



MIT Open Access Articles

Rational engineering of Escherichia coli strains for plasmid biopharmaceutical manufacturing

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation	Goncalves, Geisa A. L. et al. "Rational engineering of Escherichia coli strains for plasmid biopharmaceutical manufacturing." <i>Biotechnology Journal</i> 7, 2 (February 2012): 251-261 © 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
As Published	http://dx.doi.org/10.1002/biot.201100062
Publisher	Wiley
Version	Author's final manuscript
Citable link	https://hdl.handle.net/1721.1/121771
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/4.0/

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*}

¹Department of Bioengineering, Instituto Superior Técnico (IST), Lisbon, Portugal

²IBB-Institute for Biotechnology and Bioengineering, Center for Biological and Chemical
Engineering, IST, Lisbon, Portugal

³Department of Chemical Engineering, Massachusetts Institute of Technology,
Cambridge, MA

† 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

1 **Abstract**

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

15

16 **1. Introduction**

17 For many years, bacterial plasmid DNA (pDNA) has been used as a biological tool
18 for cloning and expression of recombinant proteins. Recently, plasmid DNA has been
19 considered as a potential biopharmaceutical, particularly for applications in gene
20 therapy and DNA vaccination. Naked plasmid DNA was first observed as a viable vector
21 for expression of heterologous genes in live mammals approximately two decades ago
22 [1]. After this discovery, pDNA became very attractive as a potential biopharmaceutical
23 since its production seems to be cheaper and faster when compared with the
24 production of proteins, viruses or cells [2]. Despite its potential, non-viral gene therapy
25 and DNA vaccines are still under development and a therapeutic product for humans has
26 not yet reached the market. Recent work on new pDNA delivery methods, such as
27 electroporation, has helped mitigate challenges related to low immunogenicity and
28 transgene expression, reinvigorating the field as a whole. Currently there are some
29 products licensed for veterinary applications [3-4] and numerous clinical trials in phase I,
30 II and III [4-5]. The DNA vaccination market is growing and with it a technological and
31 economical need to improve production methods [6]. Careful design and selection of the
32 host strain is one way to improve the yield and quality of a plasmid product.

33 Of note is the fact that most fermentation process development for plasmid
34 DNA was built upon the foundations developed for the production of recombinant
35 protein. However, significant differences exist between these processes. For
36 recombinant protein production, high level expression of genes is required to produce
37 large amounts of proteins during the fermentation process, which consist of three
38 different steps: bacterial growth, multicopy plasmid amplification and protein induction.
39 On the other hand, plasmid DNA production occurs in a different scenario, where it is

40 necessary to optimize the allocation of cellular resources between biomass formation
41 and plasmid DNA synthesis. Significant technological advances have been made in
42 downstream processing because of the recognition of the key physico-chemical
43 differences between proteins and nucleic acids [7]. In contrast, there are still many
44 opportunities to improve the upstream stages of plasmid biopharmaceutical
45 manufacturing such as development of host strains [8].

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

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

63

64 2. Effect of plasmid DNA synthesis on *E. coli* central carbon metabolism

65 Plasmid DNA synthesis can disturb *E. coli* gene regulation, altering levels of gene
66 expression and carbon flux. For the last 3 decades, many researchers observed that
67 plasmid maintenance retards host growth [10-14]. References in the literature correlate
68 the low growth imposed by plasmids with metabolic burden, since plasmid replication
69 and expression of the antibiotic resistance marker requires additional nutrients and
70 energy. Many studies have been done in order to understand this phenomenon using
71 recombinant cells to analyze the effect of plasmid amplification and gene expression on
72 growth rate [10-11, 15]. However, the relationship between plasmid DNA content and
73 growth rate has not been fully explained, as evidenced by recent work on the subject
74 [16-17].

75 The advent of technologies such as DNA microarrays, metabolic flux analysis,
76 and quantitative real-time PCR has allowed researchers to probe more deeply into the
77 perturbations caused by plasmid maintenance and replication. Specifically, recent
78 reports have shown the effects of plasmid DNA on the central metabolism of *E. coli*,
79 namely glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway [12,
80 14].

81 Glycolysis is the main pathway for glucose utilization and energy generation. It
82 is composed of ten reactions catalyzed by specific enzymes that are coded by individual
83 genes. Previous studies reported different levels of glycolytic gene expression in
84 plasmid-bearing cells versus plasmid-free cells. Some results showed most of the
85 glycolytic genes as down regulated [12], while other results have demonstrated up-
86 regulation of the same genes, such as pyruvate kinase I (*pykF*) for cells carrying plasmids
87 [14].

88 The pentose phosphate pathway (PP) can metabolize different sugars like xylose
89 and ribose, but is considered the second main destination for glucose. The PP pathway is
90 also one of the pathways responsible for biosynthesis of the nucleotide precursors
91 ribose-5-phosphate (R5P) and erythrose-4-phosphate (E4P). Another important product
92 from the PP pathway is NADPH, synthesized by glucose 6-phosphate-1-dehydrogenase
93 (Zwf) and 6-phosphogluconate dehydrogenase (Gnd). NADPH and nucleotides are
94 required for biomass and plasmid DNA production and they are intrinsically correlated in
95 the PP pathway, composing the oxidative and non-oxidative phases respectively.
96 Cunningham et al. [18] demonstrated through a mathematical model that increasing the
97 availability of NADPH via transhydrogenase activity has a positive impact on plasmid
98 DNA production by increasing the reducing power available for pDNA and antibiotic
99 resistance marker synthesis. Cells carrying high copy plasmid DNA require extra
100 synthesis of nucleotides and in this case the carbon flux directed to the PP pathway may
101 be insufficient to cover the cell's metabolic needs [19].

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

108

109 **3. Relevant genes for strain engineering**

110 Many cell line engineering efforts have sought to improve plasmid DNA
111 production by knockout or overexpression of rationally-selected genes (Table 1). One

112 main area of focus is modification of central carbon metabolism genes to increase flux
113 toward nucleotide and amino acid precursor synthesis and reduction of byproducts,
114 such as acetate. Genes related to improving pDNA quality have also been common
115 targets, as have genes that are involved in various other cellular processes relevant to
116 pDNA production such as the stringent response and DNA replication.

117

118 **3.1 Central carbon metabolism genes**

119 Central carbon metabolism is a logical target for genetic engineering strategies
120 to increase plasmid DNA yield because of the potential to increase carbon flux to
121 nucleotide precursors (Figure 1). Altering central metabolism to produce more energy
122 and reducing power could also potentially improve plasmid yields. The *pykF* and *pykA*
123 genes encode pyruvate kinase (Pyk) I and II, respectively, and the production of the two
124 isoenzymes is independent. However, both enzymes have a cooperative effect on
125 transforming phosphoenolpyruvate (PEP) into pyruvate (PYR) at the final stage of
126 glycolysis. Previous studies investigated the effect of *pykF* knockout on plasmid-free *E.*
127 *coli* cells. The growth rate of the mutant cells was slightly lower than wild type, and
128 acetic acid formation was smaller in mutant cells [20-21]. Other studies demonstrated
129 similar behavior in an *E. coli pykF pykA* double knockout bearing plasmid [22]. Less
130 production of acetate was expected based on metabolic deduction since the synthesis of
131 pyruvate would decrease, and phosphoenolpyruvate would be converted to
132 oxaloacetate (OAA) without pyruvate kinase activity. However, PEP, in the presence of
133 glucose, can be converted to pyruvate via the phosphotransferase (PTS) system.
134 Acetate secretion could be disadvantageous for cells and plasmid DNA production
135 because it can shuttle carbon away from nucleotide synthesis. However, the exact effect

136 of acetate on plasmid production is not yet clear. Carnes et al.[23], have demonstrated a
137 minor negative effect of acetate on plasmid replication. Wang et al. [14] showed that *E.*
138 *coli* BL21 cells bearing plasmids produce more acetate than plasmid-free cells during a
139 2L fermentation, but it is unclear whether increased acetate is a cause or effect of
140 plasmid production. Further support for the *pykF pykA* knockout strategy comes from a
141 mathematical model developed by Cunningham et al. [18] which demonstrated that low
142 or zero pyruvate kinase flux along with generation of NADPH by transhydrogenase
143 activity offers positive effects on plasmid DNA production. These findings were
144 experimentally verified by the authors by showing a nine-fold higher copy number in
145 JM101Δ*pykF*Δ*pykA* mutant cells bearing temperature-inducible pUC-*ori* plasmids at
146 42°C, at shake flask scale using defined medium with glucose as the carbon source [22].

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

154 Overexpression of *zwf* has also been investigated as a strategy to increase flux to
155 the pentose phosphate pathway. Williams et al. [13] verified that simultaneous
156 overexpression of *zwf* and *rpiA* in DH5α appeared to increase plasmid amplification rate
157 (mg pDNA/L/OD₆₀₀/hr) but not final specific yield (mg/L/OD₆₀₀), in fed-batch
158 fermentations using complex medium with glycerol as the carbon source. However,
159 overexpression of *zwf* alone as well as *zwf* and thioredoxin (*trxA*) to enhance reducing

160 power did not show any effect on plasmid DNA yield [13]. In a separate study, *zwf*
161 overexpression increased *E. coli* growth rate [19].

162 Mutations in FruR – a transcriptional regulator that acts on many of the genes in
163 central carbon metabolism – have also been investigated to increase plasmid yield.
164 Knocking out *fruR* in DH5 α showed an increase in the maximum specific growth rate of
165 the mutant strain during batch fermentation, but did not increase plasmid final yield
166 [24]. However, later it was shown that the same *fruR*-deficient DH5 α strain improved
167 plasmid DNA yield under exponential feeding in fed-batch conditions [25]. These results
168 demonstrate that fermentation strategy directly affects plasmid amplification behavior.
169 It was also observed that the glycolytic genes *pykF*, *zwf*, and 6-phosphofructokinase I
170 and II (*pfkA*, *pfkB*) were upregulated when *fruR* was deleted [24].

171

172 **3.2 Genes related to plasmid properties**

173 Two genes that are often knocked out to improve plasmid yield and quality are
174 *endA* and *recA*. The *endA* gene encodes DNA-specific endonuclease 1 and *recA* codes for
175 a protein essential for the *recBCD* pathway of homologous recombination. Deletion of
176 *endA* can improve the quality of plasmid preparations by eliminating non-specific
177 degradation of DNA by the endonuclease [26]. However, plasmid nicking and
178 degradation can also be caused by other, non-EndA-mediated factors [27]. *recA*
179 mutants, on the other hand, have less undesirable homologous recombination than
180 wild-type cells. Homologous recombination can lead to both changes in the plasmid DNA
181 as well as formation of plasmid multimers, which leads to an increase in plasmid-free
182 cells [26, 28]. Singer et al. [16] have observed a positive impact of *recA* mutation on

183 plasmid DNA yield. However, Yau et al. [8] observed that the effect of some mutations,
184 such as $\Delta endA$ and $\Delta recA$, are very strain and/or plasmid dependent.

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

198

199 **3.3 Other targets for genetic engineering**

200 The *relA* and *spoT* genes encode the enzymes ppGpp synthetase I and ppGpp
201 synthetase II that catalyze the synthesis of guanosine-5'-diphosphate-3'-diphosphate
202 (ppGpp) as part of the stringent response of *E. coli* to amino acid starvation. The ppGpp
203 nucleotide interacts with RNA polymerase and can cause inhibition of rRNA and tRNA
204 synthesis, affecting bacterial chromosome and plasmid origin replication [31], [32]. *relA*
205 and *spoT* knockout strains do not produce ppGpp during amino acid starvation or
206 nutrient limitation, and the response to starvation for such strains is called the relaxed

207 response [33]. *relA1* mutant strains such as DH5 α and JM108 have been successfully
208 used for amplification of plasmid DNA [34]. Appropriate amino acid composition in the
209 medium, such as extra isoleucine, has demonstrated a positive effect on pDNA
210 production in a Δ *relA* strain and seems to be an efficient strategy for high yield pDNA
211 production [34].

212 A study by Williams et al. [13] examined the effects of a series of gene mutations
213 on plasmid yield. Many of the targeted genes play a role in DNA synthesis, topology,
214 and repair (Table 1). For example, the genes *polA* and *ligA* produce DNA polymerase I
215 (Pol I) and DNA ligase respectively. Pol I is a multifunctional enzyme required for
216 numerous types of DNA repair [30]. Particularly relevant to ColE1 plasmid replication,
217 Pol I extends RNA primers and removes RNA primers postreplication, while LigA seals
218 nicks during pDNA synthesis. Overexpression of *polA* and *ligA* genes improved plasmid
219 DNA yields in shake flask cultures, but not in the bioreactor. The authors also observed
220 that *gyrAB* overexpression had a negative effect on plasmid DNA production, decreasing
221 the yield at both shake flask and bioreactor scales [13]. *gyrAB* encodes the DNA gyrase
222 subunits GyrA and GyrB that are related to ATP-dependent supercoiling of DNA [30].
223 The differences in the impact of a given mutation observed in some cases between
224 shake flask and bioreactor-scale cultures underscores the need to fully evaluate a strain
225 engineering strategy using process-relevant conditions.

226

227 **4. Host strains for plasmid DNA production**

228 Most *E. coli* strains used for plasmid DNA production were originally developed
229 for cloning or recombinant protein production [35]. The ideal host strain should be able
230 to grow to high cell density with high plasmid copy number, maintain genetic stability

231 and be amenable to the downstream purification process [36]. However, *E. coli* host
232 strains are often selected based on commercial availability or previously-established
233 laboratory-scale protocols [8]. The disadvantage of using common laboratory strains is
234 the high degree of mutations that is found in them [37].

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

247 Phue et al. [26] showed that BL21 Δ recA Δ endA, a derivative of *E. coli* B, is a
248 better producer of plasmid DNA. A previous study comparing one strain derivative of *E.*
249 *coli* B (BL21) and one strain of *E. coli* K-12 (JM109), at high glucose concentration,
250 showed that BL21 grew faster and accumulated less acetate than JM109. Microarray
251 and northern blot analyses demonstrated higher activities of the TCA cycle, glyoxylate
252 shunt, gluconeogenesis and anaplerotic pathways for BL21 [42]. Such metabolic
253 characteristics of BL21 can be advantageous for plasmid DNA production.

254 To improve safety, *E. coli* strains that allow antibiotic-free plasmid selection
255 have been developed [43-47]. Also with safety in mind, pDNA vectors with minimal or
256 no prokaryotic genetic elements are also being developed to minimize the potential for
257 adverse effects [47-48]. Sequences in the bacterial backbone can interfere with gene
258 expression, and plasmid mini-circles can offer many advantages over conventional
259 plasmids, since they do not contain this backbone sequence [49]. A genetically-
260 modified *E. coli* strain has also been developed to improve plasmid mini-circle
261 production [50].

262 Another regulatory concern of plasmid DNA production is that plasmid genetic
263 stability can directly impact safety and transfection efficiency. Structural instability of
264 plasmid DNA has been recently reviewed by Oliveira et al. [51]. One of the major
265 concerns related to the host are mobile elements, such as insertion sequences (IS) that
266 can transpose from the bacterial chromosome to plasmid DNA. Insertion sequence
267 transposition was observed in the neomycin resistance gene (IS1) or upstream of the
268 gene (IS2) in different plasmids amplified by DH5 α in an industrial process (IS1) [52] and
269 by DH5 α , JM109, TOP10F and HB101 at the laboratory scale (IS2) [53]. To overcome the
270 IS-mediated instability problem, multiple-deletion series (MDS) strains were developed,
271 removing all mobile elements of this *E. coli* genome. However, plasmid yield from MDS
272 strains has yet to be evaluated [54].

273

274 **5. Effects of fermentation strategy on strain behavior**

275 As shown above, rationally-designed *E. coli* strains and plasmids have been
276 demonstrated to directly affect fermentation yield, purification, and transfection
277 processes. However, other factors such as medium composition, carbon source,

278 fermentation strategy (*e.g.*, feeding and temperature shift), and process scale seem to
279 have an influence on strain/plasmid behavior. Recent studies have begun to elucidate
280 this influence by examining the genetic changes that occur in response to various
281 process conditions.

282 To investigate the effect of carbon source, Oh et al. [55] analyzed expression of
283 111 *E. coli* genes in glycolysis, PP and TCA among other pathways, during growth on
284 different carbon sources (glucose, glycerol and acetic acid). The gene expression profile
285 changed more significantly from glucose to acetate media than from glucose to glycerol.
286 Most biosynthetic genes were down-regulated in both glycerol and acetate media
287 compared to glucose, while TCA cycle genes were up-regulated in both media. The
288 glycolytic genes were roughly unchanged in the glycerol medium but decreased in the
289 acetate medium. *pykA* was observed as up regulated in glycerol compared to glucose.
290 Since the phosphotransferase system (PTS) is not active in the glycerol-based medium,
291 *pykA* would be essential to generate pyruvate for the TCA cycle [55]. On the other hand,
292 *PykF* has higher activity on glucose in aerobic conditions [21]. The *pykF pykA* double
293 knockout in JM101 was shown to improve plasmid copy number on glucose medium at
294 laboratory scale [22]. The effect of this mutation for plasmid production under different
295 conditions, such as glycerol as the carbon source and large scale process, is not known.
296 The choice of carbon source can affect plasmid DNA yield depending on the selected
297 strain, since *E. coli* gene regulation varies under different conditions [55].

298 Feeding strategy and fermentation scale also seem to influence plasmid DNA
299 production. Exponential feeding was demonstrated to be very effective for *fruR* mutant
300 DH5 α [25] and distinct differences in yields and gene expression were observed
301 between culture flasks and reactor fermentations [13]. The differences observed across

302 scales and fermentation strategies can sometimes be correlated to strain growth rate.
303 The feeding strategy in fed-batch fermentations has the main objective of controlling
304 specific growth rate while in shake flasks the growth rate is not actively regulated.
305 O'Kennedy et al. verified an increase of 8-fold in specific yield and 25-fold in volumetric
306 yield of plasmid DNA using exponential fed-batch in comparison with shake flask
307 cultures [56].

308 Discovery of a temperature effect on replication of some plasmids was a crucial
309 step forward in recombinant protein production [57-58] and it has been intensively used
310 in plasmid DNA production as well [13, 16, 22, 59]. However, changing temperature also
311 impacts different strains in different ways during plasmid DNA synthesis [16] and the
312 results can be seen even in the cell morphology of some strains, such as DH5 α [60]. TCA
313 cycle genes were observed as down-regulated at 42°C, possibly because of lower
314 dissolved oxygen concentration resulting in a higher concentration of acetic acid at 42°C
315 than at 37°C [61]. A recent proteomic study also observed increased acetate production
316 at higher temperatures while examining the steady-state heat-shock response of
317 MG1655 using continuous cultures. Of possible relevance to pDNA production, the
318 authors observed upregulation of the DNA-binding protein Dps upon temperature shift
319 from 37°C to 47.5°C. This protein has been shown to protect DNA from thermal stress
320 and other types of damage [62]. Another recent study by Caspeta et al. showed that
321 heating rate affects both host gene expression and organic acid accumulation in a BL21-
322 based protein production process, but their findings could be readily applied to a
323 plasmid production process as well. In particular, their data suggest that slower heating
324 rates reduce imbalances between glycolysis and the TCA cycle [63]. Low temperature
325 (30°C) at the beginning of the process is also an alternative to reduce metabolic burden

326 and reduce growth rate during biomass formation, eliminating the need to engineer the
327 host cells [38]. However, development of vectors and strains to obtain maximum yield at
328 relatively low temperatures cannot be discarded since it can greatly contribute to the
329 energy balance at the industrial scale.

330

331 **6. Impact of strain on downstream processing and transfection**

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

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

350 gyrase and sigma factor σ^S (*rpoS*) genes were observed to play an important role in the
351 regulation of plasmid topology and could be potential gene mutation targets to increase
352 the fraction of supercoiled plasmids produced [67].

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

360

361 **7. Conclusion**

362 Host strain optimization is a valuable tool for improving plasmid DNA
363 manufacturing processes. The early literature focused on studying the mechanism of
364 plasmid replication inside an *E. coli* host as well as observing physiological changes in the
365 host in response to plasmid maintenance. This review focused on more recent work
366 that has exploited the advent of genomic and transcriptomic profiling tools to observe
367 plasmid-bearing cells in more detail and compare them to their plasmid-free
368 counterparts. Both up- and down-regulation of key genes were observed in glycolysis,
369 the pentose phosphate pathway and the tricarboxylic acid cycle in response to plasmid
370 DNA production. By analyzing the metabolism of *E. coli*, strategies for systematic
371 knockout or overexpression of rationally-selected genes have been used to improve
372 plasmid DNA yield. However, the effects of a given gene mutation on plasmid yield
373 were sometimes different than expected and often varied depending on the strain's

374 genetic background and/or culture method. Culture conditions, such as fermentation
375 medium, feeding strategy and process scale, seem to be strong factors affecting plasmid
376 yield, and the mechanism and role of these factors has yet to be completely elucidated.

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

386 Creation of a pDNA production strain could potentially start with a wild-type
387 genetic background with introduction of only the mutations that have been shown to
388 enhance pDNA yield and/or quality. This strain may have advantages such as more
389 robust growth and lack of auxotrophies that are often found in other laboratory strains.
390 Another possibility would be to develop a strain that is resistant to bacteriophage
391 infection while retaining desirable production characteristics. Bacteriophage infection
392 can be catastrophic at the commercial scale, and there are currently limited methods
393 available to combat it. Most of these methods are based on altering culture conditions
394 because the phage-resistant strains created thus far often suffer from reduced growth
395 and production capabilities [69]. It would also be possible to approach strain design
396 with the goal of enhancing purification or transfection efficiency. There have been
397 several innovative studies that seek to address these downstream processes at the

398 strain level, but there are still many avenues to explore. These studies will be
399 particularly important as the regulatory structure surrounding gene-based therapeutics
400 continues to evolve and possibly become more stringent.

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

405 **References**

- 406 [1] Wolff, A. J., Malone, R. W., Williams, P., Chong, W., *et al.*, Direct gene transfer into
407 mouse muscle *in vivo*. *Science* 1990, *247*, 1465-1468.
- 408 [2] Carvalho, J. A., Prazeres, D. M. F., Monteiro, G. A., Bringing DNA vaccines closer to
409 commercial use. *IDrugs* 2009, *12*, 642 - 647.
- 410 [3] Liu, M. A., DNA vaccines: an historical perspective and view to the future. *Immunol*
411 *Rev.* 2011, *239*, 62 - 84.
- 412 [4] Kutzler, M. A., Weiner, D. B., DNA vaccines: ready for prime time? *Nat Rev Genet*
413 2008, *9*, 776-788.
- 414 [5] Manthorpe, M., Hobart, P., Hermanson, G., Ferrari, M., *et al.*, Plasmid vaccines and
415 therapeutics: from design to applications. *Adv Biochem Eng Biotechnol* 2005, *99*, 41 - 92.
- 416 [6] Leitner, W. W., Ying, H., Restifo, N. P., DNA and RNA-based vaccines: principles,
417 progress and prospects *Vaccine* 2000, *18*, 765 - 777.
- 418 [7] Prazeres, D. M. F., Monteiro, G. A., Ferreira, G. N. M., Diogo, M. M., *et al.*,
419 Purification of plasmids for gene therapy and DNA vaccination. *Biotechnol Annu Rev.*
420 2001, *7*, 1 - 30.
- 421 [8] Yau, S. Y., Keshavarz-Moore, E., Ward, J., Host strain influences on supercoiled
422 plasmid DNA production in *Escherichia coli*: implications for efficient design of large-
423 scale processes. *Biotechnol Bioeng* 2008, *101*, 529-544.
- 424 [9] Glenting, J., Wessels, S., Ensuring safety of DNA vaccines. *Microb Cell Fact* 2005, *4*,
425 26.
- 426 [10] Bailey, J. E., Silva, N. A. D., Peretti, S. W., Seo, J.-H., Srienc, F., Studies of host-
427 plasmid interactions in recombinant microorganisms. *Ann N Y Acad Sci* 1986, *469*, 194-
428 211.

429 [11] Ricci, J. C. D., Hernandez, M. E., Plasmid effects on *Escherichia coli* metabolism. *Crit*
430 *Rev Biotechnol* 2000, 20, 79 - 108.

431 [12] Ow, D. S.-W., Nissom, P. M., Philp, R., Oh, S. K.-W., Yap, M. G.-S., Global
432 transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia*
433 *coli* DH5 α during batch fermentation. *Enzyme Microb Technol* 2006, 39, 391-398.

434 [13] Williams, J. A., Luke, J., Langtry, S., Anderson, S., *et al.*, Generic plasmid DNA
435 production platform incorporating low metabolic burden seed-stock and fed-batch
436 fermentation processes. *Biotechnol Bioeng* 2009, 103, 1129 - 1143.

437 [14] Wang, Z., Xiang, L., Shao, J., Wegrzyn, A., Wegrzyn, G., Effects of the presence of
438 ColEI plasmid DNA in *Escherichia coli* on the host cell metabolism. *Microb Cell Fact* 2006,
439 5.

440 [15] Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., Kampala, D. S., Plasmid-
441 encoded protein: the principal factor in the “metabolic burden” associated with
442 recombinant bacteria. *Biotechnol Bioeng* 1990, 35, 668 - 681.

443 [16] Singer, A., Eiteman, M. A., Altman, E., DNA plasmid production in different host
444 strains. *J Ind Microbiol Biotechnol* 2009, 36, 521 - 530.

445 [17] Bohle, K., Ross, A., Plasmid DNA production for pharmaceutical use: role of specific
446 growth rate and impact on process design. *Biotechnol Bioeng* 2011, DOI
447 10.1002/bit.23138.

448 [18] Cunningham, D. S., Koepsel, R. R., Ataa, M. M., Domach, M. M., Factors affecting
449 plasmid production in *Escherichia coli* from a resource allocation standpoint. *Microb Cell*
450 *Fact* 2009, 8, 1 - 27.

451 [19] Flores, S., Anda-Herrera, R., Gosset, G., Bolívar, F. G., Growth-rate recovery of
452 *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-
453 phosphate pathway *Biotechnol Bioeng* 2004, *87*, 485 - 494.

454 [20] Siddiquee, K. A. Z., Arauzo-Bravo, M. J., Shimizu, K., Metabolic flux analysis of *pykF*
455 gene knockout *Escherichia coli* based on ¹³C-labeling experiments together with
456 measurements of enzyme activities and intracellular metabolite concentrations. *Appl*
457 *Microbiol Biotechnol* 2004, *63*, 407 - 417.

458 [21] Ponce, E., Flores, N., Martinez, A., Valle, F., Bolivar, F., Cloning of the two pyruvate
459 kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these
460 enzymes in pyruvate biosynthesis. *J Bacteriol* 1995, *177*, 5719-5722.

461 [22] Cunningham, D. S., Liu, Z., Domagalski, N., Koespsel, R. R., *et al.*, Pyruvate kinase-
462 deficient *Echerichia coli* exhibits increased plasmid copy number and cyclic AMP levels. *J*
463 *Bacteriol* 2009, *191*, 3041 - 3049.

464 [23] Carnes, A. E., Luke, J. M., Vincent, J. M., Shukar, A., *et al.*, Plasmid DNA fermentation
465 strain and process-specific effects on vector yield, quality, and transgene expression.
466 *Biotechnol Bioeng* 2010, *108*, 354-363.

467 [24] Ow, D. S.-W., Lee, R. M.-Y., Nissom, P. M., Philp, R., Inactivating FruR global
468 regulator in plasmid-bearing *Escherichia coli* alters metabolic gene expression and
469 improves growth rate. *J Biotechnol.* 2007, *131*, 261 - 269.

470 [25] Ow, D. S.-W., Yap, M. G.-S., Oh, S. K.-W., Enhancement of plasmid DNA yields during
471 fed-batch culture of a fruR- knockout *Escherichia coli* strain *Biotechnol. Appl. Biochem.*
472 2009, *52*.

473 [26] Phue, J.-N., Lee, S. J., Trinh, L., Shiloach, J., Modified *Escherichia coli* B (BL21), a
474 superior producer of plasmid DNA compared with *Escherichia coli* K (DH5 α). *Biotechnol*
475 *Bioeng* 2008, *101*, 831 - 836.

476 [27] Taylor, R. C., Walker, D. C., McInnes, R. R., *E. coli* host strains significantly affect the
477 quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.*
478 1993, *21*, 1677 - 1678.

479 [28] Summers, D., Timing, self-control and a sense of direction are the secrets of
480 multicopy plasmid stability. *Mol. Microbiol* 1998, *29*, 1137 - 1141.

481 [29] Collier, J., Epigenetic regulation of the bacterial cell cycle. *Current Opinion in*
482 *Microbiology* 2009, *12*, 722 - 729.

483 [30] Keseler, I. M., Collado-Vides, J., A., S.-Z., Peralta-Gil, M., *et al.*, EcoCyc: a
484 comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res* 2011, *39*.

485 [31] Wegrzyn, G., Replication of plasmids during bacterial response to amino acid
486 starvation. *Plasmid* 1999, *41*, 1-16.

487 [32] Wick, L. M., Egli, T., Molecular components of physiological stress responses in
488 *Escherichia coli*. *Adv Biochem Eng Biotechnol* 2004, *89*, 1 - 45.

489 [33] Magnusson, L. U., Farewell, A., Nystrom, T., ppGpp: a global regulator in *Escherichia*
490 *coli*. *Trends Microbiol* 2005, *13*, 236 - 242.

491 [34] Carnes, A. E., Williams, J. A., Plasmid DNA manufacturing technology. *Recent Pat*
492 *Biotechnol.* 2007, *1*, 1-16.

493 [35] Cai, Y., Rodriguez, S., Hebel, H., DNA vaccine manufacture: scale and quality. *Expert*
494 *Rev Vaccines* 2009, *8*, 1277-1291.

495 [36] Durland, R. H., Eastman, E. M., Manufacturing and quality control of plasmid-based
496 gene expression systems. *Adv Drug Deliv Rev.* 1998, *30*, 33 - 48.

497 [37] Bower, D. M., Prather, K. L. J., Engineering of bacterial strains and vectors for the
498 production of plasmid DNA. *Appl Microbiol Biotechnol* 2009, 82, 805 - 813.

499 [38] Williams, J. A., Carnes, A. E., Hodgson, C. P., Plasmid DNA vaccine vector design:
500 impact on efficacy, safety and upstream production. *Biotechnol Adv.* 2009, 27, 353 - 370.

501 [39] Bachmann, B. J., Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol*
502 *Rev* 1972, 36, 525 - 557.

503 [40] Grant, S. G. N., Jessee, J., Bloom, F. R., Hanahan, D., Differential plasmid rescue from
504 transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl.*
505 *Acad. Sci. USA* 1990, 87, 4645 - 4649.

506 [41] Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K. B., *et al.*, *Escherichia coli* K-12: a
507 cooperatively developed annotation snapshot—2005. *Nucleic Acids Res.* 2006, 34, 1 - 9.

508 [42] Phue, J., Noronha, S. B., Hattacharyya, R., Wolfe, A. J., Shiloach, J., Glucose
509 metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic
510 pathways are responsible for efficient glucose utilization in *E. coli* B as determined by
511 microarrays and northern blot analyses. *Biotechnol Bioeng* 2005, 90, 805 - 820.

512 [43] Cranenburgh, R. M., Hanak, J. A. J., Williams, S. G., Sherratt, D. J., *Escherichia coli*
513 strains that allow antibiotic-free plasmid selection and maintenance by repressor
514 titration. *Nucleic Acids Res* 2001, 29, 1-6.

515 [44] Mairhofer, J., Cserjan-Puschmann, M., Striedner, G., Nobauer, K., *et al.*, Marker-free
516 plasmids for gene therapeutic applications - Lack of antibiotic resistance gene
517 substantially improves the manufacturing process. *J Biotechnol.* 2010, 146, 130 - 137.

518 [45] Soubrier, F., Cameron, B., Manse, B., Somarriba, S., *et al.*, pCOR: a new design of
519 plasmid vectors for nonviral gene therapy. *Gene Ther.* 1999, 6, 1482 - 1488.

520 [46] Luke, J., Carnes, A. E., Hodgson, C. P., Williams, J. A., Improved antibiotic-free DNA
521 vaccine vectors utilizing a novel RNA based plasmid selection system. *Vaccine* 2009, 27,
522 6454-6459.

523 [47] Carnes, A. E., Luke, J. M., Vincent, J. M., Anderson, S., *et al.*, Critical design criteria
524 for minimal antibiotic-free plasmid vectors necessary to combine robust RNA Pol II and
525 Pol III-mediated eukaryotic expression with high bacterial production yields. *J Gene Med.*
526 2010, 12, 818 - 831.

527 [48] Faurez, F., Dory, D., Le Moigne, V., Gravier, R., Jestin, A., Biosafety of DNA vaccines:
528 new generation of DNA vectors and current knowledge on the fate of plasmids after
529 injection. *Vaccine* 2010, 28, 3888-3895.

530 [49] Gill, D. R., Pringle, I. A., Hyde, S. C., Progress and prospects: the design and
531 production of plasmid vectors. *Gene Ther.* 2009, 16, 165 - 171.

532 [50] Kay, M. A., He, C., Chen, Z., A robust system for production of minicircle DNA
533 vectors. *Nat Biotechnol* 2010, 28, 1287 - 1291.

534 [51] Oliveira, P. H., Prather, K. J., Prazeres, D. M. F., Monteiro, G. A., Structural instability
535 of plasmid biopharmaceuticals: challenges and implications. *Trends Biotechnol* 2009, 27,
536 503 - 511.

537 [52] Prather, K. L. J., Edmonds, M. C., Herod, J. W., Identification and characterization of
538 IS1 transposition in plasmid amplification mutants of *E. coli* clones producing DNA
539 vaccines. *Appl Microbiol Biotechnol* 2006, 73, 815 - 826.

540 [53] Oliveira, P. H., Prazeres, D. M. F., Monteiro, G. A., Deletion formation mutations in
541 plasmid expression vectors are disfavored by runaway amplification conditions and
542 differentially selected under kanamycin stress. *J Biotechnol.* 2009, 143, 231 - 238.

543 [54] Posfai, G., Plunkett, G., Feher, T., Frisch, D., *et al.*, Emergent properties of reduced-
544 genome *Escherichia coli*. *Science* 2006, 312, 1044-1046.

545 [55] Oh, M., Liao, J. C., Gene expression profiling by DNA microarrays and metabolic
546 fluxes in *Escherichia coli*. *Biotechnol Prog.* 2000, 16, 278 - 286.

547 [56] O'Kennedy, R. D., Ward, J. M., Keshavarz-Moore, E., Effects of fermentation strategy
548 on the characteristics of plasmid DNA production. *Biotechnol Appl Biochem* 2003, 37, 83-
549 90.

550 [57] Lin-Chao, S., Chen, W. T., Wong, T. T., High copy number of the pUC plasmid results
551 from a Rom/Rop-suppressible point mutation in RNA II. *Mol Microbiol* 1992, 6, 3385-
552 3393.

553 [58] Fitzwater, T., Zhang, X., Elble, R., Polisky, B., Conditional high copy number ColE 1
554 mutants: resistance to RNA1 inhibition *in vivo* and *in vitro*. *EMBO J* 1988, 7, 3289 - 3297.

555 [59] Carnes, A. E., Hodgson, C. P., A. Williams, J., Inducible *Escherichia coli* fermentation
556 for increased plasmid DNA production. *Biotechnol. Appl. Biochem.* 2006, 45, 155 - 166.

557 [60] Silva, F., Passarinha, L., Sousa, F., Queiroz, J. A., Domingues, F. C., Influence of
558 growth conditions on plasmid DNA production. *J Microbiol Biotechnol* 2009, 19, 1408-
559 1414.

560 [61] Hasan, C. M. M., Shimizu, K., Effect of temperature up-shift on fermentation and
561 metabolic characteristics in view of gene expressions in *Escherichia coli*. *Microb Cell Fact*
562 2008, 7, 1 - 13.

563 [62] Luders, S., Fallet, C., Franco-Lara, E., Proteome analysis of the *Escherichia coli* heat
564 shock response under steady-state conditions. *Proteome Sci* 2009, 7, 36.

565 [63] Caspeta, L., Flores, N., Perez, N. O., Bolivar, F., Ramirez, O. T., The effect of heating
566 rate on *Escherichia coli* metabolism, physiological stress, transcriptional response, and

567 production of temperature-induced recombinant protein: a scale-down study.
568 *Biotechnol Bioeng* 2009, 102, 468-482.

569 [64] US Dept of Health and Human Services, F. D. A., Guidance for industry:
570 considerations for plasmid DNA vaccines for infectious disease indications. 2007.

571 [65] Pillai, V. B., Hellerstein, M., Yu, T., Amara, R. R., Robinson, H. L., Comparative
572 studies on *in vitro* expression and *in vivo* immunogenicity of supercoiled and open
573 circular forms of plasmid DNA vaccines. *Vaccine* 2008, 26, 1136 - 1141.

574 [66] Li, H., Bo, H., Wang, J., Shao, H., Huang, S., Separation of supercoiled from open
575 circular forms of plasmid DNA, and biological activity detection. *Cytotechnology* 2011,
576 63, 7 - 12.

577 [67] Reyes-Dominguez, Y., Contreras-Ferrat, G., Ramirez-Santos, J., Membrillo-
578 Hernandez, J., Gomez-Eichelmann, M. C., Plasmid DNA supercoiling and gyrase activity in
579 *Escherichia coli* wild-type and *rpoS* stationary-phase cells. *J Bacteriol* 2003, 185, 1097 -
580 1100.

581 [68] Marie, C., Vandermeulen, G., Quiviger, M., Richard, M., *et al.*, pFARs, Plasmids free
582 of antibiotic resistance markers, display high-level transgene expression in muscle, skin
583 and tumour cells. *J Gene Med.* 2010, 12, 323 -332.

584 [69] Los, M., Czyz, A., Sell, E., Wegrzyn, A., *et al.*, Bacteriophage contamination: is there
585 a simple method to reduce its deleterious effects in laboratory cultures and
586 biotechnological factories? *J. Appl. Genet.* 2004, 45, 111 - 120.

587 [70] Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., *et al.*, Highly accurate genome
588 sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* 2006, 2,
589 2006 0007.

590 [71] Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, *et al.*, The complete genome
591 sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse.
592 *J Bacteriol* 2008, *190*, 2597-2606.

593 [72] Yanisch-Perron, C., Vieira, J., Messing, J., Improved M13 phage cloning vectors and
594 host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 1985, *33*,
595 103-119.

596
597

598

599

600

601

602

603

604

605

606

607

608

609

610 **Acknowledgments**

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

613 The authors have declared no conflict of interest.

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628 **Figure legends**

629 **Figure 1.** Gene knockout and overexpression strategies to improve plasmid DNA
630 production in *E. coli*. (A) Overexpression of the *zwf* and *rpiA* genes is proposed as a
631 means to increase fluxes in the pentose phosphate pathway and nucleotide synthesis.
632 (B) Knockouts of genes *pykF* and *pykA* are suggested to reduce acetate formation,
633 increase fluxes in the pentose phosphate pathway and TCA cycle. Abbreviations: G6P,
634 glucose 6-phosphate; F6P, fructose 6-phosphate; FDP, fructose 1,6-diphosphate; G3P,
635 glyceraldehyde 3-phosphate; PEP, phosphoenolpyruvate; OAA, oxaloacetate; MAL,
636 malate; 6GPC, 6-phosphogluconate; RU5P, ribulose 5-phosphate; XU5P, xylulose 5-
637 phosphate; R5P, ribose 5-phosphate; S7P, sedoheptulose 7-phosphate; E4P, erythrose 4-
638 phosphate; PTS, phosphotransferase system. This figure represents a simplified version
639 of *E. coli* central metabolism. Dark arrows represent high carbon flow in pentose
640 phosphate pathway and light arrows represent less formation of pyruvate and acetate.

641

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

651 elsewhere ([http://openwetware.org/wiki/E. coli_genotypes#Commonly used strains](http://openwetware.org/wiki/E._coli_genotypes#Commonly_used_strains)).

652 Abbreviation: Methylation-dependent restriction system (MDRS).

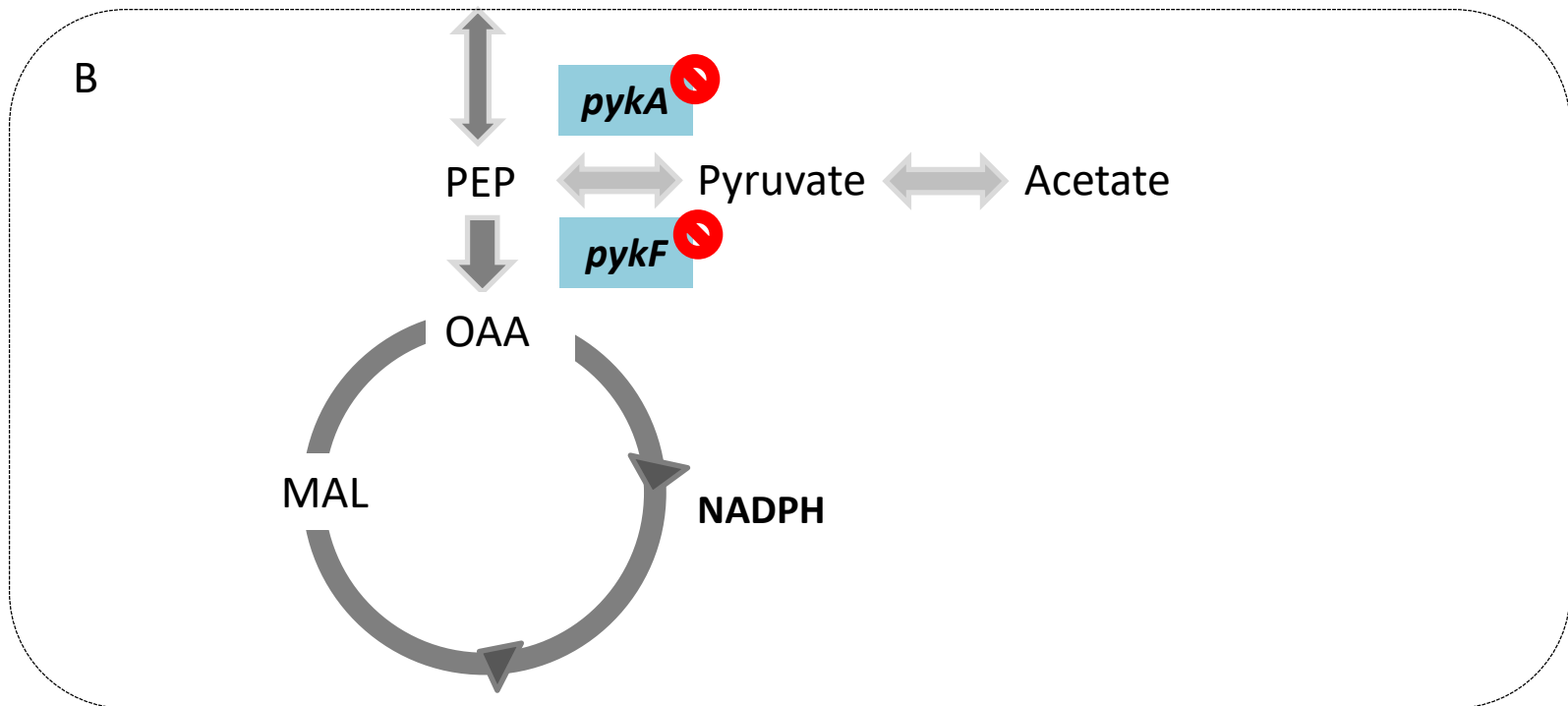
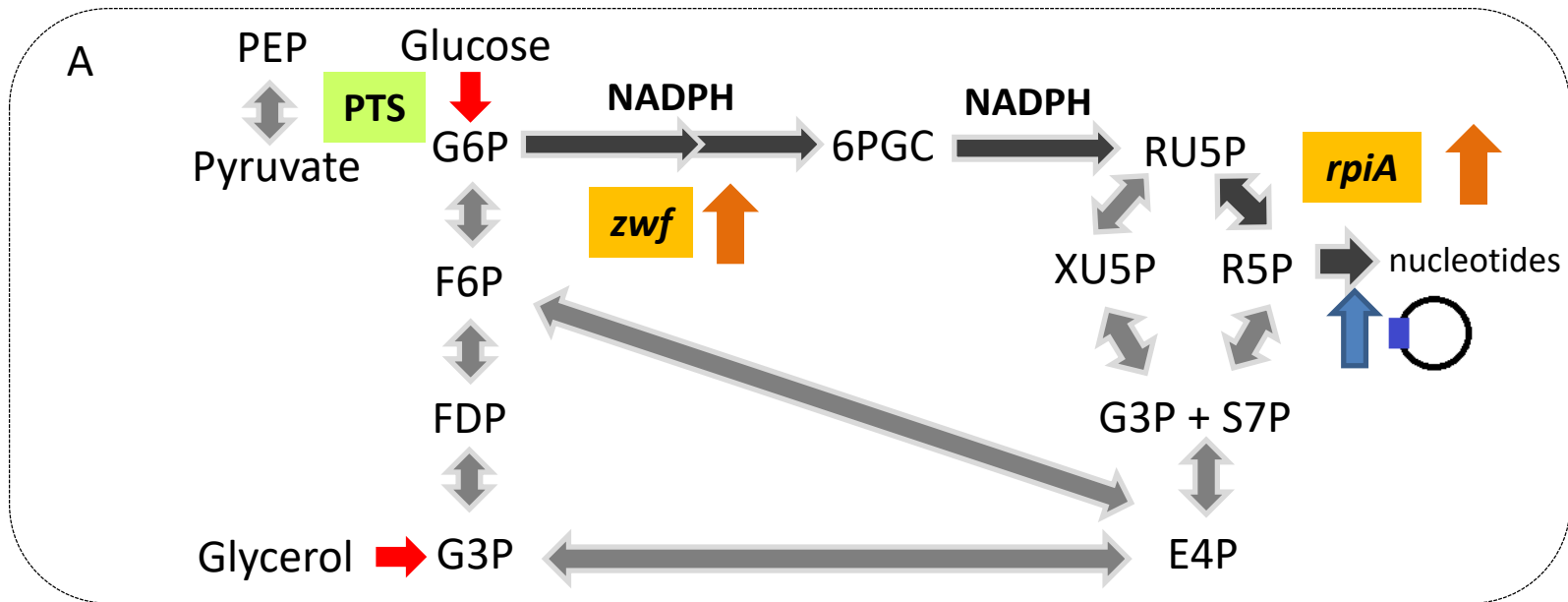
653

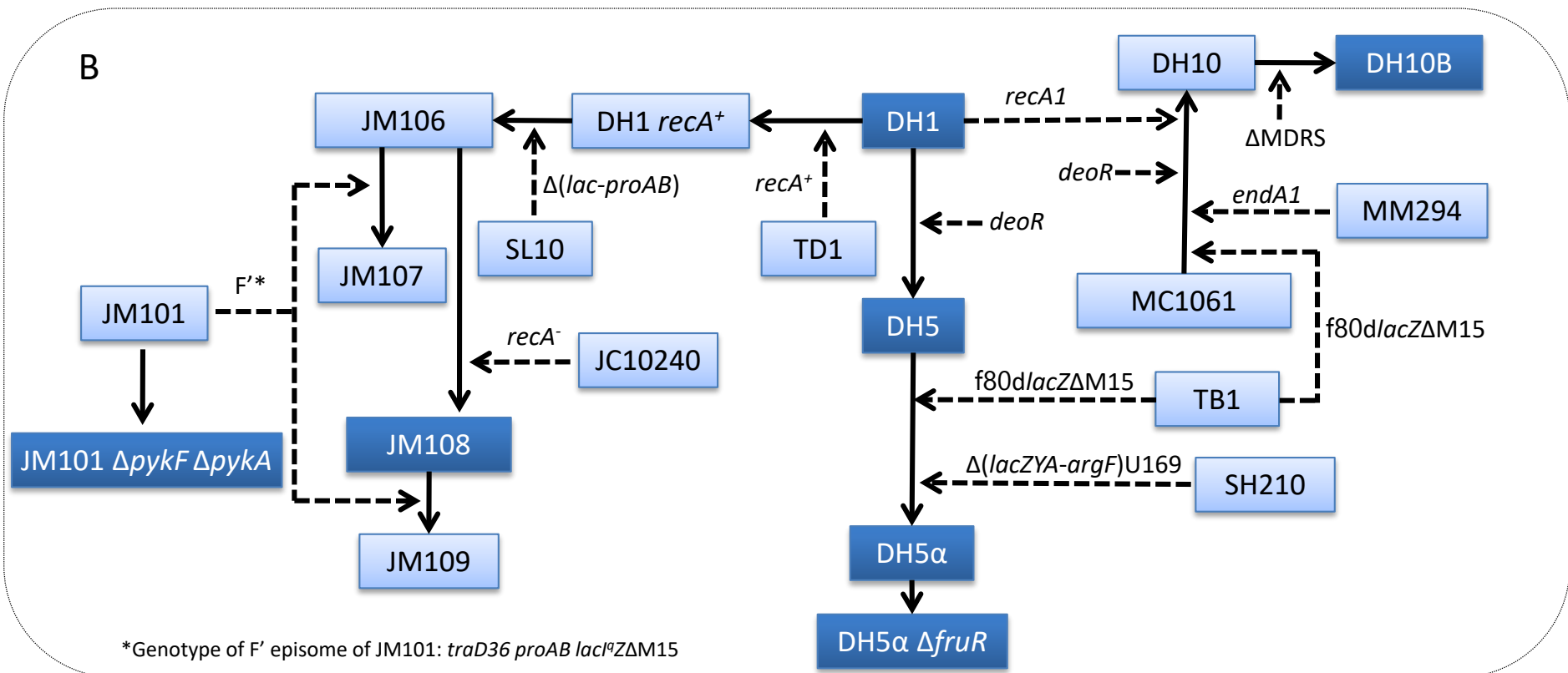
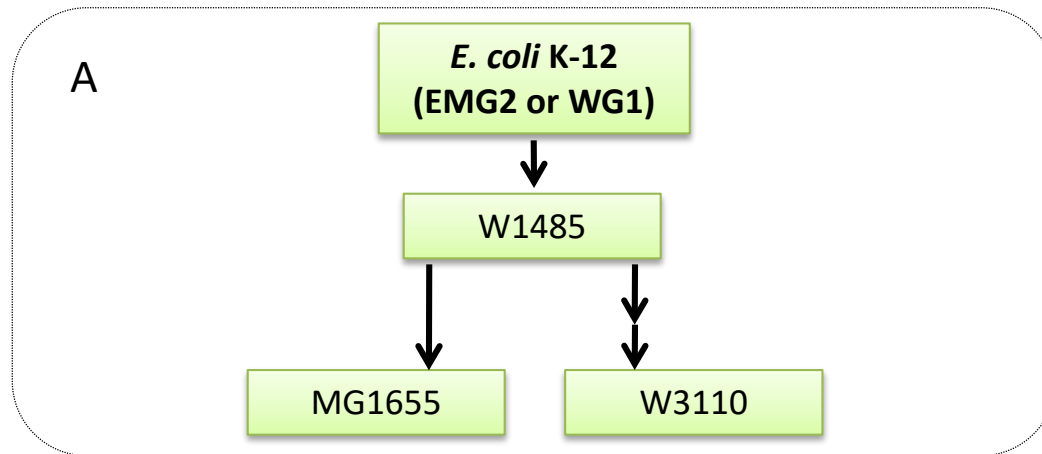
654

655

656 **Table 1.** *E. coli* genes targeted for mutation to improve plasmid DNA production. Δ
657 indicates gene knockout and \uparrow indicates gene overexpression. Information about gene
658 products and their functions were compiled from Ecocyc [30] and the indicated
659 references.

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





*Genotype of *F'* episome of JM101: *traD36 proAB lacI^qZΔM15*