



Denaturing HPLC of Nucleic Acids

Andreas Premstaller and Peter J. Oefner,

Genome Technology Center, Stanford University, Palo Alto, California, USA.

Elucidation of the genetic factors contributing to natural and disease phenotypes will require the resequencing of parts or entire genomes thousands of times. Because most genomes are highly conserved, conventional sequencing is too expensive and laborious for merely confirming sequence identity. Screening methods that only establish the presence of a mutation but provide incomplete information about the nature and location of the variation are an inexpensive but, in most instances, insufficiently sensitive alternative. Denaturing high performance liquid chromatography has changed that and new developments promise to put this technology on the same instrumental footing in terms of throughput and information content as capillary array sequencers.

Introduction

With an increasing number of genome sequences completed, denaturing high performance liquid chromatography (DHPLC) has established itself as one of the most powerful tools for DNA variation screening and allele discrimination (<http://insertion.stanford.edu/pub.html>). Aside from its obvious application in identifying mutations in human diseases,¹ it has provided unprecedented insight into human evolution and prehistoric migrations based on the screening of Y chromosomal and autosomal DNA sequences.^{2,3} It has been applied successfully to the mapping and cloning of genes in yeast,⁴ *arabidopsis thaliana*,⁵ fruit fly⁶ and mouse,⁷ and the screening of chemically induced mutations in a reverse genetic strategy towards understanding the function(s) of genes in *arabidopsis*⁸ and fruit fly.⁹ It has also been implemented in the quantitative measurement of gene expression¹⁰ and the analysis of single nucleotide extension products.¹¹

The first step towards the development of DHPLC was taken ten years ago with the synthesis of alkylated non-porous poly(styrene-divinylbenzene) particles 2 microns in diameter.^{12,13} In combination with a hydroorganic eluent containing alkylammonium ions, this stationary phase material enabled, for the first time, the size-dependent liquid chromatographic separation of double-stranded DNA

fragments with a degree of resolution only previously attained by capillary electrophoresis.¹⁴ In 1995, it was discovered that double-stranded DNA fragments up to 1500 base pairs in length containing a single-base mismatch, (i.e., a base pair other than one of the two typical Watson-Crick base pairs A-T and G-C, respectively), will be retained less on the aforementioned chromatographic support than a perfectly matched DNA homoduplex at column temperatures sufficiently high to result in partial denaturation of the analytes.^{15,16} This approach requires that the two sequences to be compared are denatured and reannealed prior to HPLC analysis to allow the formation of the original homoduplexes as well as that of heteroduplexes between the sense and antisense strands of either homoduplex. Formation of heteroduplexes is not required for a more recent embodiment of DHPLC, referred to as completely denaturing HPLC.¹⁷ It is based on the ability of the stationary phase to resolve isomeric single-stranded nucleic acids up to 100 nucleotides in length from each other even when they differ only in a single base.

Partially Denaturing HPLC

Partially denaturing HPLC typically compares two or more chromosomes as a mixture of denatured and reannealed polymerase chain reaction (PCR) amplicons. In the presence of

a mutation in one of the two chromosomal fragments, not only the original homoduplexes are formed again upon reannealing but, simultaneously, the sense and antisense strands of either homoduplex form heteroduplexes that are thermally less stable (Figure 1). The more extensive but still partial denaturation of the heteroduplexes at elevated temperatures, typically in the range of 50–70 °C depending on the GC-content of the DNA fragment under investigation, results in their reduced retention on the chromatographic separation matrix. As a consequence, one or more additional peaks appear in the chromatogram, with different mutations yielding, in most but not all instances, distinctively different peak profiles.^{18,19}

Temperature is the most important experimental parameter affecting mutation detection sensitivity and its optimum can be predicted by computation at the publicly available website <http://insertion.stanford.edu/melt.html>.²⁰ Single-nucleotide substitutions, deletions and insertions have been detected successfully within 2 to 3 minutes in unpurified amplicons typically 200–1000 bp in length, with sensitivity and specificity of DHPLC consistently approaching 100%. Prerequisite, however, is the proper preconditioning of the DNA sample. This is accomplished by placing a heat exchanger made of 80 cm of 0.01 in i.d. PEEK tubing encased in a tin alloy block before the sample loop, both of which are kept in the oven. Alternatively, the sample can be injected at ambient temperature. In this instance, the heat exchanger must be placed between the injection valve and the column. The concomitant increase in extra-column volume will decrease resolution. In practice, however, the effect is too small to impact the ability of DHPLC to resolve homo- and heteroduplexes.

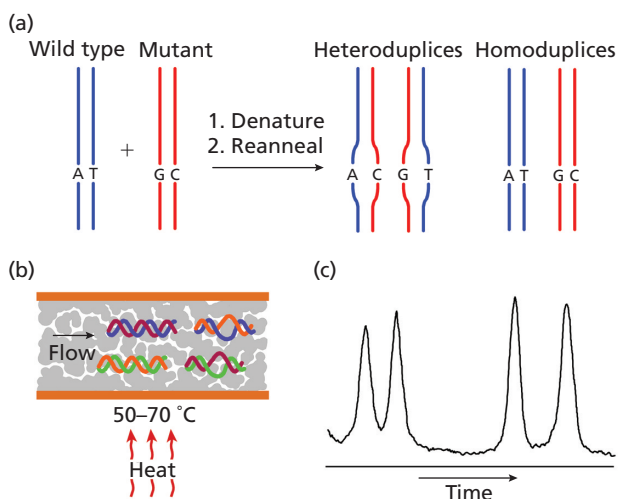
Sensitivity and Specificity of DHPLC

As mentioned earlier, choice of column temperature is critical to ensure high sensitivity of DHPLC in mutation detection. Originally, the optimum temperature at which to screen a particular DNA sequence was determined empirically by repeatedly injecting a test sample at gradually increasing column temperatures until the duplex product peak was retained about a minute less than at 50 °C. At this point, the presence of a single mismatch will usually be detected by the appearance of one to three additional peaks. Whether one or two heteroduplex and homoduplex peaks, respectively, are observed depends on several factors. They include the influence of nearest neighbour sequence on the stability of base pair mismatches²¹ and hydrogen bonding between non Watson-Crick base oppositions such as G-T and G-A.²² In addition, temperature may affect the chromatographic profile with as little difference as 2 °C, resulting in either the separation of all four species, (i.e., both the two hetero- and the two homoduplexes), or just the separation of the two hetero- or the two homoduplexes, while the corresponding homo- and heteroduplexes elute as one peak. In the presence of more than one mismatch, the number of heteroduplex peaks observed may be greater than two based on the extent of denaturation.

The empirical approach of determining the appropriate temperature of analysis, however, harbours the risk that mismatches in low-melting AT-rich domains may go undetected because of complete denaturation. Further, it impedes automation and sample throughput. For these reasons, an algorithm was developed that calculates, for every site in a known sequence, the temperature at which 50% of the

fragments are closed. Analysis is routinely performed at the highest of all site temperatures. Analysis is repeated in 4 °C decrements if the melting temperatures predicted span more than 4 °C or multitudes thereof.²⁰ Using such strategy, 165 out of 166 different polymorphic DNA fragments that had been part of a total of 476 fragments screened with an average length of 563 bp were readily detected.²³ The ability of DHPLC to detect sequence variation appeared independent of the number, nature and sequence flanking the mismatches. In one instance, the two chromosomal fragments compared differed by as many as 20 base substitutions and insertions/deletions. The only polymorphic fragment that had escaped detection contained a single mismatch, the resultant homo- and heteroduplexes of which resolved only over a temperature range as narrow as 2 °C. This is rather unusual based on a study of 103 mutations in 42 different sequence contexts.²⁰ The median number of temperatures at which heterozygosity could be detected was 8, and the range was 4–11 °C. The values represent minimum values, as the samples were not analysed beyond 5 °C on either side of the temperature recommended by the algorithm.

Figure 1: Principle of partially denaturing HPLC. (a) Prior to chromatography, a 200–1000 base-pair fragment is amplified by means of polymerase chain reaction from at least two chromosomes. Subsequently, the amplified fragments are denatured at 95 °C for 3 min, before they are allowed to reanneal by gradually lowering the temperature from 95 °C to 65 °C over 30 min. In the presence of a mutation in one of the chromosomes, not only the original homoduplexes are formed again but, simultaneously, the sense and anti-sense strands of either homoduplex form two heteroduplexes. (b) Depending on the 'GC' content of the double-stranded DNA fragments to be compared, a column temperature typically in the range of 50–70 °C is maintained to induce partial denaturation. The thermally less stable heteroduplexes denature more extensively and, consequently, are retained shorter on the stationary phase. (c) Depending on several factors, including, among others, size of the fragments, influence of nearest neighbour on the stability of both matched and mismatched base pairs, and column temperature, either all four species may be resolved completely or only in part.



The predictive power of the melting algorithm is excellent. A blind analysis of 103 mutations showed that all but four could be readily detected at the temperatures recommended by the algorithm.²⁰ The mutations missed, however, could be detected successfully by increasing column temperature by 2 °C. At present, the number of mutations whose temperature for detection has been predicted incorrectly is still too small to deduce conclusively a denominator that could lead to an appropriate adjustment of the algorithm. Although a recent study of 18 mutations in exon 1 of the *VHL* tumour suppressor gene²⁴ suggested that the WAVEMaker™ software, which is commercially available from Transgenomic Inc. (Omaha, Nebraska, USA), might predict melting temperatures more accurately (there were no discrepancies between the predictions made by the two algorithms for two other exons of the gene), this software has never been subjected to the same rigorous blind analyses as the publicly available software. Hence, it remains to be evaluated whether improved prediction for certain domains is accomplished at the expense of less accurate predictions for other domains. Finally, for reasons unknown, the study analysed the samples at a temperature that was 2 °C lower than the one actually recommended by the publicly available algorithm. In the meantime, it is recommended to determine empirically whether a higher temperature is required. This is accomplished by injecting the sample at a temperature 2 °C higher than the recommended one; if the DNA fragment elutes only slightly earlier (~0.5 min) than at the highest temperature recommended, the analysis should be repeated at the higher temperature. If the shift in elution time is greater, the analysis was already performed at the optimum temperature.

The necessity to use different column temperatures to obtain identical chromatographic profiles is the result of differences in retention of nucleic acids between columns.

The temperature displayed on the column oven does not necessarily correspond to the actual temperature in the column compartment. Measurements with certified temperature probes have shown deviations of up to 2 °C, particularly with ovens calibrated at only one temperature. Moreover, in systems that use a mere coil of PEEK tubing rather than one embedded in a tin alloy block, the actual temperature may drop ≥ 1 °C when the flow-rate is increased from 0.05 to 0.9 mL/min. It is, therefore, recommended that checks are made not only on separation efficiency but also on actual column oven temperature once a week. Especially after blackouts, it may be necessary to recalibrate the column oven. It is also not unusual to observe an increasing discrepancy between actual and displayed temperature over a period of months.

Recent experiments with arrays of capillary columns²⁵ have revealed another potential reason for discrepancies in mutation detection sensitivity between laboratories, as well as the reproducibility of chromatographic profiles from column to column within a laboratory. If all columns were indeed identical, they should yield identical chromatograms under identical conditions. However, as can be seen in Figure 2,

monolithic capillary columns run under identical conditions yielded different chromatograms. At a uniform temperature of 55 °C, only one of the three columns yielded a four-peak profile on injection of a 209 bp amplicon representing two Y-chromosome alleles differing in a single base: an A to G transition, at nucleotide position 168. The other two columns, however, yielded a single peak at 55 °C. Using individual column thermostats, the temperatures of columns 2 and 3 were increased to 56 °C and 57 °C, respectively, to obtain chromatographic profiles similar to those observed with column 1 at 55 °C. The necessity to use different column temperatures to obtain identical chromatographic profiles is the result of differences in retention of nucleic acids between columns. Differences in retention translate into different concentrations of acetonitrile required for elution of nucleic acids. Experiments with nucleic acids labelled with different fluorophores have shown that a ~0.8% increase in acetonitrile concentration required for elution corresponds to an increase in column temperature of 1 °C.¹ Consequently, if a given DNA fragment is eluted from a column at a lower acetonitrile concentration, a higher column temperature is required to achieve the same degree of partial denaturation. This is also evident from Figure 2. At a temperature of 55 °C, the amplicon eluted from column 3 at a concentration of ~13% acetonitrile, whereas the same amplicon eluted from column 2 at 14% acetonitrile (Figure 2(a)). Consequently, the temperature of column 3 had to be raised 1 °C higher than that of column 2 to obtain a similar degree of denaturation and, hence, similar separation of the homo- and heteroduplex species. At present, we believe that differences in retention are the result of differences in porosity between columns brought about by the instability of divinylbenzene. Differences in the ratio of styrene to divinylbenzene, in turn, affect the degree of polymerization and, consequently, pore size. As shown in Figure 2, modulation of column temperature offers a convenient approach to harmonizing chromatographic separation. This is by no means restricted to DHPLC. Rather, it can be applied to any chromatographic separation sensitive to temperature.

The aforementioned observations underscore the significance of proper standardization using different heteroduplex standards that yield chromatographic profiles highly sensitive to minute changes in column temperature. Such standards are now commercially available from Transgenomic Inc.; they yield typical chromatographic profiles at 56 °C, 64 °C and 70 °C, respectively. It is also obvious that column thermostats need to be highly precise and adjustable in at least 0.5 °C increments to fine tune the performance between columns.

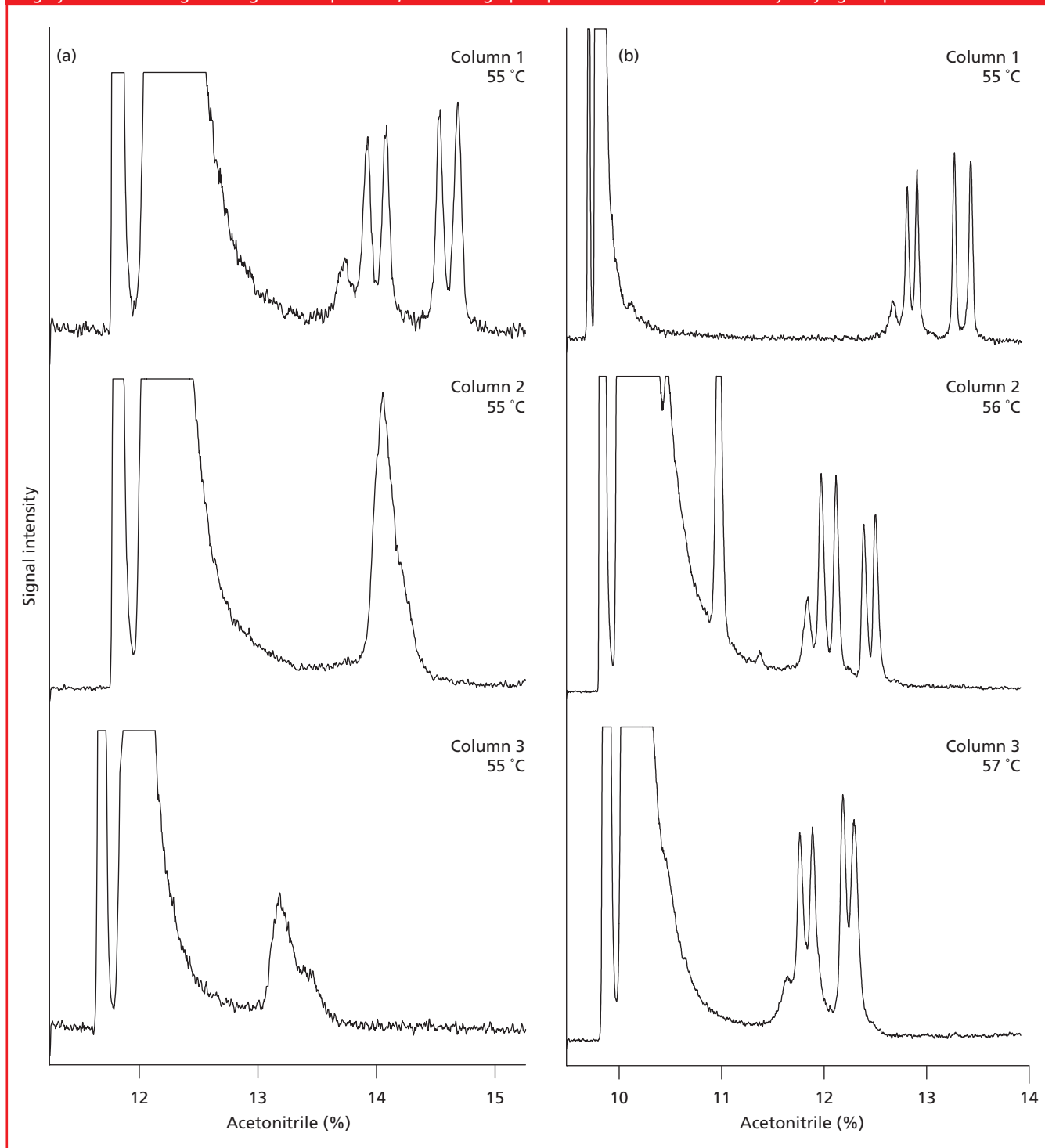
DHPLC in Comparison with Other Methods Commonly Used in Mutation Analysis

The predominant application of DHPLC is that of the mutational analysis of genes. The first gene subjected to DHPLC analysis was the calcium channel gene *CACNL1A4*.²⁶ Mutations in the gene were found to be the cause of either familial hemiplegic migraine or episodic ataxia type-2. To date, the experimental conditions for the mutational analysis of >100 human genes by DHPLC have been established as highlighted on the website (<http://insertion.stanford.edu/pub.html>). Among the most extensively studied genes are the so-called breast cancer genes *BRCA1* and *BRCA2*, and more recently *ATM*.¹ The three genes are not only large, with the number of

protein-encoding bases ranging from 5592 to 10 257 base pairs, but also lack, with few ethnic exceptions, mutational hotspots accounting for the majority of affected cases studied.

Hence, it is necessary to sequence the entire genes which renders their analysis laborious and expensive. Although direct sequence analysis of PCR products by dye-terminator

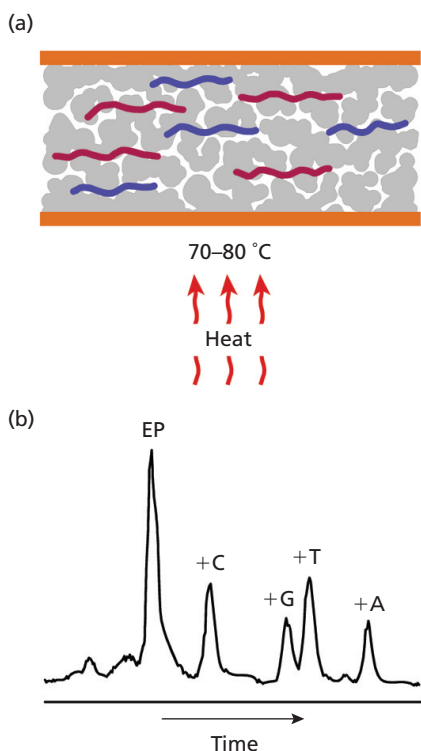
Figure 2: Harmonization of chromatographic elution profiles of different columns by means of modulation of individual column temperature. (a) The same sample was injected onto three different columns kept at the same temperature. Because of differences in porosity and, hence, surface area of the columns, retention of DNA varies between columns. Concomitantly, the concentration of acetonitrile required for elution of the same fragment will vary. Because acetonitrile acts as a denaturant itself, fragments will be denatured more extensively at a given column temperature the longer they are retained. This leads to differences in elution profiles for identical mismatches between different columns. (b) Using individual column thermostats and mutation standards highly sensitive to slight changes in temperature, chromatographic profiles can be harmonized by varying temperature.



chemistry is commonly regarded as the benchmark against which alternative strategies for mutation detection should be measured, the method does not have the same sensitivity and specificity established for the sequencing of M13 clones by dye-primer chemistry when the mutation to be detected is only present in one out of two or more chromosomes as is the situation with somatic mosaicism that is a frequent phenomenon in disorders exhibiting a high mutation rate such as tuberous sclerosis and mitochondrialopathies. In the latter instances, DHPLC has been shown to be superior in the detection of mutations compared with direct sequence analysis.^{27,28}

One of the most widely used methods for mutation screening has been single-strand conformation analysis (SSCA). Its popularity stems mainly from low cost as mutation detection is accomplished in its simplest embodiment by separating the radioactively labelled single-stranded components of the DNA fragment of interest electrophoretically in a non-denaturing polyacrylamide gel.²⁹

Figure 3: Principle of completely denaturing HPLC. (a) At column temperatures $>70^{\circ}\text{C}$ double stranded DNA fragments will denature completely. (b) The single stranded components can then be resolved from each other even if they differ only in sequence and not in size. The chromatogram depicts the separation of an extension primer (EP) and the four possible isomeric products generated by single nucleotide extension sequencing. Resolution for any given pair of alleles can be optimized by varying column temperature.¹⁷ The order of elution of the alleles depends on the stationary phase used. On poly(styrene-divinylbenzene) monoliths, as depicted, extension products elute in the order $\text{C}<\text{G}<\text{T}<\text{A}$. On micropellicular alkylated poly(styrene-divinylbenzene) particles, in contrast, the elution order is $\text{G}<\text{C}<\text{A}<\text{T}$.¹⁷



Mutation-induced changes in the sequence cause alterations in the tertiary structure of the single-stranded DNA molecules and, consequently, differences in mobility. Hence, the presence of mutations is revealed as the appearance of new bands in the electropherograms. Sensitivity of SSCA has been reported to vary from about 50–95%,^{20,30–35} with mutations located within hairpin-like structures posing the greatest challenge.³⁶ It is also recommended to keep DNA fragment size shorter than 300 base pairs.^{30,31} Although sensitivity of SSCA appears higher, at least for DNA fragments <300 bp, when performed by capillary rather than slab gel-based electrophoresis^{37,38} the need for fluorescent-dye labelling and solid-phase extraction prior to loading adds significantly to the amount of labour and cost involved in mutational analysis.

Heteroduplex analysis using conformation-sensitive gel electrophoresis (CSGE) in native gels has a sensitivity similar to or slightly inferior than SSCA.^{20,33,35} Significantly higher sensitivity has been obtained by means of denaturing gradient gel electrophoresis (DGGE) using a linear gradient of urea and formamide to induce partial melting of homo- and heteroduplexes.³⁵ At least one study reported a poorer performance of DGGE in higher melting domains compared to DHPLC.³⁹ Although GC-clamps have been widely used in DGGE to improve detection sensitivity of mutations located in high-melting domains,⁴⁰ only one study has reported their use in DHPLC to date.⁴¹

...DHPLC has emerged as the most sensitive physical mutation screening method that requires the least amount of sample manipulation.

In conclusion, DHPLC has emerged as the most sensitive physical mutation screening method that requires the least amount of sample manipulation. In situations of somatic mosaicism and mitochondrial heteroplasmy, DHPLC appears to be even superior to direct sequencing of PCR amplicons. Hence, in situations in which DHPLC reproducibly yields an altered chromatographic profile, cloning is warranted to definitely establish or rule out the presence of a mutation.

Completely Denaturing HPLC

The second mode of DHPLC is performed under completely denaturing conditions and is used for the analysis of shorter nucleic acid fragments typically 50 to 100 nucleotides in size that differ in single or multiple bases (Figure 3). The high resolving power of the chromatographic separation system makes it possible to discriminate two single-stranded nucleic acids of identical size with a difference in base composition as small as a single base out of one hundred bases.¹⁷ In this technique, the alleles of a given polymorphic locus can be resolved without the addition of a reference chromosome. The only exceptions to this rule have been C to G transversions. Although differences in retention time between different alleles have been shown to be highly reproducible, they become increasingly smaller the longer the DNA fragment. Hence, under mass screening conditions, errors in calling homozygotes accurately may occur. Moreover, the presence of additional

polymorphisms within the same fragment may complicate the analysis. Hence, direct confirmation of the chemical nature of the polymorphic allele of interest is desirable. This can be accomplished either enzymatically by performing a mini-sequencing reaction or physically by hyphenating DHPLC with electrospray ionization mass spectrometry (ESI-MS).

DHPLC of Primer Extension Reactions

Completely denaturing HPLC has been applied to the separation of mini-sequencing or primer extension products.⁴² After amplification of a fragment of DNA containing a single nucleotide polymorphism, an oligonucleotide primer is annealed immediately upstream or downstream from the polymorphism. In the presence of the appropriate dNTPs and ddNTPs, the primer is extended by one or more bases depending upon the sequence at the polymorphic site. The alleles are then distinguished on the basis of either size of the extended product or differences in retention depending on the ddNTP incorporated. Potential advantages of HPLC over other methods described for the analysis of extension reactions include low cost as there is no need for labelled oligonucleotides or dye terminators and the ability to analyse the products directly without further purification as required for mass spectrometric analysis. Another advantage of DHPLC is its inherently quantitative nature that should allow allele ratios to be determined precisely and accurately in pools of DNA samples and polyploid genomes, respectively. The pooling of DNA samples is an elegant way of accelerating the determination of allele frequencies. This is particularly true for marker-trait association studies in which the number of candidate polymorphisms and, particularly, that of samples tends to be extremely large, ranging from hundreds to thousands of affected and unaffected individuals that need to be analysed. A combination of single-base extension and DHPLC showed that absolute minor allele frequencies as small as 5% could be determined with a mean experimental error (i.e., the discrepancy between the estimated and real frequencies) of 0.014 in pools of approximately 100–800 chromosomes.^{43,44} This requires each extension reaction to be calibrated using a single heterozygote with equal allele representation to determine the ratio of the two allele-specific extension products. Ideally, the ratio should have a value of 1 as both alleles are represented equally. However, on no occasion was this observed, with deviations from expectation as great as twofold having been observed.^{43,45} Non-equal representation of alleles in the analysis of heterozygotes can be caused by differential PCR amplification of alleles, differential efficiency of incorporation of nucleotides for each allele-specific reaction, and minor differences in the UV absorbancies of the two allele-specific products, which differ in 3' base composition. Such differences are reproducible and, therefore, can be corrected. A recent investigation of SNPs in tetraploid potato confirmed the precision of single nucleotide extension analysis in combination with DHPLC, but failed to support earlier claims of its accuracy.⁴⁵ PCR was excluded as a culprit as extension reactions performed on synthetic templates yielded the same discrepancy from the expected ratios as reactions performed on amplified DNA. Absolute allele frequency estimates in the case of 1:3 and 3:1 heterozygotes were off by as much as 12%. Interestingly, similar discrepancies were observed using pyrosequencing that employs Klenow DNA

polymerase for primer extension. Only conventional dye-terminator sequencing with AmpliTaq Polymerase FS yielded allele frequency estimates that were within a few per cents of expectations. This was particularly true when estimates were generated from combined peak height measurements performed on both the forward and the reverse strand sequencing trace. Despite extensive variation of the experimental conditions, the reason for the observed discrepancy remains unknown.

DHPLC and Mass Spectrometry of Nucleic Acids

The hyphenation of capillary liquid chromatography with ESI-MS provides a wealth of experimental data that can be evaluated to gain information about the sequence of nucleic acids. Ion-pair reversed-phase liquid chromatography in monolithic poly(styrene-divinylbenzene) capillary columns allows the analytes of interest to be separated from the by-products of their production with high efficiency, by chemical synthesis or by PCR, and its low flow-rate, volatile mobile phase components and the on-line removal of cations from the DNA backbone facilitate the interfacing to the mass spectrometer.⁴⁶ The sequence information that can be obtained differs with the size of the targeted DNA, the applied mass spectrometric technique and the performance characteristics of the mass spectrometer used (Table 1).

The exact mass measurement is a simple means for verifying the composition of double-stranded or single-stranded DNA. Very similar DNA fragments that cannot be completely resolved by the chromatographic process and elute as one single chromatographic peak can be deconvoluted and their components identified because of their separate characteristic mass signals. Depending on the accuracy of the mass measurement and the size of the examined nucleic acid, not only deletions and insertions, but also base substitutions can be detected. The formation of series of multiply charged ions for nucleic acids greatly enhances the amenable mass range, and the mass measurement of enzymatic digests containing DNA up to 434 bp, corresponding to a molecular mass of over 268 000, was possible on an ion trap mass spectrometer with sufficient accuracy to call single-base deletions or insertions.⁴⁷

Shorter DNA sequences up to 100 bp show differences in chromatographic retention under completely denaturing conditions even if they differ only in one base. The exact mass measurement of the single strands enables the characterization of single base substitution polymorphisms. Two factors are important in this instance: First, the accuracy of the mass measurement is higher for smaller nucleic acids because the same relative error in the mass determination causes a smaller absolute deviation from the expected mass, and second, the

Table 1: Information about nucleic acids gained from mass spectrometry.

Mass range	Information	Technique
<120 bp	Complete sequence	IP-RP-HPLC–ESI-MS/MS comparative sequencing
<120 bp	Base substitution, insertion, deletion	IP-RP-HPLC–ESI-MS completely denaturing HPLC
<500 bp	Insertion, deletion	IP-RP-HPLC–ESI-MS non-denaturing HPLC

completely denaturing conditions during the chromatographic analysis allow the detection of A→T, T→A, C→G and G→C transversions, respectively, which do not alter the mass of the double-stranded DNA. On an ion trap, relative deviations in the mass determination between 0.2 and 0.3% were obtained. These translate into differences in the absolute mass of between 3 and 5. Because the smallest mass difference between two bases is 9 for the difference between A and T, all base changes can be determined with high accuracy. Even a C to G transversion that cannot be resolved using completely denaturing HPLC can be detected easily in fragments shorter than 100 bp if DHPLC is combined with mass spectrometry, as the transversion is associated with a shift in mass of the DNA single strands of 40 mass units.

One shortcoming of exact mass measurement alone is that it can only determine the correctness of a sum formula, but cannot verify an expected sequence. Mutations in which two bases switch sequence position are quite common and highly informative and this change in the phase of the nucleic acid does not alter the molecular mass. This information can be obtained by fragmenting the nucleic acids in tandem mass spectrometry, recording the ions corresponding to the formed fragments and evaluating the obtained data with a new mathematical algorithm denominated comparative sequencing. In this method the sequence of a nucleic acid is verified by comparing the experimental data set to generated expected values for a reference sequence and all possible single point mutations of the reference sequence.⁴⁸ A cumulative fitness parameter that takes into account found fragments, missed sequence positions and mass deviations is calculated for every single probed sequence position. If the reference sequence yields a minimum fitness value for every sequence position, then a perfect match with the experimental data is given. In the instance of single-base mismatches, the lowest value for the fitness parameter indicates the position of the mismatch with respect to the reference sequence. In combination with capillary liquid chromatography this technique has been successfully applied to the haplotyping of two and three sequence polymorphisms contained within low femtomol amounts of unpurified PCR products up to 114 base pairs long. Keys to attaining such a low detection limit have been the use of acetonitrile as sheath liquid⁴⁹ and the replacement of triethylammonium bicarbonate with butyldimethylammonium bicarbonate as ion pairing reagent.⁴⁷ Being a less basic amine, butyldimethylammonium bicarbonate improves, in concordance with an earlier study⁵⁰ the signal-to-noise ratio by a factor of 1.5. In addition, because of its higher affinity to the stationary phase, a higher concentration of acetonitrile is required to elute the nucleic acids from the column. The concomitant decrease in surface tension and the increase in volatility of the eluent cause an improvement in electrospray ionization efficiency. Finally, the successful implementation of the algorithm for comparative sequencing demonstrates that information about the entire nucleotide sequence is present in the tandem mass spectrum. Hence, more sophisticated mathematical approaches may eventually allow the *de novo* reconstruction of nucleic acid sequences with a length of up to 100 bases.

Multiparallel DHPLC

Today, DHPLC is performed predominantly on commercially available chromatographic columns in the conventional

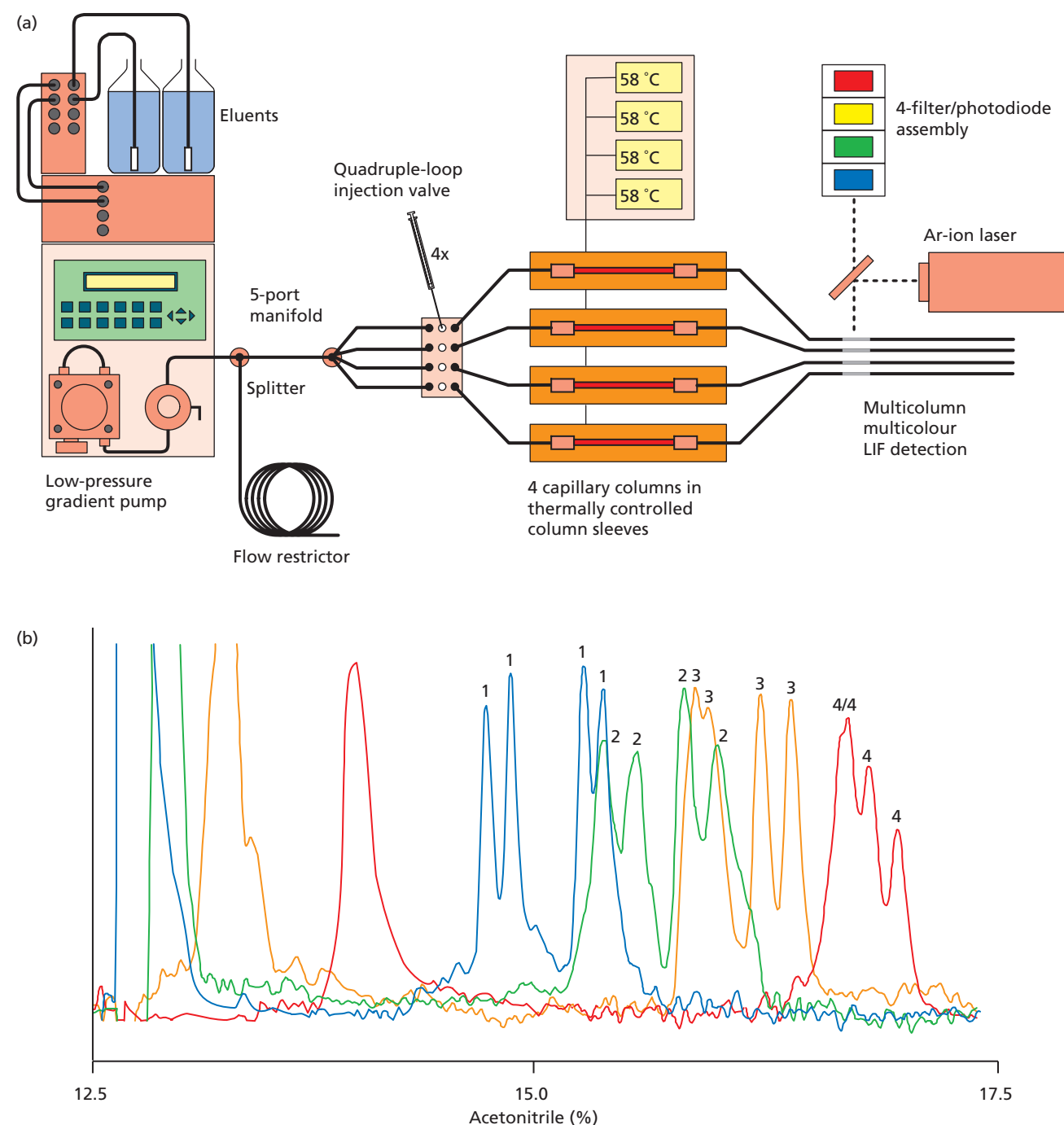
4.6 mm i.d. format packed with non-porous alkylated poly(styrene-divinylbenzene) beads.^{1,13} Despite its high sensitivity and good productivity, the serial nature of the chromatographic analysis does not easily allow the implementation of DHPLC in true genome-scale projects because of its limitations of ample throughput and cost effectiveness. Furthermore, the use of optical detectors limits the generated information content. The solution to these demands comes with the introduction of monolithic poly(styrene-divinylbenzene) columns in the capillary format with 200 µm i.d.^{46–51} These columns have proved to be at least as efficient and durable as conventional packed columns, and their low flow-rate and small sample consumption on the order of a few hundred nanolitres have opened up the way for exciting new developments.^{25,47}

The greater concentration sensitivity of the capillary format makes it possible to combine DHPLC with laser-induced fluorescence detection for the analysis of single nucleotide polymorphisms (SNPs) in analogy to capillary electrophoresis in DNA sequencing (Figure 4(a)). In this technique, higher throughput is enabled by colour multiplexing.⁵² Different amplicons are labelled with different fluorescent dyes during PCR using dye-labelled primers. The samples are then pooled and analysed simultaneously in one chromatographic column, and are monitored separately by observing their characteristic emission wavelengths. Figure 4(b) shows the multiplex detection of four 209 bp PCR products labelled with four different fluorescent dyes. The four products were generated separately and mixed together before injection, separated under partially denaturing conditions in the capillary column, and detected by a fluorescence scanner. The characteristic peak patterns allow the designation of the samples as heterozygotes. However, despite the fact that all four chromatographic traces were obtained with amplicons of the same DNA fragment, profiles vary as a function of the fluorophore attached. With increasing hydrophobicity of the fluorophores, amplicons are retained longer on the stationary phase necessitating increasingly higher concentrations of acetonitrile to elute them from the column. Because organic solvents act as denaturants themselves, DNA fragments tagged with more hydrophobic fluorescent dyes, such as rhodamine derivatives will be denatured more extensively at the same column temperature compared with amplicons labelled with fluorescein derivatives. It has been shown that an increase in acetonitrile concentration of approximately 0.8% equals an increase in column temperature of about 1 °C.²⁵ Given that most mutations can be detected over a fairly wide range of temperatures,²⁰ changes in retention brought about by different fluorophores may exert little effect on detection sensitivity. However, because the melting behaviour of DNA fragments can be predicted reliably by computation, it may be prudent to multiplex amplicons with different melting characteristics, (i.e., to label with rhodamine a fragment whose recommended temperature of analysis is 2 °C higher than that of a fragment labelled with fluorescein.)

A further increase in throughput is made possible by bundling monolithic capillary columns into arrays similar to those already used in capillary electrophoresis (Figure 4(a)). Multiple samples can be analysed at a time in different columns at the same or at different temperatures using only one pump, injection and scanning detection device.²⁵ The throughput increases with the number of columns used in the system and a

Figure 4: Parallel HPLC analysis of DNA fragments on an array of monolithic poly(styrene-divinylbenzene) capillary columns.

(a) Scheme of a four-column array. A single low-pressure gradient-mixing pump is used to generate a primary flow of 150 $\mu\text{L}/\text{min}$, which is subsequently split to create a secondary low-flow stream of $\sim 2.5 \mu\text{L}/\text{min}$ per column. With the majority of the primary flow stream going to waste, it is conceivable to operate ≥ 48 columns in parallel with a single pump. Individual column thermostats, which had been custom-made of brass tube ($15 \times 1.2 \text{ cm}$, i.d., 0.2 mm thick) wrapped with $500 \times 0.24 \text{ mm}$ o.d. nickel-chromium heating wire over a layer of Mylar tape, that provided electric insulation, and were controlled by a reading type T thermocouple connected to a control unit that operated in on-off mode with a dead-band of 0.07°C , were used to modulate column temperature. A confocal fluorescence capillary array detector allowed sample multiplexing by tagging amplicons with different fluorophores commonly used in sequencing. Injection of samples was performed manually using a custom-made electrically actuated injector with four internal $1 \mu\text{L}$ sample loops. (b) Multiplex analysis of the hetero- and homoduplexes of a 209 bp amplicon containing a single nucleotide polymorphism. Amplification was performed in separate reactions with one of the primers tagged with either FAM (peak 1), HEX (peak 2), NED (peak 3) or ROX (peak 4). Retention increases with increasing hydrophobicity of the fluorophore attached resulting in different chromatographic profiles for the same mismatch as the degree of denaturation increases with the increasing concentration of acetonitrile required to elute the hetero- and homoduplexes.



further enhancement in the number of samples analysed simultaneously is achieved by the combination of this technology with fluorescent colour multiplexing. Both multiplexing techniques can be applied to partially and completely denaturing HPLC. The ultimate goal of this development is to analyse a full PCR plate in one single chromatographic run, transforming DHPLC into a true genome-scale technology.

Conclusions

The high sensitivity and specificity of DHPLC have significantly facilitated the mutational analysis of candidate genes in inherited diseases, the mapping and cloning of mono- and polygenic traits, and the elucidation of human origins. Recent advances in column technology have enabled the construction of capillary arrays that allow increased throughput of samples and multiplexing in conjunction with laser-induced fluorescence detection. Combined with electrospray ionization mass spectrometry, nucleic acids can be resequenced successfully up to 100 bases and the phase of multiple single nucleotide polymorphisms can be determined directly without the need for prior cloning or allele-specific PCR.

References

- W. Xiao and P.J. Oefner, *Hum. Mutat.*, **17**, 439 (2001).
- L. Jin et al., *Proc. Natl. Acad. Sci., USA*, **96**, 3796 (1999).
- P.A. Underhill et al., *Nat. Genet.*, **26**, 358 (2000).
- L.M. Steinmetz et al., *Nature*, **4**, **16**, 326 (2002).
- J.I. Spiegelman et al., *Plant Cell*, **12**, 2485 (2000).
- R.A. Hoskins et al., *Genome Res.*, **11**, 1100 (2001).
- L.M. Schriml and M. Dean, *Genomics*, **64**, 24 (2000).
- A. Bentley et al., *Genetics*, **156**, 1169 (2000).
- C.M. McCallum et al., *Nat. Biotechnol.*, **18**, 455 (2000).
- P.A. Doris et al., *J. Chromatogr. A*, **806**, 47 (1998).
- B. Hoogendoorn et al., *Hum. Genet.*, **104**, 89 (1999).
- C.G. Huber, P.J. Oefner and G.K. Bonn, *Anal. Biochem.*, **212**, 351 (1993).
- C.G. Huber, P.J. Oefner and G.K. Bonn, *U.S. Patent No. 5,585,236* (1996).
- C.G. Huber, P. Oefner and G. Bonn, *Anal. Chem.*, **67**, 578 (1995).
- P.J. Oefner and P.A. Underhill, *Am. J. Hum. Genet.*, **57**, A266 (1995).
- P.J. Oefner and P.A. Underhill, *U.S. Patent No. 5,795,976* (1998).
- P.J. Oefner, *J. Chromatogr. B*, **739**, 345 (2000).
- M.C. O'Donovan et al., *Genomics*, **52**, 44 (1998).
- Gross et al., *Hum. Mutat.*, **16**, 345 (2000).
- A.C. Jones et al., *Clin. Chem.*, **45**, 1133 (1999).
- S. H. Ke and R.M. Wartell, *Nucl. Acids Res.*, **21**, 5137 (1993).
- F. Abou-ela et al., *Nucl. Acids Res.*, **13**, 4811 (1985).
- J.I. Spiegelman, M.N. Mindrinos and P.J. Oefner, *BioTechniques*, **29**, 1084 (2000).
- B. Klein, G. Weirich and H. Brauch, *Hum. Genet.*, **108**, 376 (2001).
- W. Premstaller et al., *Genome Res.*, **11**, 1944 (2001).
- R.A. Ophoff et al., *Cell*, **87**, 543 (1996).
- B.J.C. van den Bosch et al., *Nucl. Acids Res.*, **28**, e89 (2000).
- A.C. Jones, J.R. Sampson and J.R. Cheadle, *Hum. Mutat.*, **17**, 233 (2001).
- M. Orita et al., *Genomics*, **5**, 874 (1989).
- K. Hayashi and D.W. Yandell, *Hum. Mutat.*, **2**, 338 (1993).
- V.C. Sheffield et al., *Genomics*, **16**, 325 (1993).
- Q. Liu, J. Feng and S.S. Sommer, *Hum. Mol. Genet.*, **5**, 107 (1996).
- Y.S. Choy et al., *Ann. Hum. Genet.*, **63**, 383 (1999).
- E. Gross et al., *Hum. Genet.*, **105**, 72 (1999).
- C. Eng et al., *J. Med. Genet.*, **38**, 824 (2001).
- M.B. White et al., *Genomics*, **12**, 301 (1992).
- C. Dobson-Stone et al., *Eur. J. Hum. Genet.*, **8**, 24 (2000).
- L.A. Ellis, C.F. Taylor and G.R. Taylor, *Hum. Mutat.*, **15**, 556 (2000).
- T.R. Skopek et al., *Mutat. Res.*, **430**, 13 (1999).
- R. Myers et al., *Nucl. Acids Res.*, **13**, 3111 (1985).
- G. Narayanaswami and P.D. Taylor, *Genet. Test.*, **5**, 9 (2001).
- B. Hoogendoorn et al., *Hum. Genet.*, **104**, 89 (1999).
- B. Hoogendoorn et al., *Hum. Genet.*, **107**, 488 (2000).
- M. Giordano et al., *J. Biochem. Biophys. Meth.*, **47**, 101 (2001).
- A.M. Rickert et al., *BioTechniques*, **32**, 592 (2002).
- A. Premstaller, H. Oberacher and C.G. Huber, *Anal. Chem.*, **72**, 4386 (2000).
- H. Oberacher et al., *Angew. Chem. Int. Ed.*, **40**, 3828 (2001).
- H. Oberacher, B. Wellenzohn and C.G. Huber, *Anal. Chem.*, **74**, 211 (2002).
- C.G. Huber and A. Krajete, *J. Chromatogr. A*, **870**, 413 (2000).
- M.J. Greig and R.H. Griffey, *Rapid Commun. Mass Spectrom.*, **9**, 97 (1995).
- C.G. Huber et al., *J. Biochem. Biophys. Meth.*, **47**, 5 (2001).
- W. Xiao et al., *BioTechniques*, **30**, 1332 (2001).

Drs Premstaller and Oefner hold US patents related to micropellicular (5,585,236) and monolithic (09/770,410) poly(styrene-divinylbenzene) stationary phases licensed to Transgenomic, Omaha, Nebraska, USA and denaturing high performance liquid chromatography (5,795,976) licensed to Agilent (Palo Alto, California, USA), Qiagen (Valencia, California, USA), Transgenomic and Varian (Walnut Creek, California, USA). **Andreas Premstaller** received his PhD from the University of Innsbruck, Austria, for the synthesis of monolithic poly(styrene/divinylbenzene) capillary columns and their use in the characterization of nucleic acids and proteins by electrospray ionization mass spectrometry. He is a recipient of an Erwin-Schrödinger-Stipend from the Austrian Science Foundation and, currently, a postdoctoral fellow at the Stanford Genome Technology Centre.

Peter Oefner is an Associate Professor of Bioanalytical Chemistry and Associate Director of the Stanford Genome Technology Centre. His nearly 150 publications deal with the analysis of biomolecules by high performance liquid chromatography, capillary electrophoresis, mass spectrometry, and high-density DNA arrays, population genetics, human evolution, the mapping and cloning of Mendelian and quantitative traits, and functional genomics.

