Mol Biotechnol (2008) 40:252–260
DOI 10.1007/s12033-008-9082-3

RESEARCH

High Frequency Plasmid Recombination Mediated by 28 bp Direct Repeats

Sofia C. Ribeiro · Pedro H. Oliveira · Duarte M. F. Prazeres · Gabriel A. Monteiro

Published online: 8 July 2008

Abstract The stability in *Escherichia coli* of a mammalian expression vector (pCIneo) and its derivative candidate DNA vaccine (pGPV-PV) is described. These multicopy pMB1-type plasmids were found to recombine in several *rec*A *E. coli* strains due to the presence of two 28 bp direct repeats flanking intervening sequences of 1.6 kb (pCIneo) and 3.2 kb (pGPV-PV). In this recombination event, one of the direct repeats and the intervening sequence were deleted or duplicated, originating monomeric or/and hetero-dimeric plasmid forms, respectively. Additionally, the plasmid rearrangement led to the acquisition of a kanamycin resistance phenotype. Recombination frequencies between 7.8×10^{-7} and 3.1×10^{-5} were determined for DH5α and JM109(DE3) strains, respectively. Higher recombination frequencies were found in cells previously grown up to stationary growth phase being the monomeric plasmid form the prevalent one. Real-time PCR quantification revealed the presence of approximately 1.5×10^{4} recombined molecules *per* 2×10^{5} cells pre-kanamycin exposure. Under selective pressure of this antibiotic, the number of recombined molecules increased approximately 2,000-fold in a 48-h period replacing the original plasmid form. The high frequency, at which deletion-duplication occurred in the absence of kanamycin selective pressure, should be regarded as a safety concern. This work highlights the impact of mutational hot spots on expression and cloning plasmid vectors and the need to carefully design plasmid vectors.

Keywords Plasmid · Direct repeats · Deletion-duplication · Intramolecular recombination · Slipped misalignment · *Escherichia coli*

Introduction

Plasmid vectors are widely used for recombinant protein expression in prokaryotic and eukaryotic cells. Furthermore, their significance increased with the recent developments in non-viral gene therapy and DNA vaccination. The increasing number of these novel human therapeutic applications that are moving from the laboratory to the clinic is creating a need for large amounts of highly stable and purified plasmid DNA. This demand, which is expected to increase as the first products reach the market, requires a high-yielding host and downstream purification process under rigorous good manufacturing practice (GMP) guidelines [1–3]. *Escherichia coli* cells have been extensively used to propagate plasmids, being the most popular ones based on high-copy number pMB1 vectors.

Plasmid instability is a major problem in large-scale industrial manufacturing of plasmid vectors. Some of the solutions proposed to tackle structural instability rely on using recombinant-deficient (*rec*) hosts and on plasmid minimization (reviewed by [2, 4]). A variety of stressing agents such as nutrient limitation, oxidative damage, UV radiation, antibiotic and acid exposure have been accepted to result in increased instability and mutation rates [5–11]. Yet, not much attention has been given to the presence of certain labile nucleotide sequences (hot spots) in plasmids, which

Sofia C. Ribeiro and Pedro H. Oliveira contributed equally to this work.

S. C. Ribeiro · P. H. Oliveira · D. M. F. Prazeres · G. A. Monteiro (✉)
Centre for Biological and Chemical Engineering, Institute for Biotechnology and Bioengineering (IBB), Instituto Superior Técnico, 1049-001 Lisbon, Portugal
e-mail: gabmonteiro@ist.utl.pt

are known for a long time to be a major cause of mutability [12]. For instance, rearrangements between repetitive sequence units are a major cause of genome instability among eukaryotes and prokaryotes. In *E. coli*, several deletion-formation events in plasmids engineered to contain direct or inverted repeats have been described [13–18]. These events take place in the absence of general homologous recombination functions and have been proposed to proceed by mechanisms involving slipped misalignment accompanied by sister-chromosome exchange (SCE) [15, 19, 20]. The essence of these recombination models is that at a stalled replication fork, either the nascent leading and lagging strands are displaced and annealed with each other [20] or, during leading strand synthesis, the strands of the repeats misalign to form a recombinogenic intermediate either by melting and slip-pairing or by replication slippage [15]. Indeed, this type of mutations seems to be quite frequent and stimulated by failure of the replication machinery [21]. According to the proposed models, the recombination products obtained from a parental substrate harbouring repeated sequences are mainly in the monomeric (M) or hetero-dimeric forms (named 1 + 2 and 1 + 3). The M form is structurally identical to the parental plasmid but lacks the intervening sequence and one direct repeat. On the other hand, the 1 + 2 and 1 + 3 forms are head-to-tail dimers composed of one parental plasmid and one monomeric product, with the 1 + 3 form presenting an additional direct repeat and intervening sequence.

Plasmid pGPV-PV is a candidate DNA vaccine against rabies [22, 23] constructed by inserting the rabies virus (PV strain) glycoprotein gene (GPV) in the commercial pCIneo plasmid backbone (pMB1-derived). In this work, we have investigated the stability of both plasmids in *E. coli* cells through the distribution and frequency of the recombinant plasmid forms arising under different growth conditions.

Materials and Methods

Media and Enzymes

Cells were grown in Luria-Bertani (LB) medium (Sigma). When needed, ampicillin (Roche) and kanamycin (Calbiochem) were added at final concentrations of 100 and 30 μg/mL, respectively. Restriction enzymes and phage T4 DNA ligase were from Promega (Madison, USA).

Bacterial Strains and Plasmid DNA

The *E. coli* strains DH5α (F$^-$ *ϕ*80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1), HB101 (F$^-$ *mcr*B *mrr hsd*S20 (r_B^-, m_B^-) *rec*A13 *sup*E44 *ara*14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20(Smr) *xy15* λ^- *leu mtl*1) and TOP10F' (F′{*lac*I^q Tn*10* (TetR)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *ϕ*80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L *end*A1 *nup*G) were purchased from Invitrogen. *E. coli* JM109(DE3) (*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (r_K^-, m_K^+), *rel*A1, *sup*E44, λ^- Δ(*lac-pro*AB), [F′, *tra*D36, *pro*AB, *lac*I^{q}Z ΔM15], *λ*(DE3) was purchased from Promega.

The plasmids used in this work were pCIneo, a 5,472 bp mammalian expression vector (Promega), and pGPV-PV (Institute Pasteur), a 7,067 bp candidate DNA vaccine against rabies. The latter was obtained by inserting the rabies virus (PV strain) glycoprotein gene (GPV) in the pCIneo vector under the control of the cytomegalovirus promoter/enhancer [22, 23].

pCIneo3 construction started with the creation of SacII and ApaI restriction sites, respectively, at positions 1,990 and 2,455 bp of pCIneo, using the QuickChange Site-Directed Mutagenesis Kit (Stratagene). The sequence between the two sites, which contains the SV40 eukaryotic promoter (2,000–2,418) and one of the 28 bp direct repeats, was excised from the plasmid by digestion with SacII/ApaI and purified from a low melting agarose gel with the Wizard Plus Minipreps DNA Purification System (Promega). A prokaryotic promoter similar to the one that controls the expression of the amp^R gene was then inserted upstream the neo^R gene in the pCIneo fragment obtained above as follows: the corresponding sense and antisense oligonucleotides were designed to contain three extra nucleotides at each extremity to allow subsequent digestion and the restriction sites of the enzymes SacII and ApaI (at the 5′ and 3′ ends, respectively, as underlined in the sense sequence 5′ GCACCGCGGTTCAAATATGTATCCGCTC ATGAGACAATAACCCTGATAAATGCTTCAATAATA TTGAAAAAGGGCCCTAC 3′). The sense and antisense fragments were annealed using a thermocycler TGradient (Biometra, Göttingen, Germany) with a temperature ramp ranging from 70 to 25 °C. After double digestion with SacII/ApaI the fragment was gel purified and ligated with pCIneo at room temperature with T4 DNA ligase for 18 h (pCIneo2). Next, a 28 bp sequence identical to the one previously deleted was inserted upstream the new prokaryotic promoter in plasmid pCIneo2. First, an EcoRV site was created at position 1,975 bp and the vector was digested with EcoRV and SacII. The digested vector was purified (3 μg) and dephosphorylated with 0.02 U CIAP (Promega) with two cycles of 15 min at 37 °C and 15 min at 56 °C. The resulting fragment was then purified after excision from a low melting agarose gel. The 28 bp sequence was synthesized with three extra nucleotides at the extremities to allow subsequent digestion plus the restriction sites of the enzymes EcoRV and SacII (at the 5′ and 3′ ends, respectively, as underlined in the sense sequence 5′ TCAGATATCCTTCTGACACAACAGTCTCGAACTT

AAGCCGCGGTAC 3′). Hybridization of the corresponding sense and antisense oligonucleotides was carried as described above. After double digestion with SacII/EcoRV, the restriction enzymes were inactivated by heating (15 min at 65 °C) since the fragment was too small to be handled by a standard purification procedure. pCIneo3 was obtained by ligation of the resulting fragment with the linearized pCIneo2 at 15 °C for 16 h. All primers were obtained from Thermo Electron, Germany. Plasmids were sequenced by MWG-Biotech, Germany.

Plasmid DNA Analysis

Individual colonies that arose at 24 and 48 h under selective kanamycin pressure were grown (12–15 h at 37 °C and 250 rpm) in 15 mL polypropylene tubes containing 5 mL of LB medium supplemented with ampicillin. pDNA was then extracted using the High Pure Plasmid Isolation Kit (Roche) and identified by subsequent digestion with BamHI and/or EcoRI and agarose gel electrophoresis analysis. Plasmid fragments obtained upon digestion were visualized in 1% agarose gels under UV light after ethidium bromide (1 μg/mL) staining.

Cell Viability and Recombination Frequency

Cell viability was defined as the ability of a cell to multiply and was analysed by plate counting in LB solid medium. Briefly, agar plugs containing no visible colonies were excised from kanamycin Petri dishes at 0, 24 and 48 h. The agar plugs were then resuspended in sterile 0.9% NaCl, vortexed and the cell suspension was appropriately diluted and plated in LB solid medium. Recombination frequency was determined as the ratio of the average number of kan^R colonies appearing in the first 24 h to the total number of viable cells plated.

Quantitative Real-Time PCR

Quantitative real-time PCR was used to detect and quantify the recombination event in *E. coli* DH5α cells harbouring pCIneo. This was performed with the amplification of a 369 bp fragment using primers PH1F (5′ GCTTT ATTGCGGTAGTTTATCACAG 3′) and KanR (5′ CAA TAGCAGCCAGTCCCTTC 3′) flanking the direct repeats. All PCR reactions were performed in a 20 μL final volume containing 1.6 μL of $MgCl_2$ solution (2.0 mM final concentration), 0.4 μL of each primer (0.2 μM final concentrations), 11.6 μL PCR grade water, 2.0 μL of 10 × SYBR Green I mixture and 4.0 μL of sample. Reactions were carried out in a Roche Light Cycler™ detection system with the following program: 10 min at 95 °C followed by 45 cycles of 10 s at 95 °C, 5 s at 55 °C and 5 s at 72 °C. Reactions were finally kept at 70 °C for 30 s and heat-denatured over a temperature gradient of 0.1 °C/s from 70 to 95 °C. The temperature transition rate was 20 °C/s in all segments. Threshold cycle (C_T) values were calculated by the LightCycler software Version 3.4 (Roche Diagnostics) using the Fit Points method. Positive standards were prepared by spiking non-transformed 2.0×10^5 DH5α cells with increasing amounts of pure monomeric pDNA ranging from 2.0×10^1 to 2.0×10^8 molecules. The same cell number was used in each PCR reaction in order to keep potential cell inhibitors constant. Appropriate cell and pDNA dilutions were performed with milliQ water and 2 μL of each were mixed with the remaining PCR reagents as described above. Negative controls without pDNA were also prepared. Mean C_T values and standard deviations were calculated from at least three independent assays. PCR efficiency (E) was calculated by $E = 10^{-\frac{1}{m}} - 1$, where m is the slope of the linear regression equation.

Results and Discussion

Identification of the Mutational Hot Spot

Plasmid pGPV-PV (Fig. 1a) is a candidate DNA vaccine against rabies [22, 23] which was constructed using pCIneo (Fig. 1b) plasmid (pMB1-derived) as a backbone. Both plasmids contain an ampicillin resistance (amp^R) gene for selection in bacteria. The plasmids also contain a neomycin phosphotransferase (neo^R) gene under the regulation of an SV40 enhancer and early promoter region [24], which is used as a selectable marker in mammalian cells. Serendipitously, we found that several *rec*A *E. coli* strains (DH5α, TOP10F', HB101 and JM109(DE3)) harbouring these plasmids were able to grow in the presence of the aminoglycoside antibiotic kanamycin after approximately 20–40 h. No growth was observed in control experiments carried out with the same strains devoid of either pGPV-PV or pCIneo plasmids.

A subsequent restriction analysis of plasmids extracted from cells which had been subjected to kanamycin stress revealed the presence of some unexpected low and high molecular-weight restriction fragments (Fig. 2) derived from recombined plasmid molecules. These patterns were confirmed in recombinant colonies of the four *E. coli* strains, harbouring either pGPV-PV or pCIneo, whenever growth was conducted in the presence of kanamycin (data not shown). A 4-kb BamHI (hereafter called bam2) plasmid restriction fragment was common to all the recombination patterns (Fig. 2). We subsequently found that *E. coli* cells transformed with the bam2 self-ligated form were able to grow in the presence of kanamycin (and

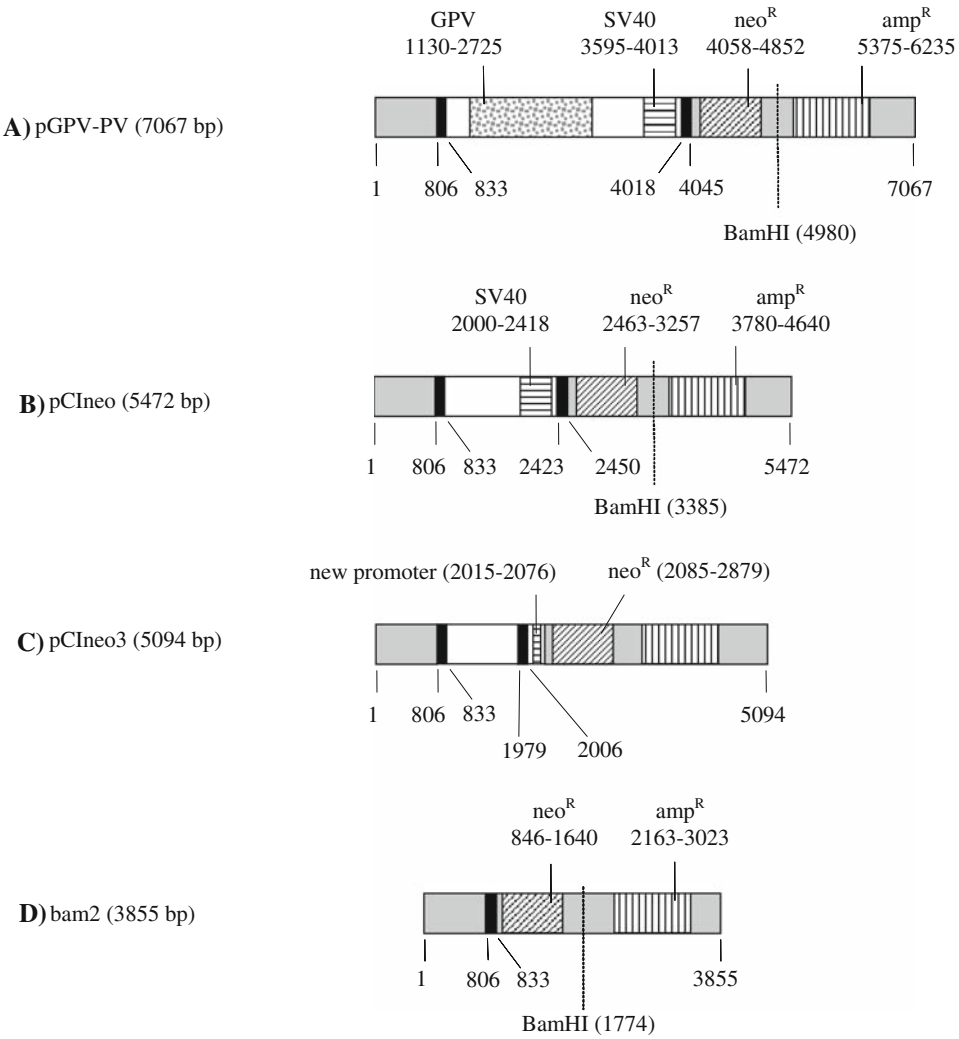

Fig. 1 Schematic view of plasmids used in this work: pGPV-PV (**a**), pCIneo (**b**), pCIneo3 (**c**) and bam2 fragment (**d**). When *E. coli* cells are grown in a kanamycin-containing medium, recombination events take place involving the 28 bp direct repeats (shown in black at positions 806–833 and 4,018–4,045 bp (**a**) and at positions 806–833 and 2,423–2,450 bp (**b**)). Digestion of pDNA extracted from these cells with the single cutting enzyme BamHI revealed an extra 3,856 bp fragment called bam2. A deletion monomeric form derived from pCIneo and pGPV-PV was obtained by ligating the extremities in bam2 (**d**). Sequencing analysis of this form revealed that it matched the pGPV-PV sequence devoid of a 3,212 bp fragment ranging from 806 to 4,017 bp or from 834 to 4,045 bp. pCIneo 3 was constructed by replacing the SV40 promoter by a new prokaryotic promoter (**c**). amp^R-ampicillin resistance gene; neo^R-neomycin resistance gene; CMV-cytomegalovirus; GPV-rabies virus (PV strain) glycoprotein gene

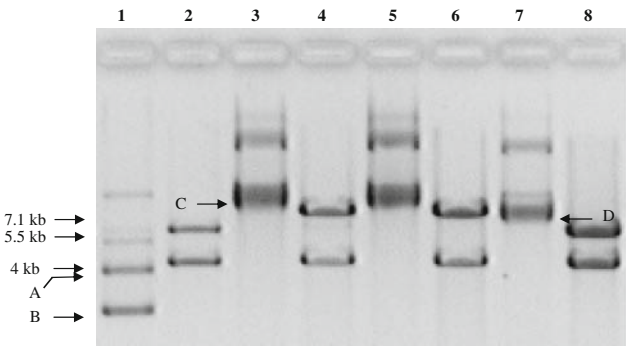

Fig. 2 BamHI restriction analysis of pDNA extracted from colonies of *E. coli* DH5α cells harbouring pCIneo and grown under kanamycin stress (30 μg/mL). Individual kan^R colonies were picked from LB plates and grown (12–15 h, 37 °C) in 5 mL LB supplemented with ampicillin. Plasmids were extracted with the HighPure Plasmid Isolation Kit (Roche). Lanes 1, 3, 5 and 7 contain non-digested plasmid DNA, whereas lanes 2, 4, 6 and 8 contain BamHI digested plasmid. The agarose gel shows the monomeric (lanes 1 and 2), 1 + 3 (lanes 3–6) and 1 + 2 (lanes 7 and 8) forms derived from pCIneo. Letters A, B, C and D correspond, respectively, to the supercoiled isoforms of the parental, monomeric, 1 + 3 and 1 + 2 forms

also in ampicillin). This indicates that the bam2 fragment holds the origin of replication and the ability to express the aminoglycoside phosphotransferase that gives the resistant phenotype to kanamycin. Sequence analysis of the bam2 fragment revealed that its 3,855 bp (Fig. 1d) matched the pGPV-PV and pCIneo sequences lacking either a 3.2 kb fragment ranging from 806 to 4,017 bp (pGPV-PV) or a 1.6 kb fragment ranging from 806 to 2,423 bp (pCIneo). An in silico analysis with the GCG software package v.9.1 (Genetics Computer Group, Madison, Wis.) further showed that both plasmids contain two 28 bp direct repeats (5′-CTTCTGACACAACAGTCTCGAACTTAAG-3′) in positions 806–833 and 4,018–4,045 (pGPV-PV) and in positions 806–833 and 2,423–2,450 (pCIneo). This means that the bam2 fragment lacks the entire sequence located between the direct repeats, including one of the repeats. Furthermore, restriction analysis of the high molecular-weight fragments obtained revealed a head-to-tail type dimeric structure composed of the bam2 fragment and parental plasmid.

Altogether, this data points to deletion-duplication events triggered by the presence of directly repeated sequences. This mutational hot spot led to the formation of the typical monomeric and hetero-dimeric products by slipped misalignment accompanied by sister chromosome exchange, in a similar way as described previously [15, 19, 20]. In our case, the obtained deletion-duplication products and the molecular weight of the fragments generated by BamHI digestion are shown in Fig. 3. Notice that the bam2 fragment actually corresponds to the monomeric deletion product and that the high molecular-weight products previously shown (Fig. 2) correspond to the heterodimeric 1 + 2 and 1 + 3 forms. Additionally, the replacement of the SV40 eukaryotic promoter with a prokaryotic one (see Fig. 1c) led to the non-detection of kanamycin resistant cells, which indicates that kanamycin stress is involved in the selection of cells which harbour the recombinant plasmid forms.

Several reports related with plasmid instability mediated by direct repeat recombination make use of an internal duplication in a gene of interest which can only be phenotypically detected after the recombination event [13, 15–19, 25, 26]. This is not our case. Rather, we have strong experimental evidences showing that the present recombination event brings the neomycin phosphotransferase gene closer to a promoter-like sequence located in the origin of replication, which allows the transcription of the neo^r gene (unpublished results).

Effect of Cell Growth Phase on the Kinetics of Mutant Emergence

The effect of cell growth phase in the kinetics of resistant emergence under kanamycin stress was analysed in *E. coli* DH5α and *E. coli* JM109(DE3) harbouring pCIneo. Cells were previously grown in LB liquid media containing ampicillin, harvested at mid-exponential, late-exponential, early-stationary and late-stationary growth phases and then seeded on LB plates containing kanamycin. The cumulative number of kan^R colonies increased for 72 h in both strains, and was about 2.3-fold higher for cells collected at early- or late-stationary growth phase in comparison with mid-exponential phase cells (for both strains). As for the number of resistant colonies obtained after plating cells collected at late-stationary phase, it was higher for *E. coli* JM109(DE3) in comparison with *E. coli* DH5α. The above variations could be explained if the potential acquirers of resistance to kanamycin had differential viability after being plated in medium with kanamycin. However, both strains, when grown till different stages, have shown a similar pattern of decrease in cell viability of about 2 orders of magnitude times after 48 h (data not shown).

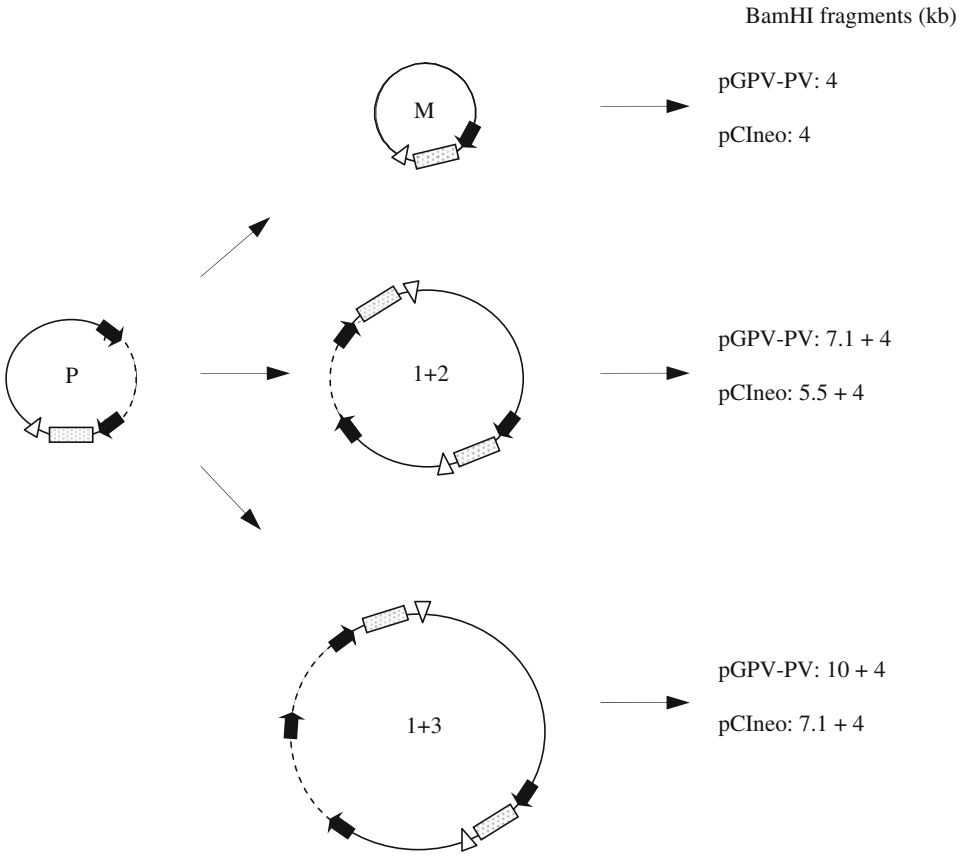

Fig. 3 Products of DNA rearrangement mediated by direct repeats in plasmids pGPV-PV and pCIneo. The 28 bp direct repeats are shown as black arrows, BamHI single cutting site is shown as open triangle and the neo^R gene is shown as a stippled box. The size of fragments obtained upon BamHI digestion is also shown. P, parental pCIneo or pGPV-PV; M, monomeric product; 1 + 2 and 1 + 3, heterodimeric products with deleted and duplicated regions

More reliable conclusions can be drawn by using recombination frequency (number of kan^R colonies/number of cells plated) instead of number of recombinants. In our system, however, recombination frequency could only be determined in cells arising during the first 24 h because, unlike cells arising later at 48 h (and beyond), they contain mostly recombined plasmid (see Fig. 4). The mean recombination frequency observed in *E. coli* DH5α was estimated to be around 1×10^{-6} while it was about 21-fold higher for JM109(DE3) (Table 1). Despite the differences between the two strains these values are well within the range of recombination frequencies described for similar systems [13, 15–18, 27, 28]. The main difference between the genotypes of the two strains is that JM109(DE3) contains a copy of the T7 RNA polymerase gene, which is under tight control of the inducible lacUV5 promoter and is used for high-level expression of recombinant proteins. We suggest that in JM109(DE3) some leakage of that promoter can occur leading to replication stalling due to hindering of DNA polymerase by an excess of T7 RNA polymerase molecules bound to plasmid T7 promoter, which in plasmid pCIneo is 234 bp downstream the first 28 bp direct repeat and 1,138 bp from the origin of replication. This transcription-associated recombination [29] could explain the higher recombination frequency observed in strain JM109(DE3). To strengthen this hypothesis, JM109(DE3) cells were grown in the presence or absence of IPTG (an inducer of the expression of T7 RNA polymerase gene). Preliminary real-time PCR results have shown that plasmid copy number upon growth in the presence of IPTG decreased substantially (from around 108 copies *per* cell to around 4 copies *per* cell), while the number of recombinant molecules was kept approximately constant. This means that the percentage of recombinant molecules in the presence of IPTG was more than 27-fold higher.

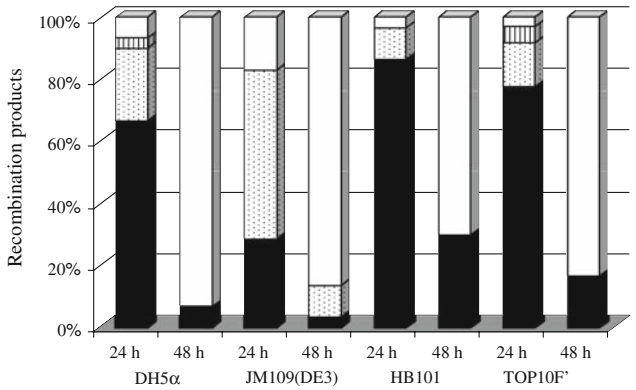

Fig. 4 Distribution of recombinant products obtained from colonies grown in LB solid media supplemented with kanamycin (30 μg/mL) as a function of *E. coli* strain (DH5α, JM109(DE3), HB101 and TOP10F') and incubation time (24 and 48 h). Individual kan^R colonies emerging after 24 or 48 h were picked from LB plates and grown (12–15 h, 37 °C) in 5 mL LB supplemented with ampicillin (100 μg/mL). Plasmids were extracted with the HighPure Plasmid Isolation Kit (Roche) and analysed by agarose gel electrophoresis. Open bars: non-recombined plasmid; stippled bars: 1 + 2 form; hatched bars: 1 + 3 form; filled bars: monomeric product. Each bar represents 30 colonies

Table 1 Effect of cell growth phase on the kinetics of mutant emergence in *E. coli* DH5α and JM109(DE3) cells harbouring plasmid pCIneo

Strain	Growth phase	Frequency of kan^R cells appearance
DH5α	Mid exponential	$(1.04 \pm 0.92) \times 10^{-6}$
	Late exponential	$(1.03 \pm 0.72) \times 10^{-6}$
	Early stationary	$(0.78 \pm 0.70) \times 10^{-6}$
	Late stationary	$(1.32 \pm 0.99) \times 10^{-6}$
JM109(DE3)	Mid exponential	$(1.06 \pm 0.46) \times 10^{-5}$
	Late exponential	$(1.93 \pm 0.57) \times 10^{-5}$
	Early stationary	$(2.75 \pm 1.00) \times 10^{-5}$
	Late stationary	$(3.10 \pm 0.82) \times 10^{-5}$

The frequency of appearance of kan^R cells was measured 24 h after plating 5×10^7 (DH5α) or 1×10^7 (JM109(DE3)) cells under kanamycin (30 μg/mL) stress

Plasmid Recombination Forms

Single colony plasmid DNA analysis of the four *E. coli* strains revealed that deletion-duplication products from pCIneo arose mainly in the first 24 h of growth, with high preponderance of monomer (M) and low percentage of the heterodimeric 1 + 2 and 1 + 3 forms (Fig. 4). This distribution of plasmid forms is in agreement with the results published by several authors [15, 17, 19]. Curiously, JM109(DE3) revealed a higher content of heterodimeric 1 + 2 at 24 h (also seen at 48 h) in comparison with the remaining strains (Fig. 4). Again, this points towards the possibility that JM109(DE3) strain has some unique genotype feature which affects the recombination mechanism. Although fewer experiments were performed, a similar distribution of plasmid forms was observed when pGPV-PV was harboured in each of the four strains (data not shown).

Another interesting result was the fact that colonies arising later (48 h), which were typically smaller, contained mainly non-recombined parental plasmid (Fig. 4). This means that in cells containing plasmid pCIneo (or pGPV-PV), some incipient and slow growth occurs even in the absence of recombination, which is responsible for the appearance of the late arising kan^R colonies. This slow growth was also confirmed in LB liquid medium supplemented with kanamycin. In these experiments, cell cultures starting with an $OD_{600\ nm}$ of 0.1 grew very slowly, reaching an OD_{600} of 0.6 only after 48–60 h. These results show the

importance of monitoring the plasmid DNA forms distribution at different stages of colony materialization, in order to distinguish colonies which contain recombined plasmid from the ones harbouring non-recombined plasmid that are "naturally resistant" and grow slowly in kanamycin.

Quantification of the Recombination Event

After a literature search, we failed to find any quantification of the number of recombined plasmid molecules existing pre-stress exposure in similar systems. To ascertain this point, the number of pre-existing recombined plasmids in cells grown in ampicillin- or kanamycin-containing media was determined by real-time quantitative PCR. For this purpose, a pair of primers flanking the direct repeats was used, and the PCR procedure was set up in order to amplify a single 369 bp amplicon from the plasmid recombinants. No amplification was detected when the parental plasmid was used as the template, since the corresponding 1,986 bp long amplicon is not obtained under the conditions used. This methodology allowed us to identify and quantify the overall number of recombined plasmids. Because it is important to accurately determine the number of plasmid molecules *per* cell, the quantitative PCR was performed directly in cells. In this way we avoid plasmid purification steps, which might not give reproducible and 100% plasmid yields when using cells harvested at different growth phases and thus with a different plasmid to impurities ratio. The calibration curve obtained with standards prepared by spiking the purified monomeric form with 2.0×10^5 DH5α cells was linear in the range of 2.0×10^3 to 2.0×10^8 molecules and PCR efficiency in this range was 79.7%. When DH5α cells were grown in liquid medium with ampicillin, the number of recombined plasmid molecules remained approximately constant at 1.5×10^4 *per* 2×10^5 cells for a period of 48 h (Fig. 5). This means that 1 out of 13 cells would have a molecule of recombined plasmid (in ca. 200–300 molecules of parental plasmid). This would be a contamination of about 0.02% in a purified plasmid batch. Considering that one recombined plasmid is sufficient to confer the kan^R phenotype, the recombination frequency would be 7.5×10^{-2}, which is almost 5 orders of magnitude higher than the frequency determined experimentally. This discrepancy can be explained if we consider that: (i) there is a decrease in cell viability after kanamycin exposure, (ii) more than one recombined plasmid *per* cell is required to give the resistance phenotype and (iii) there is competition between recombined and parental plasmids. In the presence of kanamycin, protein synthesis is inhibited, reducing the cell metabolism to a minimum or even halting it altogether, with the consequent cell death. Only cells that have a high enough copy number of recombined plasmids are able to express aminoglycoside phosphotransferase, survive and multiply to form a detectable colony. Nevertheless, the stable number of recombined plasmids in cells grown without the selective pressure is in agreement with previous studies relating to the establishment of a recombinant plasmid and its ability to overcome the inhibitory effect of the parental plasmid [30]. Because recombinant plasmids give a competitive advantage to cells, when growth was conducted under selective pressure, the number of recombined molecules reached a maximum of 3.0×10^7 *per* 2×10^5 cells in the same period of time (Fig. 5).

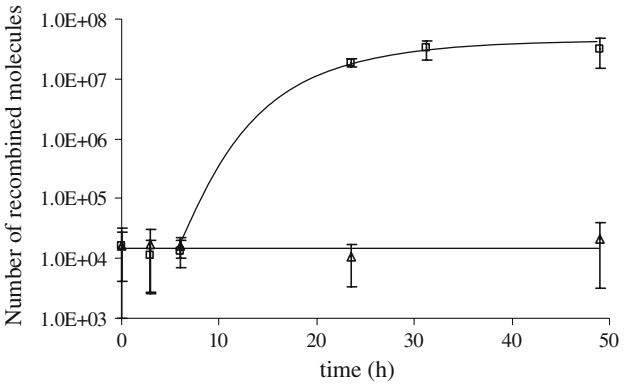

Fig. 5 Time-course evolution of recombined pCIneo in DH5α cells grown in LB liquid medium supplemented with 100 μg/mL ampicillin (open triangles) or 30 μg/mL kanamycin (open squares). The number of recombined molecules was determined by quantitative real-time PCR. Error bars represent standard deviations of at least three independent measures

Conclusions

Upon the development of an effective plasmid-based biopharmaceutical, the nature of the extra-functional genes represents a safety concern regarding plasmid stability. In this work, we have found an approximate constant number of 1.5×10^4 recombined molecules dispersed in a population of 2×10^5 cells due to the presence of two 28 bp direct repeats in the plasmid vectors under study. Such a perceptible number (0.02%) of pre-existing recombined pDNA molecules poses an obvious safety concern. Moreover, when conducting a large-scale fermentation process, the occurrence of nutrient or oxygen limitations due to high cell densities might, exacerbate mutational events in those plasmids which harbour instability regions. Furthermore, this type of recombination appears to be enhanced in JM109(DE3), a bacterial strain engineered for protein production. This raises new concerns about plasmid instability when plasmid vectors are designed as T7 promoter-based expression plasmids.

A BLAST against the NCBI nucleotide databases using the 28 bp sequence of pCIneo revealed full identity with regions present in several well-known expression and cloning vectors. This 28 bp sequence was found to be present as a single copy in vectors like pCI and pSI, often used as backbones for vaccine development. But, in some expression and cloning plasmids (e.g. pTarget, pALTER, pBIND and pACT), this sequence was found to be repeated and, therefore, might pose a concern in terms of instability. The vector pUNIV (pCIneo derivative) has recently been used for protein expression in mammalian cells and *Xenopus* sp. [31], pRIG1 (pTarget derivative) used for creation of genome-wide protein expression libraries [32], while pALTER and pCIneo have been recently used as backbones for the development of DNA vaccines against bursal disease virus [33] and leishmaniasis [34], respectively.

It is therefore vital to carefully analyse the plasmid sequence in order to unveil the presence of small populations of unwanted plasmid molecules. In the absence of this control, there is a risk that the use of therapeutic DNA for non-viral gene therapy or DNA vaccination could lead to the injection of numerous undetectable mutated plasmids with unknown biological properties.

Acknowledgements We acknowledge N. Tordo and P. Perrin (Institute Pasteur, Paris) for providing plasmid pGPV-PV. The work presented here was financially supported by Fundação para a Ciência e Tecnologia, Portugal through project grants (POCTI/BIO/32875/2000 and POCI/BIO/55799/2004) and PhD grants awarded to S. C. Ribeiro (BD/2957/2000) and P. H. Oliveira (BD/22320/2005).

References

1. Prather, K. J., Sagar, S., Murphy, J., & Chartrain, M. (2003). Industrial scale production of plasmid DNA for vaccine and gene therapy: Plasmid design, production, and purification. *Enzyme and Microbial Technology, 33*(7), 865–883. doi:10.1016/S0141-0229(03)00205-9.
2. Prazeres, D. M. F., Ferreira, G. N. M., Monteiro, G. A., Cooney, C. L., & Cabral, J. M. S. (1999). Large-scale production of pharmaceutical-grade plasmid DNA for gene-therapy: Problems and bottlenecks. *Trends in Biotechnology, 17*(4), 169–174. doi:10.1016/S0167-7799(98)01291-8.
3. Stadler, J., Lemmens, R., & Nyhammar, T. (2004). Plasmid DNA purification. *The Journal of Gene Medicine, 6*(Suppl. 1), S54–S66. doi:10.1002/jgm.512.
4. Glenting, J., & Wessels, S. (2005). Ensuring safety of DNA vaccines. *Microbial Cell Factories, 4*, 26. doi:10.1186/1475-2859-4-26.
5. Bjedov, I., Tenaillon, O., Gérard, B., Souza, V., Denamur, E., Radman, M., et al. (2003). Stress-induced mutagenesis in bacteria. *Science, 300*(5624), 1404–1409. doi:10.1126/science.1082240.
6. Chopra, I., O'Neill, A. J., & Miller, K. (2003). The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Resistance Updates, 6*(3), 137–145. doi:10.1016/S1368-7646(03)00041-4.
7. Giraud, A., Matic, I., Radman, M., Fons, M., & Taddei, F. (2002). Mutator bacteria as a risk factor in treatment of infectious diseases. *Antimicrobial Agents and Chemotherapy, 46*(3), 863–865. doi:10.1128/AAC.46.3.863-865.2002.
8. Moran, A. P., Knirel, Y. A., Senchenkova, S. N., Widmalm, G., Hynes, S. O., & Jansson, P. E. (2002). Phenotypic variation in molecular mimicry between *Helicobacter pylori* lipopolysaccharides and human gastric epithelial cell surface glycoforms. Acid-induced phase variation in Lewis(x) and Lewis(y) expression by *H. pylori* lipopolysaccharides. *The Journal of Biological Chemistry, 277*(8), 5785–5795. doi:10.1074/jbc.M108574200.
9. Negri, M. C., Morosini, M. I., Baquero, M. R., Campo Rd, R., Blazquez, J., & Baquero, F. (2002). Very low cefotaxime concentrations select for hypermutable *Streptococcus pneumoniae* populations. *Antimicrobial Agents and Chemotherapy, 46*(2), 528–530. doi:10.1128/AAC.46.2.528-530.2002.
10. Oliver, A., Cantón, R., Campo, P., Baquero, F., & Blázquez, J. (2000). High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science, 288*(5469), 1251–1253. doi:10.1126/science.288.5469.1251.
11. Ouchane, S., Picaud, M., Vernotte, C., & Astier, C. (1997). Photooxidative stress stimulates illegitimate recombination and mutability in carotenoid-less mutants of *Rubrivivax gelatinosus*. *The EMBO Journal, 16*(15), 4777–4787. doi:10.1093/emboj/16.15.4777.
12. Albertini, A. M., Hofer, M., Calos, M. P., & Miller, J. H. (1982). On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. *Cell, 29*(2), 319–328. doi:10.1016/0092-8674(82)90148-9.
13. Bi, X., & Liu, L. F. (1994). *recA*-independent and *recA*-dependent intramolecular plasmid recombination: Differential homology requirement and distance effect. *Journal of Molecular Biology, 235*(2), 414–423. doi:10.1006/jmbi.1994.1002.
14. Bi, X., & Liu, L. F. (1996). DNA rearrangement mediated by inverted repeats. *Proceedings of the National Academy of Sciences of the USA, 93*(2), 819–823. doi:10.1073/pnas.93.2.819.
15. Bi, X., & Liu, L. F. (1996). A replicational model for DNA recombination between direct repeats. *Journal of Molecular Biology, 256*(5), 849–858. doi:10.1006/jmbi.1996.0131.
16. Dianov, G. L., Kuzminov, A. V., Mazin, A. V., & Salganik, R. I. (1991). Molecular mechanisms of deletion formation in *Escherichia coli* plasmids. I. Deletion formation mediated by long direct repeats. *Molecular and General Genetics, 228*(1–2), 153–159. doi:10.1007/BF00282460.
17. Lovett, S. T., Gluckman, T. J., Simon, P. J., Sutera, V. A., Jr., & Drapkin, P. T. (1994). Recombination between repeats in *Escherichia coli* by a *recA*-independent, proximity sensitivity mechanism. *Molecular and General Genetics, 245*(3), 294–300. doi:10.1007/BF00290109.
18. Mazin, A. V., Kuzminov, A. V., Dianov, G. L., & Salganik, R. I. (1991). Molecular mechanisms of deletion-formation in *Escherichia coli* plasmids II. Deletion formation mediated by short direct repeats. *Molecular and General Genetics, 228*(1–2), 209–214. doi:10.1007/BF00282467.
19. Feschenko, V. V., & Lovett, S. T. (1998). Slipped misalignment mechanisms of deletion formation: Analysis of deletion endpoints. *Journal of Molecular Biology, 276*(3), 559–569. doi:10.1006/jmbi.1997.1566.
20. Lovett, S. T., Drapkin, P. T., Sutera, V. A., Jr, & Gluckman-Peskind, T. J. (1993). A sister-strand exchange mechanism for *recA*-independent deletion of repeated DNA sequences in *Escherichia coli*. *Genetics, 135*(3), 631–642.
21. Lovett, S. T. (2004). Encoded errors: Mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Molecular Microbiology, 52*(5), 1243–1253. doi:10.1111/j.1365-2958.2004.04076.x.
22. Bahloul, C., Jacob, Y., Tordo, N., & Perrin, P. (1998). DNA-based immunization for exploring the enlargement of

immunological cross-reactivity against the lyssaviruses. *Vaccine, 16*(4), 417–425. doi:10.1016/S0264-410X(97)00204-1.
23. Jallet, C., Jacob, Y., Bahloul, C., Drings, A., Desmezieres, E., Tordo, N., et al. (1999). Chimeric lyssavirus glycoproteins with increased immunological potential. *Journal of Virology, 73*(1), 225–233.
24. Brondyk, W. H. (1995). The pCI-neo mammalian expression vector. *Promega Notes Magazine, 51*, 10–14.
25. Mazin, A. V., Timchenko, T. V., Saparbaev, M. K., & Mazina, O. M. (1996). Dimerization of plasmid DNA accelerates selection for antibiotic resistance. *Molecular Microbiology, 20*(1), 101–108. doi:10.1111/j.1365-2958.1996.tb02492.x.
26. Morag, A. S., Saveson, C. J., & Lovett, S. T. (1999). Expansion of DNA repeats in *Escherichia coli*: Effects of recombination and replication functions. *Journal of Molecular Biology, 289*(1), 21–27. doi:10.1006/jmbi.1999.2763.
27. Chédin, F., Dervyin, E., Dervyin, R., Ehrlich, S. D., & Noirot, P. (1994). Frequency of deletion formation decreases exponentially with distance between short repeats. *Molecular Microbiology, 12*(4), 561–569. doi:10.1111/j.1365-2958.1994.tb01042.x.
28. Oliveira, P. H., Lemos, F., Monteiro, G. A., & Prazeres, D. M. F. (2008). Recombination frequency in plasmid DNA containing direct repeats—Predictive correlation with repeat and intervening sequence length. *Plasmid*, accepted.
29. Mirkin, E. V., & Mirkin, S. M. (2007). Replication fork stalling at natural impediments. *Microbiology and Molecular Biology Reviews, 71*(1), 13–35. doi:10.1128/MMBR.00030-06.
30. Bierne, H., Ehrlich, S. D., & Michel, B. (1995). Competition between parental and recombinant plasmids affects the measure of recombination frequencies. *Plasmid, 33*(2), 101–112. doi:10.1006/plas.1995.1012.
31. Venkatachalan, S. P., Bushman, J. D., Mercado, J. L., Sancar, F., Christopherson, K. R., & Boileau, A. J. (2006). Optimized expression vector for ion channel studies in *Xenopus* oocytes and mammalian cells using alfalfa mosaic virus. *Pflügers Archiv European Journal of Physiology, 454*(1), 155–163.
32. Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., et al. (2001). Creation of genome-wide protein expression libraries using random activation of gene expression. *Nature Biotechnology, 19*(5), 440–445. doi:10.1038/88107.
33. Sun, J. H., Yan, Y. X., Jiang, J., & Lu, P. (2005). DNA immunization against very virulent infectious bursal disease virus with VP2-4-3 gene and chicken IL-6 gene. *Journal of Veterinary Medicine Series B, 52*(1), 1–7. doi:10.1111/j.1439-0450.2004.00813.x.
34. Gomes, D. C. O., Pinto, E. F., Barbosa de Melo, L. D., Lima, W. P., Larraga, V., Lopes, U. G., et al. (2006). Intranasal delivery of naked DNA encoding the LACK antigen leads to protective immunity against visceral leishmaniasis in mice. *Vaccine, 25*(12), 2168–2172. doi:10.1016/j.vaccine.2006.11.060.