coated capillary was used. The capillary (100 μm i.d., 375 μm o.d.) was 27 cm long with a detection window at 7 cm from the outlet, from which the injections were done. As separation buffer of 200 mmol/L His was used, and the capillary was filled with 6% or 10% polyacrylamide polymerized in situ. The Marker V and the 10-bp marker were run in a 6% polyacrylamide, injected for 5 s electrokinetically, and detected by ultraviolet light at 254 nm. After desalting with the Centricon 10 membrane, the sample was injected electrokinetically for 30 s at 8 KV, and separation was made at 150 V/cm. The running temperature was set at 25 °C.

**Results and Discussion**

We evaluated the poly-T distribution in cystic fibrosis (CF) and unaffected chromosomes in 76 families having a CF child by dot-blot hybridization with radiolabeled allele-specific oligonucleotides (Fig. 1). Among CF chromosomes, the 5T, 7T, and 9T variants had a 1%, 51%, and 48% frequency, respectively, and among unaffected chromosomes, 3%, 84%, and 13%, respectively. Among the 19 CBAVD subjects investigated (of the 38 chromosomes analyzed), 8 (21%) carry the 5T variant, 19 (50%) the 7T variant, and 11 (29%) the 9T variant.

Typing of the poly-T tract in intron 8 of the CFTR gene in CBAVD subjects showed an increased proportion of the 5T variant in these subjects. The frequency of the 5T allele in CBAVD subjects (21%) is significantly different from that found in unaffected chromosomes (3%; Fisher’s exact test: P = 0.0008, not corrected) and in CF chromosomes (1%; Fisher’s exact test: P = 0.001, not corrected), thus confirming the implication of CFTR abnormalities in the pathogenesis of CBAVD.
OPTIMIZATION OF PCR CONDITIONS

Because of the extremely high sensitivity of the CZE system coupled with LIF detection, the greatest problem we encountered in setting up the poly-T amplification was to avoid the appearance of spurious bands in the amplified product because of primer mispairing, which can seriously affect genotypic interpretation. To analyze the poly-T tract in intron 8 of the \textit{CFTR} gene, we used a nested PCR procedure because the (GT) tract immediately preceding is also polymorphic with respect to the length (Fig. 1). To optimize specificity, we used two strategies: (a) we used 1 μL of a 300-fold dilution of the first PCR reaction as template for the nested PCR reaction and (b) we used a simplified version of the "Touchdown" PCR strategy (51), named "Stepdown" PCR (52), by performing an initial set of a few cycles at an annealing temperature 4–6 °C higher than the targeted annealing temperature calculated from the melting temperature (Tm) of the primer pairs used (denaturation at 94 °C for 45 s, annealing at 62° for 45 s, and extension at 72 °C for three cycles for the first PCR reaction; denaturation at 94 °C for 45 s, annealing at 62° for 45 s, and extension at 72 °C for three cycles, denaturation at 94 °C for 45 s, annealing at 60° for 45 s, and extension at 72 °C for an additional three cycles for the nested PCR reaction) and subsequently decreasing the annealing temperature to the targeted one for each set of primers for the remaining cycles (denaturation at 94 °C for 45 s, annealing at 60° for 45 s, and extension at 72 °C for 19 cycles for the first PCR reaction; denaturation at 94 °C for 45 s, annealing at 58 °C for 45 s, and extension at 72 °C for 18 cycles for the nested PCR reaction). A final cycle including prolonged extension (10 min at 72 °C) was added to both PCR assays. This strategy, coupled with the use of nested primers, gave the expected 59-, 61-, and 63-bp products for the 5T, 7T, and 9T alleles, respectively, indicating that the appropriate target was amplified, without any aspecific additional band. As shown in Figs. 5 and 6, products within the allelic size difference of 5T-9T can be determined unambiguously on all CZE tracings.

OPTIMIZATION OF CZE CONDITIONS

The use of isoelectric buffers in CZE was first described by Hjertén et al. (53), and the application to oligonucleotide separation was largely investigated by our group (45, 46). By using amphoteric buffers, we could assess the purity of antisense oligonucleotides in 5 minutes with high resolution by combining very low conductivity together with satisfactory buffering capacity. For oligonucleotide separation, His showed the best resolution compared with lysine or ampholytes, even if its buffering power was lower. His is an amino acid having three dissociative groups with the following pKs: 1.8 (R-COOH), 6.1 (side chain), and 9.2 (R-NH$_3^+$). Because the two highest pKs are rather close, His has reasonably good buffering capacity at its pI. Considering the pK of the side chain and NH$_3^+$, the pI of His is 7.65, well compatible with the range of stability and full dissociation of DNA.
voltages. The pI of 7.65 reduces the hydrolysis rate of the polyacrylamide, used both for coating and for sieving, thus allowing a repetitive use of the matrix and consequently increasing the number of assays with the same capillary without loss of reproducibility. These properties make His a good candidate for DNA separations, and we decided to investigate its performance in double-stranded DNA separations. In fact, although it is well known that single-stranded DNA separations allow single base resolution up to at least 300 bases (as in DNA sequencing), it must be emphasized that (a) sieving liquid polymers, as adopted in CZE, do not permit such a high resolution as that obtained in cross-linked polyacrylamide gel matrices and (b) double-stranded DNA separations allow the use of fluorescent intercalators without any need of covalently labeling the primers, as typical of single-stranded DNA analysis.

We investigated the increase in buffer concentration related to the separation of marker V, which presents fragments ranging from 8 to 587 bp. By using buffer concentrations ranging from 50 to 300 mmol/L His we observed, as shown in Fig. 2, that the best resolving power was at 200 mmol/L His with an increase in resolution from 8 to 130 bp and the concomitant separation of the 123/124-bp doublet, which is usually achieved only in the presence of ethidium bromide as intercalating agent. In all cases, however, we completely lost resolution above 130 bp. To determine the optimal His concentration, we focused on optimizing the separation of the 57-bp vs 64-bp peaks of the previously shown marker V, because this is our range of interest. By plotting the resolution ($R_s$) vs His molarity, as shown in Fig. 3, we observed that the best resolving power was at 200 mmol/L His. Moreover, we confirmed the resolving power increment, within the region of interest, by analyzing a 10-bp marker in the presence of 8% linear polyacrylamide and 200 mmol/L His; as shown in Fig. 4, we had high resolution, even of the products of spurious markers, because of incomplete synthesis, up to 100 bp.

Concerning the resolution increment in double-stranded DNA up to the 130-bp length, we think that the mechanism could be an interaction between His and DNA via coulombic attractions and consequent modification of the DNA radius of gyration (46). Recent data (Gelfi et al., unpublished) have shown that the addition of competing cations (e.g., K⁺ in KBr and Na⁺ in NaCl), even at the concentration of 10 mmol/L, completely obliterates the effects of the His buffer, suggesting that the smaller cations can effectively compete on binding sites on the DNA double helix. We hypothesize that His could bind via coulombic interaction with the phosphate groups by forming a bidentate ionic complex with the two positive groups (imidazole and alpha-amino, which both bear half-positive charge at pH = pI). His could thus be oriented so as to offer the two half-positive charges to the phosphate, while keeping the fully dissociated carboxyl group in a distal position.

Taking into account these findings, we developed a protocol that presents the advantage of rapidly genotyping short DNA-amplified fragments as in the poly-T tract by using CZE in polymer networks, His as separating buffer, and LIF detection. In comparison with the other techniques used thus far to analyze the poly-T tract variants, this protocol is more rapid and nonradioactive.
Fig. 5 shows the separation of an amplified fragment with genotype 5T/9T in 10% polyacrylamide and 200 mmol/L His at 200 V/cm using a 7-cm portion of the capillary. As one can see, the separation occurs in <10 min. Fig. 6 shows the analysis of all the genotypes we identified in our sample population; superimposition of the tracings demonstrates the reproducibility of the system. To control the system reproducibility, we analyzed 50 samples, previously genotyped by dot-blot hybridization, and we observed excellent assay-to-assay reproducibility; the SD was ±2%. This analysis also confirmed that double-stranded DNA amplified with stepdown PCR did not contain any artificial change of repetitive number as compared with traditional PCR carried out previously on the same samples.

It might sound odd that CZE in standard TBE buffer should not be able to separate, e.g., the 5T-7T or 7T-9T doublets, both having length differences of ~4%, when conventional sequencing gels have a resolution even <1% size difference. It should be noted, however, that (a) standard sequencing gels have a much higher resolution than sieving polymer solutions because of a fixed-pore geometry and (b) the 5T to 9T DNA fragments have a relatively small size (59 to 63 bp) and thus a considerably higher diffusion rate in loose polymer networks. The unique resolving power obtained in our isoelectric His buffer might be due to a combination of the following factors: (a) stiffening of the double helix by potential interactions with His, and (b) reduced transit times due to relatively higher voltages and shorter capillary lengths.

Compared with the previous report on the separation of single-stranded DNA fragments amplified by asymmetric PCR (44), our method is a model of a difficult separation on double-stranded DNA and offers the advantage of much reduced transit times in PCR product analysis, which is one of the important factors to apply CZE technique to automated DNA diagnosis.

Our data show that CZE in liquid polymers in the presence of isoelectric His is the technique of choice to analyze short, double-stranded, PCR-amplified DNA fragments, differing by as little as even the same 2-bp repeated unit.

This work was partially funded by Ministero della Sanità, Progetto Fibrosi Cistica, and Bio-Rad Labs. P.G.R. is supported by grants from Telethon-Italy (N. E.555), from the European Community (Biomed II, grant BMH4-CT97-2627), and from Associazione Italiana Ricerca sul Cancro (AIRC-Milano).

References


15. Chu CS, Trapnell BC, Curristin S, Cutting GR, Crystal RG. Genetic


21. Chu CS, Trapnell BC, Curristin S, Cutting GR, Crystal RG. Genetic


52. Hecker KH, Roux KH. High, low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. BioTechniques 1996;20:478–85.