DOI 10.1002/biot.201000372 Biotechnol. J. 2011, 6, 378–391

Review

Mutation detection in plasmid-based biopharmaceuticals

Pedro H. Oliveira[1], Kristala L. J. Prather[2], Duarte M. F. Prazeres[1] and Gabriel A. Monteiro[1]

[1] Institute for Biotechnology and Bioengineering, Center for Biological and Chemical Engineering, Instituto Superior Técnico, Lisbon, Portugal
[2] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

As the number of applications involving therapeutic plasmid DNA (pDNA) increases worldwide, there is a growing concern over maintaining rigorous quality control through a panel of high-quality assays. For this reason, efficient, cost-effective and sensitive technologies enabling the identification of genetic variants and unwanted side products are needed to successfully establish the identity and stability of a plasmid-based biopharmaceutical. This review highlights several bioinformatic tools for *ab initio* detection of potentially unstable DNA regions, as well as techniques used for mutation detection in nucleic acids, with particular emphasis on pDNA.

Received 21 October 2010
Revised 9 December 2010
Accepted 15 December 2010

Keywords: DNA vaccine · Genetic instability · Mutation · Plasmid DNA · Quality control

1 Introduction

The global rise in the number of applications involving therapeutic plasmid DNA (pDNA) is creating a demand for large amounts of highly stable and purified molecules. These advances in the development of novel plasmid-based biopharmaceuticals for DNA vaccination and gene therapy should also conform to current Good Manufacturing Practice (cGMP) guidelines and comply with the regulatory policies created by the Center for Biologics Evaluation and Research of the US Food and Drug Administration (CBER/FDA) or by the European Medicines Agency (EMA). In 1996, the FDA published several guidelines outlining recommendations to assist DNA vaccine developers during the manufacturing, preclinical and clinical stages of product development/implementation [1]. In that document, safety concerns were raised regarding the need to tightly control the native DNA sequence and potential integration events into host genomic DNA. These guidelines were revised in 2007 to reflect the accumulated preclinical and clinical experience acquired during the development of several DNA vaccines [2, 3]. Of particular relevance, before entering phase I clinical trials, are issues dealing with product identity and stability. In other words, to assure lot-to-lot consistency, there is the challenge to adopt or develop appropriate assays to monitor the occurrence of spontaneous changes in the parental sequence of a plasmid. According to the World Health Organization (WHO) guidelines, any modification of the native DNA sequence should be immediately detected and explained, as rearrangements of the plasmid within the host bacterial cell are not acceptable [3]. Typically, agarose gel electrophoresis of pDNA digested with one or multiple enzymes is used as a first approach to monitor identity and detect structural variants appearing at higher concentrations. It is also recommended that manufacturers provide the complete annotated sequence of the plasmid, avoid unstable sequences, and check its identity with the host genome by using international databases such as NCBI (www.ncbi.nlm.nih.gov/).

According to the International Conference on Harmonization (ICH) guideline Q5C, regarding the

Correspondence: Dr. Pedro H. Oliveira, IBB, Center for Biological and Chemical Engineering, Instituto Superior Técnico, Av. Rovisco Pais 1049-001 Lisbon, Portugal
E-mail: pcoliveira@ist.utl.pt
Fax: +351-218419062

Abbreviations: DGGE, denaturing gradient gel electrophoresis; **DHPLC**, denaturing high-performance liquid chromatography; **HRM**, high-resolution melting; **indel**, insertion/deletion; **IS**, insertion sequence; **pDNA**, plasmid DNA; **TGGE**, temperature gradient gel electrophoresis

testing of biotechnological and biological products [4], there is no unique stability-indicating assay or parameter able to describe the stability characteristics of a given biotechnological product. This implies that depending on the type and frequency of mutations eventually present in a plasmid population (point mutations, large deletions or insertions), the manufacturer should propose a battery of high-quality and sensitive assays able to detect and quantify any structural change.

The next sections deal in more detail with available bioinformatic tools for *ab initio* detection of potentially unstable DNA regions and strategies currently used to detect and/or quantify mutations in nucleic acids. Practical examples are given from studies performed with pDNA.

2 Bioinformatic and mathematical tools for *ab initio* detection of potentially unstable DNA regions and quantification of mutant forms

In a plasmid-based drug, the primary sequence assumes paramount importance, and prior knowledge about the presence of sequences able to induce genetic instability is crucial from an analytical perspective. Higher order non-B conformations (e.g., Z-DNA, slipped DNA, hairpins/cruciforms, triplexes, quadruplexes) usually arise from the presence of repeated DNA motifs, and have been shown to cause genetic rearrangements such as deletions, duplications, inversions or translocations (see recent reviews [5, 6]). In a previous study, Cooke and co-workers [7] evaluated the impact of certain non-B structures on the stability, yield and topology of a pBluescript vector, and found left-handed Z-DNA (resulting from alternating purine-pyrimidine tracts) to be clearly more unstable than triplex, bends and quadruplex structures. Moreover, the presence of triplex structures also led to a decrease in the amount of supercoiled plasmid produced by the *E. coli* host. The latter data, together with additional studies [8–12], highlight the need to pinpoint all DNA repeats (and other unstable motifs) having high recombination potential, preferentially at early stages of plasmid design. In practice this can be achieved using publicly available *ab initio* generic repeat detection tools that allow the user to search for exact or degenerate repeats (see Table 1). One of these tools is Reputer, which computes all repeats and palindromes with minimal length *l* (8 bp through web interface), also allowing the introduction of a Hamming or Levenshtein (edit) distance [13]. Others, such as RepeatScout [14] and Repeatoire [15] allow an efficient identification of repeat families via local multiple alignment heuristics, or, such as Repseek, account for nucleotide compositional biases and provide a statistical basis to evaluate the significance of a repeat [16]. Alternatively, if one wishes to focus on particular repeat classes, several programs are available, namely Tandem Repeats Finder (TRF) [17] and TRED [18] for tandem repeats, and PILER [19] for several specific classes of repeats. Other computational tools for *ab initio* detection of particular repeat elements typically found in larger genomes are also available and have been properly reviewed [20]. The presence of tandemly repeated G quartets may lead to the formation of G-quadruplex DNA motifs, which are known to be genetically unstable and prone to recombination. Kikin et al. [21] have developed a web-based server, QGRS Mapper, able to predict quadruplex-forming G-rich sequences (QGRS) in nucleotide sequences provided by the user. This tool enables the composition of the quadruplex to be defined and provides several options for data output, including an interactive graphical interface. Similar tools for quadruplex prediction and analysis include Quadparser [22], Quadfinder [23] and Quadpredict [24]. Changes in structural topology (e.g., curvature induced by properly phased AT tracts) play an important role in the control of gene expression but might also potentiate illegitimate recombination. In this sense, researchers interested in looking at the flexibility/curvature and other physicochemical properties of particular DNA regions might find the tools developed by Vlahovicek et al. [25] useful (see Table 1 for more details).

Another problem possibly arising during the manufacture of a plasmid biopharmaceutical concerns the transposition of mobile elements such as insertion sequences (ISs) from the bacterial genome to the vector. A recent work reports an extreme case of toxic plasmid clone instability mediated by multiple IS insertions [26]. The authors suggest that these events may prevent the maintenance of a clone developed for vaccine production and claim that IS transposition occurs more frequently than commonly anticipated (combined transposition rate of 10^{-8}/gene/generation for resident IS elements). Other examples found in the literature include the transposition of IS*1* into the vicinity of the neomycin resistance gene (neo^r) in a HIV DNA vaccine [27] and of IS*2* upstream the neo^r gene of the DNA vaccine backbone pCIneo [28]. Putative IS elements appearing as unexpected additional fragments in a vector can be readily identified by submitting its sequence to specialized databases such as IScan [29], ISfinder [30] or ISbrowser [31].

Table 1. Computational tools for *ab initio* detection of interspersed repeats, tandem repeats, insertion sequences (IS), G-quadruplexes, and prediction of physicochemical properties in DNA

Structural feature	Tool	URL/Observations	Ref.
Exact or degenerate direct and inverted repeats	Reputer	– http://bibiserv.techfak.uni-bielefeld.de/reputer – Allows the definition of Hamming or edit distance. Web search is limited to a minimal repeat length of 8 bp – Runs on Linux/Mac OS X/Solaris/Irix/Alpha platforms and through web interface – Includes a visualization tool that allows identifying the position of repeats	[13]
	RepeatScout	– http://bix.ucsd.edu/repeatscout – Runs on Linux/Mac OS X/Free BSD/DEC Tru64 platforms – Identifies families by string extension techniques	[14]
	Repeatoire	– http://wwwabi.snv.jussieu.fr/public/Repeatoire – Available to Mac OS X platforms	[15]
	Repeat Finder	– http://www.cbcb.umd.edu/software/RepeatFinder – Runs on Linux/Solaris/Alpha/UNIX platforms – Uses *k*-mer substrings identified by Reputer	[96]
	Repseek	– http://wwwabi.snv.jussieu.fr/public/RepSeek – Uses a statistic of extremes to evaluate the significance of repeats – Runs on Linux/Mac OS X platforms	[16]
Tandem Repeats	Tandem Repeats Finder (TRF)	– http://tandem.bu.edu/trf/trf.html – Finds all repeats with period size between 1 and 2000 bp – Runs on Solaris/Linux/Mac OS X and Windows platforms as well as through web interface. Older versions for other platforms are also available	[17]
	TRED	– http://tandem.sci.brooklyn.cuny.edu – Available through web interface. Source code also available upon request	[18]
Insertion Sequences (IS)	ISFinder	– http://www-is.biotoul.fr – Database of IS's isolated from eubacteria and archae	[30]
	IScan	– http://www.bioc.uzh.ch/wagner/publications-software.html – Runs on UNIX platforms – Requires an IS database such as ISFinder	[29]
	ISBrowser	– http://www-genome.biotoul.fr/ISbrowser.php – Extension of ISFinder that allows visualization of the position, orientation and distribution of complete and partial IS's in individual prokaryotic genomes	[31]
G-quadruplexes	QGRS Mapper	– http://bioinformatics.ramapo.edu/QGRS/index.php – Available through web interface – Capable of retrieving sequences from public databases. Allows interactive graphic representation of data	[21]
	Quadparser	– http://www.quadruplex.org – Runs on Mac OS X and Windows platforms as well as through web interface. Source code available upon request	[22]
	Quadfinder	– http://miracle.igib.res.in/quadfinder – Available through web interface – The user may download the results and use the output to search homologous sequences or di-nucleotide frequencies	[23]
	Quadpredict	– http://www.quadruplex.org – Available through web interface – Determines the thermodynamic stability of previously unmeasured quadruplexes from the sequence information alone	[24]
Curvature, bendability and other physicochemical properties	plot.it bend.it model.it	– http://hydra.icgeb.trieste.it/dna – plot.it generates parametric plots of 45 physicochemical and statistical parameters of DNA – bend.it calculates DNA curvature and bendability – model.it creates 3D models of DNA and outputs it in pdb format	[25]

Besides bioinformatic tools, mathematical models have also been devised aiming at describing repeat-mediated recombination events in terms of their frequency. Fujitani and Kobayashi [32] have explained the dependence of recombination frequency on repeat length, for two double-stranded DNA molecules, through a mechanistic model based on a time-dependent random-walk of a branch point connecting the two homologous segments. This study however, only analyzes the effect of repeat length on recombination frequency and not the distance between homologues. Our group developed a meta-analysis-driven non-linear mathematical function able to predict recombination frequency in pDNA harboring direct repeats [33]. This function takes the repeat and spacer lengths as input variables, and also takes into consideration the strain genotype in terms of its *rec*A background. To ascertain the real abundance of repeats in plasmids, we have also used Repseek and TRF to detect all direct, inverted and tandem repeats with high recombination potential in both natural plasmids of bacteria and commonly used vectors for protein expression and DNA vaccine development [12]. Through this study, a general overrepresentation of repeats located in eukaryotic elements and non-coding regions of plasmid vectors was found, which strongly encourages manufacturers to consider the inclusion of alternative promoter elements and to adopt strategies for plasmid minimization [6, 34–36]. Some of these unstable repeats have in fact been shown to give rise to deletion formation events in candidate molecules for DNA vaccination [8, 28].

We anticipate that greater availability of even more robust, potent and dedicated bioinformatic and mathematic tools will make it easier for the researcher to predict the full recombinogenic potential of a plasmid more rapidly.

3 Direct DNA sequencing

Significant advances in the field of DNA sequencing technology have been spurred to a large extent by the increasing efforts to analyze the complete sequences of the human genome and other organisms. In fact, given that the purpose of DNA sequencing is to determine the original sequence of a molecule and not to detect minor mutations that may be present, its use for quality control of plasmid preparations can be considered a paradox, as it will not detect genetic alterations present in small percentages [37]. Even so, DNA sequencing is commonly regarded as the gold standard for the detection and analysis of sequence variation. Issues associated with a lack of sensitivity can be ameliorated through the use of target-enrichment strategies such as uni- or multiplex PCR, molecular invasion probes or hybrid capture (reviewed in [38]). Moreover, significant progress has been made towards developing second-generation high-throughput low-cost sequencing strategies as an alternative to the traditional capillary-based Sanger biochemistry (reviewed in [39, 40]). Some examples include microelectrophoretic sequencing [41], sequencing by hybridization [42], cyclic-array sequencing [43], and single-molecule sequencing [44].

Assessment of the overall performance of a sequencing platform/methodology is also of vital importance, and should involve standardized benchmarks such as the use of external reference controls [45, 46], the adoption of quality metrics such as *phred* or *phred*-like scores or other quality control algorithms [39]. An example of the weaknesses associated with a supposedly standardized method such as DNA sequencing was brought to light by the EQUALseq program funded by the European Union between 2003 and 2005 [47]. Within the frame of this program, 60 laboratories from 21 countries were subjected to an external quality assessment (EQA) to evaluate their accuracy in determining the correct sequence stretch of pDNA molecules and PCR products. After 17 laboratories dropped out the EQA for various reasons, the sequencing results sent by the remaining laboratories (the majority using capillary electrophoresis) revealed considerable variations in the quality of primary data and evaluation [47]. The authors also suggest that EQAs should be made mandatory for laboratories providing sequencing services, an idea that was later reinforced by Márki-Zay and coworkers [48]. Table 2 briefly summarizes some advantages and disadvantages of DNA sequencing and other techniques described in this review.

4 Conformation-based electrophoretic methods for DNA separation

Electrophoresis refers to the migration of charged macromolecules through a porous matrix as a result of an applied electric field. It is commonly performed in a slab gel-based format (agarose or polyacrylamide) or in capillaries (capillary electrophoresis, CE). For routine analysis, gel inspection of a DNA molecule is usually the first and often sole approach to rapidly screen for putative large genetic variations. The DNA molecule can be digested with one or more restriction enzymes and the resulting fragments separated and identified in

Table 2. Brief summary of the advantages and disadvantages of several techniques used for mutation detection in nucleic acids

Technique(s)	Advantages	Disadvantages	Ref.
Direct DNA sequencing	– Provides the whole context of a mutation (type and location) – Although limited in terms of read-length and precision (error rates between 0.001%–1%), Sanger sequencing is more appropriate to sequences in the kb-Mb range	– Target enrichment is often needed to increase the amount of starting material – Next-generation platforms reduce the per-base cost of sequencing but are more appropriated to larger genomes	[38, 39, 97]
Single-strand conformation polymorphism (SSCP)	– Simple, relatively inexpensive – Throughput is increased in capillary or microcapillary format – Sensitivity may reach 90–100% under ideal running conditions	– Fragment sizes are limited (<0.3 kb) to ensure high sensitivity – Sensitivity is easily affected by fragment size, temperature and gel composition	[98]
Heteroduplex analysis (HA)	– Simple, relatively inexpensive – Throughput is increased in capillary or microcapillary format	– Depending on the platform, fragment sizes are limited (<0.6 kb)	[98]
Temperature or denaturing gradient gel electrophoresis (TGGE or DGGE)	– Simple, relatively inexpensive – Throughput is increased in capillary or microcapillary format (particularly for TGCE)	– Fragment size is limited (usually <1 kb) – Considerable gel-to-gel variation	[98]
Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS)	– Short analysis times – High throughput potential – Allows SNP detection	– Preparation of sample prior to analysis is costly and time-consuming – Fragments should be small (<100 nucleotides) to ensure high sensitivity	[60]
Electrospray ionization mass spectrometry (ESI-MS)	– Larger DNA molecules can be analyzed (e.g., pDNA, viral DNA) – May be used for screening unknown polymorphisms	– Sample throughput is lower than for MALDI-MS – Provides no information about the exact location of a mutation	[60]
Denaturing high performance liquid chromatography (DHPLC)	– High specificity and sensitivity (~100%)	– Relatively long assay times – Throughput is limited – Expensive equipment	[98, 99]
Chemical or enzymatic cleavage of mismatched nucleotides	– Analysis of larger (<3 kb) DNA fragments is possible – Mutation position can be known through cleavage pattern	– Ability to multiplex is reduced – Enzymes are expensive – May involve toxic chemicals – Sensitivity is poor (e.g., 1 mutant in 32 molecules for Surveyor; 50% of heteroduplex for carbodiimide)	[57–59, 99]
PCR-based	– Often needed as target-enrichment strategy for downstream assays – Real-time PCR allows quantification and analysis in a closed-tube environment	– Probability of contamination increases if nesting is needed – Errors may be introduced during amplification due to enzyme slippage or stuttering	[80]
High-resolution melting (HRM) analysis	– Low-cost, highly sensitive and specific closed-tube method – Does not require post-PCR separation	– Sensitivity is higher when small fragments are used (usually between 0.1 and 0.3 kb)	[89, 100]

agarose or polyacrylamide gels. However, only aberrant forms generated at large amounts either spontaneously [8, 9, 11] or under stress selection [8, 10], have the chance to be detected in gels since this format only detects those forms showing up at percentages typically higher than 5% [49]. This lower limit of detection is obviously a function of the gel type, of the size of the target fragment and of staining and destaining procedures adopted by the user. While agarose gel electrophoresis is best suited to resolve DNA fragments in the size range between 0.1 and 20 kb, diffusion effects within the matrix

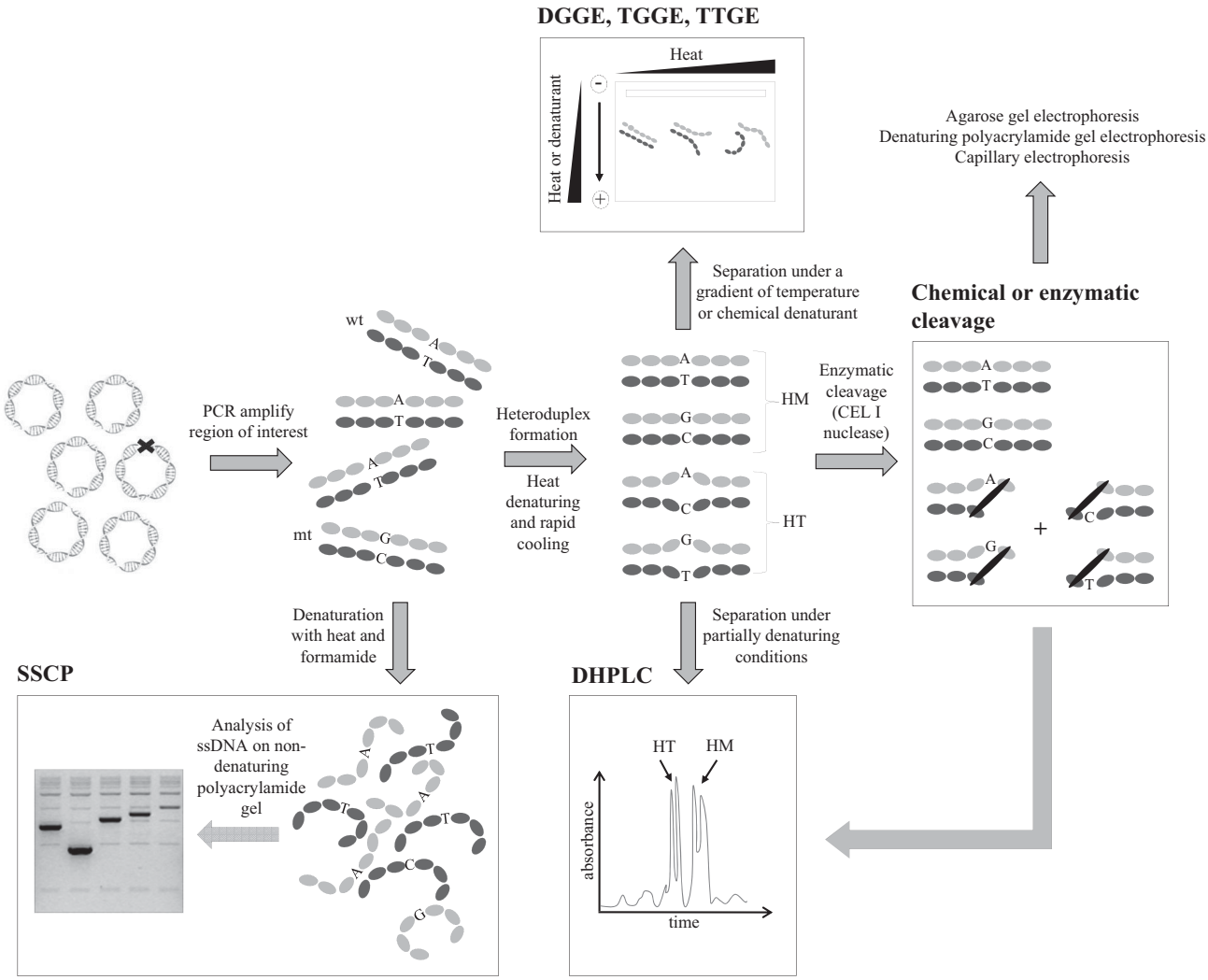


Figure 1. Schematic overview of heteroduplex-based methods used for detection of small mutations in DNA. To evaluate the presence of a low-frequency plasmid mutation (black cross), the region of interest is PCR-amplified, giving rise to wild-type (wt) and mutant (mt) amplicons. In the case of single-strand conformation polymorphism (SSCP) the PCR product is denatured with heat and formamide, and electrophoresed in a nondenaturing polyacrylamide gel. The relative electrophoretic mobility is therefore a function of the conformation of the single-stranded molecules. Alternatively, the PCR product may be denatured and subjected to rapid cooling to form homo- (HM) and heteroduplexes (HT). The latter form the basis of heteroduplex analysis (HA), which consists of analyzing their different mobilities in nondenaturing polyacrylamide gels or higher resolution improved matrixes such as mutation detection enhancement (MDE) gels. DGGE and TGGE are alternative methods for HA, based on the mobility of double-stranded molecules under increasing concentrations of denaturant (formamide or urea for DGGE) and temperature (for parallel and perpendicular TGGE). On the other hand, temporal temperature gel electrophoresis (TTGE), exploits a linear and uniform increase of temperature with time together with a constant chemical denaturant. In the case of denaturing high performance liquid chromatography (DHPLC), homo- and heteroduplexes are separated at partially denaturing temperatures within a hydrophobic matrix cartridge containing a mobile phase of acetonitrile and triethylammonium acetate buffer (TEAA). Chemical and enzymatic cleavage of mismatched nucleotides offers an alternative to the former methods, which allows pinpointing the position of the mutation. The illustration refers to the enzymatic cleavage of a mismatch by the single-strand specific plant endonuclease CEL I. The cleavage products can be subsequently analyzed by one of the several methods already described (e.g., agarose or denaturing polyacrylamide gel electrophoresis, capillary electrophoresis, DHPLC).

hamper the visualization of smaller fragments. In this case, the use of native (non-denaturing) polyacrylamide gel electrophoresis (PAGE) is more appropriate as it can give a resolution down to the single base. Typically, for detection of small mutations, several slab gel-based methods are frequently employed including single-stranded conformational polymorphism (SSCP), heteroduplex analysis (HA), and temperature or denaturing gradient gel electrophoresis (TGGE or DGGE). SSCP exploits the differences in electrophoretic mobility of wild-type and mutant ssDNA species in a non-denaturing polyacrylamide gel, as they naturally fold according to their primary sequence (see Table 2 and

Fig. 1). On the other hand, HA involves the denaturation and re-annealing of a mixture of wild-type and mutant fragments, formation of a mismatch (heteroduplex), and subsequent electrophoretic separation of homoduplexes from heteroduplexes in nondenaturing polyacrylamide gels or higher resolution improved matrixes such as mutation detection enhancement (MDE) gels. Finally, TGGE and DGGE exploit the differences in electrophoretic mobility associated with the migration of dsDNA into a gradient of increasingly denaturing conditions (heat, formamide or urea are commonly used). In other words, the existence of DNA polymorphisms will cause the corresponding fragments to melt at different positions in the gradient, thus affecting migration on the gel (Fig. 1). An alternative temporal temperature gel electrophoresis (TTGE) can also be used for HA, in which a linear and uniform increase of temperature is used during separation. The successful detection of single-base changes using these methods will vary not only with the electrophoretic conditions used, but also with fragment size, composition and type of mutation [50].

All of these approaches have also been exploited in capillary format, which surpasses gel separation in analysis time and resolving power. Capillary array instruments such as the MegaBACE 1000 (GE Healthcare) and the ABI Prism 3730*xl* (Applied Biosystems) have been designed in a 96-capillary format thus allowing faster sequencing runs and a throughput of several thousand genotypes per day. Nevertheless, the tendency observed over the past few years has been to adapt electrophoretic-based methods to microchip platforms. This greatly reduces the cost of the devices, and further increases automation, sample throughput and hyphenation with other analysis steps [51–53]. One example is that of Zhang et al. [54], who have constructed a slantwise radiative heating system for chip-based temperature gradient capillary electrophoresis (TGCE). This system allowed an effective detection of point mutations in fragments amplified from plasmid templates using a wide temperature gradient of 10°C and a short separation length of only 3 cm.

We next discuss how electrophoretic separation can be coupled to chemical or enzymatic modification of mismatched nucleotides for mutation detection.

5 Methods based on chemical or enzymatic cleavage of mismatched nucleotides

Other methods are available that involve chemical or enzymatic cleavage of mismatches in heteroduplexes (reviewed in [55]). Chemical cleavage traditionally involves treating the heteroduplex with osmium tetroxide (or with the less-toxic potassium permanganate) for modification of mismatched thymine residues, and hydroxylamine for modification of mismatched cytosine residues. Both steps are followed by piperidine cleavage at the modified mismatched base and subsequent electrophoretic analysis. Although able to detect point mutations, insertions and deletions (indels), chemical cleavage typically requires time-consuming ethanol precipitation steps, toxic reagents, and takes approximately 10 h to be completed [56]. The ability of carbodiimide to establish covalent bonds with the free imino moiety of thymine and guanine residues in ssDNA was later exploited as a non-toxic and inexpensive alternative chemical method to detect mismatched DNA [57]. This method involves the denaturation and cooling of a PCR-amplified target region for heteroduplex formation, addition of carbodiimide to PCR products, and a final carbodiimide-induced interrupted extension with primers labeled with flurophores. The latter are then subjected to electrophoretic separation. This method was able to detect all four classes of mismatches in engineered pBR322 molecules as well as several new polymorphisms in genes coding for inherited renal diseases [57]. An alternative variant involving enzymatic modification of heteroduplexes with a plant DNA endonuclease from the CEL I family of enzymes was employed [58, 59]. After heteroduplex formation, an enzyme named Surveyor™ (Transgenomic, Gaithersburg, MD, USA), cleaves the 3′ extremity of any mismatch site in both DNA strands including base substitutions and indels up to 12 nucleotides (Fig. 1). This technology was used to detect multiple desired and undesired variations present in pET-derived amplicons as large as 3 kb, with a sensitivity as low as 3–6% using agarose gel electrophoresis [58, 59].

6 Mass spectrometry-based methods

Mass spectrometry (MS) is known to be one of the most sensitive and accurate methods for identification and structural characterization of biological substances [60]. In the particular case of nucleic acids, analysis by MS was for many years hampered by their polarity, high molecular mass and thermal lability. Only with the advent of soft ionization techniques such as the matrix-assisted laser desorption ionization (MALDI) [61] and electrospray ionization (ESI) MS [62], has the analysis of large biomolecules (100–500 kDa) become possible. Briefly, MALDI involves a first step of incorporation of the

DNA-based analyte into a matrix of glycerol [63], succinic acid [64] or 3-hydroxypicolinic acid [65], followed by disintegration of the solid through irradiation with a short UV or IR laser pulse that leads to a rapid solid-to-gas phase transition (MALDI plume) harboring analyte ions and molecules as well as some clusters or aggregates from the matrix. The ions generated are commonly analyzed with a time-of-flight (TOF) mass analyzer that separates particles in a field-free drift region according to their mass-to-charge ratio (m/z). One advantage of MALDI is the usual short analysis time and potential for high sample throughput [66]; however, the occurrence of depurination and gas-phase fragmentation as well as adduct formation with potassium and sodium decreases the resolution and imposes an upper limit for analysis of oligonucleotides of about 50–100 nucleotides. ESI on the other hand, involves ion transfer from solution to gas phase, after which analysis proceeds by MS (ESI-MS). Because it is a very gentle ionization technique, DNA molecules usually do not undergo structural degradation, thus allowing the analysis of larger molecules such as pDNA.

Cheng et al. [67] have successfully determined the molecular weight of pGEM-5S (3126 bp) using ESI with Fourier transform ion cyclotron resonance (FTICR) with an average accuracy of 0.2%. The authors also describe the simultaneous presence of concatemer-like higher mass plasmid species, although no substantial ion population with exactly twice the plasmid mass was detected to support the concatemer hypothesis. Interestingly, these unexpected plasmid forms strongly resemble those generated upon rearrangement of direct repeats, leading to duplications and/or triplications [8]. Approximately at the same time, Fuerstenau and Benner [68] used an ESI-TOF approach to obtain molecular mass information for DNA samples ranging from 2.8 to 31 MDa (4.41–48.5 kb). The authors used this technique to determine the molecular weight of several plasmid samples, namely pBR322 (4363 bp) and pMSG-Cat (8405 bp). An alternative laser-induced liquid bead ion desorption (LILBD) MS method was also shown to successfully evaluate the molecular weight of pUC19 (2686 bp) [69]. The latter method makes use of micro-droplets as carriers for the analyte, which are subjected to individual irradiation under vacuum by mid-IR laser pulses. With LILBD, only a few microliters of a micromolar solution are needed and the analyte is studied in a more or less native environment [69]. Instead of analyzing the whole plasmid molecule, alternative strategies focusing on enzymatic digestion of pDNA prior to MS [63, 70], MS analysis of PCR products [63, 70–72] and synthetic oligonucleotides [63, 66, 69, 73] have also been described. MS has been used for detection of DNA adducts [74], and particularly for genotyping of indels and single nucleotide polymorphisms (SNPs) [71, 75]. In fact, the combination of Sanger dideoxy chemistry (or other) with MALDI-TOF-MS, serves as a rapid and efficient analytical tool for analysis of short DNA stretches, making it appropriate for resequencing strategies.

7 Denaturing HPLC

Denaturing HPLC (DHPLC) is an ion-pair reverse-phase chromatographic technique with high sensitivity (>96%) and specificity (>99%) for detection of single-base mismatches and small, unknown (and known) deletions/insertions in DNA fragments as large as 1 kb [76]. These fragments are previously PCR amplified from a target molecule (genomic DNA or pDNA) in which mutations are to be screened, and mixed with a wild-type reference PCR product. This mixture is then denatured, allowed to reanneal with the purpose of generating heteroduplex products and applied to a hydrophobic matrix cartridge containing a mobile phase of acetonitrile and triethylammonium acetate buffer (TEAA). Under partially denaturing conditions, homo- and heteroduplex fragments are separated according to their relative thermal stability. The positive triethylammonium ions bind to DNA and to the hydrophobic groups of the matrix, which are then released by an increasing acetonitrile gradient. Heteroduplexes are more weakly bound to the matrix and are thus released earlier than homoduplexes (Fig. 1).

A multiplex PCR approach followed by DHPLC analysis has been employed previously for the detection of genetic changes in plasmid-mediated *amp*C β-lactamase genes from Gram-negative bacteria [77]. The authors highlight the time reduction and increased sensitivity of the method in comparison with standard electrophoretic and agarose analysis. DHPLC has also been used to monitor partial denaturation induced by unstable AT-rich domains in pBR322 restriction fragments [78]. Avoiding such domains during plasmid design is essential, as they are known to be a major cause of DNA unwinding and bending, polymerase slippage and illegitimate recombination. An important increase in DHPLC throughput was achieved by the group of P. Oefner, through the use of monolith poly(styrene/divinylbenzene) capillary columns and laser-induced fluorescence detection of samples labeled with up to four different fluorophores [76, 79].

Currently commercially available DHPLC instruments include the WAVE DNA Fragment Analysis System from Transgenomic (www.transgenomic.com) and the Helix system from Varian (www.varianinc.com). Both equipments operate on 96- or 384-well plate formats and allow the incorporation of fluorescence detectors, enabling fragment labeling with distinct fluorophores.

8 PCR-based techniques

In most of the detection methods seen so far, PCR is usually needed to generate an amplicon that is then subjected to analysis by some other technique that searches for structural variation. However, in some cases, PCR is used as the sole or primary technique for mutation detection. One such example is real-time PCR in which primers are specifically designed to target one or more sequences of interest, which are then detected by means of emitted fluorescence. This fluorescence results from the use of nonspecific binding dyes (e.g., SYBR Green I) or labeled probes such as TaqMan probes, molecular beacons, Scorpion Primers or FRET probes (reviewed in [80]). Real-time PCR has the advantage of being a simple closed-tube method of analysis capable of simultaneously processing single or multiplexed samples in a 96- or 384-well format. The major drawbacks of the method come from the high cost of the equipment and labeled probes, on the likely generation of unwanted side products when using nonspecific dyes, and as a result, on the need to carefully interpret melting curves [80]. One example of the use of real-time PCR using SYBR Green I technology describes the detection and quantification of a 28-bp direct repeat-mediated deletion-formation event in the DNA vaccine backbone pCIneo [8]. Because the repeats were promptly identified and the spacer sequence between them was large (1.6 kb), the use of a single pair of flanking primers and short amplification times were enough to selectively amplify the deleted form of the plasmid. In this study, approximately 1.6×10^4 aberrant molecules were found to be spontaneously generated in a population of 2.0×10^5 *E. coli* cells [8]. This example shows how prior knowledge concerning the location of a recombination hotspot or putative fragile region allows the most adequate set of primers or probes to be designed to target novel aberrations. As an alternative to quantitative real-time PCR, nested or hemi-nested strategies could be adopted instead. The latter provides a high sensitivity of detection, although at the cost of an increased probability of sample contamination.

When designing specific PCR strategies to target recombination events in shorter repeats [e.g., simple sequence repeats (SSRs) or microsatellites], a common problem is stuttering caused by slipped-strand mispairing during polymerase amplification. This involves partial dissociation of the extended primer from the template followed by erroneous reannealing upstream or downstream of the correct target sequence. This problem seems to depend on the number of nucleotides with which the polymerase interacts [81], and can be diminished using thermolabile polymerases [82] or proofreading enzymes fused to nonspecific dsDNA binding domains [81]. An example of polymerase-induced indel mutations occurs within homopolymeric A/T tracts, such as the polyA tails, used in plasmids for mRNA maturation/stability and nuclease resistance [83]. These sequences have been shown to recombine in pUC18 during PCR amplification at a rate/repeat/PCR cycle of 1.5×10^{-2} with contractions outnumbering expansions by fivefold [84].

Commonly referred to as high-resolution melting (HRM) analysis, this technique is an extension of the classical melting curve analysis, and a low-cost, post-PCR, highly sensitive closed-tube method for genotyping and mutation screening (down to the SNP level). It involves PCR amplification of a region of interest, in the presence of a suitable dye, followed by a step of gradual heating and fluorescence monitoring. HRM can be performed in fully dedicated equipments such as the Light Scanner (Idaho Technology Inc., Salt Lake City, UT), or alternatively, in combination with real-time PCR [85]. Because it is a closed-tube method, both the risk of contamination and analysis time are greatly reduced. The overall amplification and melting acquisition steps require about 30 min if performed in capillaries or about 1.5 h if 96- or 384-well plates are used [86]. Use of SYBR Green I is not recommended as it completely inhibits PCR at >50% saturation and may redistribute during melting curve analysis [87]. For this reason, specifically developed HRM saturating dyes have emerged, that include LCGreen [87], LCGreen Plus (Idaho Technology Inc.), Light Cycler 480 ResoLight Dye (Roche, Indianapolis, IN) and Syto9 (Invitrogen, Carlsbad, CA). Although HRM profiling has been comprehensively used over the years to screen mutations in genomic and mitochondrial DNA, the group of C. Wittwer has applied it to the detection of unique A, T, G or C mutations alone or in mixed formulations in engineered plasmids with sensitivities and specificities above 95% [88, 89].

Rather than using it as the main tool to detect structural variation, PCR is commonly used for the amplification of unknown low-prevalence variants

for downstream analysis by some other technique (e.g., Sanger sequencing, DHPLC, MS, and others). However, standard PCR strategies rely on an equal selectivity for both wild-type and mutant molecules. A COLD-PCR strategy has been proposed [90] that aims at enriching minor unknown DNA variants and further improving the sensitivity of downstream techniques. It exploits the differences between melting temperatures of homo- and heteroduplexes by performing denaturation steps at a critical temperature below the melting temperature. By replacing PCR with COLD-PCR the authors obtained mutation enrichments of 10–100-fold, thus allowing a downstream MALDI-TOF analysis to further detect mutations with a minimum prevalence of 0.1–0.5% instead of the traditional 5–10% [90]. Recently, the same group presented an improved and complete enrichment (*ice*-COLD-PCR) that makes use of a synthetic reference sequence that is able to bind the wild-type sequence during amplification, leading to a preferential amplification of the mutant form [91]. Massive parallelization of enrichment reactions (1.5 million amplifications) has also been performed [92] using a microdroplet-based PCR approach in which a population of small droplets each containing a pair of distinct primers is subjected to thermal cycling prior to sequencing. Other target-enrichment strategies have been recently reviewed [38].

9 Alternative techniques for mutation detection

A number of alternative techniques have been proposed for monitoring pDNA degradation and mutation occurrence. One approach is based on the change in surface charge density that accompanies DNA degradation, and involves the use of direct current conductivity and dynamic dielectric relaxation techniques [93]. These authors found static conductivity to semiquantitatively characterize degradation (nicks as well as single- and double-strand breaks) in 20-kb plasmids upon heating to 80°C over 1 h, whereas dielectric relaxation was found to be less sensitive.

Chakrabarti and co-workers [94] developed a mismatch-capture methodology for detection of T→G transversions in plasmid-based heteroduplex fragments based on the ability of MutY glycosylase to excise mismatched adenines, generating an unsaturated open chain aldehyde. The latter is covalently linked to a biotinylated hydroxylamine and subsequently separated from non-biotinylated DNA using microspheres coated with streptavidin. This technique does not require previous knowledge of the sequence and, according to the authors, can be performed in non-uniform large (≥7 kb) DNA fragments and in the presence of 1000-fold excess of non-mutants molecules [94].

An alternative mismatch repair detection (MRD) method was developed for in vivo detection of DNA sequence variation, which exploits the methyl-directed mismatch repair system of *E. coli* [95]. It involves the cloning of DNA fragments to be screened into two MRD plasmids, followed by *E. coli* transformation with heteroduplexes generated from these constructs. This allowed a blue/white screening of mismatches in DNA fragments of up to 10 kb in size and has potential for high-throughput genotyping and mutation detection in large DNA regions. Many alternative in vivo techniques used to screen for genetic instability have been developed and recently reviewed [49].

10 Conclusions

Despite the heralded potential of non-viral DNA vaccines, much is yet to be accomplished to ensure their clinical safety. Currently, there is a pressing need to understand how frequently these molecules acquire spontaneous mutations, integrate into the host genome, and ultimately, how these events might affect the efficacy of the vaccine and the patient itself. However, the lack of standardized quality parameters able to define plasmid structural (in)stability places a serious roadblock towards DNA vaccine development. First, in silico DNA repeat mapping is not routinely implemented during the early stages of plasmid design. This would provide a snapshot of potential hotspots for recombination, thus allowing manufacturers to predict with some degree of confidence the likelihood and outcome of a mutation. Second, plasmid deletion-formation, transposition and genome integration events are generally observed at low frequency. Unless these minor variants are preferentially selected during plasmid amplification by means of some selective pressure, they will not be detected by routine gel inspection. Therefore, techniques such as the ones described in this review should be adopted as benchmarks for plasmid mutation detection, although choosing the most appropriate one(s) is not always straightforward. This selection is often hampered by costly instrumentation or lack of skilled human resources and depends on the type/abundance of the mutation, size of fragment of interest, desired throughput and required sensitivity and specificity. We anticipate that fast developments in the field of mutation detection will most certainly occur in the near future, in view of the

comprehensive increase in the number of applications involving therapeutic pDNA. As more data on plasmid mutation becomes available, the FDA/EMA's guidance will also evolve towards more clear and precise recommendations on the safety boundaries for clinical applications.

This work was supported by Fundação para a Ciência e a Tecnologia (POCI/BIO/55799/2004 through Ph. D. grant BD/22320/2005 to P.H.O.) and by the MIT-Portugal Program.

The authors have declared no conflict of interest.

11 References

[1] CBER/FDA, Points to consider on plasmid DNA vaccines for preventive infectious disease indications. *Food and Drug Administration, Center for Biologics Evaluation and Research, Office of Vaccine Research and Review*. 1996, Docket No. 96-N-0400.

[2] CBER/FDA, Guidance for industry: considerations for plasmid DNA vaccines for infectious disease indications. Guidance document. 2007.

[3] WHO expert committee on biological standardization. Guidelines for assuring the quality and nonclinical safety evaluation of DNA vaccines. *WHO Tech. Rep. Ser.* 2007.

[4] ICH guideline Q5C. Quality of biotechnological products: stability testing of biotechnological/biological products. 1995, 1–10.

[5] Zhao, J., Bacolla, A., Wang, G., Vasquez, K. M., Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* 2010, *67*, 43–62.

[6] Oliveira, P. H., Prather, K. J., Prazeres, D. M., Monteiro, G. A., Structural instability of plasmid biopharmaceuticals: Challenges and implications. *Trends Biotechnol.* 2009, *27*, 503–511.

[7] Cooke, J. R., McKie, E. A., Ward, J. M., Keshavarz-Moore, E., Impact of intrinsic DNA structure on processing of plasmids for gene therapy and DNA vaccines. *J. Biotechnol.* 2004, *114*, 239–254.

[8] Ribeiro, S. C., Oliveira, P. H., Prazeres, D. M., Monteiro, G. A., High frequency plasmid recombination mediated by 28-bp direct repeats. *Mol. Biotechnol.* 2008, *40*, 252–260.

[9] Borgi, I., Gargouri, A., A spontaneous direct repeat deletion in the pGEX fusion vector decreases the expression level of recombinant proteins in Escherichia coli. *Protein Expr. Purif.* 2008, *60*, 15–19.

[10] Kawe, M., Horn, U., Pluckthun, A., Facile promoter deletion in *Escherichia coli* in response to leaky expression of very robust and benign proteins from common expression vectors. *Microb. Cell Fact.* 2009, *8*, 8.

[11] Sun, Y., He, X., Liang, J., Zhou, X. et al., Analysis of functions in plasmid pHZ1358 influencing its genetic and structural stability in *Streptomyces lividans* 1326. *Appl. Microbiol. Biotechnol.* 2009, *82*, 303–310.

[12] Oliveira, P. H., Prather, K. J., Prazeres, D. M., Monteiro, G. A., Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. *Appl. Microbiol. Biotechnol.* 2010, *87*, 2157–2167.

[13] Kurtz, S., Schleiermacher, C., REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* 1999, *15*, 426–427.

[14] Price, A. L., Jones, N. C., Pevzner, P. A., De novo identification of repeat families in large genomes. *Bioinformatics* 2005, *21 Suppl 1*, i351–358.

[15] Treangen, T. J., Darling, A. E., Achaz, G., Ragan, M. A. et al., A novel heuristic for local multiple alignment of interspersed DNA repeats. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2009, *6*, 180–189.

[16] Achaz, G., Boyer, F., Rocha, E. P., Viari, A. et al., Repseek, a tool to retrieve approximate repeats from large DNA sequences. *Bioinformatics* 2007, *23*, 119–121.

[17] Benson, G., Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999, *27*, 573–580.

[18] Sokol, D., Benson, G., Tojeira, J., Tandem repeats over the edit distance. *Bioinformatics* 2007, *23*, e30–35.

[19] Edgar, R. C., Myers, E. W., PILER: Identification and classification of genomic repeats. *Bioinformatics* 2005, *21 Suppl 1*, i152–158.

[20] Treangen, T. J., Abraham, A. L., Touchon, M., Rocha, E. P., Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.* 2009, *33*, 539–571.

[21] Kikin, O., D'Antonio, L., Bagga, P. S., QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 2006, *34*, W676–682.

[22] Huppert, J. L., Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 2005, *33*, 2908–2916.

[23] Scaria, V., Hariharan, M., Arora, A., Maiti, S., Quadfinder: server for identification and analysis of quadruplex-forming motifs in nucleotide sequences. *Nucleic Acids Res.* 2006, *34*, W683–685.

[24] Stegle, O., Payet, L., Mergny, J. L., MacKay, D. J. et al., Predicting and understanding the stability of G-quadruplexes. *Bioinformatics* 2009, *25*, i374–382.

[25] Vlahovicek, K., Kajan, L., Pongor, S., DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* 2003, *31*, 3686–3687.

[26] Umenhoffer, K., Feher, T., Baliko, G., Ayaydin, F. et al., Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb. Cell. Fact.* 2010, *9*, 38.

[27] Prather, K. L., Edmonds, M. C., Herod, J. W., Identification and characterization of IS1 transposition in plasmid amplification mutants of *E. coli* clones producing DNA vaccines. *Appl. Microbiol. Biotechnol.* 2006, *73*, 815–826.

[28] Oliveira, P. H., Prazeres, D. M., Monteiro, G. A., Deletion formation mutations in plasmid expression vectors are unfavored by runaway amplification conditions and differentially selected under kanamycin stress. *J. Biotechnol.* 2009, *143*, 231–238.

[29] Wagner, A., Lewis, C., Bichsel, M., A survey of bacterial insertion sequences using IScan. *Nucleic Acids Res.* 2007, *35*, 5284–5293.

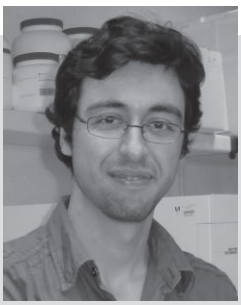
[30] Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. et al., ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 2006, *34*, D32–36.

[31] Kichenaradja, P., Siguier, P., Perochon, J., Chandler, M., ISbrowser: an extension of ISfinder for visualizing insertion sequences in prokaryotic genomes. *Nucleic Acids Res.* 2010, *38*, D62–68.

[32] Fujitani, Y., Kobayashi, I., Asymmetric random walk in a reaction intermediate of homologous recombination. *J. Theor. Biol.* 2003, *220*, 359–370.

[33] Oliveira, P. H., Lemos, F., Monteiro, G. A., Prazeres, D. M., Recombination frequency in plasmid DNA containing direct repeats – predictive correlation with repeat and intervening sequence length. *Plasmid* 2008, *60*, 159–165.
[34] Marie, C., Vandermeulen, G., Quiviger, M., Richard, M. et al., pFARs, plasmids free of antibiotic resistance markers, display high-level transgene expression in muscle, skin and tumour cells. *J. Gene Med.* 2010, *12*, 323–332.
[35] Mairhofer, J., Cserjan-Puschmann, M., Striedner, G., Nobauer, K. et al., Marker-free plasmids for gene therapeutic applications – lack of antibiotic resistance gene substantially improves the manufacturing process. *J. Biotechnol.* 2010, *146*, 130–137.
[36] Endmann, A., Baden, M., Weisermann, E., Kapp, K. et al., Immune response induced by a linear DNA vector: influence of dose, formulation and route of injection. *Vaccine* 2010, *28*, 3642–3649.
[37] Gluck, L., Alazay, L., Ratel, D., Puget, S. et al., Quality control of plasmid preparations. *Nat. Biotechnol.* 2001, *19*, 715.
[38] Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I. et al., Target-enrichment strategies for next-generation sequencing. *Nat. Methods* 2010, *7*, 111–118.
[39] Shendure, J., Ji, H., Next-generation DNA sequencing. *Nat. Biotechnol.* 2008, *26*, 1135–1145.
[40] Shendure, J., Mitra, R. D., Varma, C., Church, G. M., Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.* 2004, *5*, 335–344.
[41] Blazej, R. G., Kumaresan, P., Mathies, R. A., Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. *Proc. Natl. Acad. Sci. USA* 2006, *103*, 7240–7245.
[42] Gresham, D., Dunham, M. J., Botstein, D., Comparing whole genomes using DNA microarrays. *Nat. Rev. Genet.* 2008, *9*, 291–302.
[43] Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X. et al., Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005, *309*, 1728–1732.
[44] Healy, K., Nanopore-based single-molecule DNA analysis. *Nanomedicine (Lond)* 2007, *2*, 459–481.
[45] Rousseau, F., Gancberg, D., Schimmel, H., Neumaier, M. et al., Considerations for the development of a reference method for sequencing of haploid DNA – an opinion paper on behalf of the IFCC Committee on Molecular Diagnostics, International Federation of Clinical Chemistry and Laboratory Medicine. *Clin. Chem. Lab. Med.* 2009, *47*, 1343–1350.
[46] van der Straaten, T., Swen, J., Baak-Pablo, R., Guchelaar, H. J., Use of plasmid-derived external quality control samples in pharmacogenetic testing. *Pharmacogenomics*. 2008, *9*, 1261–1266.
[47] Ahmad-Nejad, P., Dorn-Beineke, A., Pfeiffer, U., Brade, J. et al., Methodologic European external quality assurance for DNA sequencing: the EQUALseq program. *Clin. Chem.* 2006, *52*, 716–727.
[48] Marki-Zay, J., Klein, C. L., Gancberg, D., Schimmel, H. G. et al., European external quality control study on the competence of laboratories to recognize rare sequence variants resulting in unusual genotyping results. *Clin. Chem.* 2009, *55*, 739–747.
[49] Wang, G., Zhao, J., Vasquez, K. M., Methods to determine DNA structural alterations and genetic instability. *Methods* 2009, *48*, 54–62.
[50] Highsmith, W. E., Nataraj, A. J., Jin, Q., O'Connor, J. M. et al., Use of DNA toolbox for the characterization of mutation scanning methods. II: Evaluation of single-strand conformation polymorphism analysis. *Electrophoresis*. 1999, *20*, 1195–1203.
[51] Chen, L., Ren, J., High-throughput DNA analysis by microchip electrophoresis. *Comb. Chem. High Throughput Screen*. 2004, *7*, 29–43.
[52] Buch, J. S., Kimball, C., Rosenberger, F., Highsmith, W. E. Jr. et al., DNA mutation detection in a polymer microfluidic network using temperature gradient gel electrophoresis. *Anal. Chem*. 2004, *76*, 874–881.
[53] Kailasa, S. K., Kang, S. H., Microchip-based capillary electrophoresis for DNA analysis in modern biotechnology: A review. *Sep. Purif. Rev.* 2009, *38*, 242–288.
[54] Zhang, H. D., Zhou, J., Xu, Z. R., Song, J. et al., DNA mutation detection with chip-based temperature gradient capillary electrophoresis using a slantwise radiative heating system. *Lab Chip* 2007, *7*, 1162–1170.
[55] Taylor, G. R., Enzymatic and chemical cleavage methods. *Electrophoresis* 1999, *20*, 1125–1130.
[56] Tabone, T., Sallmann, G., Chiotis, M., Law, M. et al., Chemical cleavage of mismatch (CCM) to locate base mismatches in heteroduplex DNA. *Nat. Protoc.* 2006, *1*, 2297–2304.

Gabriel Monteiro is Assistant Professor at Instituto Superior Técnico (IST), and a researcher at the Institute for Biotechnology and Bioengineering (IBB), Lisbon, Portugal. He received his undergraduate training at the University of Coimbra (1988), with an MSc in Biochemical Engineering (1991) and a PhD in Biotechnology (1998) both from IST. His main research area is in the development, production and purification of plasmids for DNA vaccination. He is also interested in the design, production and secretion of recombinant human proteins. He has authored more than 60 scientific publications including papers and book chapters.



Dr. Pedro Oliveira graduated from Instituto Superior Técnico (IST) in Lisbon, Portugal with an MSc in Biological Engineering (2005). He subsequently received his PhD (2010) from IST under the supervision of Professors Gabriel Monteiro and Miguel Prazeres with a thesis entitled "Structural instability in plasmid DNA". During this period he was also a visiting student (2008–2009) at the Massachusetts Institute of Technology (with Professor Kristala Prather) and Harvard Medical School (with Professor George Church). Currently, Dr. Oliveira is a post-doctoral fellow at IST, and is working on projects within the field of genetic instability, namely (i) evaluation of human stem cell genome integrity during their ex vivo expansion for clinical applications, (ii) computational analysis of human mitochondrial pathogenic deletions, and (iii) quality control of biopharmaceuticals.

[57] Tabone, T., Cotton, R., Mathew, N., Des Parkin, J., A novel detection method for heteroduplex DNA using carbodiimide-induced interrupted primer extension. *J. Nucleic Acids Inv.* 2010, *1*, e5.

[58] Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K. et al., Mutation detection using Surveyor nuclease. *Biotechniques*. 2004, *36*, 702–707.

[59] Qiu, P., Shandilya, H., Gerard, G. F., A method for clone sequence confirmation using a mismatch-specific DNA endonuclease. *Mol. Biotechnol.* 2005, *29*, 11–18.

[60] Oberacher, H., On the use of different mass spectrometric techniques for characterization of sequence variability in genomic DNA. *Anal. Bioanal. Chem.* 2008, *391*, 135–149.

[61] Karas, M., Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal. Chem.* 1988, *60*, 2299–2301.

[62] Whitehouse, C. M., Dreyer, R. N., Yamashita, M., Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.* 1985, *57*, 675–679.

[63] Berkenkamp, S., Kirpekar, F., Hillenkamp, F., Infrared MALDI mass spectrometry of large nucleic acids. *Science* 1998, *281*, 260–262.

[64] Nordhoff, E., Cramer, R., Karas, M., Hillenkamp, F. et al., Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Res*. 1993, *21*, 3347–3357.

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

[66] Nordhoff, E., Luebbert, C., Thiele, G., Heiser, V. et al., Rapid determination of short DNA sequences by the use of MALDI-MS. *Nucleic Acids Res.* 2000, *28*, E86.

[67] Cheng, X., Camp, D. G., 2nd, Wu, Q., Bakhtiar, R. et al., Molecular weight determination of plasmid DNA using electrospray ionization mass spectrometry. *Nucleic Acids Res.* 1996, *24*, 2183–2189.

[68] Fuerstenau, S. D., Benner, W. H., Molecular weight determination of megadalton DNA electrospray ions using charge detection time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 1995, *9*, 1528–1538.

[69] Hoffmann, J., Schmidt, T. L., Heckel, A., Brutschy, B., Probing the limits of liquid droplet laser desorption mass spectrometry in the analysis of oligonucleotides and nucleic acids. *Rapid. Commun. Mass. Spectrom.* 2009, *23*, 2176–2180.

[70] Bai, J., Liu, Y. H., Lubman, D. M., Siemieniak, D., Matrix-assisted laser desorption/ionization mass spectrometry of restriction enzyme-digested plasmid DNA using an active Nafion substrate. *Rapid Commun. Mass Spectrom.* 1994, *8*, 687–691.

[71] Elso, C., Toohey, B., Reid, G. E., Poetter, K. et al., Mutation detection using mass spectrometric separation of tiny oligonucleotide fragments. *Genome Res.* 2002, *12*, 1428–1433.

[72] Ding, C., Cantor, C. R., A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 3059–3064.

[73] Oberacher, H., Niederstatter, H., Parson, W., Characterization of synthetic nucleic acids by electrospray ionization quadrupole time-of-flight mass spectrometry. *J Mass Spectrom*. 2005, *40*, 932–945.

[74] Singh, R., Farmer, P. B., Liquid chromatography-electrospray ionization-mass spectrometry: the future of DNA adduct detection. *Carcinogenesis* 2006, *27*, 178–196.

[75] Sasayama, T., Kato, M., Aburatani, H., Kuzuya, A. et al., Simultaneous genotyping of indels and SNPs by mass spectroscopy. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 3–8.

[76] Xiao, W., Stern, D., Jain, M., Huber, C. G. et al., Multiplex capillary denaturing high-performance liquid chromatography with laser-induced fluorescence detection. *Biotechniques* 2001, *30*, 1332–1338.

[77] Perez-Perez, F. J., Hanson, N. D., Detection of plasmid-mediated AmpC beta-lactamase genes in clinical isolates by using multiplex PCR. *J. Clin. Microbiol*. 2002, *40*, 2153–2162.

[78] Huber, C. G., Berti, G. N., Detection of partial denaturation in AT-rich DNA fragments by ion-pair reversed-phase chromatography. *Anal. Chem.* 1996, *68*, 2959–2965.

[79] Premstaller, A., Xiao, W., Oberacher, H., O'Keefe, M. et al., Temperature-modulated array high-performance liquid chromatography. *Genome Res.* 2001, *11*, 1944–1951.

[80] Frayling, I. M., Monk, E., Butler, R., *Molecular Diagnostics: For the Clinical Laboratorian*, Humana Press, Totowa, NJ 2005.

[81] Fazekas, A., Steeves, R., Newmaster, S., Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* 2010, *48*, 277–285.

[82] Hite, J. M., Eckert, K. A., Cheng, K. C., Factors affecting fidelity of DNA synthesis during PCR amplification of d(C-A)n.d(G-T)n microsatellite repeats. *Nucleic Acids Res.* 1996, *24*, 2429–2434.

[83] Azzoni, A. R., Ribeiro, S. C., Monteiro, G. A., Prazeres, D. M., The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. *J. Gene Med.* 2007, *9*, 392–402.

[84] Shinde, D., Lai, Y., Sun, F., Arnheim, N., Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res.* 2003, *31*, 974–980.

[85] Rouleau, E., Lefol, C., Bourdon, V., Coulet, F. et al., Quantitative PCR high-resolution melting (qPCR-HRM) curve analysis, a new approach to simultaneously screen point mutations and large rearrangements: application to MLH1 germline mutations in Lynch syndrome. *Hum. Mutat.* 2009, *30*, 867–875.

[86] Montgomery, J., Wittwer, C. T., Palais, R., Zhou, L., Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nat. Protoc.* 2007, *2*, 59–66.

[87] Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G. et al., High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin. Chem.* 2003, *49*, 853–860.

[88] Liew, M., Pryor, R., Palais, R., Meadows, C. et al., Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin. Chem.* 2004, *50*, 1156–1164.

[89] Reed, G. H., Wittwer, C. T., Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clin. Chem.* 2004, *50*, 1748–1754.

[90] Li, J., Wang, L., Mamon, H., Kulke, M. H. et al., Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nat. Med.* 2008, *14*, 579–584.

[91] Milbury, C. A., Li, J., Makrigiorgos, G. M., Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations. *Nucleic Acids Res*. 2010, DOI: 10.1093/nar/gkq899.

[92] Tewhey, R., Warner, J. B., Nakano, M., Libby, B. et al., Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat. Biotechnol.* 2009, *27*, 1025–1031.
[93] Sheu, J. I., Sheu, E. Y., Characterization of DNA degradation using direct current conductivity and dynamic dielectric relaxation techniques. *AAPS PharmSciTech*. 2006, *7*, E36.
[94] Chakrabarti, S., Price, B. D., Tetradis, S., Fox, E. A. et al., Highly selective isolation of unknown mutations in diverse DNA fragments: toward new multiplex screening in cancer. *Cancer Res.* 2000, *60*, 3732–3737.
[95] Faham, M., Cox, D. R., A novel in vivo method to detect DNA sequence variation. *Genome Res.* 1995, *5*, 474–482.
[96] Volfovsky, N., Haas, B. J., Salzberg, S. L., A clustering method for repeat analysis in DNA sequences. *Genome Biol.* 2001, *2*, RESEARCH0027.
[97] Hoff, K. J., The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics* 2009, *10*, 520.
[98] Hestekin, C. N., Barron, A. E., The potential of electrophoretic mobility shift assays for clinical mutation detection. *Electrophoresis* 2006, *27*, 3805–3815.
[99] Peters, T., Sedlmeier, R., Current methods for high-throughput detection of novel DNA polymorphisms. *Drug Discov. Today* 2006, *3*, 123–129.
[100] Vossen, R. H., Aten, E., Roos, A., den Dunnen, J. T., High-resolution melting analysis (HRMA): More than just sequence variant screening. *Hum. Mutat.* 2009, *30*, 860–866.

