Review

Structural instability of plasmid biopharmaceuticals: challenges and implications

Pedro H. Oliveira[1], Kristala Jones 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, Av. Rovisco Pais 1049-001 Lisbon, Portugal
[2] Department of Chemical Engineering, Massachusetts Institute of Technology, Room 66-458, Cambridge, MA 02139, USA

The global increase in the number of applications involving therapeutic plasmid DNA (pDNA) is creating a need for large amounts of highly stable and purified molecules. One of the main obstacles during the developmental stages of a new therapeutic DNA molecule involves tackling a wide array of structural instability events occurring in/with pDNA and therefore assuring its structural integrity. This review focuses on major instability determinants in pDNA. Their elimination could be considered an important step towards the design of safer and more efficient plasmid molecules. Particular emphasis is given to mutations triggered by the presence of repeated sequences, instability events occurring during plasmid intracellular routing, instability mediated by insertion sequences and host genome integration.

Introduction

Following the first reports in the early 1990s describing plasmid-encoded antigen-induced immune responses in mice [1], the concept of DNA vaccination has evolved to a point where it is now often considered a valid approach to deal with several pathologies, including infectious diseases [2] and tumors [3]. Although often hampered by low immunogenicity, naked DNA vaccines do not present the risks typically associated with live-attenuated vaccines (e.g. problematic manufacture because of biosafety concerns, reversion or unwanted spread to other individuals). Moreover, the typical production and downstream processing steps needed to generate sufficiently large quantities of pharmaceutical-grade pDNA in a relatively short time frame are well established [4–6].

Although DNA vaccines are more stable than microorganism- or protein-based vaccines, one of the major problems encountered during the design of a DNA vaccine is assurance of its structural integrity. Indeed, the type and overall organization of the genetic elements present in pDNA vaccines directly impact not only its bulk production, but also its shelf stability, efficacy and ultimately its clinical approval. These plasmid vectors need to contain a number of basic elements, such as a eukaryotic expression unit (encompassing the enhancer/promoter region, signal sequence, gene of interest and a polyA transcriptional terminator), a prokaryotic origin of replication and a selection marker for propagation in bacteria (see recent reviews in Refs [7–9]). Further elements, in particular those prone to introduction of any type of structural instability, should therefore be avoided if possible. In fact, major bottlenecks related to plasmid instability often arise during the manufacturing process of a plasmid-based biopharmaceutical that could have been circumvented by *a priori* efficient and rational design of the plasmid structure.

Dynamic structure of DNA

DNA structure is intrinsically dynamic and plasmids can assume several conformations apart from the most typical negatively supercoiled B-form. Higher-order, non-Watson–Crick conformations usually stem from the presence of repeated DNA motifs, which are particularly prone to subsequent genetic rearrangements, such as deletions, duplications, inversions, translocations and insertions. Examples of such motifs include direct repeats, inverted repeats, purine•pyrimidine $(R•Y)_n$ mirror repeats and runs of closely spaced guanine residues. Much of the interest in studying some of these motifs, particularly dinucleotide, trinucleotide and pentanucleotide repeats, arose after reports of their involvement in the onset of several human neurological diseases and cancer [10]. Moreover, there is a real concern that unusual secondary structures could arise from sequences taken from viruses or from the human genome for use in DNA vaccination and gene therapy. Bearing this in mind, Cooke and co-workers [11] evaluated the influence of specific non-Watson–Crick DNA structures on the stability, yield and topology of the general cloning pBluescript vector. The authors found that Z-DNA was markedly more unstable than other structures, such as DNA triplexes, bends and quadruplexes. In addition, the triplex structure also led to decreased amounts of supercoiled plasmid.

In this article we do not address tandem repeat instability (reviewed in Ref. [12]), but instead focus on other structural instability phenomena that have been shown to be of particular concern for plasmid manufacture, DNA vaccination and gene therapy. Table 1 summarizes several determinants for pDNA instability and their main

Corresponding author: Monteiro, G.A. (gabmonteiro@ist.utl.pt).

0167-7799/$ – see front matter doi:10.1016/j.tibtech.2009.06.004 Available online 3 August 2009

Table 1. Overview of major factors affecting plasmid structural stability

Structural feature	Consequences	Refs
Plasmid size	• Large plasmids have a higher probability of containing potentially unstable regions or integrating into the host genome • Transfection efficiency decreases with plasmid size	[13,14]
Purine–pyrimidine/oligopurine–oligopyrimidine tracts; G-rich sequences	• Formation of Z-DNA, triplexes or quadruplexes that are involved in deletions and reduced plasmid yields	[11]
Chi sequences (5′-GCTGGTGG-3′)	• Loss of plasmid sequences and dimerization via the RecBCD pathway	[89]
Direct repeats	• Formation of DNA hairpin/loops • Loss of plasmid sequences or dimerization with concomitant loss or gain of genetic material	[18,22–26]
Inverted repeats	• Generation of cruciform structures • Formation of circular inverted head-to-head plasmid dimers	[27–34]
polyA sequences and other nuclease-sensitive regions	• Increased degradation by nucleases	[46,47]
Regions similar to gDNA	• Plasmid integration into host genome with disruption of potentially critical regions	[48–53]
Insertion sequences	• Gene activation • Gene down-regulation or inactivation • Deletions • Deleterious effects on plasmid replication and amplification	[36–39]

consequences, including direct and inverted repeats, polyA sequences, insertion sequences (IS) and regions similar to genomic DNA (gDNA). Plasmid size is not an instability determinant *per se*, but tends to correlate with a higher propensity for intramolecular recombination and/or genome integration events [13]. Larger plasmids also show lower transfection efficiency [14] and are responsible for a higher metabolic burden [15]. The presence of a high content of unmethylated CpG motifs in some genes (e.g. the β-lactamase gene) might also be deleterious in a candidate DNA vaccine because these might trigger an exacerbated inflammatory reaction in the host [16,17]. The next sections discuss in more detail potentially compromising structural instability phenomena and potential strategies to minimize or eliminate their occurrence.

Instability mediated by direct repeats

Repetitive DNA sequences are known to be the cause of a wide array of mutations and genetic rearrangements in both genomic and plasmid DNA. These repeat sequences can range from just a few nucleotides in length to several kilobasepairs and often lead to deletion or amplification of genetic material. Although pairing and exchange between repeated DNA sequences, which typically occur in the presence of the RecA protein, require a minimum degree of sequence similarity to proceed, DNA rearrangements induced by the presence of direct repeats can also occur independently of this protein (reviewed in Ref. [18]). This RecA-independent mechanism seems to be exclusively involved in rearrangements between short similar regions. Moreover, the frequency of this type of rearrangement increases with repeat length but exponentially decreases with the length of intervening sequences and with the use of non-perfectly similar sequences [18].

A direct consequence of recombination between two direct repeats in pDNA is the formation of monomeric and/or heterodimeric products (termed 1+2 and 1+3;

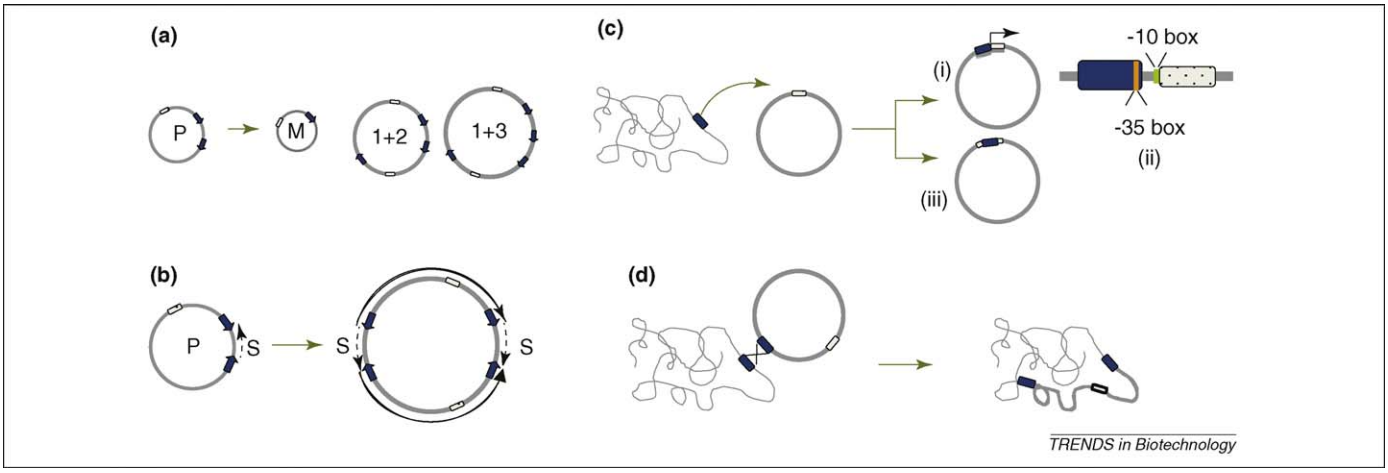

Figure 1. Possible recombination events involving plasmid DNA. (a) Recombination between directly repeated sequences (blue arrows) in a parental (P) pDNA can give rise to monomeric (M) and heterodimeric (1+2, 1+3) deletion products. The monomeric form lacks the entire spacer region and one direct repeat, whereas both heterodimeric forms are head-to-tail dimers composed of one parental plasmid and one monomeric product, with the 1+3 form bearing one additional direct repeat and an intervening sequence. (b) Recombination between inverted repeats (blue arrows) in pDNA results in head-to-head dimers bearing two pairs of interposed inverted repeats, inverted spacer sequences (S) and two giant inverted repeats (large arrows). (c) Transposition of IS elements (blue box) from gDNA (shown as coiled molecule) to pDNA can be detected as up-regulation of gene (symbolized by the white box) expression (i). This occurs via the generation of a hybrid promoter composed of a -35 region from the IS extremity (orange box) and a resident -10 hexamer (green box) present in pDNA (ii). Alternatively, IS elements can promote gene inactivation or suppression by interrupting the coding region (iii). (D) Recombination between similar regions (blue boxes) present in pDNA and in gDNA can lead to integration of the plasmid into the host genome.

Figure 1a). The monomeric product results from deletion of one direct repeat and the intervening sequence, whereas the 1+2 form is a head-to-tail dimer composed of one parental plasmid and one monomeric product. The 1+3 form has an additional direct repeat and intervening sequence. The pathways proposed for the formation of these products are based on simple slip-pair or slipped misalignment mechanisms, accompanied by sister-chromosome exchange (SCE) occurring at stalled replication forks, and have been described elsewhere [18]. Ultimately, establishment of either of these plasmid recombination products within a cell population depends on its fitness relative to that of other mutant forms and on its ability to displace parental molecules [19–21]. Recent evidence showed that this fitness can be altered by changes in antibiotic concentrations, leading to different ratios between the monomer and heterodimers (Oliveira *et al.*, manuscript submitted for publication).

In the past few years, some findings regarding the direct repeat-mediated instability in those vectors, which are commonly used as backbones for DNA vaccine development or protein expression, have raised concerns about their biological safety and applicability. For instance, Ribeiro and co-workers [22] described a spontaneous recombination event in the mammalian expression vector pCIneo and in its derivative pGPV-PV, a candidate DNA vaccine against rabies. In this system, recombination led to the formation of typical monomeric and heterodimeric forms, which enabled *E. coli* cells to acquire resistance to kanamycin. The authors also found that identical instability regions are widespread among other commonly used expression and cloning vectors, including pTarget, pALTER, pBIND and pACT (Promega, www.promega.com). Other examples for which direct repeat-mediated and RecA-independent recombination events have recently been demonstrated include the pGEX series of plasmids (GE Healthcare, www.gehealthcare.com), leading to decreased levels of glutathione S-transferase (GST) or GST-fusion protein expression [23], and pHZ1358, a vector widely used for targeted gene disruption and replacement experiments in *Streptomyces* hosts [24]. Interestingly, only the monomeric form was reported to occur in both these cases. Kawe and colleagues [25] observed a promoter deletion in *lac*-controlled expression vectors hosted in *rec*A$^-$ *E. coli* strains leading to abolishment of protein expression. This mutation event depended on the presence of repeated regions located in two *lac* operators and only occurred in systems lacking the *lac*I gene coding for the *lac* repressor. The authors thus recommend the use of fully repressible systems in conditions involving massive protein overproduction. Although the majority of the above studies that involved direct repeat-mediated recombination in pDNA have been conducted *in vivo*, an *in vitro* deletion occurring between direct repeats in hairpin-like PCR amplicons of the gene for human blood platelet glycoprotein Ib-α has been recently reported [26]. This immediately implies that occurrence of recombination events between repeated sequences might start in the early steps of plasmid assembly before amplification in the bacterial host.

Instability mediated by inverted repeats

Although comparably fewer studies have investigated recombination events between inverted repeats in pDNA compared to those analyzing direct repeats, similar conclusions could be drawn; recombination decreased exponentially with distance between repeats and was able to proceed in the absence of RecA and RecBCD [27]. Studies of inverted repeats present in *E. coli* plasmids [27–29] have shown that circular inverted head-to-head dimer molecules are frequently formed (Figure 1b). Two models have been proposed to explain the mechanisms leading to this rearrangement: the reciprocal strand-switching (RSS) model [27] and the cruciform–dumbbell (CD) model [28,30]. According to the RSS model, reciprocal switching occurs during DNA replication between leading and lagging strands within the inverted repeats. In the CD model, the inverted repeats are involved in the formation of a cruciform on the DNA molecule. In both models, dumbbell-like DNA intermediates originate, which lead to large inverted repeats after replication (Figure 1b). An alternative DNA rearrangement consisting of both direct and inverted repeats (DIR) was reported to occur in *E. coli* cells harboring pUC18. In this case, the mutation arises by strand slippage from the leading to the lagging strand of the replication fork as a result of the presence of long palindromic sequences [31].

Currently, there are no reports of inverted repeat-mediated instability in vectors used for DNA vaccination or gene therapy. Ongoing work in our laboratory has shown that the density of inverted repeats in pDNA is highly underrepresented in comparison with direct repeats, which might explain why the former have not been frequently associated with instability events. This does not mean that we should underestimate the potential for inverted repeat-mediated instability, as shown in recent examples reported for non-therapeutic plasmids [32,33] Moreover, plasmids used for therapeutic *in vivo* gene silencing via short-hairpin RNA sequences necessarily contain inverted repeats and are therefore particularly susceptible to recombinations [34].

Instability mediated by IS

IS are the most simple and common transposable elements among prokaryotes and code only for the information needed for their transposition. Often regarded as parasitic elements, IS are nevertheless known to play an important evolutionary role by promoting gene activation or inactivation, as well as genome plasticity. Their abundance in prokaryotes seems to be primarily based on horizontal gene transfer and genome size [35]. It has been shown that IS transposition in plasmids leads to activation of neighboring gene expression [36], to down-regulation or inactivation of genes [37] and to deletions [38] and changes in pDNA replication and amplification [39]. Gene activation is often caused by IS elements that have outwardly directed –35 promoter hexamers near their terminal inverted repeats (IRs). These pseudo-promoter elements can effectively turn on gene expression if placed after transposition at an accurate distance upstream of a resident –10 hexamer (reviewed in Ref. [40]) (Figure 1c).

In an attempt to consistently annotate IS data, a publicly available tool able to detect IS elements and corresponding sequence elements, such as direct and inverted repeats, was recently developed (http://www.bioc.uzh.ch/wagner/software/IScan) [41]. In addition, an extensive and comprehensive list of IS that have been isolated from gDNA and pDNA of Eubacteria and Archaea is available in the newly developed IS finder database (www-is.biotoul.fr/) [42]. A major step towards complete eradication of stress-induced IS transposition was recently achieved after a partial genome reduction of the closely related *E. coli* K12 strains MG1655 and W3110 [43]. The regions removed from the genome included recombinogenic or mobile DNA and cryptic virulence genes. Compared to the parental strains, the reduced-genome strains showed identical growth rates and recombinant protein expression yields, but could be more efficiently transformed by electroporation and were able to propagate plasmids such as pT-ITR that were unstable in other strains.

Because these safer genome-engineered strains have hitherto been seldom used, IS-mediated instability events have been reported to occur in plasmids used as backbones for DNA vaccine development. Prather and co-workers [39] found several IS*1* insertions near the gene for neomycin resistance in a HIV DNA vaccine, which ultimately affected replication and amplification in a position-dependent manner. Another example, this time involving IS*2*, was recently found in the mammalian expression vector pCIneo (Oliveira *et al.*, manuscript submitted for publication). The authors suggest that IS*2* transposition might be a widespread event in other vectors that share the same insertion site and adjacent regions.

Instability during intracellular routing

Efficient gene expression can be only achieved if the majority of plasmid molecules reach the nucleus intact and preferentially in the supercoiled isoform. The most frequent approaches to promote pDNA uptake involve its protection from nuclease degradation using adjuvants such as cationic polymers, cationic lipids or a mixture of both (reviewed in Ref. [44]). After internalization, the pDNA–adjuvant complex must escape endo-lysosomal entrapment and be efficiently dismantled in the nucleus to be transcriptionally active. However, at this stage pDNA is again exposed to nuclease degradation. The use of different transfection methods and/or adjuvants should ideally be coupled to a rational plasmid design aimed at improved resistance to nuclease degradation. This implies that 'hot-spot' regions, which are particularly prone to nuclease digestion, should be removed or modified. A recently proposed method for plasmid condensation and efficient protection against nuclease degradation uses cationic pentablock copolymers based on poly(2-diethylaminoethyl-methacrylate) and Pluronic F127 [45]. This approach was found to have lower toxicity than the widely used ExGen 500® polymer, while maintaining identical transfection levels and supercoiled content. On the other hand, Azzoni and co-workers [46,47] have demonstrated that S1 nuclease mapping of pVAX1/LacZ led to the identification of two sequences highly susceptible to nuclease digestion located in the polyA stretch of bovine growth hormone (BGH) and in the pMB1 origin of replication. In fact, changing the native BGH polyA tail in the plasmid pVAX1/LacZ using either SV40 or a synthetic polyA sequence led to an increase in half-life by up to two-fold in the presence of endonucleases [46].

Taken together, these results can be further exploited to generate improved nuclease-resistant plasmids, which could subsequently increase the number of molecules capable of reaching the cell nucleus in an active form.

Evidences for pDNA integration into the host genome

One of the major concerns regarding the use of plasmid vectors as DNA vaccines is their possible integration into the host genome. From the point of view of gene therapy this integration could be viewed as profitable for stable gene expression, but it could have serious and even deleterious effects on host genome integrity. In fact, only a single study involving the use of a PCR-based assay able to detect rare integration events allowed the detection of independent genome integration events following electroporation of mice with plasmids containing the mouse gene for erythropoietin [48]. The authors showed that a highly efficient DNA delivery method (such as electroporation) combined with a highly sensitive detection method was crucial to truly ascertain the number of plasmid integrations into host gDNA. The majority of recent studies that involved a number of different DNA vaccines and distinct organisms have found that integration occurs at a very low frequency, if at all [49–53]. In recent years, advances in electroporation technology and better control over the electric pulses delivered to the tissues have mitigated some of the risks and concerns for genome integration [54,55]. Moreover, the FDA requires rigorous analysis for biodistribution and integration for approval of clinical trials involving pDNA [56]. Among the DNA vaccines already approved for animal administration, no evidence has been found for genome integration events [57–60].

Given that pDNA regions that share homology with host genome sequences might increase the potential insertion frequency (even at a low frequency), the most straightforward approach to reduce and/or circumvent such events would involve minimization of the vector sequence by eliminating all non-essential regions. This would also reduce the host metabolic load during plasmid propagation, increase transfection efficiency and minimize potential inflammatory reactions. *De novo* DNA synthesis is also a promising approach to minimize any homology between pDNA and gDNA because the DNA sequence of coding regions can be altered without affecting the amino acid sequence of the translated protein.

Towards the ideal nonviral vector

The presence of bacterial elements in vectors intended for use in therapeutics might have detrimental effects on their performance. Deletion of bacterial sequences will eliminate some potential instability hot-spots and cryptic sequences, and several strategies for this, which are based on plasmid minimization, have been proposed (reviewed in Ref. [61]). One approach is to generate so-called minimalistic immunogenic defined gene expression (MIDGE) vectors that lack any non-essential and potentially

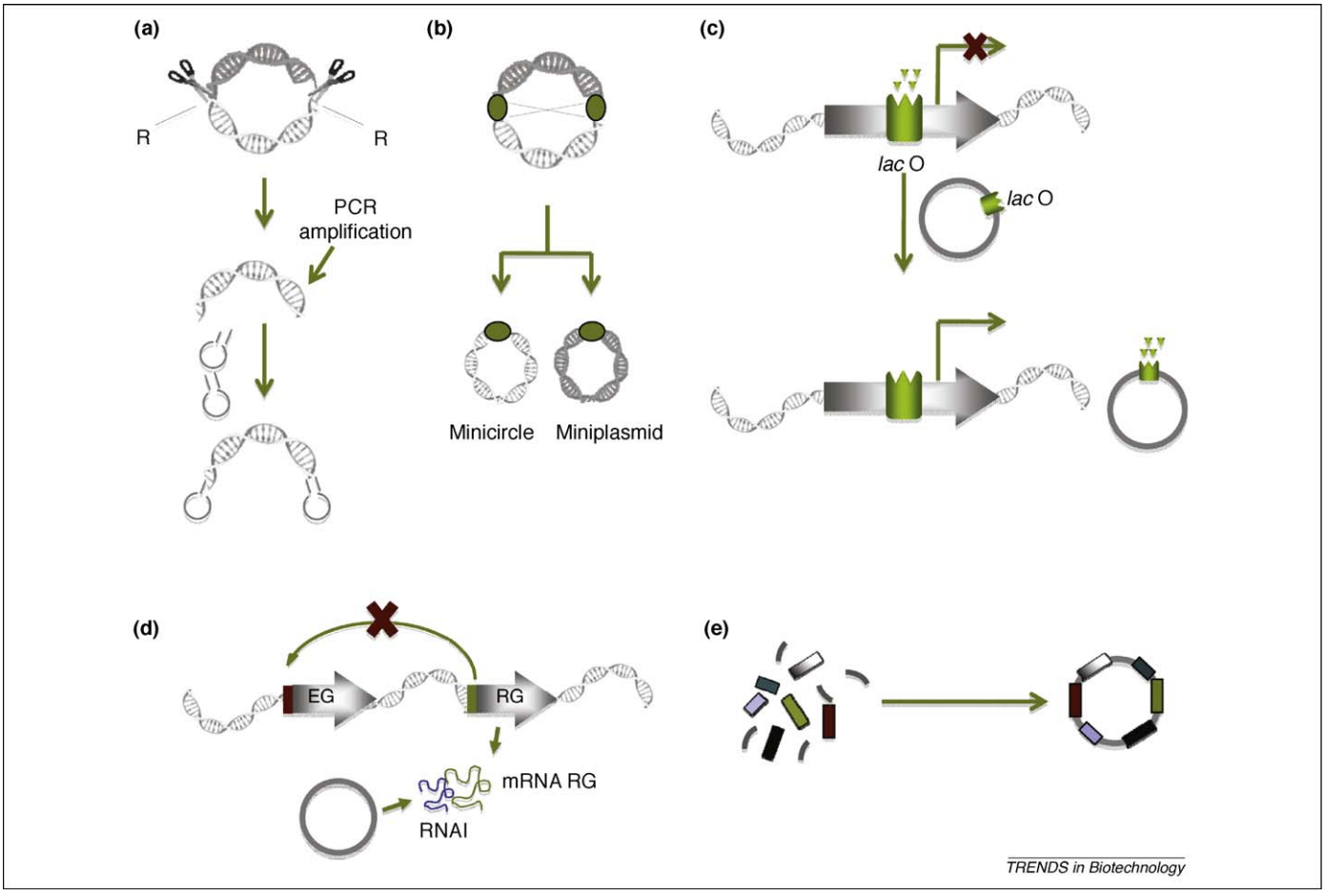

Figure 2. Approaches to minimize the content of non-essential DNA regions in plasmids. Such strategies might involve either the generation of minimal plasmids or the use of plasmid selection systems that circumvent the requirement for antibiotics for selection in bacteria. (a) An example of a minimal plasmid is the minimalistic immunogenic defined gene expression (MIDGE) vector. Here, the eukaryotic expression unit (white region) is cleaved from the bacterial vector sequences (black region) in a restriction digestion (R). The resulting eukaryotic fragments are ligated with hairpin oligodeoxynucleotides to protect the resulting molecule against nucleases. Any remaining non-ligated fragments are further digested with exonucleases, whereas the MIDGE vector is further purified by chromatography (not shown here). Alternatively, generation of these covalently closed linear dumbbell-shaped molecules can be achieved by PCR amplification of the eukaryotic expression unit from a given plasmid vector. End protection with hairpin oligos is performed in a similar manner as before. (b) DNA minicircles are generated by *in vivo* excision of the minicircle DNA (white region) using site-specific recombination between *attB* and *attP* sequences (shown as green regions), which are mediated by the phage λ integrase. Alternatively, recombination can similarly occur between *loxP* sites in a reaction catalyzed by the bacteriophage P1-derived Cre recombinase. (c) In the so-called repressor titration system [70], plasmids containing copies of the *lac* operator (*lac*O) are used to competitively titrate *lac* inhibitor molecules (LacI) (green triangles) from the *lac* operator of an essential chromosomal gene. Thus, host cells are able to grow and maintain the plasmid in the absence of any antibiotic, thereby reducing the number of plasmid elements required. (d) In the RNAi-based plasmid selection system [71], the RNAi product from ColE1 plasmids can inhibit translation of a repressor gene (RG) by antisense hybridization with its mRNA. In this way, only plasmid-containing cells are able to grow normally because the essential chromosomal gene (EG) can be efficiently translated. (e) Development of minimal *de novo* plasmids [72] by self-assembly of essential pre-cloned modules requiring only a minimal content of non-essential DNA sequences.

detrimental backbone sequences. These MIDGE vectors are covalently closed linear dumbbell-shaped fragments comprising only a eukaryotic expression unit and terminal hairpin cap oligonucleotides (Figure 2a). These fragments could also be generated by PCR amplification of an expression cassette taken from a given plasmid construct, followed by an end protection step (Figure 2a). Several studies have successfully applied MIDGE vectors (e.g. to generate cellular and humoral responses in mice and cattle and to express therapeutic genes in human stromal cells), establishing their position as valuable alternatives to conventional (non-minimal) vectors [62–64].

Other approaches for the production of vectors that are devoid of bacterial-derived DNA involve the generation of circular supercoiled expression units, the so-called minicircles (Figure 2b). These molecules are usually obtained after site-specific recombination of plasmid direct sequences after expression of a specific recombinase, such as the phage λ integrase or phage P1 Cre recombinase [65,66]. The exact positioning of the direct repeats delimiting bacterial and mammalian units leads to the formation of two circular molecules, one that is devoid of bacterial elements (the minicircle) and the other containing them (a miniplasmid). Minicircles have shown better performance in terms of expression levels, bioavailability and permeability than that of their parental vectors when administered intramuscularly and intratumorally in mice [65]. The major drawback of this approach is that specific purification methods are necessary for efficient separation of minicircles from miniplasmids. Although techniques such as triple-helix affinity chromatography [67] and triple-helix affinity precipitation [68] were successful in purifying pDNA, no data on their applicability to separation of minicircles from miniplasmids are available. Only very recently, a recombination-based plasmid separation technology was proposed for the efficient generation and

purification of minicircle DNA [69]. This approach consists of two stages: a ParA resolvase recombination system induced by L-arabinose, which gives rise to the formation of a minicircle and miniplasmid, and a subsequent protein/DNA affinity system based on interaction between a column-immobilized repressor of the lactose operon (LacI) and its corresponding operator sequence (*lacO*) present in the plasmid molecule. This strategy yielded very high recombination efficiencies of >99.5% and minicircles that were >98.5% pure.

Other alternatives for plasmid minimization use plasmid-mediated repressor titration systems to effectively activate a chromosomal selectable marker, thereby eliminating the need for its presence on the plasmid itself [70] (Figure 2c) or RNA-based selection systems exploiting the RNAI/RNAII antisense regulation of pMB1-derived plasmids [71] (Figure 2d). Williams and colleagues [72] developed the pDNAVACCultra family of minimized vectors that drive higher levels of target gene expression than those of the gWIZ high-expression reporter system. Moreover, pDNAVACCultra are designed to allow simultaneous cloning of antigens into multiple vectors of this family that feature several targeting destinations for the protein product (Figure 2e).

Taken together, these approaches suggest that plasmid minimization represents a promising approach to tackle safety issues associated with the presence of bacterial elements, but the methods currently available suffer from low efficiency and a lack of cost-effective scale-up processes, which would make them suitable for clinical application.

Conclusions and future trends

With recent developments in the fields of DNA vaccination and gene therapy, maintaining the structural stability of pDNA has become of utmost importance. This aspect should thus be the focus of particular attention, not only during bacterial amplification and pDNA bioprocessing, when mechanical shear stresses, chemically dependent hydrolysis and attack by DNAses [73–75] pose serious threats to its structural integrity, but also after delivery into the target cells, where pDNA can be exposed to a large array of recombinogenic events.

In this review, particular emphasis is given to repeated and similar sequences and to IS transposition. Attempts to explain and predict possible recombination events between similar sequences have resulted in mechanistic models [76] and meta-analysis of available pDNA recombination frequency data [77]. In the latter study, two mathematical functions to predict the recombination frequency of multicopy plasmids harboring direct repeats were derived for *rec*A$^+$ and *rec*A$^-$ bacterial strains. Using these functions, the authors found that more than 92% of the predicted values of pDNA recombination frequency were within a ±5-fold interval of deviation from experimental data. Bioinformatic tools can also be used for plasmid design because they can detect the presence of repeated sequences within the vector. Examples of such tools include REPuter [78] and Repseek [79] for detection of DNA repeats with high sequence similarity in large DNA sequences, Tandem Repeat Finder [80] for identification of tandem repeats and Swelfe [81] for detection of short repeats that are closely spaced and divergent. If any unstable repeated sequences are detected in non-essential regions of the vector, they could be removed so that vector quality in terms of main function, copy number and topology is maintained. If the repeat sequences are part of an essential plasmid region, a careful search for aberrant non-standard forms could be performed.

Predicting IS transposition is considerably more difficult and might even be impossible. Although some regional preferences have been observed [82,83], the insertion specificity of many IS elements remains unknown. In general, IS elements are usually present within a plasmid population at low frequency unless they confer a selective advantage (e.g. antibiotic gene activation), but transposition frequencies can be boosted by environmental stresses such as nutrient deprivation, as shown for *E. coli* [84]. Therefore, to completely abolish IS-mediated transposition events, the use of strains in which these IS elements have been knocked out could be considered.

A number of strains are commercially available for propagation of plasmids that contain unstable sequences, including SURE cells (Stratagene, www.stratagene.com) and Stbl2 and Stbl4 (Invitrogen, www.invitrogen.com), which harbour mutations in the UV repair system and SOS repair pathways. However, although these cells have a genotype that can stabilize repeated and Z-like sequences, they cannot entirely prevent all mutational events associated with these regions.

A number of tools have been developed to detect recombinant plasmid forms, depending on the particular recombination frequency, including agarose gel visualization, standard and real-time PCR, enzyme-based probes, capillary gel electrophoresis, fluorescence, hybridization/chip-based methods and LC-MS [85,86]. Judicious use of highly sensitive and specific detection methods coupled to rational vector design based on computer-assisted detection of instability regions will determine the requirement to embrace further approaches, such as the use of genetically stable host strains, alternative growth conditions to minimize a certain instability event and genome/plasmid minimization.

From our own experience in this field we know that one of the major drawbacks preventing more frequent detection of pDNA instability events is related to the fact that they are generally maintained at a low frequency within a plasmid population. This is because the large majority of spontaneous mutations triggered by repeated sequences or IS elements do not necessarily carry any selective advantage over other non-mutated plasmid forms present. However, ongoing work has provided some evidence that plasmid mutations seem to be a much more widespread phenomenon than initially thought. In other words, there might be a considerable percentage of mutant plasmids in a population, but because each mutant occurs at a low frequency, individual detection is usually difficult by standard methods (e.g. by agarose gel electrophoresis). We thus foresee a promising future for minimal plasmids and genome-engineered strains in tackling many of these events, but not totally eliminating them. In particular, the high density of repeated sequences located within eukaryotic elements, such as the widely used cytomegalovirus and

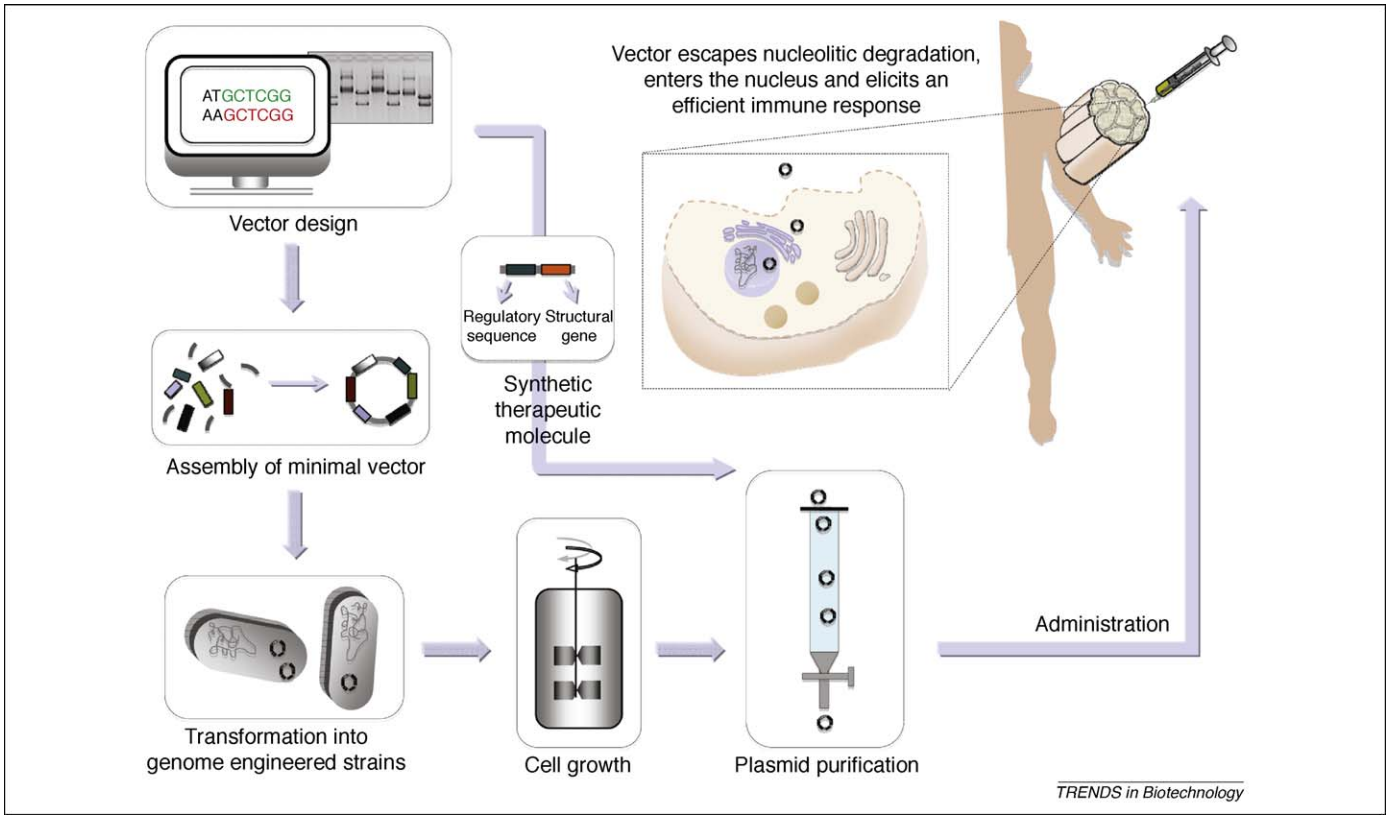

Figure 3. Approaches to minimize plasmid structural instability. One of the paths leading to structurally safer therapeutic molecules involves the generation of minimal plasmids devoid of non-essential regions. These plasmids should be preferentially amplified in genome-engineered bacteria (e.g., lacking IS elements) under tightly controlled culture conditions and must undergo a subsequent purification step (usually chromatography) before administration. Alternatively, the development of an affordable large-scale high-throughput error-free *de novo* DNA synthesis strategy to generate fully synthetic molecules (composed of regulatory sequences and structural gene) would circumvent the need for amplification in bacteria. Regardless of the synthesis pathway, the final plasmid must be able to escape degradation by nucleases to enter the nucleus and elicit an efficient immune response.

simian virus promoters, will compromise the aim of obtaining a perfectly stable plasmid.

De novo DNA synthesis of therapeutic molecules seems to be a promising approach to circumvent plasmid amplification in bacteria (Figure 3). Although several important steps have been implemented to improve gene synthesis quality by reducing the number of errors introduced [87,88], an affordable solution for large-scale high-throughput DNA synthesis is yet to be achieved.

Future challenges in the field of plasmid structural instability will involve integrated and multidisciplinary solutions such as those discussed in this review. Once these solutions are efficiently implemented, it will be possible to pave the way to safer and more efficient molecules for therapeutic use.

Acknowledgements

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.

References

1 Tang, D.C. *et al.* (1992) Genetic immunization is a simple method for eliciting an immune response. *Nature* 356, 152–154
2 Mwau, M. *et al.* (2004) A human immunodeficiency virus I (HIV-I) clade A vaccine in clinical trials: stimulation of HIV-specific T-cell responses by DNA and recombinant modified vaccinia virus Ankara (MVA) vaccines in humans. *J. Gen. Virol.* 85, 911–919
3 Liu, M. *et al.* (2004) Gene-based vaccines and immunotherapeutics. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14567–14571
4 Prazeres, D.M. *et al.* (1999) Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: problems and bottlenecks. *Trends Biotechnol.* 17, 169–174
5 Prazeres, D.M. *et al.* (2001) Purification of plasmids for gene therapy and DNA vaccination. *Biotechnol. Annu. Rev.* 7, 1–30
6 Prather, K.J. *et al.* (2003) Industrial scale production of plasmid DNA for vaccine and gene therapy: plasmid design, production and purification. *Enzyme Microb. Technol.* 33, 865–883
7 Bower, D.M. and Prather, K.L. (2009) Engineering of bacterial strains and vectors for the production of plasmid DNA. *Appl. Microbiol. Biotechnol.* 82, 805–813
8 Gill, D.R. *et al.* (2009) Progress and prospects: the design and production of plasmid vectors. *Gene Ther.* 16, 165–171
9 Williams, J.A. *et al.* (2009) Plasmid DNA vaccine vector design: impact on efficacy, safety and upstream production. *Biotechnol. Adv.* 27, 353–370
10 Bacolla, A. *et al.* (2006) The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair* 5, 1161–1170
11 Cooke, J.R. *et al.* (2004) Impact of intrinsic DNA structure on processing of plasmids for gene therapy and DNA vaccines. *J. Biotechnol.* 114, 239–254
12 Bichara, M. *et al.* (2006) Mechanisms of tandem repeat instability in bacteria. *Mutat. Res.* 598, 144–163
13 Ertl, P.F. and Thomsen, L.L. (2003) Technical issues in construction of nucleic acid vaccines. *Methods* 31, 199–206
14 Kreiss, P. *et al.* (1999) Plasmid DNA size does not affect the physicochemical properties of lipoplexes but modulates gene transfer efficiency. *Nucleic Acids Res.* 27, 3792–3798
15 Rozkov, A. *et al.* (2004) Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnol. Bioeng.* 88, 909–915
16 McMahon, J.M. *et al.* (1998) Inflammatory responses following direct injection of plasmid DNA into skeletal muscle. *Gene Ther.* 5, 1283–1290

17 Yew, N.S. *et al.* (2000) Reduced inflammatory response to plasmid DNA vectors by elimination and inhibition of immunostimulatory CpG motifs. *Mol. Ther.* 1, 255–262
18 Lovett, S.T. (2004) Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 52, 1243–1253
19 Bierne, H. *et al.* (1995) Competition between parental and recombinant plasmids affects the measure of recombination frequencies. *Plasmid* 33, 101–112
20 Mazin, A.V. *et al.* (1996) Dimerization of plasmid DNA accelerates selection for antibiotic resistance. *Mol. Microbiol.* 20, 101–108
21 Chédin, F. *et al.* (1997) Apparent and real recombination frequencies in multicopy plasmids: the need for a novel approach in frequency determination. *J. Bacteriol.* 179, 754–761
22 Ribeiro, S.C. *et al.* (2008) High frequency plasmid recombination mediated by 28 bp direct repeats. *Mol. Biotechnol.* 40, 252–260
23 Borgi, I. and Gargouri, A. (2008) A spontaneous direct repeat deletion in the pGEX fusion vector decreases the expression level of recombinant proteins in *Escherichia coli*. *Prot. Express. Purif.* 60, 15–19
24 Sun, Y. *et al.* (2009) Analysis of functions in plasmid pHZ1358 influencing its genetic and structural stability in *Streptomyces lividans* 1326. *Appl. Microbiol. Biotechnol.* 82, 303–310
25 Kawe, M. *et al.* (2009) Facile promoter deletion in *Escherichia coli* in response to leaky expression of very robust and benign proteins from common expression vectors. *Microb. Cell Fact.* 8, 8
26 Kacem, B.H. *et al.* (2008) *In vitro* direct repeats-mediated deletion during PCR amplification. *Mol. Biotechnol.* 40, 39–45
27 Bi, X. and Liu, L.F. (1996) DNA rearrangement mediated by inverted repeats. *Proc. Natl. Acad. Sci. U. S. A.* 93, 819–823
28 Lin, C.T. *et al.* (1997) A cruciform–dumbbell model for inverted dimer formation mediated by inverted repeats. *Nucleic Acids Res.* 25, 3009–3016
29 Lyu, Y.L. *et al.* (1999) Inversion/dimerization of plasmids mediated by inverted repeats. *J. Mol. Biol.* 285, 1485–1501
30 Lin, C.T. *et al.* (2001) Inverted repeats as genetic elements for promoting DNA inverted duplication: implications in gene amplification. *Nucleic Acids Res.* 29, 3529–3538
31 Pinder, D.J. (1997) DIR: a novel DNA rearrangement associated with inverted repeats. *Nucleic Acids Res.* 25, 523–529
32 Lessard, P.A. *et al.* (2004) pB264, a small, mobilizable, temperature sensitive plasmid from *Rhodococcus*. *BMC Microbiol.* 4, 15
33 Larraín-Linton, J. *et al.* (2006) Molecular and population analyses of a recombinant event in the catabolic plasmid pJP4. *J. Bacteriol.* 188, 6793–6801
34 McIntyre, G.J. and Fanning, G.C. (2006) Design and cloning strategies for constructing shRNA expression vectors. *BMC Biotechnol.* 6, 1
35 Touchon, M. and Rocha, E.P.C. (2007) Causes of insertion sequences abundance in prokaryotic genomes. *Mol. Biol. Evol.* 24, 969–981
36 Szeverényi, I. *et al.* (1996) Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes. *Gene* 174, 103–110
37 Fernández, A. *et al.* (2007) Interspecies spread of CTX-M-32 extended-spectrum β-lactamase and the role of the insertion sequence IS*1* in down-regulating bla_{CTX-M} gene expression. *J. Antimicrob. Chemother.* 59, 841–847
38 Chawla, M. and Das Gupta, S.K. (1999) Transposition-induced structural instability of *Escherichia coli*–mycobacteria shuttle vectors. *Plasmid* 41, 135–140
39 Prather, K.J. *et al.* (2006) Identification and characterization of IS*1* transposition in plasmid amplification mutants of *E. coli* clones producing DNA vaccines. *Appl. Microbiol. Biotechnol.* 73, 815–826
40 Mahillon, J. and Chandler, M. (1998) Insertion sequences. *Microbiol. Mol. Biol. Rev.* 62, 725–774
41 Wagner, A. *et al.* (2007) A survey of bacterial insertion sequences using IScan. *Nucleic Acids Res.* 35, 5284–5293
42 Siguier, P. *et al.* (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36
43 Pósfai, G. *et al.* (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046
44 Lechardeur, D. *et al.* (2005) Intracellular routing of plasmid DNA during non-viral gene transfer. *Adv. Drug Deliv. Rev.* 57, 755–767
45 Agarwal, A. *et al.* (2007) Colloidally stable novel copolymeric system for gene delivery in complete growth media. *J. Control. Release* 121, 28–37
46 Azzoni, A.R. *et al.* (2007) The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. *J. Gene Med.* 9, 392–402
47 Ribeiro, S.C. *et al.* (2004) The role of polyadenylation signal secondary structures on the resistance of plasmid vectors to nucleases. *J. Gene Med.* 6, 565–573
48 Wang, Z. *et al.* (2004) Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Ther.* 11, 711–721
49 Ledwith, B.J. *et al.* (2000) Plasmid DNA vaccines: investigation of integration into host cellular DNA following intramuscular injection in mice. *Intervirology* 43, 258–272
50 Manam, S. *et al.* (2000) Plasmid DNA vaccines: tissue distribution and effects of DNA sequence, adjuvants and delivery method on integration into host DNA. *Intervirology* 43, 273–281
51 Kang, K.K. *et al.* (2003) Safety evaluation of GX-12, a new HIV therapeutic vaccine: investigation of integration into the host genome and expression in the reproductive organs. *Intervirology* 46, 270–276
52 Coelho-Castelo, A.A.M. *et al.* (2006) Tissue distribution of a plasmid DNA encoding Hsp65 gene is dependent on the dose administered through intramuscular delivery. *Genet. Vaccines Ther.* 4, 1
53 Ramirez, K. *et al.* (2008) Preclinical safety and biodistribution of Sindbis virus measles DNA vaccines administered as a single dose or followed by live attenuated measles vaccine in a heterologous prime-boost regimen. *Hum. Gene Ther.* 19, 522–531
54 Khan, A.S. *et al.* (2005) Highly efficient constant-current electroporation increases *in vivo* plasmid expression. *DNA Cell Biol.* 24, 810–818
55 Curcio, C. *et al.* (2008) DNA immunization using constant-current electroporation affords long-term protection from autochthonous mammary carcinomas in cancer-prone transgenic mice. *Cancer Gene Ther.* 15, 108–114
56 FDA. Guidance for industry: considerations for plasmid DNA vaccines for infectious disease indications. US Food and Drug Administration, 2007
57 Powell, K. (2004) DNA vaccines – back in the saddle again? *Nat. Biotechnol.* 22, 799–801
58 Novartis. Novel Novartis vaccine to protect Canadian salmon farms from devastating viral disease. Novartis Animal Health Inc., News Release, Basel, 19 July 2005
59 Merial. USDA grants conditional approval for first therapeutic vaccine to treat cancer Merial Limited, News Release, 26 March 2007
60 Person, R. *et al.* (2008) Growth hormone-releasing hormone plasmid treatment by electroporation decreases offspring mortality over three pregnancies. *Mol. Ther.* 16, 1891–1897
61 Rodríguez, E.G. (2004) Nonviral DNA vectors for immunization and therapy: design and methods for their obtention. *J. Mol. Med.* 82, 500–509
62 Moreno, S. *et al.* (2004) DNA immunisation with minimalistic expression constructs. *Vaccine* 22, 1709–1716
63 Zheng, C. *et al.* (2006) Effect of different nuclear localization sequences on the immune responses induced by a MIDGE vector encoding bovine herpesvirus-1 glycoprotein D. *Vaccine* 24, 4625–4629
64 Mok, P.L. *et al.* (2008) *In vitro* expression of erythropoietin by transfected human mesenchymal stromal cells. *Cytotherapy* 10, 116–124
65 Darquet, A.M. *et al.* (1999) Minicircle: an improved DNA molecule for *in vitro* and *in vivo* gene transfer. *Gene Ther.* 6, 209–218
66 Bigger, B.W. *et al.* (2001) An *araC*-controlled bacterial *cre* expression system to produce DNA minicircle vectors for nuclear and mitochondrial gene therapy. *J. Biol. Chem.* 276, 23018–23027
67 Wils, P. *et al.* (1997) Efficient purification of plasmid DNA for gene transfer using triple-helix affinity chromatography. *Gene Ther.* 4, 323–330
68 Costioli, M.D. *et al.* (2003) DNA purification by triple-helix affinity precipitation. *Biotechnol. Bioeng.* 81, 535–545
69 Mairhofer, P. *et al.* (2008) Minicircle-DNA production by site specific recombination and protein–DNA interaction chromatography. *J. Gene Med.* 10, 1253–1269

70 Cranenburgh, R.M. *et al.* (2001) *Escherichia coli* strains that allow antibiotic-free plasmid selection and maintenance by repressor titration. *Nucleic Acids Res.* 29, E26
71 Mairhofer, J. *et al.* (2008) A novel antibiotic free plasmid selection system: advances in safe and efficient DNA therapy. *Biotechnol. J.* 3, 83–89
72 Williams, J.A. *et al.* (2006) pDNAVACCultra vector family: high throughput intracellular targeting DNA vaccine plasmids. *Vaccine* 24, 4671–4676
73 Monteiro, G.A. *et al.* (1999) Analysis and use of endogenous nuclease activities *in Escherichia coli* lysates during the primary isolation of plasmids for gene therapy. *Biotechnol. Bioeng.* 66, 189–194
74 Lengsfeld, C. and Anchordoquy, T. (2007) Basic issues in the manufacture of macromolecules. In *Handbook of Pharmaceutical Biotechnology* (Gad, S.C., ed.), pp. 297–317, John Wiley & Sons
75 Wu, M.L. *et al.* (2009) Stabilization of naked and condensed plasmid DNA against degradation induced by ultrasounds and high shear vortices. *Biotechnol. Appl. Biochem.* 53, 237–246
76 Fujitani, Y. and Kobayashi, I. (2003) Asymmetric random walk in a reaction intermediate of homologous recombination. *J. Theor. Biol.* 220, 359–370
77 Oliveira, P.H. *et al.* (2008) Recombination frequency in plasmid DNA containing direct repeats – predictive correlation with repeat and intervening sequence length. *Plasmid* 60, 159–165
78 Kurtz, S. and Schleiermacher, C. (1999) REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* 15, 426–427
79 Achaz, G. *et al.* (2007) Repseek, a tool to retrieve approximate repeats from large DNA sequences. *Bioinformatics* 23, 119–121
80 Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580
81 Abraham, A.L. *et al.* (2008) Swelfe: a detector of internal repeats in sequences and structures. *Bioinformatics* 24, 1536–1537
82 Sengstag, C. and Arber, W. (1987) A cloned DNA fragment from bacteriophage P1 enhances IS2 insertion. *Mol. Gen. Genet.* 206, 344–351
83 Lewis, L.A. *et al.* (1994) The non-random pattern of insertion of IS2 into the *hemB* gene of *Escherichia coli*. *Microbiol. Immunol.* 38, 461–465
84 Twiss, E. *et al.* (2005) Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress. *Mol. Microbiol.* 57, 1593–1607
85 Schmidt, T. *et al.* (1999) Quantitative analysis of plasmid forms by agarose and capillary gel electrophoresis. *Anal. Biochem.* 274, 235–240
86 Middaugh, C.R. (1998) Analysis of plasmid DNA from a pharmaceutical perspective. *J. Pharm. Sci.* 87, 130–146
87 Tian, J. *et al.* (2004) Accurate multiplex gene synthesis from programmable DNA chips. *Nature* 432, 1050–1054
88 Bang, D. and Church, G.M. (2008) Gene synthesis by circular assembly amplification. *Nat. Methods* 5, 37–39
89 Zaman, M.M. and Boles, T.C. (1996) Plasmid recombination by the RecBCD pathway of *Escherichia coli*. *J. Bacteriol.* 178, 3840–3845