This is a pre-print of an article published in Biotechnology and Bioprocess Engineering. The final authenticated version is available online at: https://doi.org/10.1007/s12257-007-0126-5
Optimization of Human Interferon Gamma Production in *Escherichia coli* by Response Surface Methodology

Victor E. Balderas Hernández, Luz M.T. Paz Maldonado, Emilio Medina Rivero, Ana P. Barba de la Rosa, Leandro G. Ordoñez Acevedo and Antonio De León Rodríguez*

Instituto Potosino de Investigación Científica y Tecnológica, División de Biología Molecular, Camino a la Presa San José 2055 Col. Lomas 4 sección CP 78216.
San Luis Potosí, S.L.P., México.

Running title: Production of human interferon γ

Submitted to: Biotechnology and Bioprocess Engineering

*Corresponding author
Tel.: +52-444-8342000
Fax: +52-444-8342010
e-mail: aleonr@ipicyt.edu.mx
ABSTRACT

The production of human interferon gamma (hIFN-γ) using a synthetic gene in *Escherichia coli* BL21-SI was optimized by response surface methodology (RSM) and a Box-Behnken design. The process variables studied were temperature, biomass concentration at induction time and the NaCl concentration as inducer. According to the Box-Behnken design, a second order response function was developed. The optimal expression conditions were a temperature of 32.6°C, induction biomass of 0.31 g/l and 0.3 M NaCl in minimal medium. The model prediction for the maximum hIFN-γ production was 77.3 mg/l, which corresponded satisfactorily with the experimental data. The hIFN-γ concentration attained under optimized conditions was 13-times higher than that obtained using the non-optimized conditions. We conclude that RSM is an effective method for the optimization of recombinant protein expression using synthetic genes in *E. coli*.

Keywords: *Escherichia coli* BL21-SI, minimum medium, therapeutic protein, synthetic gene.
INTRODUCTION

Human interferon gamma (hIFN-γ) is a cytokine involved in transcriptional control of a large number of immunological relevant genes. This glycoprotein is composed of 143 amino acid residues with a molecular mass of 20–25 kDa. The recombinant hIFN-γ produced in *Escherichia coli* has a molecular mass of 16.7 kDa and it is not glycosylated, however is physiologically active [1]. Clinical trials indicate that recombinant hIFN-γ has therapeutic efficacy on kidney cell carcinoma, colon cancer, osteopetrosis, chronic granulomatous disease and rheumatoid arthritis [2,3]. *E. coli* has many advantages as an expression system that include rapid growth, cheap cultivation, well-studied genetic background and high expression level in large-scale cultures. Thus, *E. coli* is widely used for heterologous protein expression [4–7]. However, the rapid degradation of mRNA from foreign cDNA causes poor expression in heterologous systems, which has been attributed to the difference of codon usage between eukaryotic organisms and those preferred by *E. coli* [8,9]. Codon optimization through synthetic gene design leads to improved heterologous expression, reducing translational errors [10]. On the other hand, heterologous protein production is strongly dependent on the composition of the culture medium and culture conditions such as temperature, inducer concentration, induction time and others [11,12]. Thus, optimization of fermentation conditions is further recommended before scale-up [13].

Response surface methodology (RSM) is a collection of statistical and mathematical techniques for designing experiments, building models, evaluating the effects of factors and the optimization of many chemical and biochemical processes [14,15]. In this work, we
optimized the culture conditions for the expression of a synthetic gene of hIFN-\(\gamma\) in \textit{E. coli} BL21-SI using a minimal medium and the Box-Behnken design. The temperature, biomass concentration at induction time and optimal inducer concentration were determined by RSM.

**MATERIALS AND METHODS**

\textit{Plasmid and bacterial strain}

The synthetic hIFN-\(\gamma\) gene was amplified using pCR4-516 (Entelechon, Regensburg, Germany) as template. The restriction sites \textit{NdeI} and \textit{BamHI} were added using the forward primer 5´ CAT ATG CAG GAC CCA TAC GTG AAA 3´ and reverse primer 5´ GGA TCC TTA CTG ACT TGC 3´, respectively. The amplified product was cloned in pET12a (Novagen, Darmstadt, Germany) in the same restriction sites to obtain the expression vector pBAL0, then \textit{E. coli} BL21-SI (GIBCO, Darmstadt, Germany) was transformed with pBAL0 and ampicillin was used for selection of transformants cells.

\textit{Culture Media}

LBON agar [16] containing 100 \(\mu\)g/mL ampicillin was used routinely to grow \textit{E. coli} BL21-SI/pBAL0. The minimal medium contained (per liter): 5 g glucose, 3.5 g \((\text{NH}_4)_2\text{HPO}_4\), 3.5 g KH\(_2\)PO\(_4\), 1.0 g MgSO\(_4\), 40 \(\mu\)g thiamine and 100 mg ampicillin. The pH was adjusted to 7.0 prior to autoclaving for 15 min at 121°C; glucose and MgCl\(_2\) solutions were autoclaved separately. For all experiments, pre-inocula were cultured overnight at 250
rpm and 37°C using the minimal medium plus 0.5% yeast extract (Difco, Franklin Lakes, NJ).

**Growth conditions**

Cultures were generated in 500 mL flasks containing 100 mL of minimal medium and shaken at 250 rpm. Each flask was inoculated with 3 mL of pre-inoculum to provide an initial optical density at 600 nm (OD$_{600nm}$) of 0.2. Temperature of incubation, biomass concentration at induction time and NaCl concentration for each experiment were defined according to the Box-Behnken design described below.

**Box-Behnken design and surface response methodology**

The Box-Behnken design [17] was selected to study the effect of culture conditions on the production of hINF-$\gamma$ in *E. coli* BL21-SI/pBAL0. The design evaluates three independent variables, each one at three levels with three replicates at the center points (Table 1), to fit a polynomial model:

$$ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 \quad (1) $$

Where $Y$ is the production of hIFN-$\gamma$, $X_1$, $X_2$ and $X_3$ are the independent variables for temperature, biomass concentration at induction time and NaCl concentration, respectively. $\beta_0$ is the intercept term; $\beta_1$, $\beta_2$, and $\beta_3$ are linear coefficients; $\beta_{12}$, $\beta_{13}$, and $\beta_{23}$ are interactive coefficients; and $\beta_{11}$, $\beta_{22}$, and $\beta_{33}$ are quadratic coefficients. The model was evaluated with
significance, good fit and $R^2$ values. Eq. 1 was used to build surfaces graphs for the model. The analysis of RSM, analysis of variance (ANOVA) and the optimal conditions were performed using Statgraphics Plus v5.0 software (Statistical Graphics, Herndon, VA, USA). The significant effects on dependent variables were determined by $t$-test with a probability value ($P$-value) smaller than 0.05.

**Analytical procedures**

Cell growth was monitored at OD$_{600nm}$ (Varian Cary BIO-50; Varian, Palo Alto, CA, USA) and converted to dry cell weight (DCW) by a calibration curve. Total protein concentration was analyzed by Lowry method [18] using bovine serum albumin (BioRad, Hercules, CA, USA) as standard. Expression of hIFN-$\gamma$ was determined by 15% w/v sodium doceyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [19] and visualized with 0.1% (w/v) Coomassie Brilliant Blue R250 (BioRad). For immunoblotting, proteins were transferred from the gel onto a nitrocellulose membrane (Amersham Biosciences, Piscataway, NJ, USA) using a Semi-Dry Transblot (BioRad). The primary antibody was polyclonal rabbit anti-hIFN-$\gamma$ (PreproTech, Rocky Hill, NJ, USA) and the secondary antibody was goat anti-rabbit IgG conjugated to alkaline phosphatase (BioRad). Color was developed with $p$-nitro blue tetrazolium and sodium-5-bromo-4-chloro-3-indolyl-phosphate (NBT/BCIP, Amersham Biosciences). Analysis of nitrocellulose membranes was carried out using a photo-documenter Gel-Doc 2000 (BioRad) and the Quantity One™ v4.5 software (BioRad). Recombinant hIFN-$\gamma$ (PreproTech) was used as standard. The hIFN-$\gamma$ was quantified by an enzyme-linked immunosorbant assay (ELISA) kit (R&D Systems, Minneapolis, MN, USA).
RESULTS AND DISCUSSION

Design of the synthetic hIFN-γ gene

The native hIFN-γ gene has 16 codons reported as “rare”, which cause expression problems in *E. coli* [9,20,21]. These codons are: 1 AGG and 2 AGA for arginine; 2 ATA for isoleucine; 2 CCA for proline; 3 AGT, 2 TCA and 1 TCC for serine; GGG and GGA for glycine and 1 ACA for threonine. The rare codons were replaced for those reported with more usability by *E. coli* (Table 2). The resultant synthetic hIFN-γ gene consisted in 432 bp and had 74% homology with respect to the wild-type hIFN-γ gene sequence. Sixty-three other codons with low usability by *E. coli* were also modified in one or two basis to produce codons with more frequency of use. Codon bias is one of the most important factors in the expression of eukaryotic genes in prokaryotic systems [22,23]. Therefore, many authors have proposed the use of synthetic genes to increase translational efficiency [24–29]. The optimization of the synthetic hIFN-γ gene eliminated the presence of 16 rare codons, which cause translational errors [21,30]. However, the changes did not generate codons with maximum usage in *E. coli*. This can be explained because the full optimization of the gene could be detrimental for high protein expression, due to metabolic overload in the bacteria [10]. The functionality of the synthetic hIFN-γ gene was demonstrated by its expression in *E. coli* BL21-SI.

Cultures and hIFN-γ production

In the first experiments, the production of hIFN-γ was tested in triplicate in LBON at temperature of 37°C, 0.3 g/L of biomass and 0.3 M NaCl as inducer, as previously
suggested for the BL21-SI expression system [31]. The maximum biomass concentration was 2.96±0.02 g/L at 7 h of culture and the maximum concentration of hIFN-γ attained was 5.6±0.2 mg/L at 12 h (Fig. 1a). Interestingly, using the minimal medium under the same culture conditions, maximum hIFN-γ concentration was 6-times higher than the attained in LBON, but the maximum biomass concentration was only 0.88±0.02 g/L (Fig. 1b). To improve cell growth, we tested the use of the minimal medium plus 0.5% yeast extract (Fig. 1c). The biomass concentration increased to 1.39±0.02 g/L but only 12.9±0.3 mg/L of hIFN-γ was attained. These results led us to forego the use of yeast extract in subsequent experiments. Previous reports have shown that supplementation of culture media with yeast extract or aminoacids negatively affects the production of recombinant proteins [29,32,33].

**Optimization by RSM**

The production of hIFN-γ varied in the range of 8.1–78.0 mg/l for the treatments 15 and 7 (Table 3). The analysis of variance for the adjusted model showed that the production of hIFN-γ was significantly affected by the linear and quadratic coefficients and their interaction, except the linear term of the temperature (Table 4). The mathematical model representing the production of hIFN-γ as a function of the evaluated variables in the experimental region studied is expressed by the following equation:

\[
Y = -2242.6 + 134.417 X_1 + 303.222 X_2 + 572.537 X_3 + 1.52778 X_1 X_2 - 2.37037 X_1 X_3 + 76.6667 X_2 X_3 - 2.06111 X_1^2 - 598.437 X_2^2 - 880.556 X_3^2
\]  

(2)
where \( Y \) is the hIFN-\( \gamma \) concentration, and \( X_1, X_2 \) and \( X_3 \) are temperature, biomass concentration and NaCl concentration, respectively. The standard error was 1.04091 and the \( R^2 \) value was 0.999 (Table 4). These values indicate a good fit between the model and the experimental data indicating that the treatment was highly significant. The contour plots analysis showed that the central values of the factors were optimal for the production of hIFN-\( \gamma \) (Fig. 2). According to the mathematical model, the maximum hIFN-\( \gamma \) concentration of 77.3 mg/L is attained at temperature of 32.6\( ^\circ \)C, induction biomass of 0.31 g/L and 0.3 M NaCl (Fig. 2). To verify the predicted results, additional experiments were performed using these optimized conditions. The maximum biomass attained was 0.90\( \pm \)0.02 g/L and the hIFN-\( \gamma \) concentration at 14 h was 78.3\( \pm \)0.2 mg/L (Fig. 3). This result agrees with the predicted value by the model proposed. Culture conditions such as inducer concentration, induction time and temperature affect both the cell growth and the protein production [14,29,34–36]. Zhang et al. [37] produced 70 mg/L of hIFN-\( \gamma \) using the native hIFN-\( \gamma \) gene in \( E. \ coli \) and using a complex medium. Dobrovolsky et al. [38] reported a maximum production of 0.02 mg/L of hIFN-\( \gamma \) expressed in murine mammary glands. Khalilzadeh et al. [1] reported a maximum hIFN-\( \gamma \) yield of 0.35 g hIFN-\( \gamma \)/gDCW in fed-batch cultures using \( E. \ coli \) BL21(DE3), which is four-times higher than that attained presently (0.087 g hIFN-\( \gamma \)/gDCW). Thus, fed-batch cultures using the \( E. \ coli \) BL21-SI/pBAL are further recommended to increase the hIFN-\( \gamma \) yield.

**CONCLUSION**

The use of minimal medium is shown to be most suitable for hIFN-\( \gamma \) production. This production is affected by the culture conditions. The hIFN-\( \gamma \) concentration attained under
optimized conditions was 13-times higher than the obtained using the non-optimized conditions, corroborating the usefulness of RSM in the protein expression. RSM is, thus, an effective method for the optimization of recombinant protein expression using synthetic genes in *E. coli*.

ACKNOWLEDGMENTS

This work was supported by CONACyT grant J39639 and CONACyT-FOMIX grant FMSLP-2002-4100. Victor E. Balderas Hernández acknowledges a PhD fellowship number 172257 from CONACyT.

REFERENCES


mutagenesis of recombinant human phosphatidylcholine transfer protein using a
synthetic gene: evidence for a C-terminal membrane binding domain. *Biochem.* 39:
15399–15409.

27. Li, A., Z. Kato, H. Ohnishi, K. Hashimoto, E. Matsukuma, K. Omoya, Y.
Yamamoto and N. Kondo (2003) Optimized gene synthesis and high expression of

Sallberg (2004) Codon optimization and mRNA amplification effectively enhances
the immunogenicity of the hepatitis C virus nonstructural 3/4A gene. *Gene Ther.* 11:
522-533.

De León (2007) Optimization of culture conditions for a synthetic gene expression
in *Escherichia coli* using response surface methodology: the case of human


efficient overproduction of cloned gene products with NaCl as the inducer. *J

32. Ling, H. (2002) Physiology of *Escherichia coli* in batch and fed-batch cultures with
special emphasis on amino acid and glucose metabolism. Doctoral thesis.
Department of Biotechnology, Royal Institute of Technology, Stockholm, Sweden.


**Table 1.** Values of independent variables and the levels used in the experimental design.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Independent variable</th>
<th>Levels</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>Temperature</td>
<td>28 32.5 37</td>
<td>°C</td>
</tr>
<tr>
<td>$X_2$</td>
<td>Biomass</td>
<td>0.1 0.3 0.5</td>
<td>g/L</td>
</tr>
<tr>
<td>$X_3$</td>
<td>NaCl</td>
<td>0.15 0.3 0.45</td>
<td>M</td>
</tr>
</tbody>
</table>
Table 2. Rare codon contents in the hIFN-γ genes and their frequency of use by *E. coli*.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Native hIFNγ gene</th>
<th>Synthetic hIFNγ gene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Codon</td>
<td>Frequency per 1000 codons*</td>
</tr>
<tr>
<td>Arg</td>
<td>AGG</td>
<td>1.4</td>
</tr>
<tr>
<td>Arg</td>
<td>AGA</td>
<td>2.1</td>
</tr>
<tr>
<td>Ile</td>
<td>ATA</td>
<td>4.1</td>
</tr>
<tr>
<td>Pro</td>
<td>CCA</td>
<td>8.2</td>
</tr>
<tr>
<td>Ser</td>
<td>AGT</td>
<td>7.2</td>
</tr>
<tr>
<td>Ser</td>
<td>TCA</td>
<td>6.8</td>
</tr>
<tr>
<td>Ser</td>
<td>TCC</td>
<td>9.4</td>
</tr>
<tr>
<td>Gly</td>
<td>GGG</td>
<td>9.7</td>
</tr>
<tr>
<td>Gly</td>
<td>GGA</td>
<td>7.0</td>
</tr>
<tr>
<td>Thr</td>
<td>ACA</td>
<td>6.5</td>
</tr>
</tbody>
</table>

*A cut-off of < 1% was used to arbitrarily define rare codons.*
Table 3. Box-Behnken experimental design and results for the hIFN-γ production in minimal medium.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$X_1$ Temperature (°C)</th>
<th>$X_2$ Biomass (g/L)</th>
<th>$X_3$ NaCl (M)</th>
<th>hIFN-γ (mg/L)</th>
<th>Experimental Data</th>
<th>Predicted Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.5</td>
<td>0.1</td>
<td>0.15</td>
<td>32.9</td>
<td>33.9</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>37.0</td>
<td>0.3</td>
<td>0.15</td>
<td>20.1</td>
<td>19.6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>28.0</td>
<td>0.1</td>
<td>0.30</td>
<td>9.1</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>32.5</td>
<td>0.3</td>
<td>0.30</td>
<td>76.2</td>
<td>77.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>32.5</td>
<td>0.5</td>
<td>0.15</td>
<td>35.9</td>
<td>36.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>37.0</td>
<td>0.5</td>
<td>0.30</td>
<td>16.7</td>
<td>17.1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>32.5</td>
<td>0.3</td>
<td>0.30</td>
<td>78.0</td>
<td>77.2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>37.0</td>
<td>0.3</td>
<td>0.45</td>
<td>12.9</td>
<td>13.4</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>28.0</td>
<td>0.3</td>
<td>0.15</td>
<td>15.2</td>
<td>14.7</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>32.5</td>
<td>0.5</td>
<td>0.45</td>
<td>38.6</td>
<td>37.6</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>28.0</td>
<td>0.5</td>
<td>0.30</td>
<td>12.2</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>28.0</td>
<td>0.3</td>
<td>0.45</td>
<td>14.4</td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>32.5</td>
<td>0.1</td>
<td>0.45</td>
<td>26.4</td>
<td>26.3</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>32.5</td>
<td>0.3</td>
<td>0.30</td>
<td>77.4</td>
<td>77.2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>37.0</td>
<td>0.1</td>
<td>0.30</td>
<td>8.1</td>
<td>7.6</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Analysis of variance and coefficients for the adjusted model for the predicted hIFN-γ production in the minimal medium.

<table>
<thead>
<tr>
<th>Source</th>
<th>Polynomial coefficients</th>
<th>Sum of Squares</th>
<th>D.f.</th>
<th>Mean Square</th>
<th>F-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2242.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td>134.417</td>
<td>5.95125</td>
<td>1</td>
<td>5.95125</td>
<td>5.49</td>
<td>0.0661</td>
</tr>
<tr>
<td>$X_2$</td>
<td>303.222</td>
<td>90.4512</td>
<td>1</td>
<td>90.4512</td>
<td>83.58</td>
<td>0.0003</td>
</tr>
<tr>
<td>$X_3$</td>
<td>572.537</td>
<td>17.405</td>
<td>1</td>
<td>17.405</td>
<td>16.06</td>
<td>0.0102</td>
</tr>
<tr>
<td>$X_1^2$</td>
<td>-2.06111</td>
<td>6432.07</td>
<td>1</td>
<td>6432.07</td>
<td>5936.38</td>
<td>0.0000</td>
</tr>
<tr>
<td>$X_1X_2$</td>
<td>1.52778</td>
<td>7.5625</td>
<td>1</td>
<td>7.5625</td>
<td>6.98</td>
<td>0.0459</td>
</tr>
<tr>
<td>$X_1X_3$</td>
<td>-2.37037</td>
<td>10.24</td>
<td>1</td>
<td>10.24</td>
<td>9.45</td>
<td>0.0277</td>
</tr>
<tr>
<td>$X_2^2$</td>
<td>-598.437</td>
<td>2115.71</td>
<td>1</td>
<td>2115.71</td>
<td>1952.66</td>
<td>0.0000</td>
</tr>
<tr>
<td>$X_2X_3$</td>
<td>76.6667</td>
<td>21.16</td>
<td>1</td>
<td>21.16</td>
<td>19.53</td>
<td>0.0069</td>
</tr>
<tr>
<td>$X_3^2$</td>
<td>-880.556</td>
<td>1449.36</td>
<td>1</td>
<td>1449.36</td>
<td>1337.67</td>
<td>0.0000</td>
</tr>
<tr>
<td>Total error</td>
<td></td>
<td>5.4175</td>
<td>5</td>
<td>1.0835</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (corr.)</td>
<td></td>
<td>9039.59</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$X_1$: Temperature, $X_2$: Biomass conc., $X_3$: NaCl concentration, D.f.: degrees of freedom, $F$: Fisher test, $P$-value: probability distribution value. The correlation coefficient ($R^2$) was 0.999 and the standard error was 1.04091. Statistically significant at 95% of probability level.
Figure legends.

**Figure 1.** Batch cultures of *E. coli* BL21 SI/pBAL0 at non-optimized conditions in: a) LBON medium, b) minimal medium and c) in minimal medium plus 0.5% yeast extract. (○) biomass concentration (g/L) and (●) concentration of hIFN-γ (mg/L). Arrows indicate the induction time with 0.3 M NaCl. Data represent the average of three independent experiments.

**Figure 2.** Contour plot of hIFN-γ production by *E. coli* BL21 SI/pBAL0 in minimal medium: a) temperature vs biomass concentration at optimum NaCl concentration, b) temperature vs NaCl concentration at optimum constant biomass concentration and c) biomass concentration vs NaCl concentration at optimum constant temperature.

**Figure 3.** Batch culture *E. coli* BL21 SI/pBAL0, in minimal medium using optimized conditions, (○) biomass concentration (g/L) and (●) concentration of hIFN-γ (mg/L). Arrow indicates the induction time with 0.3 M NaCl. Data represent the average of three independent experiments.