Rational engineering of *Escherichia coli* strains for plasmid
biopharmaceutical manufacturing

Geisa A. L. Gonçalves\(^1,2\), Diana M. Bower\(^3\), Duarte M. F. Prazeres\(^1,2\), Gabriel A. Monteiro\(^1,2\), Kristala L. J. Prather\(^3\)*

\(^1\)Department of Bioengineering, Instituto Superior Técnico (IST), Lisbon, Portugal
\(^2\)IBB-Institute for Biotechnology and Bioengineering, Center for Biological and Chemical Engineering, IST, Lisbon, Portugal
\(^3\)Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA

\(^\dagger\) These authors contributed equally to this paper

*Corresponding Author

77 Massachusetts Ave., Room 66-454
Cambridge, MA 02139
Phone: (617) 253-1950
Fax: (617) 258-5042
kljp@mit.edu

Keywords: plasmid biopharmaceuticals, *Escherichia coli*, strain engineering, central metabolism
Abstract

Plasmid DNA (pDNA) has become very attractive as a biopharmaceutical, especially for gene therapy and DNA vaccination. Currently, there are a few products licensed for veterinary applications and numerous plasmids in clinical trials for use in humans. Recent work in both academia and industry demonstrates a need for technological and economical improvement in pDNA manufacturing. Significant progress has been achieved in plasmid design and downstream processing, but there is still a demand for improved production strains. This review will focus on engineering of *Escherichia coli* strains for plasmid DNA production, understanding the differences between the traditional use of pDNA for recombinant protein production and its role as a biopharmaceutical. We will present recent developments in engineering of *E. coli* strains, highlight essential genes for improvement of pDNA yield and quality, and analyze the impact of various process strategies on gene expression in pDNA production strains.
1. **Introduction**

For many years, bacterial plasmid DNA (pDNA) has been used as a biological tool for cloning and expression of recombinant proteins. Recently, plasmid DNA has been considered as a potential biopharmaceutical, particularly for applications in gene therapy and DNA vaccination. Naked plasmid DNA was first observed as a viable vector for expression of heterologous genes in live mammals approximately two decades ago [1]. After this discovery, pDNA became very attractive as a potential biopharmaceutical since its production seems to be cheaper and faster when compared with the production of proteins, viruses or cells [2]. Despite its potential, non-viral gene therapy and DNA vaccines are still under development and a therapeutic product for humans has not yet reached the market. Recent work on new pDNA delivery methods, such as electroporation, has helped mitigate challenges related to low immunogenicity and transgene expression, reinvigorating the field as a whole. Currently, there are some products licensed for veterinary applications [3-4] and numerous clinical trials in phase I, II and III [4-5]. The DNA vaccination market is growing and with it a technological and economical need to improve production methods [6]. Careful design and selection of the host strain is one way to improve the yield and quality of a plasmid product.

Of note is the fact that most fermentation process development for plasmid DNA was built upon the foundations developed for the production of recombinant protein. However, significant differences exist between these processes. For recombinant protein production, high level expression of genes is required to produce large amounts of proteins during the fermentation process, which consist of three different steps: bacterial growth, multicopy plasmid amplification, and protein induction. On the other hand, plasmid DNA production occurs in a different scenario, where it is
necessary to optimize the allocation of cellular resources between biomass formation and plasmid DNA synthesis. Significant technological advances have been made in downstream processing because of the recognition of the key physico-chemical differences between proteins and nucleic acids [7]. In contrast, there are still many opportunities to improve the upstream stages of plasmid biopharmaceutical manufacturing such as development of host strains [8].

The gram-negative bacterium *Escherichia coli* is the most commonly used host for the propagation of plasmid DNA, because it is very robust, capable of fast growth with minimal nutritional requirements, and can give high pDNA yields. The genome of *E. coli* is fully sequenced and can be easily manipulated by techniques that are maturing at a rapid pace. On the other hand, *E. coli* has some disadvantages like endotoxin production and genetic instability, resulting in safety concerns surrounding its use. For this reason, there has been some work investigating other microorganisms, such as the gram-positive, food-grade organism *Lactococcus lactis*, as hosts for biopharmaceutical-grade pDNA [9]. However, taking into consideration the benefits and drawbacks, *E. coli* is currently the most suitable organism for pDNA production on the industrial scale.

This review will focus on engineering of *E. coli* strains for plasmid DNA production, describing recent developments and advances in cell line engineering. We will discuss gene mutations in *E. coli* that could have an impact on plasmid DNA production, as well as how the *E. coli* transcriptome responds to certain process conditions. We will also highlight advantages and disadvantages of particular strains, focusing on genotypic differences, fermentation yields, purification process performance, and transfection efficiency.
2. Effect of plasmid DNA synthesis on *E. coli* central carbon metabolism

Plasmid DNA synthesis can disturb *E. coli* gene regulation, altering levels of gene expression and carbon flux. For the last 3 decades, many researchers observed that plasmid maintenance retards host growth [10-14]. References in the literature correlate the low growth imposed by plasmids with metabolic burden, since plasmid replication and expression of the antibiotic resistance marker requires additional nutrients and energy. Many studies have been done in order to understand this phenomenon using recombinant cells to analyze the effect of plasmid amplification and gene expression on growth rate [10-11, 15]. However, the relationship between plasmid DNA content and growth rate has not been fully explained, as evidenced by recent work on the subject [16-17].

The advent of technologies such as DNA microarrays, metabolic flux analysis, and quantitative real-time PCR has allowed researchers to probe more deeply into the perturbations caused by plasmid maintenance and replication. Specifically, recent reports have shown the effects of plasmid DNA on the central metabolism of *E. coli*, namely glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway [12, 14].

Glycolysis is the main pathway for glucose utilization and energy generation. It is composed of ten reactions catalyzed by specific enzymes that are coded by individual genes. Previous studies reported different levels of glycolytic gene expression in plasmid-bearing cells versus plasmid-free cells. Some results showed most of the glycolytic genes as down regulated [12], while other results have demonstrated up-regulation of the same genes, such as pyruvate kinase I (*pykF*) for cells carrying plasmids [14].
The pentose phosphate pathway (PP) can metabolize different sugars like xylose and ribose, but is considered the second main destination for glucose. The PP pathway is also one of the pathways responsible for biosynthesis of the nucleotide precursors ribose-5-phosphate (R5P) and erythrose-4-phosphate (E4P). Another important product from the PP pathway is NADPH, synthesized by glucose 6-phosphate-1-dehydrogenase (Zwf) and 6-phosphogluconate dehydrogenase (Gnd). NADPH and nucleotides are required for biomass and plasmid DNA production and they are intrinsically correlated in the PP pathway, composing the oxidative and non-oxidative phases respectively. Cunningham et al. [18] demonstrated through a mathematical model that increasing the availability of NADPH via transhydrogenase activity has a positive impact on plasmid DNA production by increasing the reducing power available for pDNA and antibiotic resistance marker synthesis. Cells carrying high copy plasmid DNA require extra synthesis of nucleotides and in this case the carbon flux directed to the PP pathway may be insufficient to cover the cell’s metabolic needs [19].

The tricarboxylic acid (TCA) cycle is composed of eight reactions that oxidize the acetyl group from acetyl-CoA or from other sources. This cycle is important in energy metabolism and biosynthesis and is essential to complete the glycolysis pathway. Some TCA intermediates play an important role in amino acids synthesis like oxaloacetate (OAA) and α-ketoglutarate (AKG). In plasmid-carrying cells, most of the TCA genes were observed as up-regulated for different E. coli strains [12, 14].

### 3. Relevant genes for strain engineering

Many cell line engineering efforts have sought to improve plasmid DNA production by knockout or overexpression of rationally-selected genes (Table 1). One
main area of focus is modification of central carbon metabolism genes to increase flux toward nucleotide and amino acid precursor synthesis and reduction of byproducts, such as acetate. Genes related to improving pDNA quality have also been common targets, as have genes that are involved in various other cellular processes relevant to pDNA production such as the stringent response and DNA replication.

3.1 Central carbon metabolism genes

Central carbon metabolism is a logical target for genetic engineering strategies to increase plasmid DNA yield because of the potential to increase carbon flux to nucleotide precursors (Figure 1). Altering central metabolism to produce more energy and reducing power could also potentially improve plasmid yields. The \textit{pykF} and \textit{pykA} genes encode pyruvate kinase (Pyk) I and II, respectively, and the production of the two isoenzymes is independent. However, both enzymes have a cooperative effect on transforming phosphoenolpyruvate (PEP) into pyruvate (PYR) at the final stage of glycolysis. Previous studies investigated the effect of \textit{pykF} knockout on plasmid-free \textit{E. coli} cells. The growth rate of the mutant cells was slightly lower than wild type, and acetic acid formation was smaller in mutant cells [20-21]. Other studies demonstrated similar behavior in an \textit{E. coli pykF pykA} double knockout bearing plasmid [22]. Less production of acetate was expected based on metabolic deduction since the synthesis of pyruvate would decrease, and phosphoenolpyruvate would be converted to oxaloacetate (OAA) without pyruvate kinase activity. However, PEP, in the presence of glucose, can be converted to pyruvate via the phosphotransferase (PTS) system. Acetate secretion could be disadvantageous for cells and plasmid DNA production because it can shuttle carbon away from nucleotide synthesis. However, the exact effect
of acetate on plasmid production is not yet clear. Carnes et al. [23], have demonstrated a
minor negative effect of acetate on plasmid replication. Wang et al. [14] showed that *E.
coli* BL21 cells bearing plasmids produce more acetate than plasmid-free cells during a
2L fermentation, but it is unclear whether increased acetate is a cause or effect of
plasmid production. Further support for the *pykF pykA* knockout strategy comes from a
mathematical model developed by Cunningham et al. [18] which demonstrated that low
or zero pyruvate kinase flux along with generation of NADPH by transhydrogenase
activity offers positive effects on plasmid DNA production. These findings were
experimentally verified by the authors by showing a nine-fold higher copy number in
JM101Δ*pykFΔpykA* mutant cells bearing temperature-inducible pUC-ori plasmids at
42°C, at shake flask scale using defined medium with glucose as the carbon source [22].

Another promising target for pDNA production strain engineering is the pentose-
phosphate-pathway gene *rpiA*, which codes for ribose-5-phosphate isomerase A.
Increasing expression of *rpiA* would enhance synthesis of the nucleotide precursor
ribose-5-phosphate (R5P). As a result, an increase in nucleotide formation, and
consequently, plasmid DNA production is expected. In fact, overexpression of *rpiA* in *E.
coli* BL21 showed a 3-fold increase in plasmid copy number of a ColE1-derived plasmid
during continuous culture, using defined medium and glucose as the carbon source [14].

Overexpression of *zwf* has also been investigated as a strategy to increase flux to
the pentose phosphate pathway. Williams et al. [13] verified that simultaneous
overexpression of *zwf* and *rpiA* in DH5α appeared to increase plasmid amplification rate
(mg pDNA/L/OD<sub>600</sub>/hr) but not final specific yield (mg/L/OD<sub>600</sub>), in fed-batch
fermentations using complex medium with glycerol as the carbon source. However,
overexpression of *zwf* alone as well as *zwf* and thioredoxin (*trxA*) to enhance reducing
power did not show any effect on plasmid DNA yield [13]. In a separate study, zwf overexpression increased *E. coli* growth rate [19].

Mutations in FruR – a transcriptional regulator that acts on many of the genes in central carbon metabolism – have also been investigated to increase plasmid yield. Knocking out *fruR* in DH5α showed an increase in the maximum specific growth rate of the mutant strain during batch fermentation, but did not increase plasmid final yield [24]. However, later it was shown that the same *fruR*-deficient DH5α strain improved plasmid DNA yield under exponential feeding in fed-batch conditions [25]. These results demonstrate that fermentation strategy directly affects plasmid amplification behavior.

It was also observed that the glycolytic genes *pykF*, *zwf*, and 6-phosphofructokinase I and II (*pfkA*, *pfkB*) were upregulated when *fruR* was deleted [24].

### 3.2 Genes related to plasmid properties

Two genes that are often knocked out to improve plasmid yield and quality are *endA* and *recA*. The *endA* gene encodes DNA-specific endonuclease 1 and *recA* codes for a protein essential for the *recBCD* pathway of homologous recombination. Deletion of *endA* can improve the quality of plasmid preparations by eliminating non-specific degradation of DNA by the endonuclease [26]. However, plasmid nicking and degradation can also be caused by other, non-EndA-mediated factors [27]. *recA* mutants, on the other hand, have less undesirable homologous recombination than wild-type cells. Homologous recombination can lead to both changes in the plasmid DNA as well as formation of plasmid multimers, which leads to an increase in plasmid-free cells [26, 28]. Singer et al. [16] have observed a positive impact of *recA* mutation on
plasmid DNA yield. However, Yau et al. [8] observed that the effect of some mutations, such as ΔendA and ΔrecA, are very strain and/or plasmid dependent.

DNA methylation consists of the addition of a methyl group to a base by a DNA methyltransferase enzyme. Binding proteins can have high affinity to methylated DNA sequences and methylation sites may affect promoter activity [29]. Methylation patterns can also allow a species to distinguish its own DNA from foreign DNA. One common methyltransferase present in E. coli is the DNA-cytosine methyltransferase (Dcm) [30]. It is important to consider that plasmids produced from Δdcm versus dcm+ strains are different final products. Δdcm mutant cells did not have a strong impact on plasmid yield and quality[23]. On the other hand, plasmids produced in these Δdcm cells had a high transgene expression level in a human cell line and would therefore be recommended for gene therapy applications. In spite of improved expression, Δdcm plasmids were demonstrated to be less immunogenic, producing lower antibody responses for the influenza H5 hemagglutinin protein and would not be ideal for DNA vaccine applications [23].

3.3 Other targets for genetic engineering

The relA and spoT genes encode the enzymes ppGpp synthetase I and ppGpp synthetase II that catalyze the synthesis of guanosine-5´-diphosphate-3´-diphosphate (ppGpp) as part of the stringent response of E. coli to amino acid starvation. The ppGpp nucleotide interacts with RNA polymerase and can cause inhibition of rRNA and tRNA synthesis, affecting bacterial chromosome and plasmid origin replication [31], [32]. relA and spoT knockout strains do not produce ppGpp during amino acid starvation or nutrient limitation, and the response to starvation for such strains is called the relaxed
response [33]. *relA1* mutant strains such as DH5α and JM108 have been successfully used for amplification of plasmid DNA [34]. Appropriate amino acid composition in the medium, such as extra isoleucine, has demonstrated a positive effect on pDNA production in a Δ*relA* strain and seems to be an efficient strategy for high yield pDNA production [34].

A study by Williams et al. [13] examined the effects of a series of gene mutations on plasmid yield. Many of the targeted genes play a role in DNA synthesis, topology, and repair (Table 1). For example, the genes *polA* and *ligA* produce DNA polymerase I (Pol I) and DNA ligase respectively. Pol I is a multifunctional enzyme required for numerous types of DNA repair [30]. Particularly relevant to ColE1 plasmid replication, Pol I extends RNA primers and removes RNA primers postreplication, while LigA seals nicks during pDNA synthesis. Overexpression of *polA* and *ligA* genes improved plasmid DNA yields in shake flask cultures, but not in the bioreactor. The authors also observed that *gyrAB* overexpression had a negative effect on plasmid DNA production, decreasing the yield at both shake flask and bioreactor scales [13]. *gyrAB* encodes the DNA gyrase subunits GyrA and GyrB that are related to ATP-dependent supercoiling of DNA [30].

The differences in the impact of a given mutation observed in some cases between shake flask and bioreactor-scale cultures underscores the need to fully evaluate a strain engineering strategy using process-relevant conditions.

### 4. Host strains for plasmid DNA production

Most *E. coli* strains used for plasmid DNA production were originally developed for cloning or recombinant protein production [35]. The ideal host strain should be able to grow to high cell density with high plasmid copy number, maintain genetic stability
and be amenable to the downstream purification process [36]. However, *E. coli* host strains are often selected based on commercial availability or previously-established laboratory-scale protocols [8]. The disadvantage of using common laboratory strains is the high degree of mutations that is found in them [37].

*E. coli* K-12 strains like DH5α, DH10B and JM108 are typically used for plasmid DNA production [38]. The first *E. coli* K-12 strain was isolated in 1922, and since that time thousands of mutant strains have been produced [39-40]. The genome sequence can be useful to better understand the differences and similarities between *E. coli* strains [41]. In Figure 2, we describe the creation of *E. coli* strains and the genetic relationship between these strains. Many strains of *E. coli* have been created through a series of mutations to facilitate cloning of heterologous genes and for the purpose of stably maintaining plasmid DNA for the production of recombinant proteins. However, it is still not known whether these mutations are beneficial for plasmid DNA production or not, since different strategies are used for protein expression. As of 2007, the top three patented high-yield pDNA fermentation processes use DH5, DH5α and JM108 as the main host strain [34].

Phue et al. [26] showed that BL21ΔrecAΔendA, a derivative of *E. coli* B, is a better producer of plasmid DNA. A previous study comparing one strain derivative of *E. coli* B (BL21) and one strain of *E. coli* K-12 (JM109), at high glucose concentration, showed that BL21 grew faster and accumulated less acetate than JM109. Microarray and northern blot analyses demonstrated higher activities of the TCA cycle, glyoxylate shunt, gluconeogenesis and anaplerotic pathways for BL21 [42]. Such metabolic characteristics of BL21 can be advantageous for plasmid DNA production.
To improve safety, *E. coli* strains that allow antibiotic-free plasmid selection have been developed [43-47]. Also with safety in mind, pDNA vectors with minimal or no prokaryotic genetic elements are also being developed to minimize the potential for adverse effects [47-48]. Sequences in the bacterial backbone can interfere with gene expression, and plasmid mini-circles can offer many advantages over conventional plasmids, since they do not contain this backbone sequence [49]. A genetically-modified *E. coli* strain has also been developed to improve plasmid mini-circle production [50].

Another regulatory concern of plasmid DNA production is that plasmid genetic stability can directly impact safety and transfection efficiency. Structural instability of plasmid DNA has been recently reviewed by Oliveira et al. [51]. One of the major concerns related to the host are mobile elements, such as insertion sequences (IS) that can transpose from the bacterial chromosome to plasmid DNA. Insertion sequence transposition was observed in the neomycin resistance gene (IS1) or upstream of the gene (IS2) in different plasmids amplified by DH5α in an industrial process (IS1) [52] and by DH5α, JM109, TOP10F and HB101 at the laboratory scale (IS2) [53]. To overcome the IS-mediated instability problem, multiple-deletion series (MDS) strains were developed, removing all mobile elements of this *E. coli* genome. However, plasmid yield from MDS strains has yet to be evaluated [54].

5. Effects of fermentation strategy on strain behavior

As shown above, rationally-designed *E. coli* strains and plasmids have been demonstrated to directly affect fermentation yield, purification, and transfection processes. However, other factors such as medium composition, carbon source,
fermentation strategy (e.g., feeding and temperature shift), and process scale seem to have an influence on strain/plasmid behavior. Recent studies have begun to elucidate this influence by examining the genetic changes that occur in response to various process conditions.

To investigate the effect of carbon source, Oh et al. [55] analyzed expression of 111 E. coli genes in glycolysis, PP and TCA among other pathways, during growth on different carbon sources (glucose, glycerol and acetic acid). The gene expression profile changed more significantly from glucose to acetate media than from glucose to glycerol. Most biosynthetic genes were down-regulated in both glycerol and acetate media compared to glucose, while TCA cycle genes were up-regulated in both media. The glycolytic genes were roughly unchanged in the glycerol medium but decreased in the acetate medium. pykA was observed as up regulated in glycerol compared to glucose. Since the phosphotransferase system (PTS) is not active in the glycerol-based medium, pykA would be essential to generate pyruvate for the TCA cycle [55]. On the other hand, PykF has higher activity on glucose in aerobic conditions [21]. The pykF pykA double knockout in JM101 was shown to improve plasmid copy number on glucose medium at laboratory scale [22]. The effect of this mutation for plasmid production under different conditions, such as glycerol as the carbon source and large scale process, is not known. The choice of carbon source can affect plasmid DNA yield depending on the selected strain, since E. coli gene regulation varies under different conditions [55].

Feeding strategy and fermentation scale also seem to influence plasmid DNA production. Exponential feeding was demonstrated to be very effective for fruR mutant DH5α [25] and distinct differences in yields and gene expression were observed between culture flasks and reactor fermentations [13]. The differences observed across
scales and fermentation strategies can sometimes be correlated to strain growth rate.
The feeding strategy in fed-batch fermentations has the main objective of controlling specific growth rate while in shake flasks the growth rate is not actively regulated. O’Kennedy et al. verified an increase of 8-fold in specific yield and 25-fold in volumetric yield of plasmid DNA using exponential fed-batch in comparison with shake flask cultures [56].

Discovery of a temperature effect on replication of some plasmids was a crucial step forward in recombinant protein production [57-58] and it has been intensively used in plasmid DNA production as well [13, 16, 22, 59]. However, changing temperature also impacts different strains in different ways during plasmid DNA synthesis [16] and the results can be seen even in the cell morphology of some strains, such as DH5α [60]. TCA cycle genes were observed as down-regulated at 42°C, possibly because of lower dissolved oxygen concentration resulting in a higher concentration of acetic acid at 42°C than at 37°C [61]. A recent proteomic study also observed increased acetate production at higher temperatures while examining the steady-state heat-shock response of MG1655 using continuous cultures. Of possible relevance to pDNA production, the authors observed upregulation of the DNA-binding protein Dps upon temperature shift from 37°C to 47.5°C. This protein has been shown to protect DNA from thermal stress and other types of damage [62]. Another recent study by Caspeta et al. showed that heating rate affects both host gene expression and organic acid accumulation in a BL21-based protein production process, but their findings could be readily applied to a plasmid production process as well. In particular, their data suggest that slower heating rates reduce imbalances between glycolysis and the TCA cycle [63]. Low temperature (30°C) at the beginning of the process is also an alternative to reduce metabolic burden.
and reduce growth rate during biomass formation, eliminating the need to engineer the host cells [38]. However, development of vectors and strains to obtain maximum yield at relatively low temperatures cannot be discarded since it can greatly contribute to the energy balance at the industrial scale.

6. Impact of strain on downstream processing and transfection

The choice of the host strain to solve problems associated with the purification process is essential in order to obtain a final high yield and quality plasmid DNA process. As reviewed by Bower et al., advances have been made in engineering *E. coli* strains to reduce genomic DNA and RNA in the end of the fermentation process, thus increasing the ratio of product to impurities in the fermentation stream [37].

Other aspects that are also important for an improved purification process include segregational stability and plasmid topology. To avoid segregational instability and the resulting low yields caused by defective partitioning of plasmids during cell division, it is possible to modulate genes active during segregation, such as *parA*, also known as *gyrB*, one of the genes responsible for DNA gyrase synthesis, and *cer*, also known as *rnhA*, required for cleavage of RNAII-R loop into primer template, in the beginning of ColE1 plasmid replication [7, 30]. The supercoiled plasmid form is also another important parameter to measure in the quality and control of the final product as recommended by the FDA [64], since supercoiled plasmids seem to be more efficient in generating an immune response, as demonstrated through *in vivo* tests [65]. Significant work has been done on improving the chromatography process in order to obtain this form separated from the others (open circular and linear) [66]. On the other hand, plasmid topology issues have also been addressed using molecular biology and genetic studies. The DNA
gyrase and sigma factor σS (rpoS) genes were observed to play an important role in the regulation of plasmid topology and could be potential gene mutation targets to increase the fraction of supercoiled plasmids produced [67].

Specific gene mutations in *E. coli* strains and vectors have shown positive results on transfection efficiency. High-level transgene expression was observed in muscle, skin and tumor cells using plasmid free of antibiotic resistance markers [68]. High transgene expression levels in a human cell line were also verified using plasmids produced from *dcm*-deficient strains as discussed previously [23]. Recent emphasis on strengthening the interface between bioprocess and clinical research will likely yield more studies like these in the future.

### Conclusion

Host strain optimization is a valuable tool for improving plasmid DNA manufacturing processes. The early literature focused on studying the mechanism of plasmid replication inside an *E. coli* host as well as observing physiological changes in the host in response to plasmid maintenance. This review focused on more recent work that has exploited the advent of genomic and transcriptomic profiling tools to observe plasmid-bearing cells in more detail and compare them to their plasmid-free counterparts. Both up- and down-regulation of key genes were observed in glycolysis, the pentose phosphate pathway and the tricarboxylic acid cycle in response to plasmid DNA production. By analyzing the metabolism of *E. coli*, strategies for systematic knockout or overexpression of rationally-selected genes have been used to improve plasmid DNA yield. However, the effects of a given gene mutation on plasmid yield were sometimes different than expected and often varied depending on the strain’s
genetic background and/or culture method. Culture conditions, such as fermentation medium, feeding strategy and process scale, seem to be strong factors affecting plasmid yield, and the mechanism and role of these factors has yet to be completely elucidated.

There is still a need for improving the cell lines used for plasmid production. Most of the strains discussed in this review have a highly-mutagenized genetic background originally designed for cloning and recombinant protein expression and it is not known if this background interacts with the rationally-designed mutations added to each strain. A lot of effort has been made and impressive results were obtained for *E. coli* strain development that has contributed to improved plasmid DNA yields. However, since the effects of genotypic background on new mutations in the *E. coli* genome is still unknown, it would be interesting to investigate such aspects in the future in order to design an *E. coli* host strain specifically for plasmid DNA production.

Creation of a pDNA production strain could potentially start with a wild-type genetic background with introduction of only the mutations that have been shown to enhance pDNA yield and/or quality. This strain may have advantages such as more robust growth and lack of auxotrophies that are often found in other laboratory strains. Another possibility would be to develop a strain that is resistant to bacteriophage infection while retaining desirable production characteristics. Bacteriophage infection can be catastrophic at the commercial scale, and there are currently limited methods available to combat it. Most of these methods are based on altering culture conditions because the phage-resistant strains created thus far often suffer from reduced growth and production capabilities [69]. It would also be possible to approach strain design with the goal of enhancing purification or transfection efficiency. There have been several innovative studies that seek to address these downstream processes at the
strain level, but there are still many avenues to explore. These studies will be particularly important as the regulatory structure surrounding gene-based therapeutics continues to evolve and possibly become more stringent.

As a whole, this review intended to bring together existing advances in engineering *E. coli* strains for plasmid DNA production in order to contribute to the development of new hosts adapted to meet the upstream and downstream processing challenges associated with large scale production of plasmid DNA.
References

[1] Wolff, A. J., Malone, R. W., Williams, P., Chong, W., et al., Direct gene transfer into mouse muscle in vivo. *Science* 1990, 247, 1465-1468.

[2] Carvalho, J. A., Prazeres, D. M. F., Monteiro, G. A., Bringing DNA vaccines closer to commercial use. *iDrugs* 2009, 12, 642 - 647.

[3] Liu, M. A., DNA vaccines: an historical perspective and view to the future. *Immunol Rev.* 2011, 239, 62 - 84.

[4] Kutzler, M. A., Weiner, D. B., DNA vaccines: ready for prime time? *Nat Rev Genet* 2008, 9, 776-788.

[5] Manthorpe, M., Hobart, P., Hermanson, G., Ferrari, M., et al., Plasmid vaccines and therapeutics: from design to applications. *Adv Biochem Eng Biotechnol* 2005, 99, 41 - 92.

[6] Leitner, W. W., Ying, H., Restifo, N. P., DNA and RNA-based vaccines: principles, progress and prospects *Vaccine* 2000, 18, 765 - 777.

[7] Prazeres, D. M. F., Monteiro, G. A., Ferreira, G. N. M., Diogo, M. M., et al., Purification of plasmids for gene therapy and DNA vaccination. *Biotechnol Annu Rev.* 2001, 7, 1 - 30.

[8] Yau, S. Y., Keshavarz-Moore, E., Ward, J., Host strain influences on supercoiled plasmid DNA production in *Escherichia coli*: implications for efficient design of large-scale processes. *Biotechnol Bioeng* 2008, 101, 529-544.

[9] Glenting, J., Wessels, S., Ensuring safety of DNA vaccines. *Microb Cell Fact* 2005, 4, 26.

[10] Bailey, J. E., Silva, N. A. D., Peretti, S. W., Seo, J.-H., Srienc, F., Studies of host-plasmid interactions in recombinant microorganisms. *Ann N Y Acad Sci* 1986, 469, 194-211.
[11] Ricci, J. C. D., Hernandez, M. E., Plasmid effects on *Escherichia coli* metabolism. *Crit Rev Biotechnol* 2000, 20, 79 - 108.

[12] Ow, D. S.-W., Nissom, P. M., Philp, R., Oh, S. K.-W., Yap, M. G.-S., Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5α during batch fermentation. *Enzyme Microb Technol* 2006, 39, 391-398.

[13] Williams, J. A., Luke, J., Langtry, S., Anderson, S., et al., Generic plasmid DNA production platform incorporating low metabolic burden seed-stock and fed-batch fermentation processes. *Biotechnol Bioeng* 2009, 103, 1129 - 1143.

[14] Wang, Z., Xiang, L., Shao, J., Wegrzyn, A., Wegrzyn, G., Effects of the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism. *Microb Cell Fact* 2006, 5.

[15] Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., Kampala, D. S., Plasmid-encoded protein: the principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnol Bioeng* 1990, 35, 668 - 681.

[16] Singer, A., Eiteman, M. A., Altman, E., DNA plasmid production in different host strains. *J Ind Microbiol Biotechnol* 2009, 36, 521 - 530.

[17] Bohle, K., Ross, A., Plasmid DNA production for pharmaceutical use: role of specific growth rate and impact on process design. *Biotechnol Bioeng* 2011, DOI 10.1002/bit.23138.

[18] Cunningham, D. S., Koepsel, R. R., Ataai, M. M., Domach, M. M., Factors affecting plasmid production in *Escherichia coli* from a resource allocation standpoint. *Microb Cell Fact* 2009, 8, 1 - 27.
[19] Flores, S., Anda-Herrera, R., Gosset, G., Bolivar, F. G., Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway *Biotechnol Bioeng* 2004, 87, 485 - 494.

[20] Siddiquee, K. A. Z., Arauzo-Bravo, M. J., Shimizu, K., Metabolic flux analysis of pykF gene knockout *Escherichia coli* based on 13C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. *Appl Microbiol Biotechnol* 2004, 63, 407 - 417.

[21] Ponce, E., Flores, N., Martinez, A., Valle, F., Bolivar, F., Cloning of the two pyruvate kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these enzymes in pyruvate biosynthesis. *J Bacteriol* 1995, 177, 5719-5722.

[22] Cunningham, D. S., Liu, Z., Domagalski, N., Koespsel, R. R., et al., Pyruvate kinase-deficient *Echerichia coli* exhibits increased plasmid copy number and cyclic AMP levels. *J Bacteriol* 2009, 191, 3041 - 3049.

[23] Carnes, A. E., Luke, J. M., Vincent, J. M., Shukar, A., et al., Plasmid DNA fermentation strain and process-specific effects on vector yield, quality, and transgene expression. *Biotechnol Bioeng* 2010, 108, 354-363.

[24] Ow, D. S.-W., Lee, R. M.-Y., Nissom, P. M., Philp, R., Inactivating FruR global regulator in plasmid-bearing *Escherichia coli* alters metabolic gene expression and improves growth rate. *J Biotechnol*. 2007, 131, 261 - 269.

[25] Ow, D. S.-W., Yap, M. G.-S., Oh, S. K.-W., Enhancement of plasmid DNA yields during fed-batch culture of a fruR- knockout *Escherichia coli* strain *Biotechnol. Appl. Biochem.* 2009, 52.
[26] Phue, J.-N., Lee, S. J., Trinh, L., Shiloach, J., Modified *Escherichia coli* B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5α). *Biotechnol Bioeng* 2008, 101, 831 - 836.

[27] Taylor, R. C., Walker, D. C., McInnes, R. R., *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.* 1993, 21, 1677 - 1678.

[28] Summers, D., Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Mol. Microbiol* 1998, 29, 1137 - 1141.

[29] Collier, J., Epigenetic regulation of the bacterial cell cycle. *Current Opinion in Microbiology* 2009, 12, 722 - 729.

[30] Keseler, I. M., Collado-Vides, J., A., S.-Z., Peralta-Gil, M., *et al*., EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res* 2011, 39.

[31] Wegrzyn, G., Replication of plasmids during bacterial response to amino acid starvation. *Plasmid* 1999, 41, 1-16.

[32] Wick, L. M., Egli, T., Molecular components of physiological stress responses in *Escherichia coli*. *Adv Biochem Eng Biotechnol* 2004, 89, 1 - 45.

[33] Magnusson, L. U., Farewell, A., Nystrom, T., ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol* 2005, 13, 236 - 242.

[34] Carnes, A. E., Williams, J. A., Plasmid DNA manufacturing technology. *Recent Pat Biotechnol.* 2007, 1, 1-16.

[35] Cai, Y., Rodriguez, S., Hebel, H., DNA vaccine manufacture: scale and quality. *Expert Rev Vaccines* 2009, 8, 1277-1291.

[36] Durland, R. H., Eastman, E. M., Manufacturing and quality control of plasmid-based gene expression systems. *Adv Drug Deliv Rev.* 1998, 30, 33 - 48.
[37] Bower, D. M., Prather, K. L. J., Engineering of bacterial strains and vectors for the production of plasmid DNA. *Appl Microbiol Biotechnol* 2009, 82, 805 - 813.

[38] Williams, J. A., Carnes, A. E., Hodgson, C. P., Plasmid DNA vaccine vector design: impact on efficacy, safety and upstream production. *Biotechnol Adv.* 2009, 27, 353 - 370.

[39] Bachmann, B. J., Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev* 1972, 36, 525 - 557.

[40] Grant, S. G. N., Jessee, J., Bloom, F. R., Hanahan, D., Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. USA* 1990, 87, 4645 - 4649.

[41] Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K. B., *et al.*, *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res.* 2006, 34, 1 - 9.

[42] Phue, J., Noronha, S. B., Hattacharyya, R., Wolfe, A. J., Shiloach, J., Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and northern blot analyses. *Biotechnol Bioeng* 2005, 90, 805 - 820.

[43] Cranenburgh, R. M., Hanak, J. A. J., Williams, S. G., Sherratt, D. J., *Escherichia coli* strains that allow antibiotic-free plasmid selection and maintenance by repressor titration. *Nucleic Acids Res* 2001, 29, 1-6.

[44] 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.

[45] Soubrier, F., Cameron, B., Manse, B., Somarriba, S., *et al.*, pCOR: a new design of plasmid vectors for nonviral gene therapy. *Gene Ther.* 1999, 6, 1482 - 1488.
[46] Luke, J., Carnes, A. E., Hodgson, C. P., Williams, J. A., Improved antibiotic-free DNA vaccine vectors utilizing a novel RNA based plasmid selection system. *Vaccine* 2009, 27, 6454-6459.

[47] Carnes, A. E., Luke, J. M., Vincent, J. M., Anderson, S., et al., Critical design criteria for minimal antibiotic-free plasmid vectors necessary to combine robust RNA Pol II and Pol III-mediated eukaryotic expression with high bacterial production yields. *J Gene Med.* 2010, 12, 818 - 831.

[48] Faurez, F., Dory, D., Le Moigne, V., Gravier, R., Jestin, A., Biosafety of DNA vaccines: new generation of DNA vectors and current knowledge on the fate of plasmids after injection. *Vaccine* 2010, 28, 3888-3895.

[49] Gill, D. R., Pringle, I. A., Hyde, S. C., Progress and prospects: the design and production of plasmid vectors. *Gene Ther.* 2009, 16, 165 - 171.

[50] Kay, M. A., He, C., Chen, Z., A robust system for production of minicircle DNA vectors. *Nat Biotechnol* 2010, 28, 1287 - 1291.

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

[52] Prather, K. L. J., 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.

[53] Oliveira, P. H., Prazeres, D. M. F., 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.
[54] Posfai, G., Plunkett, G., Feher, T., Frisch, D., et al., Emergent properties of reduced-genome *Escherichia coli*. Science 2006, 312, 1044-1046.

[55] Oh, M., Liao, J. C., Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. Biotechnol Prog. 2000, 16, 278 - 286.

[56] O’Kennedy, R. D., Ward, J. M., Keshavarz-Moore, E., Effects of fermentation strategy on the characteristics of plasmid DNA production. *Biotechnol Appl Biochem* 2003, 37, 83-90.

[57] Lin-Chao, S., Chen, W. T., Wong, T. T., High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Mol Microbiol* 1992, 6, 3385-3393.

[58] Fitzwater, T., Zhang, X., Elble, R., Polisky, B., Conditional high copy number ColE1 mutants: resistance to RNA1 inhibition *in vivo* and *in vitro*. *EMBO J* 1988, 7, 3289 - 3297.

[59] Carnes, A. E., Hodgson, C. P., A.Williams, J., Inducible *Escherichia coli* fermentation for increased plasmid DNA production. *Biotechnol. Appl. Biochem.* 2006, 45, 155 - 166.

[60] Silva, F., Passarinha, L., Sousa, F., Queiroz, J. A., Domingues, F. C., Influence of growth conditions on plasmid DNA production. *J Microbiol Biotechnol* 2009, 19, 1408-1414.

[61] Hasan, C. M. M., Shimizu, K., Effect of temperature up-shift on fermentation and metabolic characteristics in view of gene expressions in *Escherichia coli*. *Microb Cell Fact* 2008, 7, 1 - 13.

[62] Luders, S., Fallet, C., Franco-Lara, E., Proteome analysis of the Escherichia coli heat shock response under steady-state conditions. *Proteome Sci* 2009, 7, 36.

[63] Caspeta, L., Flores, N., Perez, N. O., Bolivar, F., Ramirez, O. T., The effect of heating rate on Escherichia coli metabolism, physiological stress, transcriptional response, and
production of temperature-induced recombinant protein: a scale-down study.

*Biotechnol Bioeng* 2009, *102*, 468-482.

[64] US Dept of Health and Human Services, F. D. A., Guidance for industry: considerations for plasmid DNA vaccines for infectious disease indications. 2007.

[65] Pillai, V. B., Hellerstein, M., Yu, T., Amara, R. R., Robinsona, H. L., Comparative studies on *in vitro* expression and *in vivo* immunogenicity of supercoiled and open circular forms of plasmid DNA vaccines. *Vaccine* 2008, *26*, 1136 - 1141.

[66] Li, H., Bo, H., Wang, J., Shao, H., Huang, S., Separation of supercoiled from open circular forms of plasmid DNA, and biological activity detection. *Cytotechnology* 2011, *63*, 7 - 12.

[67] Reyes-Dominguez, Y., Contreras-Ferrat, G., Ramires-Santos, J., Membrillo-Hernandez, J., Gomez-Eichelmann, M. C., Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and rpoS stationary-phase cells. *J Bacteriol* 2003, *185*, 1097 - 1100.

[68] 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.

[69] Los, M., Czyz, A., Sell, E., Wegrzyn, A., *et al.*, Bacteriophage contamination: is there a simple method to reduce its deleterious effects in laboratory cultures and biotechnological factories? *J. Appl. Genet.* 2004, *45*, 111 - 120.

[70] Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., *et al.*, Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* 2006, *2*, 2006 0007.
[71] Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, et al., The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* 2008, 190, 2597-2606.

[72] Yanisch-Perron, C., Vieira, J., Messing, J., Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 1985, 33, 103-119.
Acknowledgments

This work was supported by the MIT-Portugal Program and Fundação para Ciência e a Tecnologia (PhD grant SFRH/BD/33786/2009 to Geisa A. L. Gonçalves).

The authors have declared no conflict of interest.
Figure legends

Figure 1. Gene knockout and overexpression strategies to improve plasmid DNA production in *E. coli*. (A) Overexpression of the *zwf* and *rpiA* genes is proposed as a means to increase fluxes in the pentose phosphate pathway and nucleotide synthesis. (B) Knockouts of genes *pykF* and *pykA* are suggested to reduce acetate formation, increase fluxes in the pentose phosphate pathway and TCA cycle. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; FDP, fructose 1,6-diphosphate; G3P, glyceraldehyde 3-phosphate; PEP, phosphoenolpyruvate; OAA, oxaloacetate; MAL, malate; 6GPC, 6-phosphogluconate; RUSP, ribulose 5-phosphate; XU5P, xylulose 5-phosphate; R5P, ribose 5-phosphate; S7P, sedoheptulose 7-phosphate; E4P, erythrose 4-phosphate; PTS, phosphotransferase system. This figure represents a simplified version of *E. coli* central metabolism. Dark arrows represent high carbon flow in pentose phosphate pathway and light arrows represent less formation of pyruvate and acetate.

Figure 2. *E. coli* K-12 and derivatives -- creation of new strains and relationship between different strains. (A) Lineage of MG1655 and W3110, close relatives of wild-type *E. coli* K-12 [70]. (B) Generation of strains containing multiple mutations from MC1061, DH1 and JM101 [71-72]. Dark boxes represent commonly-used *E. coli* strains for plasmid DNA production and recent developments in *E. coli* strains designed for high yield pDNA processes. Full line arrows represent the relationship between the strains and dashed line arrows represent mutations carried from one strain to the other. Better comprehension of the relationship between the strains is possible through strain genotype analysis, and full genotypes of commonly used strains have been compiled.
elsewhere (http://openwetware.org/wiki/E._coli_genotypes#Commonly_used_strains).

Abbreviation: Methylation-dependent restriction system (MDRS).
Table 1. *E. coli* genes targeted for mutation to improve plasmid DNA production. △ indicates gene knockout and ↑ indicates gene overexpression. Information about gene products and their functions were compiled from Ecocyc [30] and the indicated references.

| Gene   | Product                                           | Mutation | Expected phenotype                                                                 | Ref.   |
|--------|---------------------------------------------------|----------|----------------------------------------------------------------------------------|--------|
| recA   | DNA strand exchange and recombination protein; protease and nuclease activity | △        | Minimized recombination of cloned DNA, pairing and exchange between repeated DNA sequences | [26]   |
| endA   | DNA-specific endonuclease I                       | △        | Decreased non-specific digestion of plasmid                                         |        |
| rpiA   | Ribose-5-phosphate isomerase A                    | ↑        | Increased biosynthesis of nucleotide precursors in pentose phosphate pathway        | [14]   |
| pykA   | Pyruvate kinase I, II                             | Δ        | Increased pentose phosphate pathway and TCA cycle flux and reduced acetate synthesis | [22]   |
| pykF   |                                                   |          |                                                                                  |        |
| topA   | Topoisomerase I                                   | △        | Increased RNAII R-loop formation (ColE1 plasmids)                                  |        |
| polA   | DNA polymerase I                                  | ↑        | Increased availability of enzyme that extends the RNA primer template and removes RNA primers postreplication |        |
| ligA   | DNA ligase                                        | ↑        | Increased availability of enzyme that seals nicks during pDNA synthesis             | [13]   |
| gyrAB  | DNA gyrase                                        | ↑        | Increased availability of enzyme that negatively supercoils the covalently closed circular plasmid |        |
| priA   | Primosome components                              | ↑        | Increased availability of primosomes (one lagging strand primosome is required per plasmid per replication) |        |
| priB   |                                                   |          |                                                                                  |        |
| priC   |                                                   |          |                                                                                  |        |
| trxA   | Thioredoxin                                       | ↑        | Increased level of hydrogen donors for ribonucleotide reductase                    |        |
| grx1   | Glutaredoxin                                       | ↑        |                                                                                  |        |
| zwf    | Glucose 6-phosphate-1-dehydrogenase               | ↑        | Increased pentose phosphate pathway flux                                           | [19]   |
| fruR   | Transcriptional regulator                         | △        | Deletion of global regulator of carbon flow through the central metabolic pathways | [25]   |
| relA   | ppGpp synthetase I, II                            | △        | Abolished stringent response to amino acid starvation                              | [31]   |
| spoT   |                                                   |          |                                                                                  |        |
| dcm    | DNA-cytosine methyltransferase                    | △        | Improved transgene expression and reduced immunogenicity                            | [23]   |
**A**

- **E. coli K-12** (EMG2 or WG1)
  - W1485
  - MG1655
  - W3110

**B**

- JM101 ΔpykF ΔpykA
  - JM107
    - JM106
    - DH1 recA^+ Δ(lac-proAB)
      - DH1 recA^-
        - JC10240
  - JM108
    - JM109
  - JM102
    - DH1 recA^+ ΔMDRS deoR endA1
      - DH10
        - DH10B
    - MC1061
      - MC1062
        - MM294
          - f80lacZΔM15

*Genotype of F' episome of JM101: traD36 proAB lac^RΔM15*