A constraint-based modeling approach to reach an improved chemically defined minimal medium for recombinant antiEpEX-scFv production by Escherichia coli

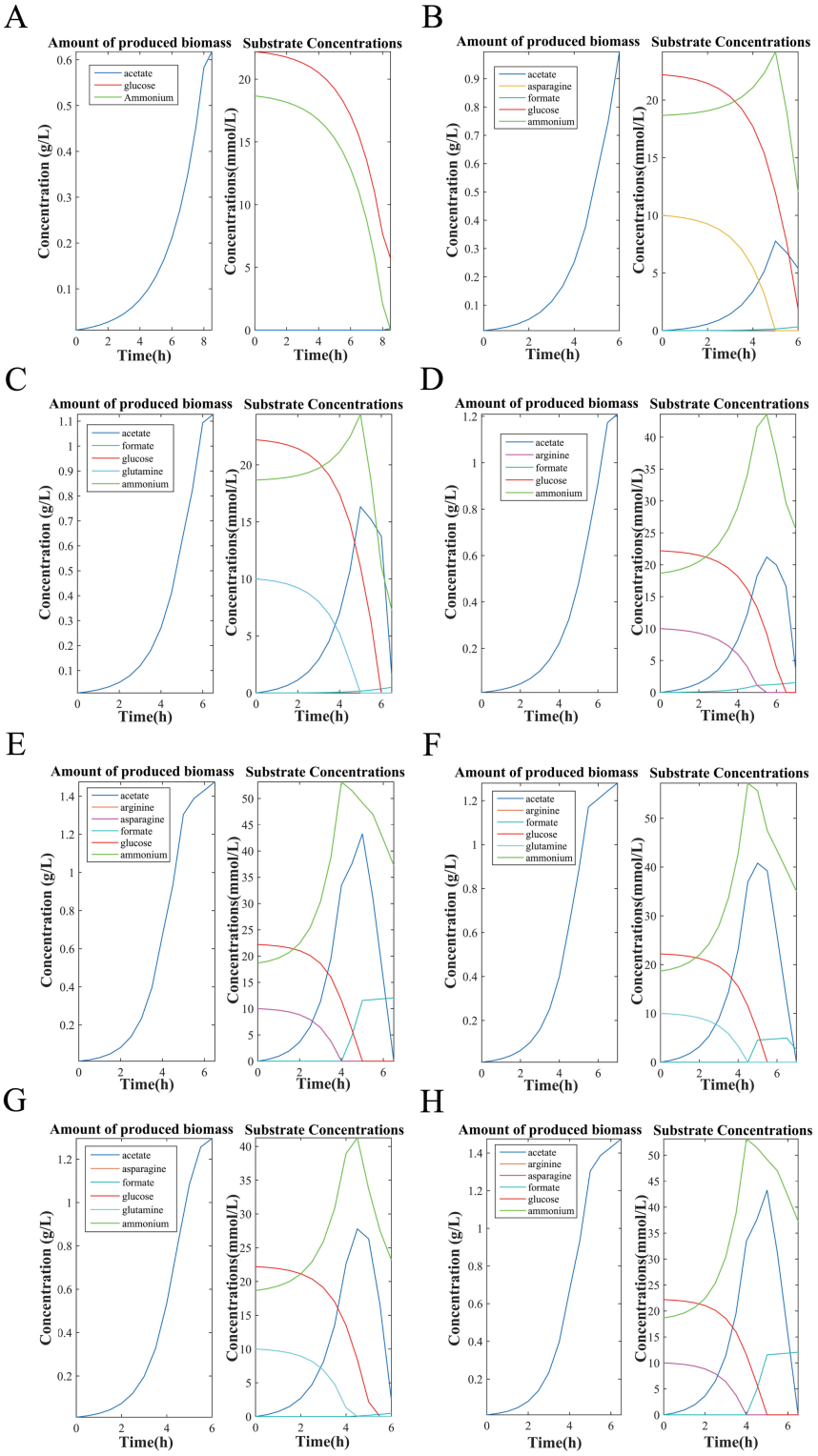
Aidin Behravan[1], atieh hashemi[1], and Sayed-Amir Marashi[2]

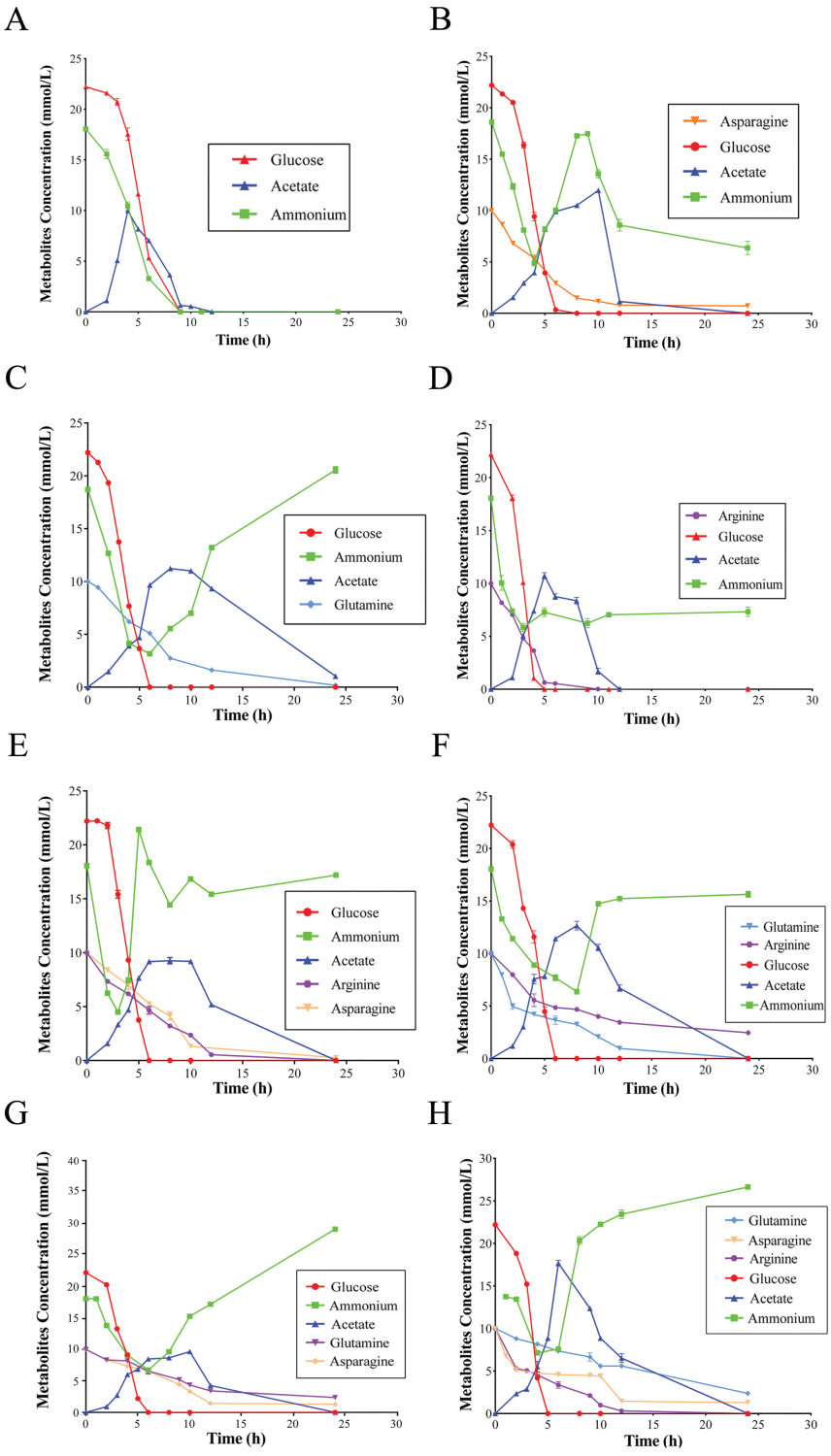
[1]Shahid Beheshti University of Medical Sciences
[2]University of Tehran College of Science

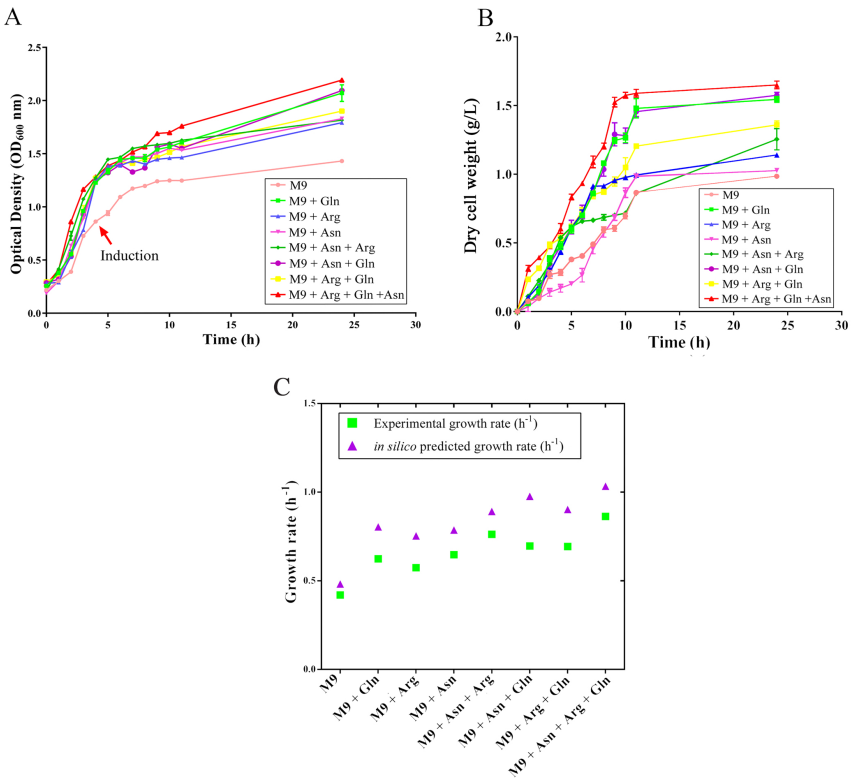
February 22, 2024

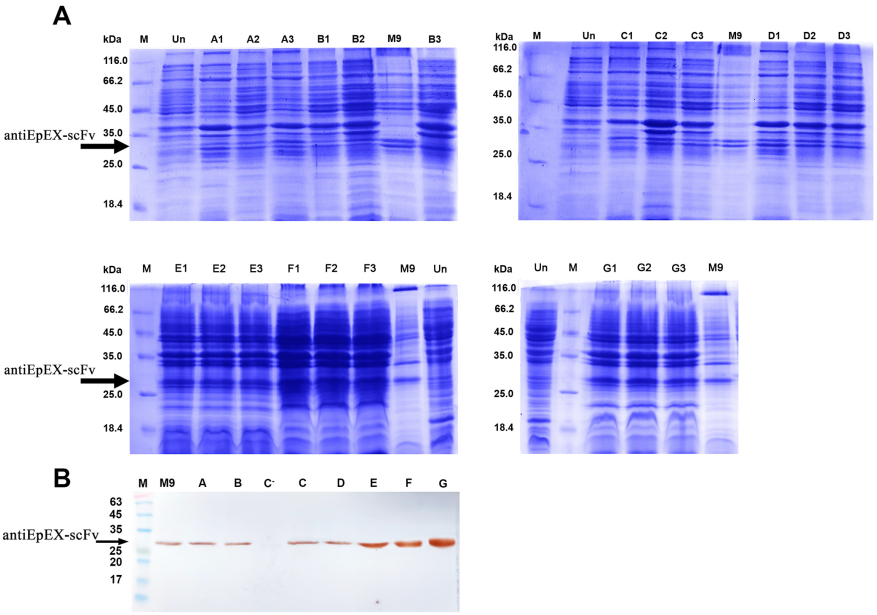
Abstract

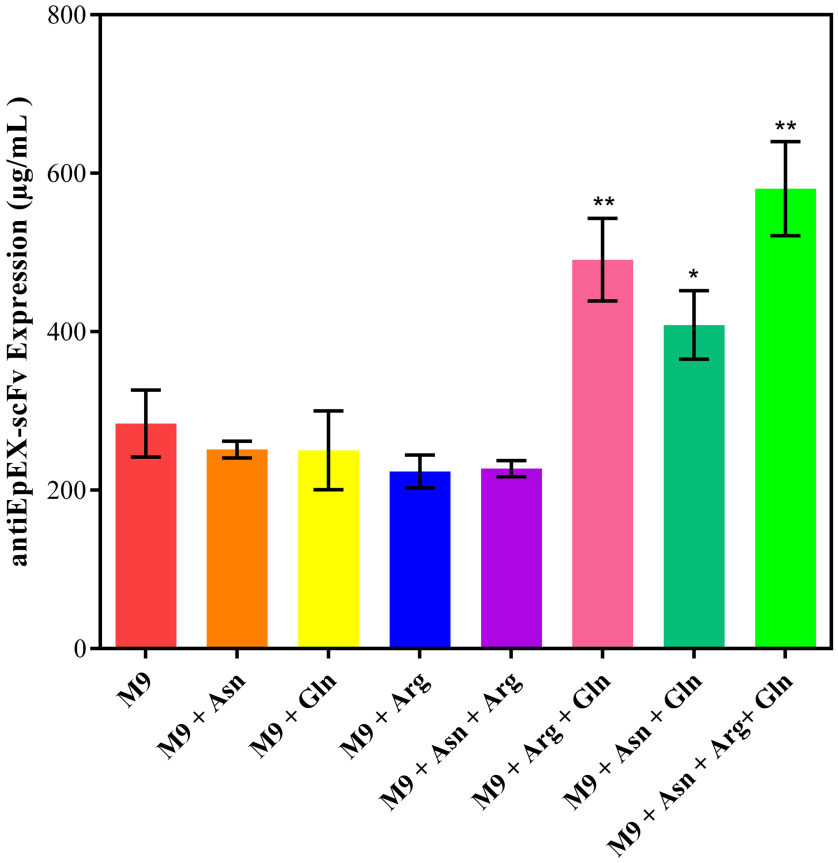
Increasing demand for recombinant therapeutic proteins highlights the necessity of their yield improvement. Culture medium formulation is a popular approach for bioprocess optimization to improve therapeutic protein production. Constraint-based modeling can empower high-precision optimization through information on how media compounds affect metabolism and cell growth. In the current study, a genome-scale metabolic model (GEMM) of Escherichia coli cells was employed to design strategies of minimal medium supplementation for higher antiEpEX-scFv production. Dynamic flux balance analysis of the recombinant E. coli cell model predicted that ammonium was depleted during the process. Based on the simulations, three amino acids (Asn, Gln and Arg) were chosen to be added to the medium to compensate for ammonium depletion. Experimental validation suggested that the addition of these amino acids (one-by-one, or in combinations) can indeed improve cell growth and recombinant protein production. Then, design of experiment was used to optimize the concentrations of amino acids in the growth medium. About two-fold increase in the growth rate and total scFv expression level was observed using this strategy. We conclude that the GEMM-based approach can provide insights into an effective feeding strategy to improve the production of recombinant protein in E. coli.



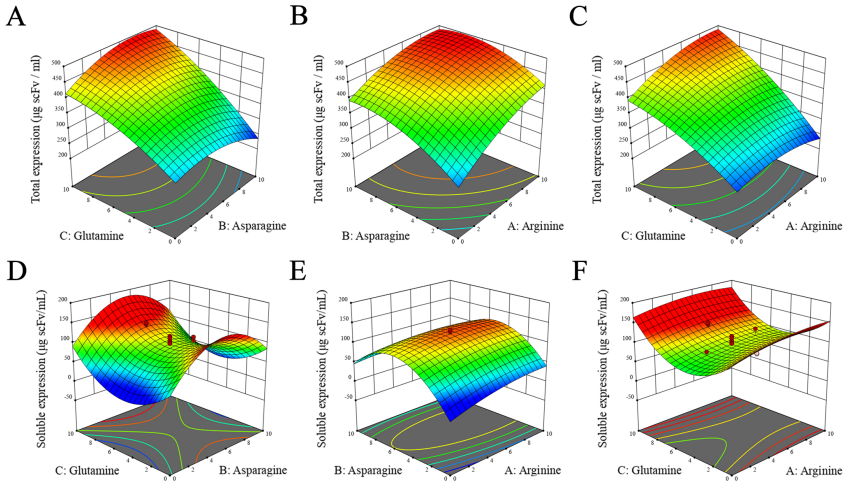


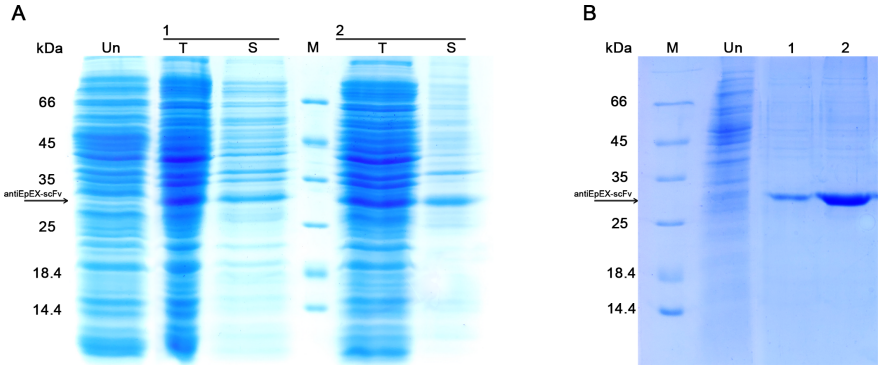






Experiments	Level of Arginine concentration	Level of Asparagine concentration	Level of Glutamine concentration
1	0	0	+α
2	0	0	0
3	-1	-1	1
4	0	+α	0
5	0	0	0
6	1	1	1
7	-1	-1	-1
8	0	0	0
9	-1	1	-1
10	0	-α	0
11	+α	0	0
12	-α	0	0
13	1	-1	-1
14	0	0	0
15	1	1	-1
16	1	-1	1
17	0	0	0
18	-1	1	1
19	0	0	-α
20	0	0	0





A constraint-based modeling approach to reach an improved chemically defined minimal medium for recombinant antiEpEX-scFv production by *Escherichia coli*

Running Title
GEMM-based strategy to improve scFv production

- Aidin Behravan

Department of Pharmaceutical Biotechnology, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran

- Atieh Hashemi *,

Department of Pharmaceutical Biotechnology, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran

- Sayed-Amir Marashi *

Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran

*Corresponding author:

- Atieh Hashemi* (https://orcid.org/0000-0001-7121-5306)

At_hashemi@sbmu.ac.ir

Address: ValiAsr Avenue, Niayesh Junction, PO Box 14155-6153, Tehran, Iran.
Tel/Fax: +98 21 88200067, +98 9123700760

- Sayed-Amir Marashi* (https://orcid.org/0000-0001-9801-7449)

marashi@ut.ac.ir

Address: Shafiei Str. 13, Qods Str., Enghelab Avenue, PO Box 1417614411, Tehran, Iran.
Tel/Fax: +98 21 66491622

Abstract

Increasing demand for recombinant therapeutic proteins highlights the necessity of their yield improvement. Culture medium formulation is a popular approach for bioprocess optimization to improve therapeutic protein production. Constraint-based modeling can empower high-precision optimization through information on how media compounds affect metabolism and cell growth. In the current study, a genome-scale metabolic model (GEMM) of *Escherichia coli* cells was employed to design strategies of minimal medium supplementation for higher antiEpEX-scFv production. Dynamic flux balance analysis of the recombinant *E. coli* cell model predicted that ammonium was depleted during the process. Based on the simulations, three amino acids (Asn, Gln and Arg) were chosen to be added to the medium to compensate for ammonium depletion. Experimental validation suggested that the addition of these amino acids (one-by-one, or in combinations) can indeed improve cell growth and recombinant protein production. Then, design of experiment was used to optimize the concentrations of amino acids in the growth medium. About two-fold increase in the growth rate and total scFv expression level was observed using this strategy. We conclude that the GEMM-based approach can provide insights into an effective feeding strategy to improve the production of recombinant protein in *E. coli*.

Keywords: antiEpEX-scFv, Constraint-based modeling, Dynamic FBA, DoE, Medium formulation

Introduction

Escherichia coli is the most popular prokaryotic expression system for industrial production of proteins. Its well-known genetics, ability to grow in inexpensive mineral media and availability of diverse cloning vectors makes *E. coli* a good choice for the overproduction of recombinant proteins (Terpe, 2006). Often, improvement of recombinant protein expression in *E. coli* is accomplished by trial and error strategies such as modification of expression vectors, host strains and media composition (Makino, Skretas, & Georgiou, 2011). Growth medium formulation is an essential part of the process development for improving recombinant protein production (Almo & Love, 2014). A variety of commercially available rich media exist which can be used for industrial protein production, but most of them are expensive. Additionally, such media contain undefined constituents, which potentially may lead to less reproducible results. Using chemically defined media makes it possible to follow and understand the metabolic processes during induction and protein expression (Rahmen et al., 2015) and increases batch-to-batch consistency and production reproducibility, which, in turn, makes the downstream processes simpler (van der Valk et al., 2010).

Despite their inherent limitations, design of experiments (DoE) methods are typically used for medium optimization (Galbraith, Bhatia, Liu, & Yoon, 2018; Savizi, Soudi, & Shojaosadati, 2019; Uhoraningoga, Kinsella, Henehan, & Ryan, 2018). On the other hand, it is well-established that metabolism can directly affect growth rate and the recombinant protein expression (Lozano Terol, Gallego-Jara, Sola Martínez, Cánovas Díaz, & De Diego Puente, 2019). Rational design of cell culture medium can redirect the metabolic fluxes toward those pathways which are required for recombinant protein overproduction (da Silva et al., 2017; Liu et al., 2016; Torres et al., 2019).

Genome-scale metabolic models (GEMMs) are useful tools for simulating flux distributions of metabolic pathways, (Sohn et al., 2010) as well as understanding the physiological behavior of a microorganism under different medium compositions. Therefore, GEMs can be used to predict the growth media in which the desired phenotype is expected (Fouladiha et al., 2020). Using accurate metabolic models, many experiments can be simulated virtually with a high level of accuracy (Bordbar, Monk, King, & Palsson, 2014). Constraint-based modeling can predict the demand of the cells for carbon and nitrogen sources to produce recombinant protein in *E. coli* (Meadows, Karnik, Lam, Forestell, & Snedecor, 2010; Zeng & Yang, 2019). In the present study, we used a GEM-based strategy for medium design to enhance the production of recombinant antiEpEX-scFv (EpCAM extracellular domain single-chain variable fragment) as a model protein. Here, we used dynamic flux balance analysis (dFBA) method (Mahadevan, Edwards, & Doyle, 2002) to make predictions about the nutritional requirements of *E. coli* BW25113 during fermentation in a chemically-defined growth medium. The results of this analysis were used to supplement the growth medium with the required nutrients, which increased cell growth rate and recombinant protein production. Then, DoE method was used for determining the best concentrations for nutrient supplementation in the medium. To our knowledge, this is the first study in which a genome-scale metabolic model of *E. coli* has been used for medium development to enhance recombinant protein production. Additionally, a culture medium with specified concentrations of components for optimal recombinant antiEpEX-scFv production is introduced.

Materials and Methods

Modeling of metabolism in recombinant *E. coli*

A GEMM of *Escherichia coli,* which will be referred to as *i*JO1366 (Orth et al., 2011), was used in this study. Modeling was accomplished using the COBRA Toolbox (Schellenberger et al., 2011) with glpk as the linear programming solver. For simulations, the metabolic reaction of antiEpEX-scFv production was added to *i*JO1366. The protein production reaction summarizes the amino acid composition of the recombinant protein, as follows:

19 Ala + 9 Arg + 9 Asn + 8 Asp + 5 Cys + 15 Gln + 7 Glu + 28 Gly + 9 His + 10 Ile + 19 Leu + 10 Lys + 5 Met + 9 Phe + 12 Pro + 35 Ser + 24 Thr + 5 Trp + 14 Tyr + 12 Val + 1136.78 H_2O + 1136.78 ATP => 1136.78 ADP + 1136.78 H^+ + 1136.78 P_i + antiEpEX-scFv

The reactions representing plasmid were added to the network separately. These reactions include the recombinant DNA sequence which produce pETDuet-antiEpEX-scFv plasmid and the amino acids sequence of antibiotic marker for the plasmid load on the host cell. The lower bound of plasmid and antibiotic marker was set to 0.00063 mmol gDW^{-1} h^{-1} and 0.000569 mmol gDW^{-1} h^{-1}, respectively (Ow, Lee, Yap, & Oh, 2009). M9 minimal medium composition was simulated in the model such that the cell could freely uptake all the minerals from the medium. For oxygen uptake the lower bound of -20 mmol gDW^{-1} h^{-1} was considered, which represented the maximum oxygen uptake rate for *E. coli* when growing aerobically (Varma, Boesch, & Palsson, 1993).

Flux balance analysis (FBA) is being widely used for the problem of finding the optimal metabolic state of a cell subject to certain metabolic constraints (Orth, Thiele, & Palsson, 2010). FBA can be written as a linear programming (LP) problem:

$$\begin{array}{lll} \text{maximize} & v_{obj} & \\ \text{Subject to :} & \mathbf{S.v = 0} & \\ & v_{i \geq} 0 & \text{for every irreversible reaction } i \\ & \mathbf{v}^{lb} \leq \mathbf{v} \leq \mathbf{v}^{ub} & \end{array}$$

Where v_{obj} is the objective of the metabolic system.

Dynamic flux balance analysis (dFBA) is a related approach to predict time-resolved metabolic profiles for a growing cell (Mahadevan et al., 2002). In the present study, dFBA was employed to predict the amount of biomass, substrate utilization and by-product secretion during the experiment. The step size of the dFBA was set to 0.5 h, and an FBA was performed at each time point to obtain time-course data of the metabolite concentrations and fluxes. A linear programming problem is solved at the beginning of the first interval and the steady-state flux distribution is determined through the maximization of the bacterial growth rate. Recombinant protein production was simulated by finding the maximum flux through the protein production reaction under experimental conditions using FBA, and then, setting the lower bound of protein production to half of its maximum possible value. The initial glucose and ammonium concentrations were set to 22.2 mM and 18.7 mM, respectively. To investigate the effect of supplementation on the results of dFBA simulations, the initial

concentration of each amino acid was set to 10 mM (Bren et al., 2016), and then, dFBA simulation was repeated.

Media and Bacterial strains

E. coli BW25113 (DE3) strain containing the pETDuet-antiEpEX-scFv plasmid (was previously constructed in our lab) was used in this study (Behravan & Hashemi, 2021). M9 medium (containing KH_2PO_4 3 g/L, Na2HPO4 6 g/L, NaCl 0.5 g/L, $MgSO_4$ 2 mM, $CaCl_2$ 0.1 mM, NH_4Cl 1 g/L, 4 g/L glucose, $FeCl_3$ 0.01 mM, 0.1 mL 1000x trace element) was purchased from Teknova (USA). Amino acids were purchased from Sigma-Aldrich (Germany). All other chemicals with analytical grade were provided from Merck.

Experimental growth conditions and Expression of antiEpEX-scFv

The glycerol stock of bacteria was streaked on Luria-Bertani (LB) agar plate and incubated for 18 h at 37 ^{o}C. To prepare preculture, a single clone of overnight incubated agar plate was placed into the sterilized LB medium (10 g/L of tryptone, 5 g/L of yeast extract, and 10 g/L of NaCl) and was incubated for 18 h at 37 ^{o}C in 200 rpm. Thee preculture was harvested by centrifugation (6000×g for 5 min at 4 °C), and the pellet was washed twice with phosphate buffered saline (PBS). Then, the cell pellet was resuspended either in M9 minimal or in M9 medium supplemented with certain amino acids. In the next step, the culture was incubated at 37 ^{o}C and 200 rpm until its optical density at 600 nm (OD_{600}) reached 0.8, and then the cells were induced by adding 0.8 mM of IPTG for 24 hours. Finally, the bacterial cells were harvested through centrifugation (10000×g for 10 min at 4 °C). For initial determination of protein expression, the cell pellets were suspended in the lysis buffer (150 mM NaCl, 1 mM EDTA, 50 mM Tris pH 7.5, 1 mg/ml lysozyme, 1% triton X100) and sonicated for 30 min (20s ON, 10s OFF at 400 W). After centrifugation of the total cell lysate (10000×g for 30 min at 4 °C), protein samples were electrophoresed on a 15% SDS-PAGE gel and visualized by Coomassie Brilliant Blue G-250 Dye. AntiEpEX-scFv expression was confirmed using western blot analysis. The proteins were electro-transferred from the gel into the polyvinylidene difluoride (PVDF). The transferred membrane was blocked in 5% nonfat milk for 1 hour, and then, was washed three times with TBST followed by incubation with His-tag antibody (Sigma, UK) overnight. After washing again, the membrane was incubated in anti-mouse HRP conjugated immunoglobulin, (Sigma, UK) as a secondary antibody for two hours, and then, detected using a solution of 3,3′-diaminobenzidine (DAB) (Sigma, UK).

Analytical methods

Cell growth profiles in different media were determined based on optical density measurements, *i.e.*, by measuring the absorbance of the cell culture every 1 h at 600 nm using a spectrophotometer (E-Chrome Tech, Taiwan). Growth rate was calculated in each case by the logarithmic derivative of the optical density curve at the mid-log phase. For determining the amount of biomass in the culture, 10 mL of culture broth was collected in a pre-weighed microtube and centrifuged (in 10000×g for 10 min at 4 ^{o}C). Then the pellet was heated in oven at 150 °C for 24 h, and the dried biomass was cooled for 1 h, and finally weighed. To ensure the repeatability of the method, the experiment was done in three replicates for each of the 12 sampling points. For determining the

consumption of ammonium, glucose, amino acids, and measuring the concentration of secreted acetate, at each time point, 1 mL aliquots of culture broth was centrifuged (10,000×g for 5 min at 4 °C), and the supernatant was collected and analyzed. The above-mentioned metabolites were analyzed enzymatically using appropriate kits (Megazyme, Ireland), according to manufacturer's protocols.

Optimization of amino acid concentrations

Total protein production, as well as soluble protein production can be considered as the measures of recombinant protein production in biotechnology. For determining the best concentration of each amino acid that should be added to M9 minimal medium (*i.e.*, for achieving maximum recombinant protein production) Central Composite Design (CCD) was performed using Design-Expert software version 12 (Stat-Ease Inc., USA). Tweny experiments were designed with three factors in five level using CCD. All the experiments were performed twice and two responses were considered: The first response was the protein expression level in the soluble form, and the second one was the total expression level of antiEpEX-scFv. In order to validate the models presented in the result section, the optimum concentration of amino acids which leads to the maximum predicted level of antiEpEX-scFv (as soluble and total expression) was experimentally examined.

Preparation and quantification of antiEpEX-scFv

In order to investigate the effect of amino acid supplementation on recombinant protein production, after centrifugation (10000×g for 10 min at 4 °C), the cell pellets were resuspended in lysis buffer (1% triton X100, 50 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl, 1 mg/mL lysozyme). The cell lysate was then disrupted with sonication on ice (400 W, 20 s ON, 10 s OFF for 30 min). Sonicated samples were centrifuged (14000×g) for 30 min at 4 °C. After centrifugation (10000×g) of the harvested cell pellets for 10 min at 4 °C followed by resuspention in solution containing 20 mM Tris pH 7.5, 50 mM NaCl, 50% glycerol and 1 mg/mL lysozyme, the soluble fraction of protein was separated. Then, after 40 min of incubation on ice, the cells were disrupted by sonication on ice (400 W, 20 s ON, 10 s OFF for 30 min), and centrifuged at 10000×g for 30 min at 4 °C. Then, the supernatant was collected as the soluble fraction followed by centrifugation in 14000×g for 30 min at 4 °C. In order to determine the concentration of recombinant protein in each sample, concentrations of total lysate and supernatant were determined separately using bicinchoninic acid (BCA) protein assay kit (Takara, Japan). Then, samples were electrophoresed on a 15% SDS-PAGE gel and the intensity ratio of recombinant protein to all the proteins existing in the sample was calculated using ImageJ software (Schneider, Rasband, & Eliceiri, 2012). Finally, by multiplying the obtained ratio to the protein concentration, recombinant protein concentration was calculated for each sample (each experiment was repeated twice).

Purification of the recombinant antiEpEX-scFv

For obtaining purified antiEpEX-scFv, bacterial cells were harvested and resuspended in the lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mg/mL lysozyme, 1% triton X100). The container was vortexed and sonicated for 30 min and then centrifuged at 10000×g for 30 min at 4 °C. The pellet was resuspended in denaturation buffer (1% triton X100, Tris 50 mM, 8 M urea, NaCl 50 mM, pH 8) and loaded to the Ni-nitrilotriacetic acid (NTA)-agarose affinity chromatography column under denaturing conditions

according to the manufacturer's instructions (Qiagen, Netherlands). The Ni–NTA column was washed using wash buffer (50 mM NaH_2PO_4, 8 M urea, 300 mM NaCl, 20 mM imidazole), and then the antiEpEX-scFv was eluted by elution buffer (50 mM NaH_2PO_4, 8 M urea, 300 mM NaCl, 250 mM imidazole) from the column. The purified protein concentration was measured by BCA protein assay kit (Smith et al., 1985).

Results

Metabolic Modeling for Designed Medium Composition

Here, we used a metabolic network of *E. coli, i*JO1366*,* for constraint-based modeling of metabolic fluxes in the bacterium. To achieve this goal, dFBA was employed to simulate the growth of the recombinant host cells, and to design a strategy for supplementation of M9 minimal medium in order for increasing antiEpEX-scFv production. The maximum protein production rate was calculated by using FBA, and then, half of the calculated maximal possible value was set as the lower bound of antiEpEX-scFv production rate as a constraint. In addition, biomass production rate was set as the objective function. Glucose and ammonium were set as substrates with their initial concentration which was set based on M9 minimal medium. Then, dFBA was used to calculate the substrate uptake, by-product secretion and biomass production rates over time. The result of dFBA suggested that ammonium was the first metabolite to be consumed. Nitrogen starvation leads to bacterial growth restriction, and consequently, protein production limitation (Figure 1A). Hence, the addition of amino acids to the medium can provide the necessary amount of nitrogen during fermentation, and consequently, improve protein production. Amino acids are the building blocks of proteins, and additionally, they can act as alternative sources of nitrogen. We simulated the addition of each of the standard amino acids to the medium to see if it can overcome the ammonium depletion. In particular, Asn, Gln and Arg were predicted to be able to prevent the predicted ammonium depletion (Figure 1). Base on this study, addition of certain amino acids (namely His, Ile, Met, Phe, Tyr, Leu, and Lys) were found not to be used by the cells. The remaining ten amino acids were predicted to be consumed, although could not prevet the predicted ammonium depletion trend (supplementary figure S1).

Experimental validation of the computational results

To validate the dFBA predictions, seven media (*i.e.*, the minimal medium with any subset of the three amino acids) were formulated and the metabolic profile of recombinant *E. coli* in each medium was investigated. During fermentation, biomass accumulation, recombinant protein production and concentration of amino acids (Asn, Gln, Arg), acetate, ammonium, glucose were measured. To this end, aliquots of supernatants at ten sequential time steps over the growth of the bacteria were collected from each culture to measure the concentration of metabolites in medium. The results of this analysis are represented in Figure 2. Ammonium depletion in M9 containing NH_4^+ as the sole source of nitrogen was observed during log phase of the growth as was predicted by the model (Figure 2A). Altogether, the results suggest that dFBA is able to successfully predict the metabolic profile of the recombinant *E. coli* cells, including the ammonium production in the supplemented media. Figure 2 suggests that glucose, as the main carbon source, was consumed at the early log phase of the bacterial growth. As the fermentation process continues, the cells may utilize other carbon sources in the medium. During the exponential growth phase, acetate was accumulated in the medium, while after the complete consumption of glucose, acetate was being consumed as a carbon source. Moreover, as shown in Table 2, during the fermentation, the acetate concentration remained lower than the inhibitory threshold (5 g/l, or equivalently, 83.26 mmol/L) for all supplemented media, and hence, it had no inhibitory effect on growth and

recombinant protein production. Additionally, Figure 2 suggests that the availability of free amino acids in the medium enables the cells to consume them efficiently and synthesize cellular materials leading to fast growth (Figure 3). The excess nitrogen, that is, the by-product of amino acid degradation pathways, was excreted as ammonium to the culture medium, which would be consumed by cells later (when free amino acids are not present in the medium). Based on the data we have obtained, when glutamine is present in the medium, the cell can produce more ammonium (22.5 ± 0.012 mmol/L) comparing to addition of other amino acids.

To determine the growth profile and the growth rates in different culture conditions, OD_{600} was measured every 1 hour. The growth rate was calculated for each condition as the derivative of the logarithmic OD curve at the mid-exponential phase. Different growth rates were observed for different culture media. The growth rate in the M9 minimal medium was the lowest value due to ammonium depletion during batch fermentation in shaking flask. On the other hand, the highest growth rate value was obtained for the medium supplemented with all three amino acids (0.878 ± 0.015 h^{-1}), with ~120% increase compared to the minimal medium (0.4 ± 0.025 h^{-1}) (Figure 3). From Figure 3, one can observe that, consistent with the prediction results, addition of certain amino acids to the medium will improve the growth profile of *E. coli*. Due to improved cell growth, we expected that protein production would also increase. SDS-PAGE was used to investigate recombinant antiEpEX-scFv expression (Figure 4A). Using western blot analysis and antibody against 6xHis-tag, the expression of scFv protein was confirmed (Figure. 4B). The amounts of antiEpEX-scFv expression level using SDS-PAGE are presented in Figure 5. By comparing the expression levels among all media, a similar trend was observed for the M9 medium and the media with one additional amino acid (Arg, Asn, or Gln). In contrast, those media which include a combination of the amino acids show different trends. Although in M9+Arg+Asn no significant improvement in recombinant protein expression was observed compared to that in the M9 minimal medium, the expression level of recombinant protein in M9+Asn+Gln (408 ± 5.42 μg/mL, or equivalently, 0.244 g/g DCW) and M9+Arg+Gln (488 ± 4.98 μg/mL, or equivalently, 0.297 g/g DCW) showed 1.4 and 1.7 fold increase compared to the expression level in M9 minimal medium (283 ± 7.32 μg/mL, or equivalently, 0.188 g/g DCW), respectively. Finally, M9+Asn+Arg+Gln achieved the highest level of recombinant protein expression (588 ± 8.6 μg/mL, or equivalently, 0.355 g/g DCW), by about two-fold compared to the M9 minimal medium.

Determining the optimal concentration of amino acids by central composite design

After studying the effect of supplementing amino acids to M9 medium on the expression level of the recombinant protein, central composite design (CCD) was used to find the optimum concentration for the three amino acids which maximizes the total (and the soluble) antiEpEX-scFv expression level. Consequently, five levels of concentration for each amino acid were tested in 20 experiments. Total and soluble antiEpEX-scFv expression levels were examined as the result of the experiments. In order to improve reliability of the results, the designed experiments were performed in two repetitions, and the averaged values are presented in Figure 6. In these experiments, a wide range of protein production values was observed, *i.e.*, 250-480 μg/mL for the total protein, and 26-130 μg/mL for soluble protein.

Quadratic regression models were developed to obtain maximum total antiEpEX-scFv expression level as well as solubility, by means of Design Expert software. Moreover, analysis of variance (ANOVA) was used to confirm the models. ANOVA values for total protein expression (μg/mL) and solubility are presented in Supplementary S2 table 1 and 2 respectively. The *F*-value of the developed model for protein solubility

optimization (13.59) proves the model's significance. Also, ANOVA results showed that the developed model for total antiEpEX-scFv expression level was significant (where *F*-value = 95.58). Mathematical analysis revealed that lack of fit values of the models (P=0.0745 for total protein expression response and P=0.059 for solubility response) were not significant which means that the models were selected properly. Perfect correlation between the values predicted by RSM models and experimental data was achieved in both models ($R^2 = 0.9663$ for total protein expression optimization model and $R^2 = 0.8030$ for antiEpEX-scFv solubility optimization model). Furthermore, adjusted R^2 in both mentioned models were in reasonable coincidence with predicted R^2. As shown in Supplementary S2 figure 1a and 2a predicted and experimental values are in good agreement for both responses studied here. Also, the linear trend of data points in normal probability plots of the Studentized residuals confirmed the significance and adequacy of the proposed models (Supplementary S2 figure 1b, 2b). As a result, these models can be utilized to navigate the design space. Except AC, BC and A2, all the quadratic (B2 and C2), linear (A, B and C), and interactive terms (AB), were significant for solubility response based on P-values of the model terms. All the linear and interactive terms of the model developed for total expression were significant in contrary with quadratic terms which were not significant. Furthermore, it was concluded that glutamine concentration had the largest effect on both responses. The interactive effects between independent variables on two responses were studied using three-dimensional response surface graphs. Based on Figures 7A-C, simultaneous increase of amino acid concentrations leads to significant improvement of total protein expression as illustrated by ANOVA results (sup2 table 1). On the other hand, while increasing arginine and asparagine alone does not have a notable effect on the total protein expression (Figures 7A-C), high concentrations of glutamine and moderate concentrations of asparagine have a great positive effect on the solubility of the protein (Figure 7D). Figure 7E suggests that maximum solubility can be achieved while both amino acid concentrations are in the midrange (about 6mM). As shown in Figure 7F low and high concentrations of the glutamine in the presence of arginine, result in more solubility while arginine concentration variation has a negligible effect on the results. The optimum concentrations of glutamine, arginine and asparagine suggested by the models to achieve maximum antiEpEX-scFv levels were 9.9, 9.5 and 6.1 mM respectively.

Experimental validation of CCD-RSM models

According to experimental validation of the quadratic models, the predicted and the experimental values are in good agreement for both responses. The maximum predicted value for the soluble fraction (177.9 μg/mL) was in good agreement with the experimentally obtained value (206.1 ± 7.3 μg/mL). Moreover, validation of the total expression model contributes to total protein expression of 445.1 ± 6.54 μg/mL, which is close to the predicted production of 493.2 μg/mL. As represented in figure 8A, the experiments were done in duplicates for validation of the quadratic models. So, based on the results obtained here, the accuracy of the CCD models was fully confirmed. Finally, recombinant antiEpEX-scFv protein expressed in the basic M9 medium and the developed one (with optimized concentrations of amino acids based on CCD-RSM model) was purified using Ni-NTA affinity chromatography column. BCA analysis revealed that antiEpEx-scFv was overexpressed in model-based optimized culture medium by 2.27 fold (355.5 ± 8.94 μg/mL, or equivalently, 0.236 g/g DCW μg/mL) in comparison to the basic M9 minimal medium (156.4 ± 6.52 μg/mL, or equivalently, 0.126 g/g DCW) μg/mL) (Figure 8B).

Discussion

In the present study, constraint-based modeling was used as a novel approach to develop a defined minimal medium for the improvement of the antiEpEX-scFv production by recombinant *E. coli*. Here, to our knowledge for the first time, *i*JO1366, the GSMM of *E. coli,* was utilized for designing the supplementation strategy. We employed the dFBA method to simulate the evolution of metabolite concentrations in the medium over time. Using the model enables us to describe accurately cellular metabolic pathways, such as amino acids consumption, utilization of glucose and acetate, and overflow of by-products like acetate, ammonium and formate. In the next step, predictions of the model were experimentally validated.

Metabolic modeling results suggested that ammonium was the limiting nutrient during the growth of bacteria in the M9 medium. It was experimentally confirmed that the ammonium concentration decreased over time. In order to supply nitrogen, the effects of different amino acids on growth rate as well as recombinant protein production were evaluated with dFBA. According to previously published data, in addition to nitrogen supply, amino acid supplementation can surpass the metabolic burden imposed by recombinant protein production (Kumar, Chauhan, Shah, Gupta, & Rathore, 2020). During the production of recombinant proteins in *E. coli*, excessive utilization of amino acids as indispensable precursor molecules for protein synthesis on the one hand, and on the other hand down-regulation of genes effective in amino acid synthesis such as *ppc* (encoding phosphoenolpyruvate carboxylase) and *glyA* (encoding serine hydroxymethyltransferase), result in amino acid starvation. Under the starvation conditions, growth-associated genes are downregulated to help the expression of genes in charge of cell maintenance and survival. Moreover, amino acid scarcity during the translation process leads to an immediate halt of protein synthesis. These alterations of cellular metabolism is a major contributor to the metabolic burden induced on the starved cells which can affect cell growth and recombinant protein production .(Sarkandy et al., 2010) Here, modeling results suggested that the addition of certain amino acids to the medium could increase the available ammonium in the cell culture and compensate for the ammonium deficiency. These results were experimentally validated and about a two-fold increase in antiEpEx-scFv expression was obtained. Consistent with our findings, the effect of adding amino acids to the culture medium on cell growth and production of recombinant protein has been investigated in various studies. Kumar et al. showed that the addition of amino acids to the cell culture medium led to a decrease in metabolic burden, and improved recombinant protein expression and cell growth rate (Kumar et al., 2020). Also, a previous study on *S. cerevisiae* has shown that supplementation of certain amino acids to the defined medium positively influences xylanase production and biomass formation (Görgens, van Zyl, Knoetze, & Hahn-Hägerdal, 2005).

In our study, when glutamine, arginine or asparagine was added to the M9 minimal medium, ammonium was produced due to amino acid metabolism in the cell and the excess amount of ammonium was secreted and accumulated in the culture medium. The effect of metabolism of these amino acids on ammonium secretion and energy metabolism has been investigated previously (Bren et al., 2016; Maser, Peebo, Vilu, & Nahku, 2020). Moreover, here, the quadratic model developed to find the optimum concentration of the three amino acids to maximizes the total (and the soluble) antiEpEX-scFv expression level showed that glutamine concentration had the largest effect on both responses. Furthermore, experimental results showed more ammonium production in the medium supplemented by glutamine compared to those supplemented by other amino acids. Our results are consistent with the previous studies in which glutamine is suggested to be the most effective supplement for

increasing recombinant protein production in *E. coli*, as it can be consumed as a source of carbon and nitrogen for this bacterium (Heyland, Fu, Blank, & Schmid, 2011) and plays a main role in reducing the metabolic burden caused by the production of recombinant proteins (Bren et al., 2016; Pérez-Rodriguez, Ramírez-Lira, Trujillo-Roldán, & Valdez-Cruz, 2020). According to figure 3 the most boosting effect on growth of recombinant *E. coli* BW25113 was observed by glutamine supplementation (2.276 ± 0.024 at 24 h) compared to arginine or asparagine addition (1.794 ± 0.013at 24 h or 1.821 ± 0.02 at 24 h). In agreement with our study, glutamine was one of the amino acids which was proved effective for increasing the growth and IL-2 expression in batch and fed-batch high cell-density cultures in Yegane-Sarkandy et al. study. This observation may be due to the downregulation of *acn* (aconitase) and *gltD* (glutamatesynthase) genes during recombinant protein production in *E. coli* (Sarkandy et al., 2010). Synthesis of glutamate, glutamine and α-oxoglutarate (α-ketoglutarate) are related to these genes. Therefore, down-regulation of these genes can decrease *E. coli* growth rate. Furthermore, arginine can increase the available energy of the cell by increasing the TCA flux (Bren et al., 2016). A recent study has shown that glutamine and arginine are the two most-consumed amino acids in *E. coli* because their metabolism is associated with the glycolysis pathway and the TCA cycle. Furthermore, bacteria also use these amino acids to generate energy or uses them as substrates in anabolic reactions (Maser et al., 2020).

In conclusion, to optimize the fermentation process with the aim of increasing the production of recombinant antiEpEX-scFv protein, we exploited genome-scale metabolic modeling. More precisely, dynamic flux balance analysis was used for predicting the changes in the culture media during the growth, and then, model predictions were experimentally validated. Based on the simulations and experimental validation, we found that ammonium depletion during the fermentation could be compensated for by addition of certain amino acids, namely Asn, Gln and Arg. In the next step, we used DoE method to determine optimum concentration of each amino acid to be added to the minimal medium in order to achieve maximal scFv production. We showed that supplementation with amino acids was able to compensate for ammonium depletion and could increase the cell growth and protein production to more than two fold. The results indicate that determination of the exact metabolic limitations and overcoming them can be successfully used for developing feeding strategies for improved protein production in minimal medium. Our study, however, pave the way for further investigations and suggest that the genome scale metabolic model can provide a valid and fast estimation of essential amino acids necessary to improve the productivity of recombinant protein in different cell factories.

Acknowledgments

This work was supported by the research deputy of Shahid Beheshti University of Medical Sciences in Tehran, Iran (IR.SBMU.PHARMACY.REC.1398.096).

Additional information

Supplementary file number 1 includes figure S1 that represents dFBA results for simulating the effect of amino acid supplementation on ammonium depletion. Supplementary file number 2 includes details of DoE analyses.

References:

Almo, S. C., & Love, J. D. (2014). Better and faster: improvements and optimization for mammalian recombinant protein production. *Current Opinion in Structural Biology*, *26*, 39–43. doi:10.1016/j.sbi.2014.03.006

Behravan, A., & Hashemi, A. (2021). Statistical optimization of culture conditions for expression of recombinant humanized anti-EpCAM single-chain antibody using response surface methodology. *Research in Pharmaceutical Sciences*, *16*, 153–164. doi:10.4103/1735-5362.310522

Bordbar, A., Monk, J. M., King, Z. A., & Palsson, B. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*, *15*, 107–120. doi: 10.1038/nrg3643

Bren, A., Park, J. O., Towbin, B. D., Dekel, E., Rabinowitz, J. D., & Alon, U. (2016). Glucose becomes one of the worst carbon sources for E.coli on poor nitrogen sources due to suboptimal levels of cAMP. *Scientific Reports*, *6*(1), 24834. doi:10.1038/srep24834

da Silva, M. R., Portela, C. A. F., Albani, S. M. F., de Paiva, P. R., Tanizaki, M. M., & Zangirolami, T. C. (2017). Experimental design and metabolic flux analysis tools to optimize industrially relevant Haemophilus influenzae type b growth medium. *Biotechnology Progress*, *33*(6), 1508–1519. doi: 10.1002/ btpr.2546

Fouladiha, H., Marashi, S. A., Torkashvand, F., Mahboudi, F., Lewis, N. E., & Vaziri, B. (2020). A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. *Bioprocess and Biosystems Engineering*, *43*(8), 1381–1389. doi:10.1007/s00449-020-02332-6

Galbraith, S. C., Bhatia, H., Liu, H., & Yoon, S. (2018). Media formulation optimization: current and future opportunities. *Current Opinion in Chemical Engineering*, *22*, 42–47. doi:10.1016/j.coche.2018.08.004

Görgens, J. F., van Zyl, W. H., Knoetze, J. H., & Hahn-Hägerdal, B. (2005). Amino acid supplementation improves heterologous protein production by Saccharomyces cerevisiae in defined medium. *Applied Microbiology and Biotechnology*, *67*(5), 684–691. doi:10.1007/s00253-004-1803-3

Heyland, J., Fu, J., Blank, L. M., & Schmid, A. (2011). Carbon metabolism limits recombinant protein production in Pichia pastoris. *Biotechnology and Bioengineering*, *108*(8), 1942–1953. doi:10.1002/bit.23114

Kumar, J., Chauhan, A. S., Shah, R. L., Gupta, J. A., & Rathore, A. S. (2020). Amino acid supplementation for enhancing recombinant protein production in E. coli. *Biotechnology and Bioengineering*, *117*(8), 2420–2433. doi:10.1002/bit.27371

Liu, P., Huang, M., Guo, M., Qian, J., Lin, W., Chu, J., … Zhang, S. (2016). Combined 13C-assisted metabolomics and metabolic flux analysis reveals the impacts of glutamate on the central metabolism of high β-galactosidase-producing Pichia pastoris. *Bioresources and Bioprocessing*, *3*(1), 47. doi:10.1186/s40643-016-0124-6

Lozano Terol, G., Gallego-Jara, J., Sola Martínez, R. A., Cánovas Díaz, M., & De Diego Puente, T. (2019).

Engineering protein production by rationally choosing a carbon and nitrogen source using E. coli BL21 acetate metabolism knockout strains. *Microbial Cell Factories*, *18*(1), 151. doi:10.1186/s12934-019-1202-1

Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical Journal*, *83*(3), 1331–1340. doi:10.1016/s0006-3495(02)73903-9

Makino, T., Skretas, G., & Georgiou, G. (2011). Strain engineering for improved expression of recombinant proteins in bacteria. *Microbial Cell Factories*, *10*(1), 32. doi:10.1186/1475-2859-10-32

Maser, A., Peebo, K., Vilu, R., & Nahku, R. (2020). Amino acids are key substrates to Escherichia coli BW25113 for achieving high specific growth rate. *Research in Microbiology*, *171*(5–6), 185—193. doi:10.1016/j.resmic.2020.02.001

Meadows, A. L., Karnik, R., Lam, H., Forestell, S., & Snedecor, B. (2010). Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. *Metabolic Engineering*, *12*(2), 150–160. doi:10.1016/j.ymben.2009.07.006

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011. *Molecular Systems Biology*, *7*, 535. doi:10.1038/msb.2011.65

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. doi:10.1038/nbt.1614

Ow, D. S.-W., Lee, D.-Y., Yap, M. G.-S., & Oh, S. K.-W. (2009). Identification of cellular objective for elucidating the physiological state of plasmid-bearing Escherichia coli using genome-scale in silico analysis. *Biotechnology Progress*, *25*(1), 61–67. doi:10.1002/btpr.51

Pérez-Rodriguez, S., Ramírez-Lira, M. de J., Trujillo-Roldán, M. A., & Valdez-Cruz, N. A. (2020). Nutrient supplementation strategy improves cell concentration and longevity, monoclonal antibody production and lactate metabolism of Chinese hamster ovary cells. *Bioengineered*, *11*(1), 463–471. doi:10.1080/21655979.2020.1744266

Rahmen, N., Fulton, A., Ihling, N., Magni, M., Jaeger, K. E., & Büchs, J. (2015). Exchange of single amino acids at different positions of a recombinant protein affects metabolic burden in Escherichia coli. *Microbial Cell Factories*, *14*(1), 10. doi:10.1186/s12934-015-0191-y

Sarkandy, S. Y., Khalilzadeh, R., Shojaosadati, S. A., Sadeghizadeh, M., Farnoud, A. M., Babaeipour, V., & Maghsoudi, A. (2010). A novel amino acid supplementation strategy based on a stoichiometric model to enhance human IL-2 (interleukin-2) expression in high-cell-density Escherichia coli cultures. *Biotechnology and Applied Biochemistry*, *57*(4), 151–156. doi:10.1042/BA20100320

Savizi, I. S. P., Soudi, T., & Shojaosadati, S. A. (2019). Systems biology approach in the formulation of chemically defined media for recombinant protein overproduction. *Applied Microbiology and Biotechnology*, *103*(20), 8315–8326. doi:10.1007/s00253-019-10048-1

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, *6*(9), 1290–1307. doi:10.1038/nprot.2011.308

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. doi:10.1038/nmeth.2089

Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., … Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*, *150*(1), 76–85. doi:10.1016/0003-2697(85)90442-7

Sohn, S. B., Graf, A. B., Kim, T. Y., Gasser, B., Maurer, M., Ferrer, P., Lee, S. Y. (2010). Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. *Biotechnology Journal*, *5*(7), 705–715. doi:10.1002/biot.201000078

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, *72*(2), 211–222. doi:10.1007/s00253-006-0465-8

Torres, M., Berrios, J., Rigual, Y., Latorre, Y., Vergara, M., Dickson, A. J., & Altamirano, C. (2019). Metabolic flux analysis during galactose and lactate co-consumption reveals enhanced energy metabolism in continuous CHO cell cultures. *Chemical Engineering Science*, *205*, 201–211. doi:10.1016/j.ces.2019.04.049

Uhoraningoga, A., Kinsella, G. K., Henehan, G. T., & Ryan, B. J. (2018). The Goldilocks Approach: A Review of Employing Design of Experiments in Prokaryotic Recombinant Protein Production. *Bioengineering*, *5*(4), 89–115. doi:10.3390/bioengineering5040089

van der Valk, J., Brunner, D., De Smet, K., Fex Svenningsen, A., Honegger, P., Knudsen, L. E., … Gstraunthaler, G. (2010). Optimization of chemically defined cell culture media--replacing fetal bovine serum in mammalian in vitro methods. *Toxicology in Vitro : An International Journal Published in Association with BIBRA*, *24*(4), 1053–1063. doi:10.1016/j.tiv.2010.03.016

Varma, A., Boesch, B. W., & Palsson, B. (1993). Stoichiometric Interpretation of Escherichia coli Glucose Catabolism under Various Oxygenation Rates. *Appl Environ Microbiol*, *59*(8), 2465–2473. doi: 10.1128/aem.59.8.2465-2473.1993

Zeng, H., & Yang, A. (2019). Quantification of proteomic and metabolic burdens predicts growth retardation and overflow metabolism in recombinant Escherichia coli. *Biotechnology and Bioengineering*, *116*(6), 1484–1495. doi:10.1002/bit.26943

Table 1. Experimental growth rate in different media.

Media	**Growth rate (h^{-1})**
M9	0.025 ± 0.4
M9 + Asn	0.016 ± 0.631
M9 + Gln	0.01 ± 0.634
M9 + Arg	0.034 ± 0.540
M9 + Asn + Arg	0.762
M9 + Arg + Gln	0.01 ± 0.682
M9 + Asn + Gln	0.031 ± 0.665
M9 + Arg + Gln + Asn	0.015 ± 0.878

Table 2. Maximum acetate concentration in different media.

Media	**Maximum acetate production (mmol/L)**
M9	0.032 ± 7.66
M9 + Asn	0.012 ± 12.98
M9 + Gln	0.025 ± 11.29
M9 + Arg	0.042 ± 8.49
M9 + Asn + Arg	0.011 ± 9.15
M9 + Arg + Gln	0.046 ± 12.35
M9 + Asn + Gln	0.02 ± 9.65
M9 + Arg + Gln + Asn	0.016 ± 17.31

Figure 1. Results of dFBA for cell growth in the minimal medium (A); and in medium supplemented with amino acids: asparagine (B); glutamine (C); arginine (D); asparagine + arginine (E); arginine + glutamine (F); asparagine + glutamine (G); and all the three amino acids (E). For all dFBA simulations, the initial concentration of glucose was 22.2 mM, ammonium was 18.6 mM and each amino acid was set to 10.0 mM.

Figure 2. Extracellular metabolite profiles for recombinant *E. coli* cells grown in the minimal medium (A), and in medium supplemented with 10 mM of amino acids: asparagine (B); glutamine (C); arginine (D); asparagine + arginine (E); arginine + glutamine (F); asparagine + glutamine (G); and all the three amino acids (H). Data are representative of two independent experiments**.** All graphs are drawn using GraphPad Prism 8 software.

Figure 3. Growth profile of recombinant *E. coli* in different media. Cell density profiles for recombinant *E. coli* grown in the minimal medium and in medium supplemented with amino acids (A). dry cell weight (g/L) for recombinant *E. coli* culture grown in the minimal medium and in medium supplemented with amino acids (B). The *in silico* predicted growth rate in comparison with the experimental data for the eight different cultivation conditions (C). In all plots, the avarage value of two independent experiments is shown.

Figure 4. SDS-PAGE and western blot analyses showing antiEpEX-scFv protein expression in the minimal medium, and the medium supplemented with amino acids. (A) SDS-PAGE analysis of total lysate. Gels were loaded with 10 µL of total lysate of bacteria and stained with Coomassie blue G250. Total lysate proteins of *E. coli* expressing antiEpEx-scFv in medium supplemented with amino acids: asparagine (A_1-A_3); glutamine (B_1-B_3); Arginine (C_1-C_3); asparagine + arginine (D_1-D_3); asparagine + glutamine (F_1-F_3); arginine + glutamine (E_1-E_3); and all the three amino acids (G_1-G_3). Expression was done in three independent experiments, UN: uninduced bacterial lysate; M9: Total lysate proteins of *E. coli* BW25113 (DE3) expressing antiEpEx-scFv in M9 minimal medium as a control; M: protein molecular weight marker (14.4-116 kDa). Arrows indicate the band of antiEpEx-scFv protein with the expected size of (~ 29 kDa). (B) western blot analysis for recombinant antiEpEX-scFv using anti-polyhistidine monoclonal antibody. Total lysate proteins of *E. coli* expressing antiEpEX-scFv in medium supplemented with amino acids: asparagine (A); glutamine (B); arginine (C); asparagine + arginine (D); asparagine + glutamine (F); arginine + glutamine (E); and all the three amino acids (G). M9: Total lysate proteins of *E. coli* expressing antiEpEx-scFv in M9 minimal medium; Uninduced bacterial lysate (C^-); prestained molecular weight marker (10 – 250 kDa) (M).

Figure 5. Effect of amino acids supplementation in M9 minimal medium on target protein production. Recombinant antiEpEX-scFv concentration (g/L). Among different media that supplemented with amino acids, the maximum expression level of antiEpEX-scFv was observed in M9 supplemented with all three amino acids. Data are expressed as the mean values± standard deviation of independent experiments (n=3). * means $p < 0.05$, and ** means $p < 0.001$ in the t-test, as compared to the protein level expressed minimal medium (control).

Figure 6. The Central Composite Design (CCD) of different concentrations of three amino acids with experimental values of total and soluble antiEpEX-scFv expression level as a response. A: 20 experiments were designed at various combinations of amino acids concentrations by CCD. B: The level of total protein and soluble protein for each experiment is shown in the bar chart. Each experiment was repeated twice.

Figure 7. 3D Response surface plots to represent the interaction of variables on the antiEpEX-scFv total and soluble expression levels using the CCD-RSM methodology. (A-C) these response surface plots illustrated that the total expression level is directionally related to the increase of concentrations of the amino acids, and especially, glutamine. (D-E) Response surface plots to illustrate the glutamine and arginine in high concentrations and asparagine in medium concentrations are effective on protein solubility.

Figure 8. SDS-PAGE analysis to characterize the optimum conditions. (A) Two repeats of antiEpEX-scFv expression in optimum compostion of amino acids (M9 minimal medium supplemented with 9.9 mM Glutamine, 6.05 Asparagine, 9.5 mM Arginine), (S: Soluble fraction, T: Total cell lysate) (B) The purified antiEpEX-scFv with Ni-NTA column, (1) Eluted proteins fractions from expression in M9 minimal medium, (2) Eluted proteins fractions from expression in M9 minimal medium that supplemented with optimum

concentration of amino acids, (Un) Uninduced total cell lysate, (M) Protein molecular weight marker (14.4–116 kDa).