Nested genetic bit analysis (N-GBA) for mutation detection in the p53 tumor suppressor gene

Steven R. Head*, Yu-Hui Rogers, Kalpana Parikh, Guang Lan, Stephen Anderson1, Philip Goelet and Michael T. Boyce-Jacino

Molecular Tool, Inc., Alpha Center, Hopkins Bayview Research Campus, 5210 Eastern Avenue, Baltimore, MD 21224, USA and 1Center for Advanced Biotechnology and Medicine, Piscataway, NJ 08854-5638, USA

Received July 24, 1997; Revised and Accepted October 28, 1997

ABSTRACT
There is a growing and significant demand for reliable, simple and sensitive methods for repeated scanning of a given gene or gene fragment for detection and characterization of mutations. Solid-phase sequencing by single base primer extension of nested GBA™ primers on miniaturized DNA arrays can be used to effectively scan targeted sequences for missense, insertion and deletion mutations. This paper describes the use of N-GBA arrays designed to scan the sequence of a 33 base region of exon 8 of the p53 gene (codons 272–282) encompassing a hot spot for mutations associated with the development of cancer. Synthetic DNA templates containing various missense, insertion and deletion mutations, as well as DNA prepared from pancreatic and biliary tumor cells, were genotyped using the exon 8 arrays.

INTRODUCTION
As the number of new gene sequences steadily increases, the importance of repeated sequencing of genes related to disease processes, such as cancer, has become apparent. The prevalence of mutations in complex genetic loci, such as p53, BRCA1 and BRCA2, exemplifies the challenges facing researchers attempting to create effective tools for analysis of variability and alterations within these genes, both in population analysis and individual patient analysis. The required sensitivity and resolution of mutation detection are important factors in determining which method of mutation detection is most cost-effective. Identification of the presence of a mutation within the p53 gene of a breast cancer biopsy, for example, may be sufficient for a clinician to better predict the clinical outcome of the patient (1). The precise location of mutations, however, can yield clues about the epidemiology of the disease and lead to a better understanding of its cause and prevention (2). For many applications, more detailed sequence information (in terms of mutation position and type of mutation) available to the researcher or clinician results in more effective diagnosis and therapy. However, acquiring more detailed information has tended to require more complex and expensive analysis tools (e.g., gel-based DNA sequencing machines), resulting in a trade-off of quality for quantity in genetic analysis. Clearly, there is a need for mutation analysis technology that provides detailed sequence information at a substantially lower cost and more convenient format than gel-based sequencing. Development of this type of technology, such as reported here, would enable more widespread use of information about genetic variation in clinical and research applications.

The application of array-based nucleic acid analysis techniques for mutation detection has been made possible by rapid technology developments in the manufacture of high-density arrays. DNA array methods have almost exclusively depended on hybridization to discriminate and detect non-wild type bases in the target sequence. The most developed array-based sequencing or sequence scanning approach, sequencing-by-hybridization (SBH), is inherently sensitive to small changes in hybridization conditions and may require sophisticated analyses of high numbers of oligonucleotide probes (20 or more per base of the target sequence), under two or more sets of hybridization conditions, in order to accurately determine the nature and location of mutations in the targeted region (3). The array redundancy needed to provide a suitable level of accuracy has led to the development of 16 000–20 000 or more probe arrays to sequence a few hundred bases of target. This complexity translates to data analysis complexity, thereby restricting the use of such approaches.

As an alternative to the SBH approach, we have developed a solid-phase sequencing method using Nested-Genetic Bit Analysis (N-GBA) for mutation scanning. GBA utilizes single base extension (with labeled ddNTPs) of solid-phase primers to determine the base composition at specific sites of hybridized template DNA (4). The use of arrayed, nested GBA primers for solid-phase sequencing allows the detection and positioning of mutations within the targeted sequence. The GBA method increases the flexibility and utility of DNA array mutation detection techniques by exploiting the advantages of primer extension biochemistry over differential hybridization for DNA sequence analysis. This results in a dramatic reduction in the number of probes required per target base when compared to standard hybridization-based methods for detection and positioning of mutations, thereby enabling a higher information content per oligonucleotide array of up to 20-fold (3,5).

*To whom correspondence should be addressed. Tel: +1 410 558 5900; Fax: +1 410 558 5910; Email: head@mtool.com
We have chosen a 33 base region (codons 272–282) of exon 8 of the human p53 gene as a model for development of a prototype N-GBA array for solid-phase sequencing (SPS). This region is one of several mutation hot spots found in p53, a tumor suppressor gene that encompasses 11 exons (393 codons) and ~19 kb of chromosome region 17p13.105–p12. Mutations in the p53 gene are the single most common genetic alteration in human cancers and generally result in loss of function of the protein. A significant fraction of colon, lung and breast cancers have been reported to contain p53 mutations (2,6). Since more than 100 000 additional cases of each of these cancers are diagnosed every year, the potential application of p53 analysis is significant, both clinically and commercially.

The majority of p53 mutations are missense (6) and tightly clustered between codons 118 and 309, the DNA binding region of the protein (Fig. 1). Amino acids 175, 248, 249, 273 and 282 account for 40% of the total reported missense mutations; the predominance of these so-called ‘hot-spots’ can vary depending on both the tissue of origin of the cancer and the geographic origin of the cancer patient (2). This diversity and dispersion of clinically relevant mutations poses a significant challenge to the development of routine detection strategies.

MATERIALS AND METHODS

Primers

PCR primers were designed to amplify a 97 bp product. The upper PCR primer hybridized to codons 261–269 of exon 8 (AGC′-TTCTCTTCC-3′). The lower PCR primer hybridized to codons 285–293 of exon 8 (AGC′-CCCTTTCTTGCGGAGAACAG-3′). Each PCR primer was synthesized both with and without four phosphorothioated bases at the 5′ end. GBA primers were disulfide modified at their 5′ ends and were designed (20–25 bases in length) to match the wild type sequence and to contain various missense, insertion and deletion mutations for both strands of codons 272–282 in the p53 gene. All oligos were synthesized and OPC column purified by Research Genetics, Inc. (Huntsville, AL) and Genosys, Inc. (Houston, TX).

Tumor DNA samples

DNA was prepared from pancreatic and biliary adenocarcinoma xenograft tumors by Dr Scott Kern in the Department of Oncology, The Johns Hopkins University School of Medicine. Gel-based sequencing was performed independently in Dr Kern’s laboratory to identify samples containing p53 mutations. The methods have been described and some results have been previously published (7–9).

PCR

PCR reactions were performed in 96-well V-bottom polycarbonate plates (Corning Costar, Acton, MA). The final concentration of the reaction mixture was 400 µM each dNTP, 50 mM KCl, 10 mM Tris–HCl (pH 8.4), 1.5 mM MgCl2, 0.5 µM each primer, 2.5 ng/µl DNA and 0.025 U/µl AmpliTaq DNA polymerase (Perkin Elmer, Foster City, CA). Each reaction was overlaid with 30 µl mineral oil and cycled in a BioIII thermocycler (Sun BioScience, Inc., Branford, CT). An initial 2 min denaturation step at 94°C was followed by 35 cycles consisting of denaturation (1 min at 94°C), annealing (2 min at 55°C) and extension (3 min at 72°C).

Generation of single-stranded template by exonuclease digestion

PCR product, amplified using one phosphorothioated PCR primer and one unmodified primer, was treated with T7 gene 6 exonuclease (Amersham, Arlington Heights, IL) at a final concentration of 0.06 U/µl PCR (diluted in buffer supplied by the manufacturer) to digest the non-phosphorothioate protected strand. After 30 min of incubation at room temperature, NaCl, EDTA and cetyl trimethyl ammonium bromide (CTAB) were added to a final concentration of 1.5 M, 10 mM and 1 mM, respectively, to stop the exonuclease digestion and aid in subsequent hybridization to arrayed GBA primers.

Attachment and patterning of GBA primers to glass

Attachment of 5′ disulfide modified oligonucleotides to glass surfaces was performed via an intermediate mercaptosilane layer using a disulfide bond exchange reaction as described elsewhere (Rogers et al., manuscript in preparation). Briefly, this attachment was obtained by a two-step process of silane treatment and oligonucleotide binding. Glass slides (Cel-Line, Inc., Newfield, NJ) were etched in 25% aqueous ammonium hydroxide, rinsed in milliQ water and then in 95% ethanol. They were treated for 30 min in 3-mercapto-propyl-trimethoxysilane (Aldrich Chemical Company, Inc., Milwaukee, WI) in an acidic buffer of aqueous ethanol (95% ethanol, pH 4.5). Slides were cured for at least 24 h under dry inert gas (Ar or N2). The 5′-disulfide modified oligonucleotides were diluted to 10 µM concentration in pH 9.0 carbonate buffer (500 mM) and arrayed on the mercaptosilane coated glass slides using a Hamilton 220 automated pipetting robot. Drop volumes of 50 nl with 1.0–1.5 mm spacing (center-to-center) were used. Arrayed primers were incubated in a humid chamber for 12–16 h, followed with a rinse in TNTw (10 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20) then in de-ionized water. Glass slides were purchased from Cel-Line pre-printed with teflon patterns (HTC coating) to contain liquids on the slides.

Both batch to batch variability in oligonucleotide attachment efficiency was observed and a manuscript is currently in preparation (Rogers et al.) that fully discusses the reproducibility and variability of the above described attachment chemistry. The effects of overall attachment variability from slide to slide was compensated for by the use of an enzymatic signal generation system (Molecular Probes’ ELF substrate) which enabled positive signal saturation by incubation of ELF substrate for 3–10 min. The variability in the relative attachment of the arrayed primers within a given slide was observed to be significantly lower than slide to slide variability (data not shown).

Hybridization

Hybridization of synthetic and PCR templates was performed on the glass slides by addition of 20–100 µl of DNA template, in a solution containing 1.5 M NaCl, 10 mM EDTA and 1% CTAB per
Figure 1. Four mutational hot-spot regions containing frequently cited p53 mutations are indicated by the bars marked A–D. A, codons 132–143; B, codons 174–179; C, codons 236–258; D, codons 272–282. Shown below the wild type sequence for codons 272–282 (region D) is a CCD image of an N-GBA array for the + strand of codons 272–282 and the array profile after image processing. Synthetic DNA template based on the wild type sequence was hybridized to the arrayed N-GBA primers, patterned in four identical arrays (each array contains duplicate points and bounded by a hydrophobic teflon coating) per glass microscope slide. Arrays were then extended by Klenow DNA polymerase (3’→5’ exonuclease free) and one of four different ddNTP mixes containing a fluorescein label on one of the four bases. Fluorescein labels on the extended arrays were detected with an anti-fluorescein alkaline phophatase conjugate and Molecular Probes’ ELF fluorogenic substrate.

array. Each template strand was allowed to hybridize to four identical arrays for 30 min at either room temperature or 37°C. After hybridization, slides were washed in TNTw buffer to remove unhybridized templates.

Extension of GBA primers
Following the hybridization step, one of four polymerase extension mixes was added to each array. In addition, arrays having no hybridized template were extended to control for template independent signals. Each extension mix contained 20 mM Tris–HCl pH 7.5, 10 mM MgCl₂, 25 mM NaCl, 10 mM MnCl₂, 15 mM sodium isocitrate, 1.5 µM of three unlabeled 2’,3’-dideoxynucleoside triphosphates (ddNTPs), 1.5 µM of one fluorescein labeled ddNTP and 0.08 U/µL of Klenow fragment (3’→5’ exonuclease free) of Escherichia coli DNA polymerase (Amersham). Extension conditions had been previously optimized (9) to extend solid-phase bound primers complementary to hybridized templates. All fluorescently labeled chain-terminating ddNTPs used in the extension reactions were purchased from DuPont NEN (Wilmington, DE). The extension reactions were carried out at room temperature for 15 min. The slides were then washed with TNTw, with 0.1 N NaOH to remove hybridized template and again by TNTw.

Signal detection and analysis
Detection of fluorescein signals on extended GBA primers was done by addition of an anti-fluorescein alkaline phosphatase conjugate (Boerhinger Mannheim Corp., Indianapolis, IN) diluted 1:5000 in TNTw with 1% fraction-V BSA (Sigma, St. Louis, MO) added. Each slide was incubated for 30 min then washed repeatedly in TNTw to remove unattached antibody. An alkaline phosphatase fluorogenic precipitating substrate (ELF), diluted 1:20 in buffer supplied by the manufacturer (Molecular Probes, Eugene, OR), was then applied. Following a 3–10 min incubation, excess substrate was washed off and the signal was visualized by illumination with either a standard UV light box or hand held UV lamp (360 nm). Images were captured using a EDC-1000L CCD camera (Electrim Corp., Princeton, NJ) equipped with a 530 nM emission filter (Fig. 1).

Image software developed at Molecular Tool was used to convert two dimensional arrays into one dimensional profiles for each labeled base by summation of pixel values in one direction across each spot (Fig. 1). The original raw 2D profiles are usually noisy and have varying background due to inhomogeneous illumination. The profiles were initially smoothed using a 7 kernel window. This step was followed by background subtraction based on local minima values using the space between the spots. Analysis of signals from arrays extended with and without hybridized template showed that the ratio of template dependent to template independent signals was ≥10 (data not shown).

RESULTS
Synthetic template sequence analysis
The results of N-GBA array analysis of both strands of a wild type synthetic DNA template (based on the sequence of codons 272–282 of exon 8 of the human p53 gene) are shown in Figure 2A. In this experiment all base positions were genotyped correctly from analysis of the profile data. Variability is seen in overall peak heights and signal-to-noise ratios. Strong positive signals (signal-to-noise ratio >4) are seen with 94% (62/66) of the primers. The remaining 6% (4/66) show weak positive signals (signal to noise ratios 1.5–4.0). Three primers show signals 3–4 times lower than the nearest neighboring peaks. We believe this variability in data quality is due to inconsistent GBA primer attachment caused by differences in primer quality or context-dependent template hybridization efficiency. Template independent
controls (arrays with no template hybridized) were run with all experiments and showed no significant signals (peak heights <10% of template dependent signals). A manuscript describing the analysis of the GBA primer attachment method is currently in preparation (Rogers et al.).

Figure 2B shows + and – strand N-GBA profiles of synthetic templates containing missense mutations. The missense mutation is identified by an unambiguous signal change (indicated by the arrow) followed by several weak or non-existent signal (to the right of the altered peak in the + strand and to the left in the – strand) consistent with a 3’ mismatch of the wild type GBA primer with the non-wild type base on the template strand. The presence of the non-wild type base at the mutation site interferes with base calling downstream from the mutation, apparently due to the mismatch of the 3′ end of the GBA primer with the template. As this mismatch moves away from the 3′ end towards the 5′ end of the GBA primer, allowing the 3′ hybrid complex to form, the signal rapidly recovers. The downstream signal is significantly reduced on the + strand for at least four bases and on the – strand for at least two bases. GBA signals are completely missing only from the primers with mismatches at their 3′ ends. The value of genotyping both strands is apparent, as approaching the mutation from both sides should result in complete sequence data in the case of missense and deletion mutations.

N-GBA array profiles of synthetic DNA templates containing five base insertions are shown in Figure 2C. The frameshift caused by the insertions results in a more significant loss of interpretable downstream signal (four bases on both the + and – strands) when compared to the missense mutation in Figure 2B (only a single base on each strand). Results of N-GBA analysis of single base insertion templates also showed 4 bases of downstream signal loss (data not shown). The profile signals recover in both the + and – strands when only four bases of complementarity exist between the 3′ end of the GBA primer and the template strand, despite the effects of the five base frameshift.

The five base deletion profiles (Fig. 2D) show a more significant loss of signal strength downstream from the deletion site due to the
Figure 3. Partial N-GBA array profiles of wild type and missense synthetic DNA templates mixed in various ratios. (A) + and – strand profiles of 100% wild type template; (B) + and – strand profiles of 95% wild type template and 5% mutant template; (C) + and – strand profiles of 80% wild type template and 20% mutant template; (D) + and – strand profiles of 50% wild type template and 50% mutant template; (E) + and – strand profiles of 100% mutant template. The position of the altered base in the mutant strand is indicated on each profile by the arrow.

combined effect of the frame-shifted template and the five deleted bases. Interestingly, for the template tested, the + strand profile correctly identifies the first three bases 3′ to the deletion before incorrect or insufficient signals are generated. The – strand profile shows correct base calling for two bases 3′ to the deletion before loss of signal for 12 consecutive bases, indicating recovery of signal after seven bases of the 3′ end of the GBA primer can hybridize to the template beyond the frameshift. The varying ability of the Klenow fragment to extend over mismatched primers is apparent in these examples. Further experiments designed to address this issue are clearly needed.

**Heterozygote genotyping**

In order to evaluate the sensitivity of the N-GBA method for detection of mutations in heterozygous samples, mixtures of wild type and mutant templates were analyzed on N-GBA arrays. Array profiles of wild type and mutant (missense) synthetic DNA templates mixed in various ratios before hybridization are shown in Figure 3. These results show relative signal strength variation dependent upon the ratio of the mutant and wild type templates. A clear signal in both strands for the mutant base is seen when the mutant template is present at 20% of total template concentration and is detectable at 5% in the + strand profile. Because the level of noise (background signals) varies with different GBA primers, the ability to detect low concentrations of non-wild type template in a background of wild type template is likely to vary from base to base.

Heterozygous mixtures (50:50) of wild type synthetic template DNA and templates containing five base insertions (Fig. 4A) and five base deletions (Fig. 4B) were also analyzed on N-GBA arrays. The loss of downstream signal (caused by insertions and deletions) is masked by the signals from the wild type template, making interpretation of the profiles more difficult. The heterozygous bases are identifiable in each example, allowing positioning of the mutations within the template sequences. The inserted bases (Fig. 4A) lie between the heterozygous signals on the + strand and – strands indicated by the arrows. The deleted bases (Fig. 4B) begin with the heterozygous signal on the + strand and extend to the heterozygous signal on the – strand (five bases total).

**Tumor sample genotyping**

N-GBA profiles of PCR-generated templates from five tumor cell lines are shown in Figure 5. One sample (Fig. 5A) is wild type in codons 272–282, three (Fig. 5B, C and D) have missense mutations in codon 273 and one (Fig. 5E) has a missense mutation in codon 280. All four mutations are clearly identified by an altered peak on both the + and – strand profiles. The general sequence quality for the PCR templates appears similar to the synthetic template data. The homzygous genotypes of the isolated DNA from these samples resulted in no significant wild type background signals.

**DISCUSSION**

**Interpretation of N-GBA solid-phase sequencing results**

Our results demonstrate that even the simplest N-GBA array, one primer per interrogated target position, can identify missense mutations within the template strand. Downstream signal loss due
Figure 4. (A) N-GBA array profiles of 50:50 mixtures of wild type and five base insertion synthetic DNA templates. The insertion test sequence is shown below each profile in red font, with the inserted bases underlined. The position of the heterozygous bases in the profiles are indicated by arrows. (B) N-GBA array profiles of 50:50 mixtures of wild type and five base deletion synthetic DNA templates. The deletion test sequence is shown below each profile in red font, with the position of the deleted bases indicated by spaces.

Figure 5. N-GBA array profiles of PCR products amplified from DNA extracted from cancer cell xenotransplants. (A) DNA from pancreatic cancer showing wild type sequence in this region; (B) DNA from pancreatic cancer showing C→T transition (codon 273); (C) DNA from pancreatic cancer showing G→A transition (codon 273); (D) DNA from pancreatic cancer showing C→G transversion (codon 273); (E) DNA from biliary cancer showing C→G transversion (codon 280).
to a mismatch between the 3′ end of a GBA primer and a missense mutation on the template is recovered by sequencing the opposite strand. The non-wild type signals generated from insertion mutations result from the detection of the first base of the inserted sequence by the GBA primers. Similarly, non-wild type signals generated from deletion mutations result from the detection of the first base on the deletion’s ‘other side’. Both insertions and deletions are generally followed by loss of signal from several downstream bases, presumably due to the disruptive effects of a frameshift on the template hybridization and/or the polymerase extension reaction. Recovery of sequence profile signals downstream from insertion and deletion mutations are seen with as little as four bases of complementarity of the 3′ end of the GBA primer to the template.

When scanning heterozygous templates, loss of downstream signal is masked by the wild type template signals. The sensitivity of this method for detection of mixed templates will vary depending on the background signals associated with each individual GBA primer. Our results show sensitivity in the range of 5–20% for the selected GBA primer. Careful comparison to wild type sequence data and template independent controls should help improve the sensitivity of mutation detection in samples of mixed genotypes.

The variability in signal strength seen in the N-GBA array profiles may be due to both variability in attachment efficiency of the GBA primers to the glass surface and relative hybridization and extension efficiencies of the template strands to each GBA primer. Hybridization experiments indicate that the relatively long (22–25 bases) GBA primers hybridize well to the template strands despite the introduction of single base mismatches (data not shown). This is demonstrated by the rapid recovery of primer extension signal as the mismatch mutation on the template strand moves towards the 5′ end of the GBA primer (Fig. 2B). The loss of signal resulting from point mutations at positions complementary to the 3′ end of the wild type GBA primer is due to interference with the extension reaction, rather than inefficiency in template hybridization, since the 3′ mismatched templates hybridize with equal efficiency under the conditions here.

The more substantial disruption of downstream signal caused by templates with insertion and deletion mutations (Fig. 2C and D) is expected due to the greater destabilization of the hybrid complex, resulting in reduction in polymerase extension efficiency. The subsequent recovery of signal downstream from the five base insertion and deletion templates indicates that the polymerase can extend a relatively unstable primer template complex consisting of as few as four hybridized bases. Experiments with GBA primers containing only five complementary bases to the template show no measurable template hybridization under conditions used in these experiments (data not shown). The extension of short, unstable primer template complexes is probably assisted by the stable hybridization of the template to the 5′ end of the GBA primer and formation of a looped structure at the site of the insertion/deletion.

The addition of the polymerase mediated extension step to solid-phase hybridization adds distinct advantages over SBH for mutation detection. Higher signal to noise ratios result in less dependence on comparison of data to wild type reference signals. Although the complete sequence usually cannot be deduced with insertion mutations when low complexity arrays are used, the location of the mutation, as defined by the boundaries of the non-wild type bases, can be clearly identified. Detection of insertions and deletions within a string of similar bases has been shown to be difficult using SBH methods (3) and may prove easier using GBA’s primer extension biochemistry. In addition, the one primer–one nucleotide approach for mutation scanning significantly simplifies data analysis and interpretation, especially useful in high volume, mutation scanning projects. N-GBA does require that arrayed probes have free 3′ ends for polymerase extension, presenting some restrictions on the use of in situ synthesis to create high density arrays

Our results suggest that the application of N-GBA to mutation scanning using miniaturized oligonucleotide arrays will allow high-resolution mutation detection in genes where repeated sequencing is required in an easy-to-use, inexpensive format. Prior knowledge of the wild type sequence allows for facile array design and precise location of mutations within the targeted region. By genotyping both strands, complete and redundant sequence data can be obtained for both homozygous and heterozygous missense and deletion mutations and positional information can be obtained for insertions.

ACKNOWLEDGEMENTS

We wish to thank Dr Scott Kern for generously supplying tumor DNA samples, Dr Jon Karn for his extensive contributions throughout this project and Tina McIntosh for assistance in preparation of this manuscript.

REFERENCES