Using Molecular Beacons to Detect Single-Nucleotide Polymorphisms with Real-Time PCR

Musa M. Mhlanga*† and Lovisa Malmberg*†

*Department of Cell Biology, New York University School of Medicine, New York, New York; †Molecular Medicine, Department of Medical Sciences, Uppsala University, Uppsala, Sweden.

Detection of single-nucleotide polymorphisms (SNPs) in high-throughput studies promises to be an expanding field of molecular medicine in the near future. Highly specific, simple, and accessible methods are needed to meet the rigorous requirements of single-nucleotide detection needed in pharmacogenomic studies, linkage analysis, and the detection of pathogens. Molecular beacons present such a solution for the high-throughput screening of SNPs in homogenous assays using the polymerase chain reaction (PCR). Molecular beacons are probes that fluoresce on hybridization to their perfectly complementary targets. In recent years they have emerged as a leading genetic analysis tool in a wide range of contexts from quantification of RNA transcripts, to probes on microarrays, to single-nucleotide polymorphism detection. The majority of these methods use PCR to obtain sufficient amounts of sample to analyze. The use of molecular beacons with other amplification schemes has been reliably demonstrated, though PCR remains the method of choice. Here we discuss and present how to design and use molecular beacons to achieve reliable SNP genotyping and allele discrimination in real-time PCR. In addition, we provide a new means of analyzing data outputs from such real-time PCR assays that compensates for differences between sample condition, assay conditions, variations in fluorescent signal, and amplification efficiency. The mechanisms by which molecular beacons are able to have extraordinary specificity are also presented.

Key Words: molecular beacons; real-time polymerase chain reaction; single-nucleotide polymorphism; genotyping; fluorescence resonance energy transfer; spectral genotyping; hybridization; fluorescence.
the need for separation of amplicons from the reaction mixture. Thus PCR confers an inherent specificity non-existent in other widely available techniques and the sensitivity allows analysis of small initial concentrations possible.

Unambiguous detection of the polymorphism and a clear distinction between heterozygotes and homozygotes is achievable only when enough material is present for analysis. At the heart of all nucleic acid-based detection systems is the capacity to assay a conformational change of the nucleic acid probe. This property of high-fidelity hybridization is unique to nucleic acid molecules. Probes using this property are able to seek out their complements among vast numbers of unrelated molecules. During hybridization, the product will form bonds that rank among the strongest in nature. Thus great specificity and ability to form strong macro-molecular associations give hybridization-based probes an as yet unsurpassed advantage in real-time PCR detection of SNPs. When working with a homogenous system, the main approaches for assaying the conformational change are enzymatic probe cleavage, resulting in fluorescence as with TaqMan probes (3), or conformational change of the probe on hybridization, resulting in fluorescence as when using molecular beacons (4). Several similar schemes have been proposed for the detection of specific nucleic acids in homogenous solutions such as “sunrise primers” (5), “scorpion probes” (6), and “cyclicons” (7).

DESCRIPTION OF METHOD

Molecular beacons are single-stranded nucleic acid molecules with a stem-and-loop structure that combine to give molecular beacons their specific properties (Fig. 1). The loop portion is complementary to a predetermined sequence in a target nucleic acid. The stem is formed by the annealing of arm sequences, complementary to each other, that are on either side of the probe sequence. A fluorescent moiety is covalently linked to the end of one arm, and a quenching moiety to the end of the other arm.

Free in solution, the stem keeps the two moieties of the molecular beacon in extremely close proximity to each other, causing the fluorescence of the fluorophore to be quenched by mechanisms more efficient than fluorescence resonance energy transfer (FRET). Hence the emission spectrum of the fluorophore does not have to be matched to the absorption of the quencher, and the need to have FRET pairs as fluorophores and quenchers is thereby obviated. This increases the number of fluorophores available, excepting limitations of the instruments used in the real-time detection of fluorescence with PCR (8).

When hybridizing to their targets, molecular beacons are forced to undergo a conformational change forming a probe–target hybrid that includes more base pairs and thus is more stable than the stem hybrid in the unhybridized molecular beacon. The rigidity and length of the probe–target hybrid precludes the simultaneous existence of the stem hybrid and contributes to the very high specificity of molecular beacons. Consequently, the conformational change causes the fluorophore and the quencher to move away from each other, restoring fluorescence (Fig. 1). Experimental results have shown molecular beacons to be more discriminating than conventional oligonucleotide probes (9).

At temperatures between the dissociation temperatures of perfectly complementary hybrids and mismatched hybrids, perfectly complementary targets can be distinguished from mismatched targets by their higher fluorescence. This is known as the “window of discrimination” (Fig. 2). The presence of the hairpin stem significantly enhances the specificity of molecular beacons in comparison to that of linear probes, such as the TaqMan probe, regardless of the target concentration used, by allowing interaction only with their perfectly complementary targets (8). This enables the detection of SNPs in an unambiguous and reliable manner.

During PCR, molecular beacons undergo a number of changes in their conformation (Fig. 3). At the denaturing step they assume a random coil configuration and are fluorescent due to the denaturing of their stems. At annealing temperatures the stem renatures and fluorescence declines. In the presence of complementary amplicons the molecular beacon is able to bind at the annealing temperature and thus generate fluorescence. At primer extension temperature or during polymerization the molecular beacon is no longer able to bind to its complementary amplicon and dissociates. Consequently, measurements of amplicon abundance are made during each thermal cycle by determining the fluorescence intensity generated by the molecular beacons at the end of the annealing stages.

The ability to perform multiplex reactions with molecular beacons can be used for real-time detection of multiple SNPs in a single sealed tube. Multiplex assays use molecular beacons designed to give fluorescence in several colors. In a duplex assay one color is visible in the presence of the major allele, a second in the presence of the minor allele (Fig. 4). This principle of spectral genotyping has been exploited to detect sequence variation at the single-nucleotide level with templates of
various origins using up to four colors (10–13). A variety of instruments have been developed that are able to read fluorescence and perform thermal cycling simultaneously, making the process of rapid screening in large populations open to automation.

Since molecular beacons are highly specific to their targets, the number of false positives can be significantly reduced. This additional property means that only perfectly complementary targets fluoresce at a given temperature. Hybridization does not occur when the target contains a mismatched nucleotide or deletion. This has made molecular beacons particularly well suited for monitoring the synthesis of specific nucleic acids in real time and the simultaneous detection of point mutations or polymorphisms. This capability has been further exploited for real-time hybridization monitoring in multiplex PCR to spectrally genotype human alleles, and most recently for high-throughput screening of SNPs in association studies (work in progress).

### Wavelength-Shifting Molecular Beacons

It is possible to construct molecular beacons that fluoresce in many colors excepting limitations of the instruments that read the fluorescence. This limitation reduces the variety of fluorophores that can be linked to molecular beacons, thereby limiting multiplex detection of targets. Instruments that can perform PCR while simultaneously monitoring fluorescence in real time use monochromatic light sources such as a laser or light-emitting diodes. This leads to efficient excitation of only a subset of available fluorophores and poor or nonexistent excitation of most others. A general solution to this problem is the wavelength-shifting molecular beacon (14), which can fluoresce in a variety of colors but is excited with monochromatic light sources. Wavelength-shifting molecular beacons absorb light energy with a harvester fluorophore and emit that energy as fluorescent light via an emitter fluorophore within efficient FRET distance of the harvester (Fig. 5). The fluorescent light from the emitter fluorophore has at its own characteristic color. By increasing the variety of fluorophores in a homogenous one-tube assay, throughput, a major demand in rapid SNP genotyping, is increased. Further, due to the monochromatic light source, simpler devices can be used to read fluorescence.

Protocols for the use of wavelength-shifting molecular beacons in real-time SNP detection are similar to those with standard molecular beacons with differences in choice of harvester/emitter pairs (14). Wavelength-shifting molecular beacons are especially well suited for the solution of complex genotyping problems in which more than a single type of target sequence may be encountered in the same reaction, for example, a reaction containing not only the normal or variant allele of an SNP, but also a completely different allele at the same position or a new or unknown polymorphism a few nucleotides apart. These polymorphisms are especially complicated to detect since allele-specific hybridization probes cannot be designed to detect each nucleotide substitution independently. The current SNP database by no means saturates all the SNPs existing at particular loci in the human genome. This sets the stage for the discovery of several unknown variants that have yet to be characterized with allele frequencies still undetermined.

### Design of PCR Primers and Molecular Beacons

Primers for PCRs with molecular beacons are designed to flank the probe sequence, with the molecular beacon sequence recommended to be 20–30 bases 3' or 5' of the forward or reverse primer, respectively. In extenuating circumstances a compromise must be found between the positions of the primer and the molecular beacon to avoid competition for similar binding sites though molecular beacons have been shown to still function. For primer design, the usual constraints and precautions should be followed. Further, the primers should be designed to produce amplicons no more than 250 bp in length (Fig. 6A). Experimental evidence has shown that longer amplicons cause signals to diminish, possibly due to greater difficulty of the molecular beacons in invading the double strands during the annealing step of PCR.

For real-time PCR detection of SNPs, molecular beacons are designed to be complementary to a region of the amplicon where the SNP of interest occurs. The loop region of the molecular beacon is designed first. The loop region is designed to contain between 18 and 25 nucleotides that give a melting temperature slightly above the annealing temperature of the PCR. This causes the probe–target hybrid to be stable during annealing when the signal detection takes place. The amplicon length sequences (usually 5–7 nucleotides) are then designed to allow the stem to dissociate about 7–10°C above the detection temperature, which is usually the annealing temperature of DNA polymerase in PCR. For example, if the annealing temperature of a reaction is 55°C, then the T_m of the loop region should be approximately 57°C and the T_m of the stem region should be between 62 and 65°C.

The melting temperatures of the loop sequence of a molecular beacon can be calculated by a variety of primer design programs that use the "percent GC" rule. These programs are available commercially or online (www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi). When designing the stem of the molecular beacon, a
FIG. 1. Structure and principle of operation of molecular beacons. The loop portion of the molecular beacon hairpin is complementary to a target sequence. The stem is formed by two sequences that are complementary to each other. In the hairpin formation, the fluorophore, attached to one end of the probe, is quenched by the quencher, attached to the other end of the probe, and no fluorescence is emitted. On hybridization of the molecular beacon to its target the stem dissociates, causing the quencher and the fluorophore to move away from each other, resulting in the restoration of fluorescence.

FIG. 2. Fluorescence of an allele-specific molecular beacon as a function of temperature in the absence of target (bottom curve), in the presence of perfectly complementary target (top curve), and in the presence of a target containing a mismatched nucleotide at the same position (middle curve). Due to their higher fluorescence perfectly complementary hybrids can be distinguished from mismatched hybrids at temperatures in the “window of discrimination.”

FIG. 3. Phase transitions of molecular beacons. When annealed, the molecular beacon–target hybrid is fluorescent. As the temperature is raised, the fluorescent probe–target hybrid denatures and a nonfluorescent molecular beacon in a closed formation is formed. On further rise in temperature, the hairpin stem denatures and the molecular beacon forms a fluorescent random coil.
DNA folding program (http://bioinfo.math.rpi.edu/~zukerm) is recommended to estimate the melting temperature of the stem and to ensure that the dominant folding pattern is the intended hairpin. Otherwise, the stem sequence should be redesigned or the loop region should be slid along the target gene keeping the SNP position as near to the center of the probe as possible (Fig. 6B). It is also important to make sure that no annealing is possible between the PCR primers and the molecular beacon. Recently commercially available software packages have become available, such as that from Premier Biosoft, (Palo Alto, CA), that integrate primer design with molecular beacon design, optimizing over a wide range of conditions. Gorilla Genomics (Alameda, CA) will soon begin supplying a large collection of predesigned molecular beacons commercially.

Thermal Denaturing Profiles

To ensure that the designed molecular beacon actually meets experimental requirements, thermal denaturing profiles are recommended. Six tubes each containing 200 nM of molecular beacon, 3.5 mM MgCl₂, and 10 mM Tris–HCl, pH 8.0, in a total volume of 50 µl should be prepared. Six tubes are used if doing a duplex reaction that is, with just two different-colored fluorophores such as FAM and TET. Tube 1 should contain a FAM-labeled molecular beacon with its perfect oligonucleotide complement, Tube 2 should contain the same molecular beacon as Tube 1 but the mismatched oligonucleotide complement. Tubes 3 and 4 should be the same for the TET molecular beacons and Tubes 5 and 6 are simply the FAM and TET molecular beacons alone. The oligonucleotide target should be added to a final concentration of 400 nM. Using a thermal cycler able to monitor the fluorescence in real-time, the temperature should be decreased from 95°C to 25°C holding for 30 s every 1°C. These results are then plotted as fluorescence versus temperature to obtain a “window of discrimination” as seen in Fig. 2. This “window of discrimination” is the annealing temperature range to utilize in PCR to obtain optimal point mutation discrimination.

Concentrations of Reagents

The assay is optimized in the same way as ordinary PCR reactions taking into account MgCl₂ and primer concentrations. It can also be helpful to alter the concentrations of molecular beacons used to reduce fluctuations and maximize the fluorescence signals obtained.

**PROTOCOL**

The protocol given below is used for estrogen receptor gene codon 325 which will be used as an example throughout the remainder of this review.

**DNA Samples**

The starting material for the reaction can be any nucleic acid otherwise used as PCR substrate. In this example DNA extracted from EDTA-treated blood samples is used.

**Primers and Probes**

The PCR primers for the estrogen receptor gene were designed based on their known cDNA sequences (GenBank Accession No. M12674):
forward primer:  
5′-CAG ATG GTC AGT GCC TTG TTG GA-3′

reverse primer:  
5′-CGA AGC TTC ACT GAA GGG TCT TG-3′

Molecular beacons probes were designed following design parameters described in the previous section, additional assistance can be obtained at the Public Health Research Institute, New York (http://www.molecular-beacons.org). Nucleotide bases given in caps represent the stem portion and the bold letter indicates the nucleotide base that corresponds to the SNP position. The fluorophores used were 6-carboxyfluorescein (FAM) and tetrachloro-6-carboxyfluorescein (TET) with 4-(4′-dimethylaminophenylazo)benzoic acid (DABCYL), functioning as the quencher moiety:

MB_C-allele:  
5′-FAM-cca agc GAG CCC CCC
ATA CTC TA gct tgg-DABCYL-3′

MB_G-allele:  
5′-TET-cca agc GAG CCC CCC
ATA CTC TA gct tgg-DABCYL-3′.

Results and Data Analysis

Data output from instruments performing real-time SNP detection shows an increase in fluorescence signal for each PCR cycle if there is a perfect complementary target present in the solution (Fig. 7). The point at which the amount of fluorescence rises to a detectable level is generally termed the threshold cycle (Cₜ). The number of cycles required for fluorescence intensity to reach a detectable level is inversely proportional to the logarithm of the number of template molecules present at the start of the PCR. Results from a multiplex real-time SNP detection assay can be plotted according to Cₜ values for each particular color or wavelength or by plotting the fluorescence values received in the last PCR cycle (Fig. 8A).

An alternative approach is recommended; to determine the ratio (R) between the fluorescence signals from one of the two allele-specific molecular beacons (F_G-allele) and the sum of the fluorescence from both molecular beacons (F_G-allele + F_C-allele), as shown in Eq. [1] (15),

$$ R = \frac{F_{G\text{-allele}}}{F_{G\text{-allele}} + F_{C\text{-allele}}} $$  \hspace{1cm} [1]

When using this approach, differences in amplification efficiency between samples with different DNA concentrations or quality do not affect interpretation or analysis of the results. Ratios fall into three separate groups, each representing one of the three possible genotypes, removing the need for tight clusters (Fig. 8B). It does, however, require a quick visual inspection of standard plots of the data to ensure that PCR has been successful;
FIG. 7. Example of fluorescence signal output from a real-time detection in a thermal cycler. The molecular beacon specific for the major (C) allele is FAM-labeled and the molecular beacon specific for the minor (G) allele is TET-labeled. At the annealing step of each PCR cycle, the fluorescence resulting from hybridization of molecular beacons to newly formed amplicons is measured. The results from running an unknown sample (patient A) are shown together with the results from known controls present in the same run. Patient A is found to be a C homozygote.

FIG. 8. (A) Scatterplot of 94 results when genotyping codon 325 of the estrogen receptor gene. The fluorescence values from the last PCR cycle of two molecular beacons representing the two alleles are plotted. The samples fall into three distinct groups, corresponding to the three possible genotypes. (B) Histogram of 94 ratios obtained when genotyping codon 325 of the estrogen receptor gene (the same samples as in Fig. 8A). Ratios are obtained by dividing the TET (G-allele) fluorescence by the sum of the fluorescence from both TET (G allele) and FAM (C allele). The ratios are plotted in a histogram, resulting in three distinct groups, which correspond to the genotypes of the samples.
i.e., true positives and not very low signals are being observed. This requirement may also be met by programming a number of basal conditions that have to be met before ratio analysis can begin. Using a ratio approach also simplifies statistical analysis of the result and enables comparisons between different PCR runs of the same assay.

CONCLUDING REMARKS

When analyzing large sets of samples the possibility of pooling is advantageous to decrease cost as well as workload. Using molecular beacons it has been shown, in the case of the estrogen receptor gene codon 325 SNP, that it is possible to detect as little as 3.1% of one allele present in a mixed sample of two alleles. In addition, the molecular beacon above displayed a linear relationship between the fluorescence signal ratio (R) and the amount of G allele over the range 100 to 3.1% (15). These results are promising for future applications of molecular beacons to large-scale screening of SNPs using pooled samples.

The focus of this review has been the use of molecular beacons in real-time PCR SNP applications. Besides PCR, several alternative methods exist for the amplification of nucleic acids. These approaches involve target/sample amplification, and have names such as rolling circle amplification (RCA) (16, 17), nucleic acid sequence-based amplification (NASBA) (18), strand displacement amplification (SDA) (19), and ligase chain reaction (LCR) (20). Advantages exist in methods such as rolling circle amplification and NASBA since they are isothermal. Greater distinctions are evident in the overall ability of all of these methods to detect point mutations or polymorphisms.

An ability to examine human genetic variation at its source, in living cells, pushes the limits of most technologies existing today. A burgeoning area of research is the introduction of molecular beacons into cells for the detection of RNA transcripts (21, 22). It may be possible to follow particular SNP variants in vivo and to gain an understanding at the molecular level of the cellular context for genetic diversity sans amplification by means such as PCR.

ACKNOWLEDGMENTS

We most sincerely thank Ann-Christine Syvanen, Fred Russell Kramer, and Sanjay Tyagi for contributing their experience and advice when reading and commenting on this review. The work at the Public Health Research Institute was supported by National Institutes of Health Grants ES-10536 and HL-43521 to F.R.K. and S.T. M.M.M. is supported by National Institutes of Health Grant ST32 GM07239-26. L.M. is supported by the Borgström Foundation.

REFERENCES