Re-engineering of an *Escherichia coli* K-12 strain for the efficient production of recombinant human Interferon Gamma

Rajat Pandey[a,b,*], Nitin Kumar[a,b], Gabriel A. Monteiro[a], Venkata Dasu Veeranki[b], D.M.F. Prazeres[a]

[a] *iBB-Institute for Bioengineering and Biosciences, Department of Bioengineering, Instituto Superior Técnico, Lisbon 1049-001, Portugal*
[b] *Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati 781039, India*

ABSTRACT

The *Escherichia coli* phosphoglucose isomerase (*pgi*) mutant strain GALG20 was developed previously from wild-type K12 strain MG1655 for increased plasmid yield. To investigate the potential effects of the *pgi* deletion/higher plasmid levels on recombinant human Interferon Gamma (IFN-γ) production, a detailed network of the central metabolic pathway (100 metabolites, 114 reactions) of GALG20 and MG1655 was constructed. Elementary mode analysis (EMA) was then performed to compare the phenotypic spaces of both the strains and to check the effect of the *pgi* deletion on flux efficiency of each metabolic reaction. The results suggested that *pgi* deletion increases amino acid biosynthesis and flux efficiency towards IFN-γ synthesis by 11%. To further confirm the qualitative prediction that the *pgi* mutation favours recombinant human IFN-γ expression, GALG20 and MG1655 were lysogenised, transformed with a plasmid coding for IFN-γ and tested alongside with BL21(DE3) for their expression capabilities in shake flask experiments using complex media. IFN-γ gene expression was analysed by quantifying plasmid and mRNA copy number per cell and IFN-γ protein production level. Specific IFN-γ yields confirmed the in silico metabolic network predictions, with GALG20(DE3) producing 3.0-fold and 1.5-fold more IFN-γ as compared to MG1655(DE3) and BL21(DE3), respectively. Most of the total IFN-γ was expressed as inclusion bodies across the three strains: 95% in GALG20(DE3), 97% in BL21(DE3) and 72% in MG1655(DE3). The copy number of mRNA coding for IFN-γ was found to be higher in GALG20(DE3) as compared to the other two strains. Overall, these findings show that GALG20(DE3) has the potential to become an excellent protein expression strain.

1. Introduction

Over-expression of therapeutically relevant proteins is an important research area for the Biotechnology and Pharmaceutical industry. The gram-negative bacterium *Escherichia coli* is one of the hosts of choice for this specific purpose [1]. Major reasons for this preference include the fact that (i) *E. coli*'s genome is completely elucidated [2,3], (ii) genetic manipulation is straightforward and (iii) large amounts of target protein (up to 50% of total cell protein) can be produced [4–6]. While most systems used to produce recombinant therapeutic proteins in *E. coli* (Table 1) rely on B-type strains, the potential of K-12 strains [7] is largely un-explored. The major reason for overlooking *E. coli* K-12 strains is associated with its propensity to accumulate high amounts of acetate due to unbalanced carbon metabolism [8,9]. This characteristic can slow down both cellular growth and heterologous protein production.

Most heterologous protein production systems in *E. coli* rely on plasmid-based expression. Several vectors with varying copy number have been developed for this purpose (Table S1). A high copy number plasmid carrying the target gene is expected to produce larger amounts of target protein when compared with a low copy number counterpart [10]. However, in some instances high plasmid copy number may lead to an increased metabolic burden, reducing bacterial growth rate and plasmid stability [11,12]. Some researchers have successfully managed to produce more recombinant protein from a high copy number plasmid by increasing plasmid segregational stability [13,14].

The performance of an expression host can be improved either by changing the growth environment or by altering its genetic architecture. Such alterations may exert positive or negative effects on the metabolic network flexibility, individual reactions, transcriptional control and flux efficiency for a desired product. Many researchers have engineered the *E. coli* genetic makeup and recombinant plasmids by

Table 1
Illustrative examples of recombinant therapeutic protein production in E. coli.

Target Protein	E. coli strain	Type	Reference
Human growth hormone	XL1-blue	K12	[56]
Human growth hormone	BL21	B	[57]
Granulocyte colony-stimulating factor	BL21DE3	B	[58]
Interferon alpha 2b	JM109DE3	K12	[59]
Interferon alpha 2b	BL21DE3	B	[60]
Interferon beta	BL21DE3	B	[61]
Interleukin 6	BL21DE3	B	[62]
Interferon gamma	BL21DE3	B	[63]
Human erythropoietin	BL21DE3	B	[64]

deleting or adding genes in order to create a superior microbial cell factory [15–24]. An interesting target for deletion is *pgi*, the gene that codes for phosphoglucose isomerase enzyme. Changes in the availability of this enzyme can potentially affect the NADH/NADPH metabolism, glyoxylate cycle and Krebs cycle significantly. For example, an *E. coli* K-12 *pgi* mutant strain that primarily uses the pentose phosphate pathway (PPP) for glucose catabolism had an active glyoxylate shunt [25]. Another report describes that a 2.6 fold increase in riboflavin production is obtained by redirecting the carbon flux into the oxidative pentose phosphate pathway via deletion of *pgi* in the wild-type *E. coli* K-12 strain MG1655 [26]. Likewise, a 2-fold increase in the production of poly (3-hydroxybutyrate) by a *pgi* knock out of an *E. coli* DF11 strain (compared to JM109) was observed as a result of the redirection of the carbon flow towards pentose phosphate pathway [27]. However, and to the best of our knowledge, the impact of *pgi* mutation in *E. coli* K-12 strains on the production of recombinant therapeutic proteins, more specifically on recombinant human IFN-γ has not been studied.

A constraint-based metabolic flux balance analysis (FBA) has been used to understand changes in phenotype associated with *pgi* deletions designed to enhance yields of natural compounds like shikimate and catechin [28,29]. Nevertheless, FBA may not be the best methodology to study the changes in the entire phenotypic space because it focuses on a single phenotype fulfilling the optimization of a given objective function, mostly maximizing growth [30]. Production of recombinant proteins depends on the availability of nucleotides for plasmid and mRNA synthesis and on amino acids for protein synthesis. The fact that the synthetic pathways responsible for nucleotide and amino acid synthesis are very intricately connected to central carbon pathways renders analysis of all possible flux distributions essential to quantify the potential of an expression host for recombinant protein production.

Plasmids are one of the key molecular tools at the heart recombinant protein production. However, in the last 20 years, plasmid-based biopharmaceuticals have also been developed for application in non-viral gene therapy and DNA vaccination [31]. Furthermore, plasmids are an important reagent in other applications like the *ex-vivo* transfection of cells for therapy [32] and the transient transfection of mammalian cells for the production of recombinant proteins [33] and viral vectors [34]. The production of large quantities of plasmid DNA is critical for the success of these applications. Metabolic engineering has proven to be a very powerful tool for the development and improvement of *E. coli* strains dedicated to plasmid production [35,36]. We have recently developed the strain GALG20 (*Δpgi, ΔendA, ΔrecA*), an *E. coli* engineered mutant of the K-12 strain MG1655, with the specific goal of increasing plasmid productivity [37,38]. The *pgi* gene was knocked out from *E. coli* MG1655 to redirect carbon flow into the pentose phosphate pathway and promote nucleotide biosynthesis, whereas *endA* and *recA* mutations were made to avoid recombination and non-specific digestion of DNA [39,40]. The resulting GALG20 strain is capable of producing significantly higher amounts of plasmid DNA (up to 25 fold) when compared with the parental strain MG1655 (*ΔendA, ΔrecA*) and with a general cloning strain DH5α [38]. Following up on this work, we hypothesized that GALG20 could be a valuable

recombinant protein producer strain by the virtue of its enhanced plasmid producing capabilities.

In this work, metabolic network analysis of GALG20 and MG1655 strains was performed first to evaluate the impact of a deletion of *pgi* on the expression of recombinant human IFN-γ. Next, an experimental plan was developed to lysogenize GALG20 and MG1655, and test the ability of the resulting GALG20(DE3) and MG1655(DE3) strains to express IFN-γ. The BL21(DE3) strain was included in the study for comparative purposes. The performance of the three strains was then tested by measuring plasmid and mRNA copy number per cell and IFN-γ expression levels.

2. Methods

2.1. Analysis of metabolic phenotypic spaces

A detailed network of the central metabolic pathways with 100 metabolites (Table S2) and 114 reactions ($S = 100(M) \times 114(N)$) was constructed, with 9 exchange reactions and 17 reversible reactions [30,41]. The complete set of reactions is shown in Table S3. Glucose, ammonium, phosphate, O_2, CO_2, ethanol [42,43] and acetate were considered as external metabolites. Glucose was considered as the sole carbon source via the phosphotransferase system (PTS) based uptake (reaction #1). For the construction of mutant phenotypes, the phosphoglucose isomerase catalysed reaction (reaction #3) was eliminated, originating a network of $S = 100(M) \times 113(N)$) reactions. Ethanol and acetate were considered as the major overflow metabolites secreted by cells to balance $NADH/NAD^+$ and ATP/ADP under high carbon source uptake rate or low oxygen availability (which is a typical situation during batch growth of *E. coli*) [41]. Once secreted by the cells ethanol and acetate can be consumed back; thus, to simulate such phenotypes, the glyoxylate cycle (reaction #24), gluconeogenesis and Entner-Doudoroff pathway (reaction #23) were also included. Phenotypes pertaining to both NADH and NADPH dependent malic enzyme reactions were also simulated (reactions #25 and #26) and model was given freedom to choose the best [45]. Reactions for the synthesis of individual nucleotides (reactions #64 to #79) and amino acids (reactions #35 to #63) from intermediary metabolites were included separately. *E. coli* is known to have transhydrogenase activity, and for this a transhydrogenation reaction (reaction #85 and 86) was incorporated with a cost of 0.25 mol ATP/mole NADH [46]. Maintenance energy requirement of cells was addressed by incorporating an ATP hydrolysis reaction (reaction #102). For ATP regeneration via oxidative phosphorylation both NADH (reaction #87) and FADH (reaction #88) were considered separately with a yield of 2 and 1 mol of ATP on one mole of NADH and FADH respectively. Biomass reaction was set with constituent amino acids, nucleotides, lipids and other requirements (reaction #100). Recombinant proteins are polymerized from amino acids by spending energy in the form of ATP. A reaction for the synthesis of human IFN-γ was included by quantifying the per mole amino acid requirement for the His-tagged IFN-γ designed (Table S4) and assuming 4 ATPs per peptide bond (reaction #101). MATLAB R2012a (MathWorks) and Excel 2013 (Microsoft office) were used for all the modelling and quantitative work over a DELL computer machine with an Intel core i5 2.4 GHz CPU, windows 7 OS and 4GB RAM. The metabolic phenotypic space was analysed by quantifying elementary modes in MATLAB using an algorithm developed by Terzer and Stelling [47]. All the elementary flux modes were represented by a $N \times M$ matrix E, where N is the number of reactions, M is the number of elementary modes and elements $e_{i,j}$ are elementary mode coefficients in the matrix E.

The phenotypic spaces of GALG20(DE3) and MG1655(DE3) were compared to check the effect of the *pgi* deletion on flux efficiency of each reaction [30,48]. Flux efficiency of each reaction was determined by first calculating the efficiency matrix $M\varepsilon$ from E, where $\varepsilon_{j,i}$ is the element of $M\varepsilon$.