Analysis of DNA Single Nucleotide Polymorphisms by Mass Spectrometry

Peter M. Vallone and John M. Butler

National Institute of Standards and Technology, Biotechnology Division, 100 Bureau Drive Mail Stop 8311, Gaithersburg, MD 20899; Email: peter.vallone@nist.gov

Abstract

Single nucleotide polymorphisms (SNP) are the most frequent form of DNA sequence variation in the human genome and are becoming increasingly useful as genetic markers for genome mapping studies, medical diagnostics, and human identity testing. The primer extension reaction is a commonly employed molecular biology assay for probing a known single nucleotide polymorphism site in genomic DNA. In the primer extension assay a short (< 30 bases) DNA oligonucleotide or "primer" is extended by a single nucleotide unit. The identity of the extended base allows the sample to be accurately genotyped.

One method for SNP detection currently employed relies on the mass resolution between a primer and its single base extension product(s) utilizing Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry or MALDI-TOF MS. The speed of data collection by this technique is on the order of 5 sec per sample and has the potential for high throughput when interfaced with a robotic system and automated data analysis. The primer extension assay and variations upon it are described here as well as MALDI-TOF MS data collection and interpretation of primer extension reactions.

Keywords: single nucleotide polymorphism, DNA, mass spectrometry, time-of-flight, matrix-assisted-laser-desorption-ionization, genetic markers, multiplex

Background on SNPs

Single nucleotide polymorphisms (SNPs) or "snips" are the most commonly found DNA sequence variation in the human genome. A SNP is simply a single base pair difference in an otherwise conserved region of genomic DNA. SNPs are thought to be present in approximately every 1000 bases in the human genome (*1,2*). **Figure 1** depicts a simple example of a SNP. The base identity at the highlighted position defines the genotype for this genetic marker. In this example a G to A transition (purine to purine base change) occurs between individual #1 and #2. The letter R can also be used to represent a G to A transition. The International Union of Biochemistry and Molecular Biology (IUBMB) has established symbols for describing base degeneracies. Table 1 shows a list of the IUBMB degenerate base codes along with their meaning.

There are typically 3 SNP possibilities or genotypes for markers found on an autosomal chromosome (a chromosome other than a sex chromosome). In the case of the G/A polymorphism an individual's DNA could contain G on both chromosomes, an A on both, or G and A with one on each chromosome. A case where either a G or A is found exclusively on both chromosomes is referred to as homozygous, and when G and A are both found exclusively on separate chromosomes is referred to as heterozygous. The possible gentotypes for this biallelic system would be homozygous G, homozygous A, and heterozygous G/A.

Various methods exist for the discovery of SNPs. Some of these include single strand conformation polymorphism (SSCP) and denaturing high performance liquid chromatography (DHPLC) (*3,4*). These techniques are useful for inferring that a SNP

may exist in the region of analyzed DNA. The concept of "scanning" for potential SNPs is important but does NOT provide the exact location of the polymorphic site or the sequence identity of the SNP itself. Direct sequencing can also be applied to SNP discovery (*5*), which provides the exact location and sequence identity of the polymorphism. SNP sequence and position are confirmed in a population by sequencing a specific region of interest and comparing differences in multiple subjects confirming the presence and frequency of SNP sites. The frequency of a SNP is considered significant if it is found in greater than 1% of the human population (*6*). Access to public databases that contain redundant sequence information and overlapping genomic clones can also be compared providing an "*in silico*" means of discovering informative SNP markers (*7*). This approach takes advantage of the large amounts of DNA sequence that is already available.

It has been suggested that an appropriate set of SNP markers could be used to identify genetic factors associated with complex disease traits (*8,9*). SNPs that vary across differing populations will also be useful for the purpose of human identification (*10*). As the sequencing of the human genome continues, resulting SNPs are deposited into databases as they are identified. Table 2 contains the Internet addresses for 5 sites containing over 2.5 million human SNPs.

Considering the large number of SNPs present in the human genome the goal of correlating specific sets of SNPs with a particular disease state is a daunting task. Although SNPs are plentiful in the human genome they are not as informative (with only 2 alleles) as other types of less common genetic variations such as short tandem repeats

that exhibit 5 or more alleles. It is therefore necessary to screen a greater number of SNPs markers to discover meaningful differences between individuals (*10*). This genotyping effort will require high throughput screening of possible SNP candidates in order to ascertain disease correlations. Once a set of SNPs is found to be useful for a specific purpose the goal will be to probe healthy versus afflicted individuals for a particular set of markers in a quick and cost effective manner. The focus here is on a basic molecular biology based assay for detection of known SNP sites and how MALDI-TOF MS can serve as quick and inexpensive way to genotype a sample (see also *11*).

Background and Issues for MALDI-TOF MS of DNA

The emergence of matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) for the rapid analysis of short DNA oligomers has made it an attractive method for implementation in nucleic acid based technologies (*12*). The desorption-ionization process utilizes a small organic matrix molecule that absorbs photons of a given energy from a pulsed laser source (commonly a N_2 laser 8= 337.1 nm) to eject DNA oligomers into the gas phase fully intact, primarily with a single charge (*12*). 3-hydroxpicolinic acid (3HPA) is a commonly used matrix for ionizing nucleic acid oligomers (*13*).

The use of MALDI-TOF MS for analyzing the products of primer extension assays has become a focal point for assay development. Primer extension assay products typically consist of a SNP primer and its corresponding extension product(s). The mass difference between the primer and extension products is normally in the range of 273 to 313 Da (**Figure 2**). However resolving the mass difference between multiple extension products

(in heterozygous samples) usually in the range of 9 to 40 Da is a greater challenge. In MALDI-TOF MS there is a decrease in resolution with increasing fragment mass that results in difficulty when examining heterozygous samples. For example with an A/T (IUBMB code W) heterozygote the mass difference between the 2 expected extension products will be 9 Da. It is essential that these peaks are well resolved in order to make an accurate determination of the genotype. For this reason the mass of SNP primers is limited (usually < 30 bases) to allow for the resolution of extension products with low mass differences. The exact mass limit of a SNP primer will vary with improvement in sample preparation and MS instrumentation. A list of mass differences for dideoxynucleotide analogs used in the primer extension assay is shown in Table 3.

Another challenge involves the removal of salts required in the primer extension reaction prior to analysis by MS. Alkali salts such as Na^+ (22 Da) and K^+ (39 Da) bind to the negative phosphodiester backbone of DNA resulting in multiple adduct peaks. The presence of salts reduces the overall sensitivity of the MS technique and can also interfere when examining multiple primer extension products (see *14*). The removal of salts is necessary to increase sensitivity and reduce salt adduct formation when detecting SNPs by MALDI-TOF MS. However, salt removal steps add extra time and cost to performing the primer extension assay. Two different common cases are when A/C (IUBMB M) or G/T (IUBMB code K) polymorphisms are probed. The mass differences between A/C and G/T heterozygotes are 24 and 25 Da respectively. However if the resolution of the mass spectrometer is not adequate and a large amount of Na+ ions are present in the sample it may be difficult to correctly distinguish between a true heterozygote and sodium adduct (22 Da).

Another MS based method of detecting SNPs involves specific hybridization of peptide nucleic acid oligomers (PNAs) to a PCR amplicon (*15,16*). PNAs contain a neutral amide backbone that does not bind salt, which makes them attractive for analysis by MALDI-TOF MS. Typically the assay consists of uniquely mass labeled PNAs that are designed to perfectly hybridize to a biotinylated PCR target. Washing conditions are controlled so that imperfectly matched PNAs are removed from the solution while the perfect matched PNAs are retained. The retained PNAs are then dissociated from the PCR template and analyzed by MALDI-TOF MS. Each uniquely mass-labeled PNA detected represents a SNP allele.

Primer Extension Assay

The primer extension assay is often referred to as the single base extension, PinPoint (*17*) or minisequencing assay (*18*). Variations upon this type of assay are popular for SNP detection, especially in conjunction with analysis by MALDI-TOF MS. Here we will describe the basics of the primer extension assay, examine the results of a primer extension reaction and how MALDI-TOF MS fits as an analysis tool. An outline of the steps required to go from template DNA to a genotype determination are described in **Figure 3**. The primary molecular biology techniques can be broken down into (1) PCR amplification, (2) phosphatase ddNTP digestion, and (3) SNP primer extension.

Components of the Primer Extension Assay

Prior to performing the primer extension reaction the region of DNA containing the SNP site is usually copied using the polymerase chain reaction (PCR). PCR amplifies a region

of the template (genomic) DNA bounded by the PCR primer set selected for a particular system. The size of the PCR product or amplicon can range in size, but is usually around 100-200 base pairs. PCR is followed by a clean up step where excess deoxynucleotides (dNTPs) and PCR primers are removed. The excess dNTPs and PCR primers will interfere with the primer extension reaction and can be removed from the reaction using a type of micro scale chromatography or enzymatic degradation with shrimp alkaline phosphatase for the dNTPs and exonuclease I for the PCR primers). Enzymatic degradation has the advantage of being performed in the same tube in essentially the same buffer as the PCR reaction.

A single primer is designed and synthetically synthesized to probe the SNP site of interest and is referred to as the SNP primer. A SNP extension primer should hybridize specifically at specified temperature conditions to the region directly adjacent to the SNP site. SNP primers should be designed to avoid overlap with positions previously occupied by the PCR primers used in amplification. A preliminary check is performed on the SNP primer sequence to predict the formation of potential inter- or intramolecular structures that would prohibit hybridization to the site of interest. The SNP primer can be designed to bind to either the "top" or "bottom" strand of the PCR amplicon in a 5' to 3' direction. Either designation will work and the primer with lowest tendency to mis-hybridize or form secondary structures is usually selected. The result of the extension reaction (identity of the base added) will differ of course between a "top" versus a "bottom" SNP primer (see **Figure 1**). This fact is taken into account during the interpretation of the data in order to provide a consistent genotype.

Reagents for the primer extension assay include the synthetically prepared SNP primer, dideoxynucleotide triphosphate analogs (ddNTPs), and a thermostable polymerase capable of efficient incorporation of ddNTPs (Thermosequenase or AmpliTaq FS) with the supplied reaction buffer. The dideoxy terminators, which are incorporated into the SNP primer, are shown in **Figure 2**. The basic structure of the 2', 3'-dideoxynucleotide analogs (ddC, ddT, ddA, ddG) with their corresponding masses are given. The absence of a hydroxyl group at the 3' position prevents additional bases from being incorporated onto the primer. Thus, the reaction terminates after the first dideoxynucleotide analog is added. The sequence identity at the SNP site dictates which dideoxynucleotide terminators are incorporated onto the primer. The temperature cycling necessary for the reaction can be performed in a standard PCR thermocycler. Temperature cycling allows for the repeated hybridization and extension of the SNP primers until a significant amount of extension product is produced (enough to be analyzed by MALDI-TOF MS).

A limitation to note for the primer extension assay is when closely spaced multiple SNP sites are present. If a SNP site lies within the region where a SNP primer must bind then a "mismatch" between the primer and template may occur prohibiting effective binding of the primer and a failure to genotype the marker. We will illustrate the basic primer extension assay using an unmodified oligonucleotide primer with unmodified ddNTPs designed to measure a SNP site on the Y chromosome.

Example with Y SNP M96

In **Figure 4** shows a DNA sequence containing the M96 SNP marker found on the Y-chromosome (*19*) with the polymorphic site exhitibitng either a G or a C (IUBMB code S). Given that the M96 marker is found on the single copy Y chromosome only two genotypes should result, either homozygous G or homozygous C. In **Figure 4**, the 2 smaller outer arrows represent the PCR primers used to amplify this region of DNA prior to the primer extension reaction. The SNP primer was designed to probe the polymorphic site via the primer extension assay. The 17-mer DNA oligonucleotide matches the sequence on the "top" strand and thus will hybridize to the complementary "bottom" strand directly adjacent to the SNP site (G/C) as shown in **Figure 4**. The SNP primer sequence with the expected primer extension products of the ddC and ddG terminator are shown in the lower portion of **Figure 4** with their expected masses. The sequence specificity of the primer extension process is the basis for determination of a genotype. The measurement of the mass difference between the SNP primer and the extension products provide the information in order to determine a genotype for a given sample. The experimental conditions for the primer extension assay and thermal cycling conditions have been included for the M96 example. These conditions provide a general guideline for performing the assay.

SNP Primer Extension Assay

Starting with a 10 μL PCR reaction that has been treated with Shrimp Alkaline Phosphatase (SAP) to remove excess dNTPs. A single strand nuclease may be added to remove excess PCR primers. The SNP primer extension reaction can be run in the "leftover" PCR buffer.
Add the following to the specified final concentration in 20μL total volume
SNP primer 1.25μM
ddNTPs 125μM
Thermosequenase 1 Unit
Volume adjustment to 20 μL

SNP Primer Extension Cycling Parameters

94 °C for 1 min
94 °C for 15 s
48 °C for 15 s] 40 cycles
68 °C for 25 s
25 °C until completed

Prior to analyzing the primer extension assay products by MS the salt level of a sample is reduced for optimal results. This sample purification may be performed with solid-phase capture and release after washing (*20*) or by passing the sample over a ZipTip micro chromatography column (*21*). In the example shown here (**Figure 5**), the salts were simply diluted out. One microliter (from the total of 20 μL) of the primer extension reaction was diluted in 9 μL of pure water containing NH_4^+ cation exchange beads. The diluted sample was vortexed for approximately one minute. One microliter of the diluted sample was placed onto the MALDI plate along with 1μL of matrix and allowed to dry. The stock matrix used consisted of 0.7 M 3-hydroxypicolinic acid (Aldrich Chemical, Milwaukee, WI) and 0.07 M ammonium citrate (Sigma Chemical, St. Louis, MO) in 1:1 water and acetonitrile (*22*). Stock matrix solution was diluted 2.5 fold prior to usage. The sample spots were allowed to air dry and the sample plate was placed into the instrument.

Mass spectra were collected on a Bruker BIFLEX III time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a pulsed nitrogen laser (337.1 nm). The following operating voltages were employed for signal collection, IS1 = 19.0 kV, IS2 = 15.0 kV, reflector = 20.0 kV, detector 1.7 kV. Spectra were taken in positive ion

mode with between 100 and 300 laser shots collected on each sample for signal averaging purposes. Laser power attenuation was adjusted with each sample to obtain optimal sensitivity and resolution. Prior to data collection the mass spectrometer was calibrated with 15 and 36 base single strand DNA oligomers, with masses of 4577.1 and 11091.3 Da respectively.

The results shown in **Figure 5** are for the detection of the M96 SNP site (S = G/C) for two different DNA samples A and B (top and bottom panels respectively). In the top panel the two main peaks with masses of 5154 and 5467 Da are the SNP extension primer along with the extension product. The difference in mass between these two peaks is 313 Da, which corresponds to the incorporation of a ddG by the polymerase. The genotype of "Individual A" is thus G for the Y chromosome M96 SNP marker. "Individual B" (**Figure 5**, lower panel) exhibited a mass difference of 273 Da between the primer and extension product and thus has a genotype of C for M96.

Variations Upon the Primer Extension Assay

The basic primer extension assay has been modified in various ways to improve resolution, sensitivity, and multiplexing capabilities. In the following section we will describe a few examples and discuss their similarities and differences. A hypothetical example of a heterozygous individual containing a C/T polymorphism will be used to illustrate each assay. The expected mass spectrum for each of the methods is depicted in **Figure 6**. The details of each assay are further discussed in the given reference articles.

Primer Extension (PinPoint) Assay

The basic primer extension assay discussed in detail previously is also referred to as the PinPoint assay (*17*). This is shown in **Figure 6A**. In the PinPoint assay a primer is designed to hybridize directly adjacent to the SNP site. A mixture of unmodified dideoxynucleotide terminators and polymerase extend the primer by a single base. The resulting reaction products analyzed by MS are the extension primer and the singly extended primers (+ddC and + ddT). The mass peak at 5828 Da represents the SNP primer while peaks at 6101 and 6116 Da are the ddC and ddT extension products respectively.

MassTag-labeled primers

The MS spectrum shown in **Figure 6B** is similar to the PinPoint assay except that the SNP primer contains "n" (where n = 1 to 20) non-complementary bases attached to the 5'-end of the primer, usually poly T tails (*23*). The purpose of the tail is to alter the mass range of the primer and extension products. Mass shifting can be useful when attempting to analyze multiple primer extension products by MALDI-TOF MS, commonly referred to as multiplexing. One can simply add a "tail" consisting of natural nucleotides to the SNP primer instead of focusing on designing SNP primers with appropriate masses that will hybridize under the same conditions (*23*). The MS schematic shows that the mass-tagged primers are shifted up to higher mass values. This allows "room" for another SNP primer to be used in the assay.

Mass-tagged ddNTPs

This assay is a variation upon the PinPoint that utilizes a mass-tagged ddNTP (*24*). Commercially available fluorescently labeled ddNTPs commonly used for DNA

sequencing can serve as the mass-tagged ddNTP. Mass-tagging is useful when detecting heterozygous SNPs in which the masses of the extension products may be difficult to resolve from one another. In the case shown in **Figure 6C** the mass difference between the two extension products is greater due to the mass tagging of ddT. This variation on the assay would be especially useful for distinguishing A/T heterozygotes whose extended primer masses will differ by only 9 Da without a mass-tag. The MS plot shows how the masses of the +ddC and mass-tagged +ddT extension products are further separated from each other.

PROBE Assay

In the PRimer Oligo-Base Extension (PROBE) assay a mixture of dideoxy- and deoxynucleotides are used in the primer extension reaction (*25*). In **Figure 6D** dideoxycytosine along with the other three deoxynucleotides are incorporated by the polymerase. The result of this mixture of triphosphates is a SNP primer that is extended by one base (+ddC) for the C allele and 4 bases for the T allele. On the T allele top strand, the extension occurs until the next ddC is present in the template (in our hypothetical example four bases +dTdGdAddC). The MS example shown in **Figure 6D** depicts how the T SNP allele is shifted in mass by the incorporation of dT-dG-dA-ddC. The PROBE assay is similar to mass-tagging of ddNTPs in that it allows better resolution of heterozygous alleles from one another. The length of the extension products will vary depending on the sequence following the SNP site and which ddNTPs are present in the reaction, which may add complexity when designing multiplex SNP assays. The SNP primer and SNP sites in a multiplex system will have to be carefully selected for optimal performance. Another variation on the PROBE assay called VSET for Very Short

ExTension has been described in which 3 ddNTPs and one dNTP are used in the extension reaction (*26*).

Cleavable Base Assay

The cleavable base assay is a variation on the primer extension assay that results in extension fragments of a lower mass (*20,27*). This allows for improved resolution and sensitivity. Changing the position of the cleavable base within the SNP primer simplifies multiplex assay design due to the ability to control the mass of the DNA fragments (*20*). The cleavable primer extension assay utilizes a SNP primer containing a biotin moiety at the 5'-end and a chemically cleavable nucleotide near the 3'-end. The biotin group allows solid-phase capture by magnetic beads coated with streptavidin. The SNP primer is bound to the bead while the unwanted components of the PCR and extension assay (salt, ddNTPs, enzyme) are washed off. A 3' portion of the SNP primer is then released from the solid-phase via chemical cleavage at the cleavable base. The shortened fragments of DNA are then spotted in matrix and analyzed in the mass spectrometer. **Figure 6E** illustrates the reduction in mass with a chemically cleaved SNP primer. The use of cleavable primers benefits multiplexing and leads to increased sensitivity due to the analysis of lower mass DNA products.

GOOD Assay

The GOOD assay is similar to the cleavage assay in the respect that it analyzes shorter length DNA products (*28,29*). The GOOD assay also benefits from the use of a charge tagged primer (*30*) and an alkylation step that improves sensitivity for MS analysis (*31*). Replacing 3HPA in the GOOD assay is a peptide matrix (α-cyano-4-hydroxy-cinnamic

acid methyl ester) that promotes homogeneous crystallization and elimination of sweetspots. The use of the peptide matrix in conjunction with alkylated DNA products provides an increase in sensitivity for MALDI-TOF MS detection. The reagents for the GOOD assay differ somewhat from those common to other primer extension assays. The SNP primer contains a modified base with a linker for coupling a positive charge tag to the primer. There are also phosphorothioate linkages near the 3'-end of the primer. These backbone linkages are relatively resistant to phosphodiesterase (PDE) degradation. PDE digestion shortens the length of the primer by degrading the 5'-end of the primer until the first phosphorothioate linkage is reached. Standard ddNTPs are replaced with thio derivative α-S-ddNTPs, which can be alkylated. The MS spectrum in **Figure 6F** depicts a PDE digested SNP primer with a mass of 885 Da and the corresponding extension products. By shifting the assay to a lower mass, heterozygous alleles are more easily resolved and assay sensitivity is improved. Some limited multiplexes have also been performed with the GOOD assay (*29*).

Issues for Multiplex SNP analysis by MALDI-TOF MS

Due to the large number of SNPs present in the human genome it is evident that there exists a need to quickly genotype many samples at multiple marker locations. The examples described here so far have been for a single marker or singleplex reaction. Analyzing multiple SNP markers in the same assay increases genotyping throughput. One could reasonably imagine probing 2, 3, 4 or more SNP sites by the primer extension assay and then analyzing the products by MS. However there are many issues inherent with primer extension assays and analysis by MALDI-TOF MS that make multiplexing SNP assays more challenging than singleplex reactions.

For example, imagine that 5 SNP markers were to be probed in a single reaction. A 5-plex PCR reaction must first be performed if the SNPs are not all within the same region. This is not trivial. PCR can be difficult and time consuming in order to optimize the amplification of 5 separate PCR products in a single tube reaction. Careful selection of PCR primer pairs is essential and multiplex PCR assays need to be optimized often through labor-intensive efforts. Prior to performing the primer extension reaction 5 compatible SNP primers need to be designed. These 5 primers should be able to hybridize efficiently at the same annealing temperature and primer extension conditions. Primers must not be involved in secondary binding reactions with each other or other regions of the DNA templates. The masses of the primer and extension product(s) cannot overlap. Overlapping or degenerate masses will complicate genotyping and in certain cases may even make it impossible. The presence of doubly and even triply charged ions of the higher mass SNP primers and extension product(s) must also be taken into account for interference with the expected masses of smaller primers. In **Figure 7** the peaks at 2577 and 2734 Da are the doubly charged species of the primer and extension product respectively. In the case of the singleplex reaction the doubly charged ions do not add to the complexity of interpreting the spectrum. However, if other primer sets were being extended and analyzed simultaneously multiple charged species can complicate interpretation of the data. There is only a limited amount of mass "space" available for multiplexing. It is difficult to optimize MALDI-TOF MS operating voltages and laser power so that ions of low and high mass are detected with similar efficiencies. A problematic characteristic of the DNA matrix 3HPA is a tendency to form "sweet spots" during crystal formation. Non-homogeneous crystal formation results in various

localized areas in the crystal where the DNA resides. Localization of the DNA in the crystal can cause difficulties for automated data collection schemes because "searching" for the optimal DNA signal in the crystal increase the time involved in data collections. The optimization of all experimental variables for a robust multiplex SNP assay can be aided by computer software and predictive algorithms but empirical tuning of the assays is invariably required. Even with all of these challenges successful multiplexes have been demonstrated such as a 12-plex (*21*), a10-plex (*32*), and a 17-plex (*33*).

Methods for SNP analysis

While this article focuses on SNP detection by means of mass spectrometry, there are other methods and assays for the detection of known polymorphisms. Table 4 summarizes various assays and technologies for SNP genotyping. The primer extension assay, described in this article, is effectively used with other technologies other than mass spectrometry including capillary electrophoresis and fluorescence polarization.

One obstacle in terms of widespread acceptance of mass spectrometry is that is not a traditional tool in the average molecular biology laboratory. Gel and capillary electrophoresis techniques are very well established in molecular biology labs so it is a natural progression to couple assays with these detection techniques. Plus there is a considerable cost associated with the average MALDI-TOF system and an experienced operator is usually required for a high throughput undertaking. The use of MALDI-TOF MS as a technique for the analysis of SNPs can be addressed in detail by considering the possibilities for high-throughput SNP genotyping.

High-throughput SNP genotyping

At this time primer extension coupled with MALDI-TOF MS is primarily useful for many samples, but only a few markers per run. This is in contrast to DNA chip technologies that are useful for many markers but fewer samples. MALDI-TOF MS has the ability to detect signal from primer extension products in a matter of seconds, enabling a high volume of sample processing. However, the limitation lies in the number of samples that can be analyzed in a parallel fashion. It may take a technique such as capillary electrophoresis 20-30 minutes to make a similar measurement, but with the use of different dyes and efficient primer design multiple measurements can be made per run. Such high throughput issues are discussed in greater detail in another chapter of this volume entitled "High-throughput genotyping of short tandem repeat DNA markers with time-of-flight mass spectrometry" by John M. Butler. The use of MALDI-TOF MS for SNP detection has inherent advantages such as rapid signal detection, amenability to automation, and the robust nature of the measurement. Continuing improvement in multiplexing capabilities for the primer extension/MALDI-TOF MS system will be essential to allow increasing numbers of SNP samples typed per instrument run.

References

(1) Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S., and Schmidtke, J. (1985) An estimate of unique DNA sequence heterozygosity in the human genome. *Hum Genet* 69:201-205

(2) Kwok, P. Y., Deng, Q., Zakeri, H., Taylor, S. L., and Nickerson, D. A. (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* 31:123-126

(3) Pushnova, E. A., Akhmedova, S. N., Shevtsov, S. P., and Schwartz, E. I. (1994) A rapid and simple DNA fingerprinting method using RFLP and SSCP analysis of the hypervariable noncoding region of human mitochondrial DNA. *Hum Mutat* 3:292-296

(4) Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L., and Oefner, P. J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996-1005

(5) Kwok, P. Y., Carlson, C., Yager, T. D., Ankener, W., and Nickerson, D. A. (1994) Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* 23:138-144

(6) Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lander, E.S. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-1082

(7) Gu, Z., Hillier, L., and Kwok, P. Y. (1998) Single nucleotide polymorphism hunting in cyberspace. *Hum Mutat* 12:221-225

(8) Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science*. 273:1516-1517

(9) Collins, F. S., Guyer, M. S., and Charkravarti, A. (1997) Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1580-1581

(10) Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y., Budowle, B. (1999) The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* 20:1682-1696

(11) Griffin, T.J., Smith, L.M. (2000) Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol* 18:77-84

(12) Fitzgerald, M. C. and Smith, L. M. (1995) Mass spectrometry of nucleic acids: the promise of matrix-assisted laser desorption-ionization (MALDI) mass spectrometry. *Annu Rev Biophys Biomol Struct* 24:117-140

(13) Wu, K.J., Steding, A., Becker, C.H. (1993) Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun Mass Spectrom* 7:142-146

(14) Shaler, T.A., Wickham, J.N., Sannes, K.A., Wu, K.J., Becker, C.H. (1996) Effect of impurities on the matrix-assisted laser desorption mass spectra of single-stranded oligonucleotides. *Anal Chem* 68:576-579

(15) Griffin, T.J., Tang, W., Smith, L.M. (1997) Genetic analysis by peptide nucleic acid affinity MALDI-TOF mass spectrometry. *Nature Biotech* 15:1368-1372

(16) Ross, P.L., Lee, K., Belgrader, P. (1997) Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Anal Chem* 69:4197-4202

(17) Haff, L.A. and Smirnov, I.P. (1997) Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res 7*:378-388

(18) Tully, G., Sullivan, K.M., Nixon, P., Stones, R.E., Gill, P. (1996) Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics* 34:107-113

(19) Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, R.W., Feldman, M.W., Cavalli-Sforza, L.L., Oefner, P.J. (2000) Y chromosome sequence variation and the history of human populations. *Nature Genetics* 26:358-361

(20) Li, J., Butler, J.M., Tan, Y., Lin, H., Royer, S., Ohler, L., Shaler, T.A., Hunter, J.M., Pollart, D.J., Monforte, J.A., Becker, C.H. (1999) Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* 20:1258-1265

(21) Ross, P., Hall, L., Smirnov, I., Haff, L. (1998) High level multiplex genotyping of MALDI-TOF mass spectrometry. *Nature Biotech* 16:1347-1351

(22) Chiu, N. H., Tang, K., Yip, P., Braun, A., Koster, H., and Cantor, C. R. (2000) Mass spectrometry of single-stranded restriction fragments captured by an undigested complementary sequence. *Nucleic Acids Res* 28: e31

(23) Haff, L.A. and Smirnov, I.P. (1997) Multiplex genotyping of PCR products with MassTag-labeled primers. *Nucleic Acids Res 25*:3749-3750

(24) Fei, Z., Ono, T., Smith, L.M. (1998) MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res* 26:2827-2828

(25) Braun, A., Little, D.P., Koster, H. (1997) Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clinical Chem* 43:1151-1158

(26) Sun, X., Ding, H., Hung, K., Guo, B. (2000) A new MALDI-TOF based mini-sequencing assay for genotyping of SNPs. *Nucleic Acids Res* 28: e68

(27) Monforte, J.A., Becker, C.H., Shaler, T.A., Pollart, D.J. (1997) Oligonucleotide sizing using immobilized cleavable primers. US Patent 5,700,642

(28) Sauer, S., Lechner, D., Berlin, K., Lehrach, H., Escary, J.L., Fox, N., Gut, I.G. (2000) A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res* 28: e13

(29) Sauer, S., Lechner, D., Berlin, K., Plancon, C., Heuermann, A., Lehrach, H., and Gut, I. G. (2000) Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res* 28: e100

(30) Gut, I.G., Jeffery, W.A., Pappin, D.J.C., Beck, S. (1997) Analysis of DNA by 'Charge Tagging' and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 11:43-50

(31) Gut, I. G. and Beck, S. (1995) A procedure for selective DNA alkylation and detection by mass spectrometry. *Nucleic Acids Res* 23:1367-1373

(32) Butler, J.M. and Becker, C.H. (2001) *Improved analysis of DNA short tandem repeats with time-of-flight mass spectrometry*. Washington, DC: National Institute of Justice. Available on-line at: http://www.ojp.usdoj.gov/nij/pubs-sum/188292.htm

(33) Mengel-Jorgensen, J., Sanchez, J.J., Borsting, C., Kirpekar, F., Morling, N. (2004) MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 Y-chromosome single-nucleotide polymorphisms. *Anal. Chem.* 76: 6039-6045.

(34) Nikiforov, T.T., Rendle, R.B., Goelet, P., Rogers, Y.H., Kotewicz, M.L., Anderson, S., Trainor, G.L., Knapp, M.R. (1994) Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res* 22:4167-4175

(35) Chen, X., Levine, L., Kwok, P.Y. (1999) Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res* 9:492-498

(36) Ahmadian, A., Gharizadeh, B., Gustafsson, A.C., Sterky, F., Nyren, P., Uhlen, M., Lundeberg, J. (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* 280:103-110

(37) Sapolsky, R.J., Hsie, L., Berno, A., Ghandour, G., Mittmann, M., Fan, J.B. (1999) High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet Anal* 14:187-192

(38) Gilles, P.N., Wu, D.J., Foster, C.B., Dillon, P.J., Chanock, S.J. (1999) Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nature Biotech* 17:365-370

(39) Livak, K.J. (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* 14:143-149

(40) Lee, L.G., Livak, K.J., Mullah, B., Graham, R.J., Vinayak, R.S., Woudenberg, T.M. (1999) Seven-color, homogeneous detection of six PCR products. *Biotechniques* 27:342-349

(41) Giesendorf, B.A., Vet, J.A., Tyagi, S., Mensink, E.J., Trijbels, F.J., Blom, H.J. (1998) Molecular beacons: a new approach for semiautomated mutation analysis. *Clinical Chem* 44:482-486

(42) Howell, W.M., Jobs, M., Gyllensten, U., Brookes, A.J. (1999) Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nature Biotech* 17:87-88

(43) Delahunty, C., Ankener, W., Deng, Q., Eng, J., Nickerson, D.A. (1996) Testing the feasibility of DNA typing for human identification by PCR and an oligonucleotide ligation assay. *Am J Hum Genet* 58:1239-1246

(44) Germer, S., Higuchi, R. (1999) Single-tube genotyping without oligonucleotide probes. *Genome Res* 9:72-78

(45) Butler, J.M., Wilson, M.R., Reeder, D.J. (1998) Rapid mitochondrial DNA typing using restriction enzyme digestion of polymerase chain reaction amplicons followed by capillary electrophoresis separation with laser-induced fluorescence detection. *Electrophoresis* 19:119-124

(46) Medintz, I., Wong, W.W., Sensabaugh, G., Mathies, R.A. (2000) High speed single nucleotide polymorphism typing of a hereditary haemochromatosis mutation with capillary array electrophoresis microplates. *Electrophoresis* 21:2352-2358

(47) Schmalzing, D., Belenky, A., Novotny, M.A., Koutny, L., Salas-Solano, O., El-Difrawy, S., Adourian, A., Matsudaira, P., Ehrlich, D. (2000) Microchip electrophoresis: a method for high-speed SNP detection. *Nucleic Acids Res* 28: e43

(48) Lyamichev, V., Mast, A.L., Hall, J.G., Prudent, J.R., Kaiser, M.W., Takova, T., Kwiatkowski, R.W., Sander, T.J., de Arruda, M., Arco, D.A., Neri, B.P., Brow, M.A. (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotech* 17:292-296

(49) Mein, C.A., Barratt, B.J., Dunn, M.G., Siegmund, T., Smith, A.N., Esposito, L., Nutland, S., Stevens, H.E., Wilson, A.J., Phillips, M.S., Jarvis, N., Law, S., de Arruda, M., Todd, J.A. (2000) Evaluation of

single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res* 10:330-343

Acknowledgments

This work was supported in part by grant number 1999-IJ-R-A094, which is an interagency agreement between the National Institute of Justice and the NIST Office of Law Enforcement Standards, awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Table 1. The International Union of Biochemists and Molecular Biologists (IUBMB) has designated single letter codes to represent the multiple nucleotide combinations listed here. The biallelic SNPs are highlighted in **bold**.

IUBMB Code	Bases
N	A,C,G,T
V	G,A,C
B	G,T,C
H	A,T,C
D	G,A,T
K	G,T
S	G,C
W	A,T
M	A,C
Y	C,T
R	A,G

Table 2. Internet-accessible public databases containing SNP information.

Database	Web Address
NCBI SNP Database	http://www.ncbi.nlm.nih.gov/SNP/
The SNP Consortium	http://snp.cshl.org/
NCI CGAP-GAI	http://lpg.nci.nih.gov/GAI/
HGBASE	http://hgvbase.cgb.ki.se/
Whitehead Institute Human SNP Database	http://www.broad.mit.edu/snp/human/

Table 3. Tabulation of mass differences between dideoxynucleotide analogs (ddC, ddT, ddA, ddG). The differences listed are due to the mass difference between the nitrogenous bases (AGCT). These mass differences provide a guideline for the required resolution in order to detect heterozygous samples unambiguously.

Δ mass	273 ddC	288 ddT	297 ddA	313 ddG
ddC		+15	+24	+40
ddT	-15		+9	+25
ddA	-24	-9		+16
ddG	-40	-25	-16	

Table 4. Various methods used for SNP genotyping. The methods can be grouped into primer extension, hybridization, allele-specific, and enzymatic digestion assays.

	Method	*Description*	**References**
Primer extension assays	Mass spectrometry	Primer extension across the SNP site with ddNTPs; mass difference between the primer and extension product is measured to reveal nucleotide(s) present	(*17*)
	Direct sequencing	PCR products are sequenced and compared to reveal SNP sites	(*5*)
	Genetic Bit Analysis	Primer extension with ddNTPs is detected with a colorimetric assay in a 96-well format	(*34*)
	Minisequencing	Primer extension with fluorescent ddNTPs followed by electrophoretic separation; primer size variation permits multiplex detection	(*18*)
	Fluorescence polarization	Primer extension across the SNP site with dye-labeled ddNTPs; monitoring changes in fluorescence polarization reveals which dye is bound to the primer	(*35*)
	Pyrosequencing	Primer extension with luminometric detection of pyrophosphate released upon nucleotide incorporation	(*36*)
Hybridization	High density arrays (Affymetrix chip)	Thousands of oligonucleotide probes are represented at specific locations on a microchip array; fluorescently-labeled PCR products hybridize to complementary probes to reveal SNPs	(*6*,*37*)
	Electronic dot blot (Nanogen chip)	Potential SNP alleles are placed at discrete locations on a microchip array; an electric field at each point in the array is used to control hybridization stringency	(*38*)
	TaqMan 5' nuclease assay	A fluorescent probe consisting of reporter and quencher dyes is added to a PCR reaction; amplification of a probe-specific product causes cleavage of the probe and generates an increase in fluorescence	(*39,40*)
	Molecular beacons	Hairpin stem on oligonucleotide probe keeps fluorophore and its quencher in contact until hybridization to DNA target, which results in fluorescence	(*41*)
	Dynamic allele specific hybridization	Oligonucleotide probe is added along with a double-strand specific intercalating dye; sample is heated above probe denaturing temperature while monitoring fluorescence	(*42*)
Allele-specific	Oligonucleotide ligation assay (OLA)	Colorimetric assay in microtiter 96-well format involving ligation of two probes if the complementary base is present	(*43*)
	T_m-shift genotyping	Allelic-specific PCR is performed with a GC-tail attached to one of the forward allele-specific primers; amplified allele with GC-tailed primer will exhibit a melting curve at a higher temperature	(*44*)
Enzymatic	PCR-RFLP with CE	Variation in restriction sites result in different sized DNA fragments that can be distinguished when separated by capillary electrophoresis or microchip CE	(*45,46*)
	Enzyme mismatch cleavage	Two DNA samples are mixed; a mismatch repair enzyme cuts heteroduplex at SNP site; size-based separation used	(*47*)
	Invader assay	Flap endonuclease cleaves overlapping oligonucleotide probes that either fully match or contain a single mismatch at the SNP site	(*48,49*)

Figure 1. Example of a single nucleotide polymorphism (SNP) occurring in two different individuals in an otherwise conserved region of DNA. This particular SNP, shown in red, is a G-to-A transition. Note that the complementary strands, shown in blue, also exhibit a transition (C-to-T).

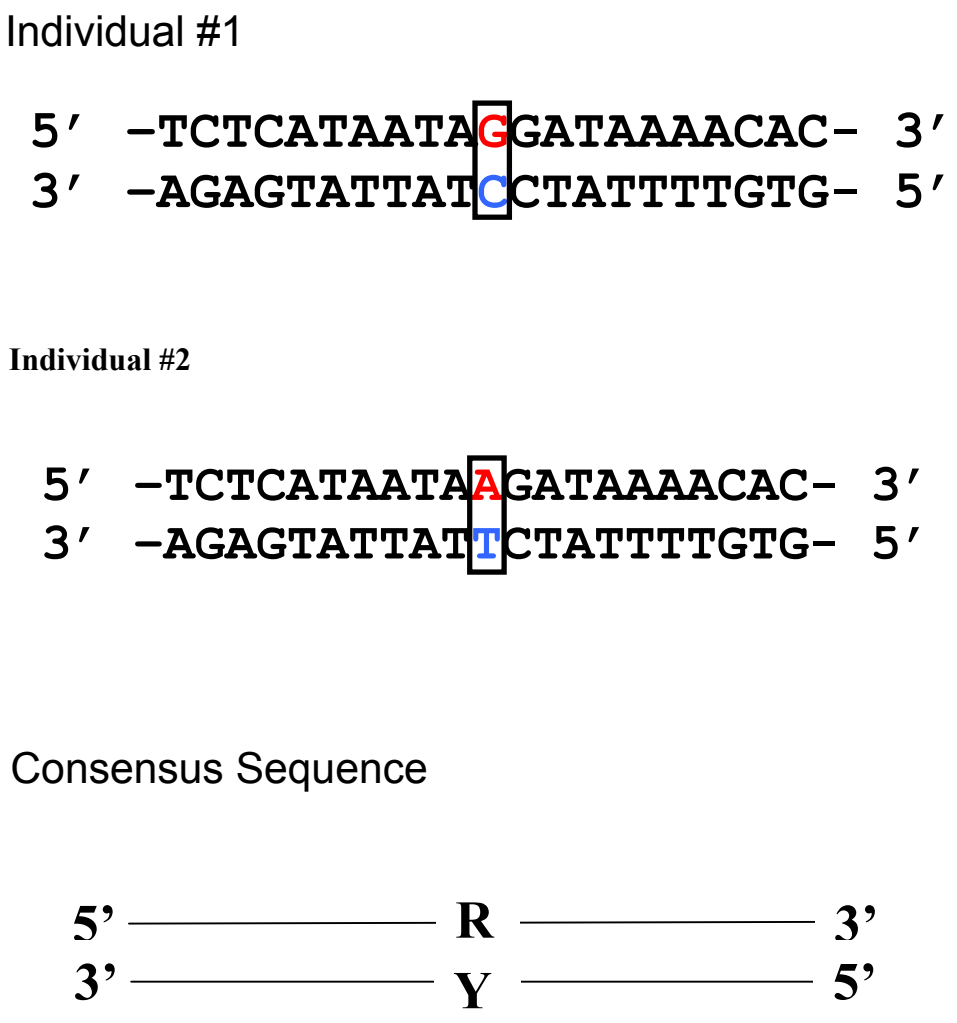


Figure 2. Structure of a 2',3'–dideoxynucleotide. The substitution of a hydrogen atom in place of a hydroxyl group at the 3' position (boxed region) of the ribose ring prohibits the addition of subsequent nucleotides and thus the termination of a primer extension reaction. The corresponding masses for the four dideoxynucleotide monophosphates are also listed.

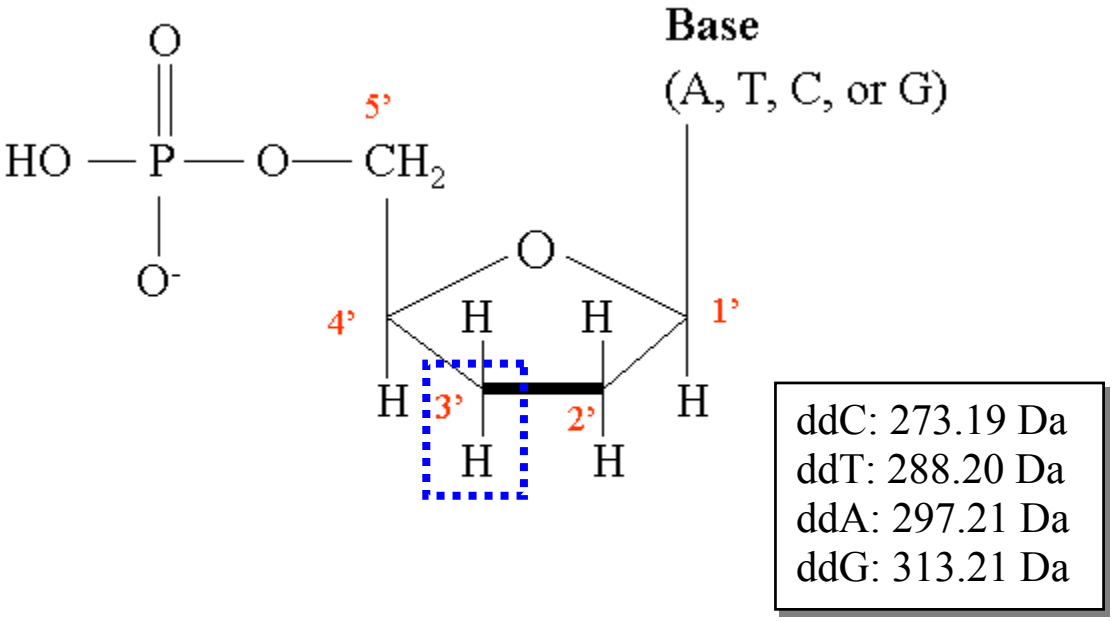


Figure 3. Summary of steps involved in a generic primer extension assay.

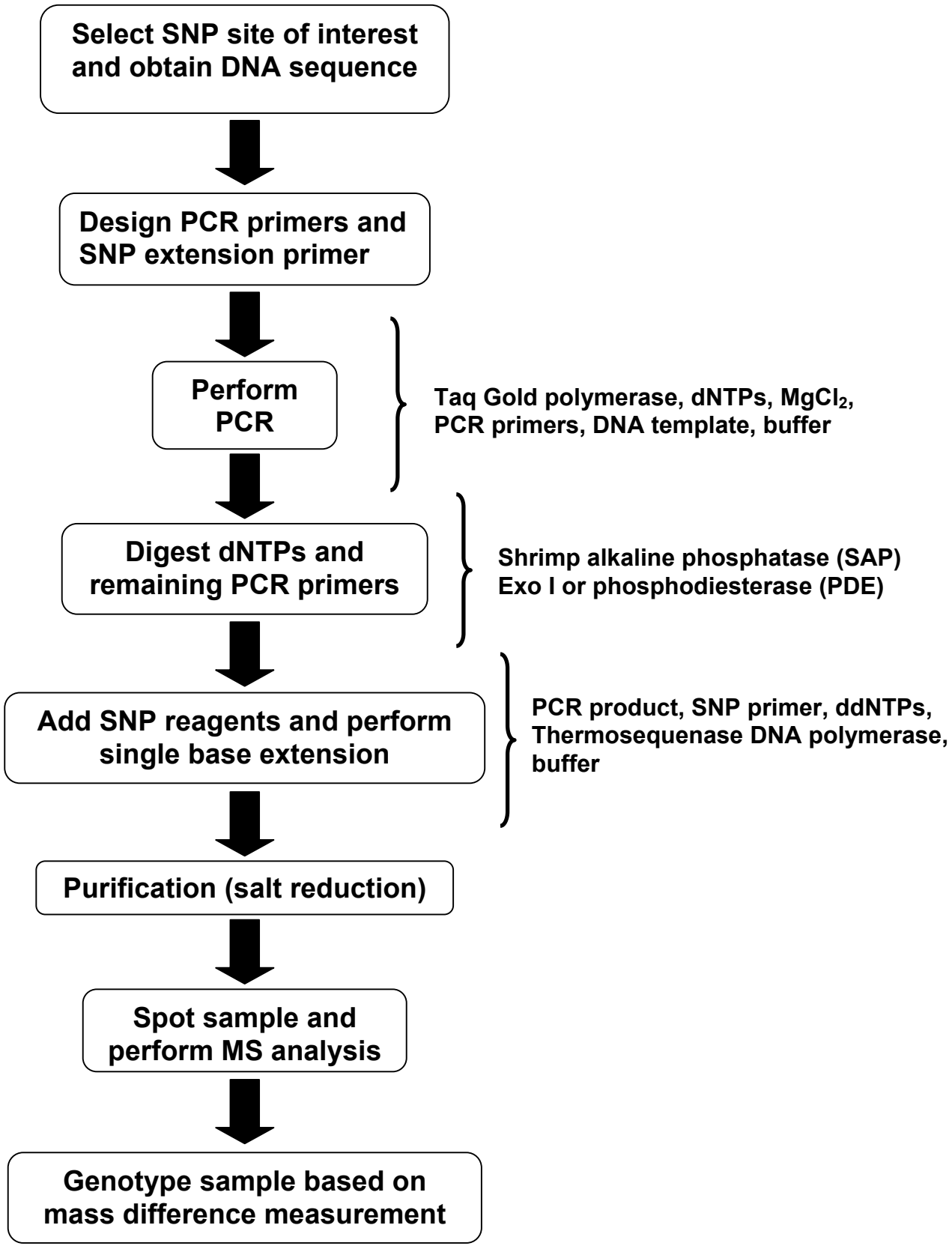
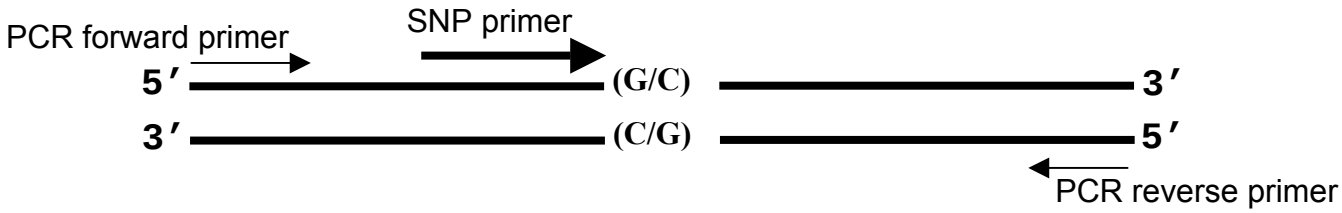


Figure 4. Schematic of the M96 SNP marker and SNP primer. Shown is a section of the PCR amplicon for the region of DNA containing the M96 SNP (S). A primer for the primer extension assay is shown which is designed to hybridize to the "bottom" strand of the amplicon. The possible products of the primer extension assay are shown below with their corresponding masses. SNP primer was designed to have an annealing temperature of 58.6 ^{o}C



Possible extension products

5′-ACAGGTCTCTCATAATAddC-3′	5154 + 273
5′-ACAGGTCTCTCATAATAddG-3′	5154 + 313

Figure 5. Single base extension mass spectra from two different individuals exhibiting different genotypes at the M96 marker. The SNP detection primer used in this assay has the sequence 5'-**ACAGGTCTCTCATAATA-**3' with an expected mass of 5154 Da (see Figure 4). The individual in the top panel has a G at the SNP site based on a mass difference of 313 Da between the primer and extension product. The sample in the lower panel exhibited a genotype of C at the same site because the primer and extension product differ by 273 Da. These results were collected on a Bruker BIFLEX III time-of-flight mass spectrometer using experimental conditions described in the text. Mass spectra were smoothed and enhanced by the application of matrix convolution filters contained in the XMASS 5.0 analysis software package (Bruker Daltonics, Germany).

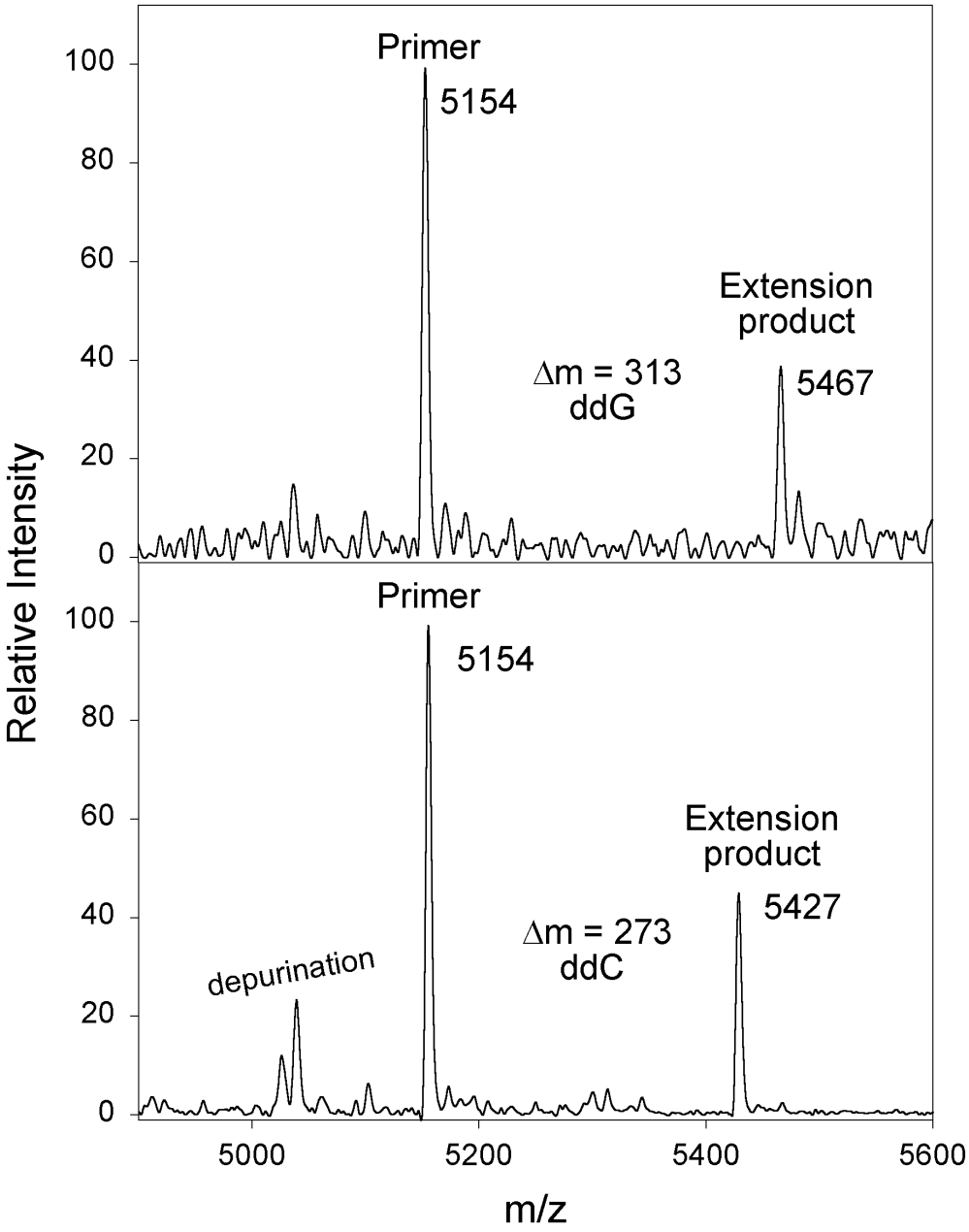
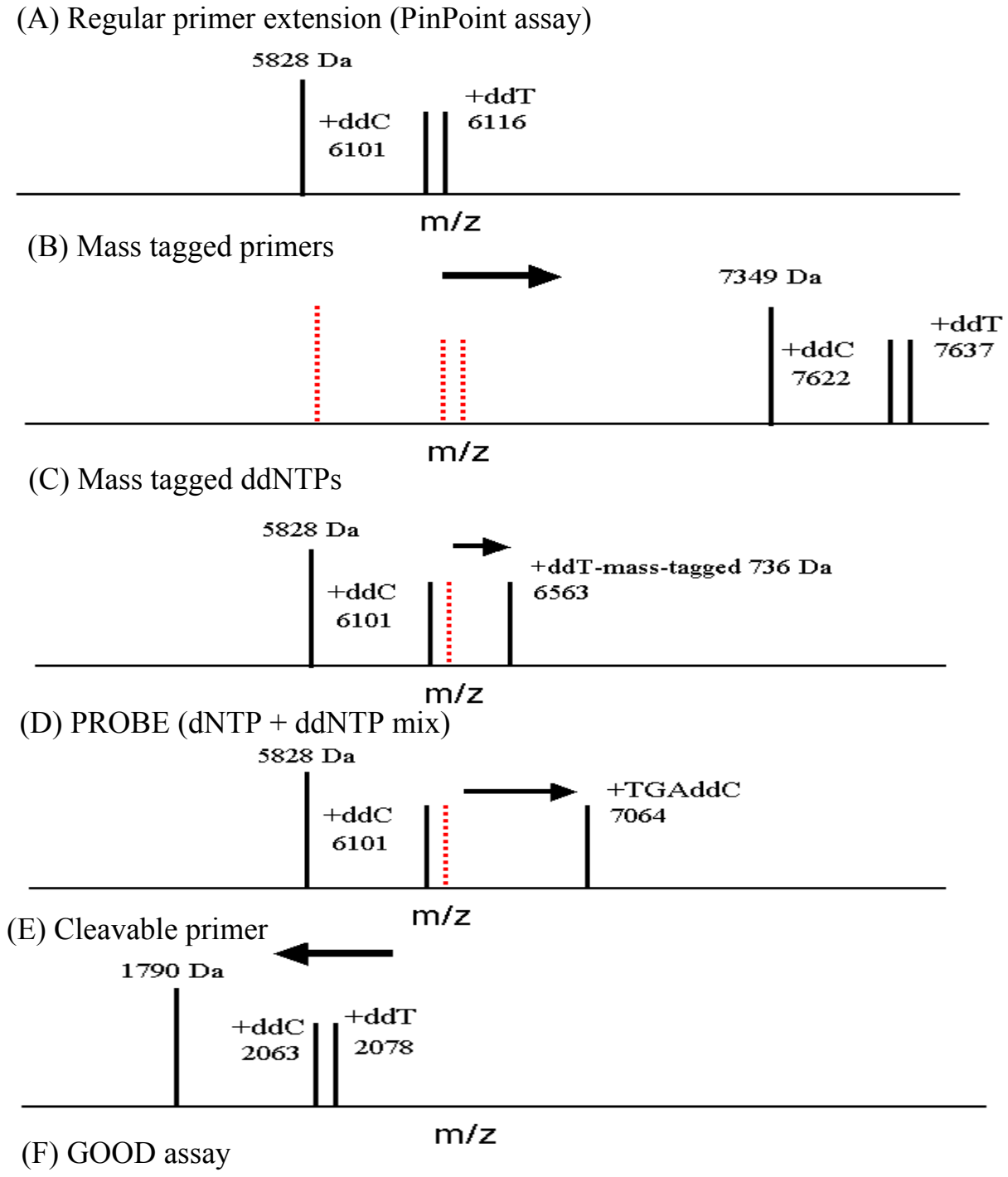


Figure 6. Mass spectra illustrating the results from variations on the primer extension assay. A hypothetical example of a C/T SNP marker from a heterozygous individual is used to demonstrate the effect on mass spectral data obtain from different assays. Modifications upon the basic primer extension assay enable improved resolution of heterozygotes, increased multiplexing potential, and greater sensitivity.




885 Da
1158 Da
+α-S-ddC
1173 Da
+α-S-ddT
m/z

Figure 7. Ions impacting multiplex SNP assay design. Doubly-charged and triply-charged ions from higher mass primers and extension products can interfere with singly-charged ions of smaller mass primers if the multiplex primer set is not well designed. Primer synthesis impurities, shown with arrow, also impact how closely together primers can be spaced. In the plot below doubly-charged ions at 2577 and 2734 Da do not interfere with determining the correct genotype in this singleplex example. However, if additional markers of a higher mass range (around 10,300 Da or so) were to be detected using primers then doubly-charged ions from the higher mass primer could obscure the signal at 5154 – 5467.