Genome-scale metabolic model-based engineering of *Escherichia coli* enhances recombinant single-chain antibody fragment production

Aidin Behravan · Atieh Hashemi · Sayed-Amir Marashi · Hamideh Fouladiha

Received: 7 May 2022 / Accepted: 29 August 2022 / Published online: 8 September 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

**Abstract**

*Purpose* *Escherichia coli* is an attractive and cost-effective cell factory for producing recombinant proteins such as single-chain variable fragments (scFvs). AntiEpEX-scFv is a small antibody fragment that has received considerable attention for its ability to target the epithelial cell adhesion molecule (EpCAM), a cancer-associated biomarker of solid tumors. Due to its metabolic burden, scFv recombinant expression causes a remarkable decrease in the maximum specific growth rate of the scFv-producing strain. In the present study, a genome-scale metabolic model (GEM)-guided engineering strategy is proposed to identify gene targets for improved antiEpEX-scFv production in *E. coli*.

*Methods* In this study, a genome-scale metabolic model of *E. coli* (*iJO1366*) and a metabolic modeling tool (FVSEOF) were employed to find appropriate genes to be amplified in order to improve the strain for increased production of antiEpEX-scFv. To validate the model predictions, one target gene was overexpressed in the parent strain *Escherichia coli* BW25113 (DE3).

*Results* For improving scFv production, we applied the FVSEOF method to identify a number of potential genetic engineering targets. These targets were found to be localized in the glucose uptake system and pentose phosphate pathway. From the predicted targets, the *glk* gene encoding glucokinase was chosen to be overexpressed in the parent strain *Escherichia coli* BW25113 (DE3). By overexpressing *glk*, the growth capacity of the recombinant *E. coli* strain was recovered. Moreover, the engineered strain with *glk* overexpression successfully led to increased scFv production.

*Conclusion* The genome-scale metabolic modeling can be considered for the improvement of the production of other recombinant proteins.

**Keywords** AntiEpEX-scFv · *Escherichia coli* · FVSEOF · Genome-scale metabolic model · *glk* · Glucokinase

**Introduction**

Nowadays, most clinically valuable proteins are recombinantly produced. For such proteins, expression is directly influenced by the metabolism of the producing cell factory. *Escherichia coli* is one of the well-established microbial cell factories characterized
by rapid growth and high-yield production of recombinant proteins (Ferrer-Miralles and Villaverde 2013). Despite continual efforts, there are still limiting barriers to the overexpression of recombinant proteins in E. coli. Metabolic burden observed in recombinant protein-producing cells leads to biomass yield reduction, which is part of the stress response triggered by protein overexpression. This stress response, in turn, can hamper the production of recombinant protein.

In the past few decades, metabolic engineering strategies have been successfully employed in developing industrial cell factories (Fernández-Cabezón and Nikel 2020). In metabolic engineering, specific biochemical reactions are modified by deletion and/or overexpression of specific genes, leading toward optimal pathway usage (Ko et al. 2020). These modifications lead to better productivity of the strains by changing the flux of certain metabolic reactions. Based on such a strategy, improved production of numerous heterologous products has been reported. However, a systematic understanding of the involved metabolic pathways, as well as the regulatory mechanisms, is highly needed for rational metabolic engineering.

Systems-level metabolic modeling strategies have been developed in recent years to guide metabolic engineering. These novel approaches such as genome-scale metabolic model (GEM)-guided engineering have shown promising results in practice. Genome-scale modeling can predict the cellular phenotype at the systems level and recommend genetic manipulations to modulate the relationship between target protein overexpression and biomass production so that strains with high growth and maximal protein productivity can be attained. High accuracy in the prediction of cell phenotype and minimization of consumed laboratory resources and time for developing productive strains are the most momentous advantages of model-guided metabolic engineering (Orth et al. 2011).

In GEM-based metabolic engineering, flux balance analysis (FBA) and flux variability analysis (FVA) can be utilized to compute the maximum production rate of biomass, when cell growth is the objective function. On the other hand, various algorithms such as FSEOF (flux scanning based on enforced objective flux) (Choi et al. 2010) as an FBA-based method and FVSEOF (flux variability scanning based on enforced objective flux) (Park et al. 2012) and OptForce (Ranganathan et al. 2010) as FVA-based methods have been employed to determine relevant genes to be amplified in order to achieve improved production by industrial cell factories. However, in practice, most of the GEM-based studies have focused on the production of small-molecule chemicals, rather than recombinant proteins (Ko et al. 2020).

The scFv (single-chain variable fragment) antibody against EpEX (EpCAM extracellular domain) has recently drawn great attention in biomedicine for dual therapeutic and diagnostic applications due to its small size and low immunogenicity (Eyvazi et al. 2018). In this study, as a model protein, scFv is chosen to be overproduced in E. coli with the help of metabolic engineering techniques. FVSEOF method was used to determine appropriate modifications of the genes in order to increase flux towards antiEpEX-scFv overproduction. By analyzing the FVSEOF results, the glk gene was chosen for overexpression. Experimental tests were then performed to show the positive effect of glk overexpression on antiEpEX-scFv overproduction, which proved the appropriateness GEM-based model for achieving a better cell factory.

Materials and methods

Metabolic modeling and target gene prediction

In this study, the E. coli metabolic model iJO1366 has been used. This model is the genome-scale reconstruction of the metabolic network of E. coli K-12 MG1655, which has 1366 genes, 2251 metabolic reactions, and 1136 metabolites (Orth et al. 2011). Modeling was done by the COBRA Toolbox v2.0 (Schellenberger et al. 2011) with ‘glpk’ as the linear programming solver. The metabolic reaction of antiEpEX-scFv production and the constraints representing the M9 medium composition were added to the model, according to the previously described method (Behravan et al. 2022).

In order to determine strategies for increasing antiEpEX-scFv production, the FVSEOF algorithm was used. In this algorithm, the flux changes of metabolic reactions in the model are calculated in response to the theoretical increase of the production rate of antiEpEX-scFv. For this purpose, first, the minimum and maximum values of the objective function
are calculated. Then, the maximum cell growth rate is calculated while the objective flux gradually increases. Finally, by using the FVA method, the maximum and minimum fluxes of every reaction are calculated. The metabolic network and FVSEOF algorithm can be used to predict the effect of an enforced increase in theoretical scFv production on the flux changes of the other metabolic reactions in *E. coli* cells. FVSEOF finds a list of metabolic reactions whose fluxes are changed as a result of a theoretical increase in scFv production. Based on FVSEOF results, the in vitro changes in metabolic reactions that had altered rates during the theoretical increase of scFv can potentially lead to increased scFv production in vitro.

**Bacterial strains, plasmids, and cultivation conditions**

*Escherichia coli* DH5α was used for recombinant plasmid cloning. *E. coli* BW25113 (DE3) (kindly provided by Prof. Leimkühler, University of Potsdam) was utilized for protein expression (Bühning et al. 2017). The plasmid pETDuet-1 (a gift from Dr. Bandehpour, Shahid Beheshti University of Medical Sciences, Tehran, Iran) was utilized as a co-expression vector. The plasmid pETDuet-antiEpEX-scFv was previously constructed in our laboratory (Behravan and Hashemi 2021). *Pfu* DNA polymerase, T4 DNA ligase, protein molecular weight markers and restriction enzymes were obtained from ThermoFisher Scientific (Waltham, Massachusetts, United States). DNA fragments were purified from agarose gel using a gel extraction kit (Roche Diagnostics GmbH, Mannheim, Germany). M9 minimal media containing (per liter) 0.5 g of NaCl, 6 g of Na₂HPO₄, 3 g of KH₂PO₄, 1 g of NH₄Cl supplemented with 5 g glucose 1⁻¹, 2 mmol MgSO₄ 1⁻¹, 0.01 mmol FeCl₃ 1⁻³, 0.1 mmol CaCl₂ 1⁻¹, and 0.1 ml 1000×trace metals element (Teknova Inc., California, United States), and LB medium composed of (per liter) 10 g of tryptone, 5 g of yeast extract, and 10 g of NaCl were used as the culture media. All other chemicals were purchased from Merck (Merck KGaA, Darmstadt, Germany) in analytical grade.

**Construction of recombinant plasmid**

The *glk* gene was amplified from the genome of *E. coli* BW25113 employing the primers glk-F, 5’-CCG GAA TTT TGA AGA ATG ACA AAG TATGC-3’ (the *EcoRI* site is marked) and glk-R, 5’-AAACTGCAG CCCGATATAAAAGGAAGGAT-3’ (the *PstI* site is marked) both primers were 29 bp long. Cycling condition for PCR was 94 °C for 3 min followed by 30 cycles of 94 °C, 30 s; 56 °C, 35 s, 72 °C 1 min and 30 s and 1 cycle of 72 °C, 10 min. To generate plasmid pETDuet-glk, the PCR product was digested with restriction enzymes *PstI* and *EcoRI*, and then ligated to pETDuet-1 treated with the same two enzymes. pETDuet-glk-antiEpEX-scFv expression plasmid was constructed by digesting the pGH vector carrying the antiEpEX-scFv gene with *Xhol* and *NdeI* to get the gene with a polyhistidine tag in its C-terminal, which was then ligated with pETDuet-glk treated by *NdeI/Xhol*. The restriction enzyme digestion assay and sequencing were used to confirm the constructs. Then, the recombinant plasmids were transformed into the chemically competent *E. coli* BW25113 (DE3) cells.

**Expression of antiEpEX-scFv**

For antiEpEX-scFv expression, a number of *E. coli* BW25113 single colonies harboring either pETDuet-glk-antiEpEX-scFv or pETDuet-antiEpEX-scFv were separately inoculated into 5 ml of LB medium supplemented with 100 μg ampicillin ml⁻¹ and incubated for 18 h at 37 °C with shaking (6000×g). After centrifugation (6000×g for 5 min at 4 °C), the pellet was resuspended into 100 ml of M9 minimal medium containing ampicillin (50 mg mL⁻¹). When cell density reached an OD₆₀₀ value of 0.8, the expression of antiEpEX-scFv was induced with 0.8 IPTG mmol l⁻¹ at 37 °C. Using centrifugation (6000×g for 5 min at 4 °C), the pellet was suspended in 30 mL of lysis buffer containing 50 mmol Tris l⁻¹ 7.5, 1 mmol· EDTA l⁻¹, 1 mg· lysozyme ml⁻¹, 150 mmol NaCl l⁻¹, 1% Triton X-100, and then, sonicated for 30 min (20 s ON, 10 s OFF at 400 W) according to the previous study (Behravan et al. 2022). After centrifugation of the cell lysate (10,000×g for 30 min at 4 °C), protein samples were electrophoresed on a 15% SDS-PAGE gel and visualized using Coomassic brilliant blue G-250 dye. By using a wet Trans-Bolt (Bio-Rad Laboratories, Inc., California, United States), the proteins were electro-transferred from the gel into the...
polyvinylidene difluoride (PVDF) membrane to perform a Western blot analysis. The transferred membrane was blocked in 5% nonfat milk for 1 h, and then, was washed three times with TBST and then incubated in His-tag antibody (Sigma, UK) overnight. After washing again, the membrane was incubated in anti-mouse HRP conjugated immunoglobulin (Merck KGaA, Darmstadt, Germany) as the secondary antibody for two hours, and then, detected by means of a solution of 3,3′-diaminobenzidine (DAB) (Merck KGaA, Darmstadt, Germany). The recombinant antiEpEX-scFv was purified using the Ni–NTA affinity chromatography column under denaturing conditions based on the manufacturer’s protocol (Qiagen Inc., Hulsterweg, Netherlands). Utilizing the bicinchoninic acid assay (BCA assay), the concentration of the purified protein was measured by Takara BCA Protein Assay Kit (Takara Bio Inc., Nojihigashi, Japan).

Growth profile and glucose analysis

To investigate the cell growth profile, OD\textsubscript{600} was determined every hour, using a spectrophotometer (E-Chrome Tech, Taipei, Taiwan). Logarithmic derivation of the optical density curve used for calculation of growth rate. In order to determine glucose concentration, one milliliter of sample from culture broth was harvested in one-hour intervals. The supernatant was collected following 10 min of centrifugation at 10,000×g. The concentration of glucose was measured using a commercial enzymatic kit (MegaZyme, Wicklow, Ireland).

RT-qPCR analysis

To compare the relative expression of glk as a target gene between \textit{E. coli} BW25113/Duet-glk-scFv and \textit{E. coli} BW25113/Duet-scFv, RT-qPCR were employed. \textit{E. coli} strains were cultured in 50 ml M9 medium and induced with 0.8 IPTG mmol l\textsuperscript{-1} in OD\textsubscript{600} = 0.8. After 3 h, samples were collected and diluted to OD\textsubscript{600} = 0.4. Based on the manufacturer’s protocol, total RNA was extracted from bacterial cells utilizing TRizol reagent (Ambion Inc., Austin, TX). The purity and quantity of the isolated RNA were measured by Synergy HTX multimode reader (Agilent, Santa Clara, United States). Subsequently, it was stored at –80 °C for further use. Then, a cDNA synthesis kit (Yekta Tajhiz Azma, Tehran, Iran) was employed to synthesize cDNA according to the instruction provided by the manufacturer. Based on the literature, primers were selected and assessed for GC content, specificity, potential secondary structures, and amplicon size. Primer sequences synthesized by Metabion (Metabion GmbH, Germany) are presented in Table 1. StepOne Real-Time PCR System (Applied Biosystems Waltham, Massachusetts, United States) was employed for SYBR-Green qPCR reactions in 48 well optical reaction plates. cDNA (0.5 ng/reaction) was used as the template for qPCR reactions with 5 µl SYBR Green PCR Master Mix (2×) (Yekta Tajhiz Azma, Tehran, Iran) and primers at 10 µM final concentration. Samples were exposed to thermal plan as follows: 95 °C, 30 s followed by 40 cycles of 95 °C, 5 s and 60 °C, 30 s. The PCR reactions were done in three technical replicates for more accuracy and ensuring reproducibility. The \(2^{-\Delta\Delta C_t}\) method was used to evaluate relative gene expression against the reference gene (Livak and Schmittgen 2001).

\begin{table}
\centering
\caption{Primers used in RT-qPCR assay}
\begin{tabular}{llll}
\hline
Name & Oligonucleotide sequences (5′→3′) & Primer’s length (bp) & Amplicon length (bp) \\
\hline
16S & F: TACCGCATAACGTCGCAAGA & 20 & 168 \\
& R: AGTCTGGACCGTGCTTCAGT & 20 & 168 \\
\hline
 glac & F: CTGTATGGCATCGCTTTGCC & 20 & 207 \\
& R: TTACCTTGCACGGTTCTGC & 20 & 207 \\
\hline
\end{tabular}
\end{table}

Results

Prediction of overexpression targets

The genome-scale metabolic model of \textit{E. coli}, iJO1366 (Orth et al. 2011), was employed for the prediction of metabolic engineering targets that can improve antiEpEX-scFv production. To this end,
the FVSEOF algorithm was exploited, as explained in the Materials and Methods section. In jO1366, the FVSEOF algorithm predicted ten metabolic reactions whose flux increase can enhance antiEpEX-scFv production. Therefore, the enzymes that catalyze these reactions were considered potential targets of overexpression. Among their associated genes, two genes were related to the import of glucose (galP and glk), four genes were related to the pentose phosphate (PP) pathway (zwf, rpe, pgl, gnd), two genes were related to the folate biosynthesis pathway (focA and purU), and two genes were related to alternative carbon metabolism subsystem (xylA, mak). These reactions and their associated metabolic genes are illustrated in Fig. 1.

The transcriptome analysis showed that glk is upregulated during the recombinant protein production in E. coli (Oh and Liao 2000). Therefore, in the present study, the glk gene was chosen as the target gene for overexpression.

Effects of glk overexpression on antiEpEX-scFv production

The glk gene was amplified from the E. coli genome by PCR, resulting in a 1004 bp glk gene which can encode 321 amino acids with a molecular weight of about 35 kDa. Then, glk and antiEpEX-scFv coding sequences were inserted into the first and second multiple cloning sites of pETDuet-1 to generate the plasmid pETDuet-glk-antiEpEX-scFv (Fig. 2). The constructed pETDuet-glk-antiEpEX-scFv plasmid was confirmed by restriction enzyme digestion (Fig. 3) and sequencing. The plasmid was then transformed into E. coli BW25113 (DE3). Recombinant E. coli BW25113/Duet-glk-scFv and E. coli BW25113/Duet-scFv were cultured in the M9 minimal medium supplemented with 100 μg ampicillin ml to an OD600 of 0.8, and then induced with 0.8 mmol·IPTG l−1 for 24 h. The SDS-PAGE analysis confirmed the presence of two separate protein bands with molecular weights of 35 kDa (glk) and 29 kDa (antiEpEX-scFv) (Fig. 4a). The expressed scFv protein was confirmed using Western blot analysis (Fig. 4b).

Fig. 1 Illustration of the set of predicted overexpression targets on the metabolic network map. The predicted targets for overexpression are indicated by their bold gene names. (PPP pathway: Pentose Phosphate Pathway, fru: Fructose, glc: Glucose, g6p: glucose-6-phosphate, f6p: fructose-6-phosphate, pgl: 6-phosphogluconolactone, Ru5P: ribulose-5-phosphate, 6pgc: 6-phosphogluconate, R5P: ribose-5-phosphate, galP: galactose permease, zwf: glucose-6-phosphate dehydrogenase, xylA: xylose Isomerase, mak: fructokinase, glk: glucokinase, gnd: 6-phosphogluconate dehydrogenase, pgl: 6-phosphogluconolactonase, rpe: ribulose-5-phosphate-3-epimerase)
According to the concentration of purified antiEpEX-scFv, the glk-overexpressing strain, after 24 h post-induction cultivation, showed an increase in antiEpEX-scFv titer (235.4 ± 9.5 µg ml⁻¹; 0.428 g gDCW⁻¹), which was approximately 110% higher than that of the strain without glk overexpression (110.2 ± 7.7 µg ml⁻¹; 0.202 g gDCW⁻¹) (Fig. 4c). The results confirmed that altered glucose metabolism by glk overexpression could improve the antiEpEX-scFv production.

Growth and glucose consumption profiles

To examine how co-expression of glk with antiEpEX-scFv can affect bacterial growth rate and glucose consumption rate, E. coli (BW25113), recombinant E. coli BW25113/Duet-glk-scFv and E. coli BW25113/Duet-scFv, were cultured in the M9 minimal medium containing 5 or 10 g glucose ·l⁻¹ at 37 °C. IPTG in final concentration of 0.8 mmol l⁻¹ was added in OD₆₀₀ = 0.8 for protein induction.
All strains in the M9 minimal medium containing 5 g glucose l⁻¹ grow logarithmically as long as glucose is available in the medium (Figs. 5a, b). When glucose is depleted, strains enter the stationary phase. As shown in Fig. 4a, maximum specific growth rate in both recombinant E. coli strains BW25113/Duet-scFv and BW25113/Duet-glk-scFv (\(\mu_{\text{max}} = 0.46 \pm 0.03\) and \(\mu_{\text{max}} = 0.55 \pm 0.003\) respectively) was lower than that in the parent strain (\(\mu_{\text{max}} = 0.63 \pm 0.01\)). Since recombinant protein and cell biomass share common precursors, increased protein expression at the expense of decreased cell density is presumably due to the alteration of intracellular fluxes of precursors towards protein synthesis in recombinant strains. However, recombinant E. coli BW25113/Duet-glk-scFv has a greater specific growth rate than the E. coli BW25113/Duet-scFv (Table 2), especially when more glucose is available in the medium. As shown in Fig. 4c and Table 2, \(\mu_{\text{max}}\) value of E. coli BW25113/Duet-glk-scFv (0.81 ± 0.04) is much higher than those of E. coli BW25113/Duet-scFv (0.59 ± 0.003) and wild type strain (0.73 ± 0.02). As expected, in the M9 minimal medium containing 10 g glucose l⁻¹, the maximum specific growth rates of all strains are higher than those in the medium supplemented with 5 g glucose l⁻¹.

As shown in Table 3, the glucose consumption rate by E. coli BW25113/Duet-glk-scFv (1.145 ± 0.01) is greater than that of the E. coli BW25113/Duet-scFv (0.968 ± 0.02) and the parent strain (0.851 ± 0.05). Also, as illustrated in Fig. 4b, the required time for the complete consumption of glucose for the parent strain and for E. coli BW25113/Duet-glk-scFv was 10 h, while E. coli BW25113/Duet-scFv needs 12 h to consume all the glucose in the medium. Interestingly, the required time for complete consumption of glucose for the parent strain and for E. coli BW25113/Duet-glk-scFv is similar (about 24 h), while E. coli BW25113/Duet-scFv does not consume all of the available glucose in 24 h (Fig. 5d).

Evaluation of glk transcription level

In order to validate the overexpression of glk, the real-time PCR experiment was performed. The E. coli BW25113 harboring pETDuet-glk-scFv showed significantly higher levels of glk transcription level compared to the E. coli BW25113 harboring pET-Duet-scFv. After 3 h of induction, the relative quantification of glk transcript revealed that glk level was overexpressed in E. coli BW25113/Duet-glk-scFv by 14.78-fold in comparison to the E. coli BW25113/Duet-scFv (Fig. 6).

Discussion

The development of hosts that have desirable metabolic characteristics, with the ability to optimally produce heterologous products is an important issue in microbial metabolic engineering. Various GEM-based approaches have enabled scientists to recognize gene deletion or overexpression targets for developing cell factories. As an example, production of \(L\)-valine was successfully improved in an engineered E. coli strain based on the MOMA simulation results (Park et al. 2007). Also, amplification of the idi gene selected by FSEOF together with the dxs gene led to lycopene overproduction (Choi et al. 2010). Predicting the metabolic consequences of gene deletion is often much simpler than those of gene overexpression. In the case of a deleted gene, its corresponding metabolic flux can simply be assumed as zero, while, due to complex regulation and control of metabolism, fluxes of overexpressed genes do not necessarily

Fig. 3 Confirmation of recombinant plasmid pETDuet-glk-antiEpEX-scFv by enzymatic digestion reaction: 1: plasmid pETDuet-glk-antiEpEX-scFv (7150 bp); 2: recombinant plasmid pETDuet-glk-antiEpEX-scFv digested with EcoRI and PstI (6347 bp and 1973 bp); 3: recombinant plasmid pETDuet-glk-antiEpEX-scFv digested with XhoI and NdeI (5184 bp and 810 bp); M: DNA marker 1 kb
increase. Moreover, even if gene overexpression results in a metabolic flux increment, the amount of this increase is hardly predictable.

In this study, in order to increase flux towards antiEpEX-scFv overproduction, among several targets predicted by FVSEOF, the *glk* gene was selected for overexpression. According to our results, recombinant expression of antiEpEX-scFv and *glk* resulted in a decrease in the maximum specific growth rate of recombinant strains compared with the parent strain. A decrease in growth rate is normally detectable in bacteria transformed with multicopy plasmids for producing recombinant proteins. One should keep in mind that plasmid DNA replication, plasmid-encoded mRNA synthesis and their translation in bacteria impose a metabolic burden on the engineered strains, which in turn, results in growth retardation (Flores et al. 2004). This metabolic burden is due to the cell’s inability to supply the extra demand of energy and building blocks required for the replication of the multicopy plasmid and expression of its genes (Li and Rinas 2020). However, a significant increase was observed in the \( \mu_{\text{max}} \) of the recombinant strains from 0.592 ± 0.003 in BW25113-Duet-scFv to 0.81 ± 0.043

---

**Fig. 4** SDS-PAGE and Western blot analyses for characterization of antiEpEX-scFv recombinant proteins in *E. coli* cell extracts. A SDS-PAGE analysis of total lysate. (1, 2): induced total lysate of *E. coli* BW25113/Duet-scFv; (M): Molecular weight protein marker (14.4–116 kDa); (3, 4): induced bacterial lysate with empty plasmid pETDuet-1, (5, 6): induced total lysate of *E. coli* BW25113/Duet-glk-scFv. B Western blot analysis for recombinant antiEpEX-scFv using the anti-polyhistidine monoclonal antibody. (C−): Uninduced bacterial lysate; (M): prestained molecular weight marker (10–250 kDa); (1): induced total lysate of *E. coli* BW25113/Duet-scFv; (2): induced total lysate of *E. coli* BW25113/Duet-glk-scFv. C SDS-PAGE analysis of the purified antiEpEX-scFv. (M): Molecular weight protein marker (14.4–116 kDa); (C−) Uninduced bacterial lysate;(1) The purified antiEpEX-scFv from *E. coli* BW25113/Duet-scFv; (2) The purified antiEpEX-scFv from *E. coli* BW25113/Duet-glk-scFv. All experiments were done in M9 minimal medium containing glucose as the carbon source. When OD₆₀₀ reached 0.8, cells were induced for 24 h using 0.8 mmol IPTG l⁻¹. The experiments were performed in duplicates. The protein bands corresponding to antiEpEX-scFv and glk are shown by arrows.
in BW25113-Duet-glk-scFv when the expression of the glk gene was overexpressed, which is comparable to the wild-type strain (0.729 ± 0.022). In other words, the cellular performance of E. coli is improved by engineering glucose uptake systems. Likewise, De Anda et al., showed that the overexpression of galP-glk in E. coli reduces acetate accumulation, and additionally, improves as well as improvement of the cellular growth rate and recombinant protein production (De Anda et al. 2006).

### Table 2
Specific growth rate of different strains in M9 minimal medium in different glucose concentrations

| Strains               | Growth rate (h⁻¹) in M9 + 5 g/L glucose | Growth rate (h⁻¹) in M9 + 10 g/L glucose |
|-----------------------|----------------------------------------|----------------------------------------|
| BW25113              | 0.637 ± 0.013                          | 0.729 ± 0.022                          |
| BW25113-Duet-scFv     | 0.462 ± 0.034                          | 0.592 ± 0.003                          |
| BW25113-Duet-glk-scFv | 0.552 ± 0.003                          | 0.81 ± 0.043                           |

**Fig. 5** Growth profiles and glucose consumption rates. When OD₆₀₀ reached 0.8, cells were induced for 24 h using 0.8 mmol IPTG l⁻¹. Growth and glucose consumption profiles of the parent strain (BW25113) and the recombinant strains in M9 medium supplemented with 5 g glucose l⁻¹ (A, B) and 10 g glucose l⁻¹ (C, D) are represented. Error bars illustrate the standard deviation of two experimental replicates. All graphs are drawn using GraphPad Prism 8 software. Data are presented as mean ± SD, n = 2.
The *glk* gene encodes the enzyme glucokinase catalyzing the ATP-dependent phosphorylation of the glucose that was imported by GalP. Overexpression of *glk* presumably leads to carbon flux redirection into the glycolysis and PP pathway, compensating for the special metabolic demands of the engineered *E. coli* strain (Hernández-Montalvo et al. 2003). The PP pathway, which is closely interconnected with glycolysis, normally provides some of the required blocks for biosynthesis of histidine, nucleotides and aromatic amino acids, e.g., erythrose-4-phosphate and ribose-5-phosphate (Stincone et al. 2015). Moreover, NADPH, a power source for biosynthetic reactions, is reduced in the oxidative branch of this pathway (Christodoulou et al. 2018). In a similar study, the PP pathway of *E. coli* has been engineered leading to a reduction of the metabolic load caused by recombinant protein production (Flores et al. 2004). In the present work, using a similar approach, a significant positive effect was observed on the productivity of the scFv producing strain. The *glk*-overexpressed strain produced approximately 2.1 times higher titer of scFv compared to the strain without *glk* overexpression. One can conclude that the metabolic engineering target predicted in our study was successfully validated via the improvement observed in the scFv production.

Using DNA microarray, Oh et al. (2000) revealed that overproduction of recombinant LuxA could lead to the downregulation of *ppc*, *fba*, *gnd*, and *atpA* genes, as well as upregulation of heat shock and *glk* genes in three *E. coli* strains, namely JM109, MC4100, and VJS676A. Based on the transcriptome profile obtained in that study, instead of the phosphotransferase system, glucose kinase was suggested to have the major role to provide glucose-6-phosphate in protein overproducing conditions in the *E. coli* cells (Oh and Liao 2000). On the other hand, overexpression of recombinant proteins was shown to induce expression of heat shock genes and rapid stress response (Gill et al. 2000). Interestingly, *glk* has been reported to play an essential role in bacterial stress responses (Zhang et al. 2020). Although *glk* protein plays a minor role in glucose metabolism, under stress conditions like heterologous protein expression or growth in acidic conditions, this enzyme becomes necessary for a sufficient supply of glucose-6-phosphate (Arora and Pedersen 1995; Zhang et al. 2020). Therefore, *glk* seems to be a relevant target gene to be overexpressed for achieving an increased recombinant protein productivity.

In conclusion, the goal of our work was to improve the maximum specific growth rate of an scFv-producing *E. coli* strain, and consequently, to increase the scFv production yield. GEM-guided metabolic engineering strategies have been

### Table 3
Glucose consumption rate in different strains in M9 minimal medium in different glucose concentrations

| Strains          | Uptake rate (g/g DCW h) in M9 + 5 g/L glucose | Uptake rate (g/g DCW h) in M9 + 10 g/L glucose |
|------------------|---------------------------------------------|-----------------------------------------------|
| BW25113         | 0.851 ± 0.05                                | 0.730 ± 0.034                                 |
| BW25113-Duet-scFv| 0.968 ± 0.02                                | 0.706 ± 0.039                                 |
| BW25113-Duet-glk-scFv | 1.145 ± 0.01                               | 1.175 ± 0.03                                 |

**Fig. 6** Gene expression analysis by real time PCR. mRNA expression level of *glk* was evaluated in *E. coli* BW25113/Duet-glk-scFv and *E. coli* BW25113/Duet-scFv, the 16srRNA gene amplification was used as an endogenous control. Data are expressed as the mean ± SEM of four experiments. ******p ≤ 0.0001; Results show a significant overexpression of *glk* at the mRNA level in *E. coli* BW25113/Duet-glk-scFv compared to *E. coli* BW25113/Duet-scFv.**
previously used for increasing recombinant protein production (Nocon et al. 2014; Fouladiha et al. 2020; Behravan et al. 2022). Here, the GEM-guided metabolic engineering strategy was used to improve the scFv production in *E. coli* BW25113 (DE3). We applied FBA and FVSEOF methods to identify potential genetic engineering targets. From the predicted targets, *glk* gene encoding glucokinase was chosen to be overexpressed. The engineered strain with *glk* overexpression successfully increased scFv production to 2.1-fold, which proves the suitability of the exploited rational design strategy for strain development. This approach can be employed to determine the bottlenecks of the intracellular metabolic pathways by modification of cellular characteristics on the basis of metabolic engineering, and therefore, can be considered for improving the production of other recombinant proteins. In other words, we believe that our method for the production of scFv is a successful example of GEM-guided metabolic engineering, and can be applied to other recombinant protein production systems to achieve higher productivity and product yields.

**Acknowledgements** We would like to thank Prof. Dr. Silke Leimkühler and Dr. Bandehpour for providing *Escherichia coli* BW25113 (DE3) and plasmid pETDuet-1, respectively.

**Author Contributions** AH and SAM: Contributed to the conception and design of the study. AB and HF: carried out the experiments and analyzed the data. The first draft of the manuscript was written by AB. AH and SAM: Reviewed and edited the article for spelling, grammar, and intellectual content. AH and SAM: Organized and supervised the whole project. AH: Provided the facilities and materials required for the project. All authors agreed to be responsible for the contents of the manuscript.

**Funding** This work was supported by the research deputy of Shahid Beheshti University of Medical Sciences in Tehran, Iran (IR.SBMU.PHARMACY.REC.1399.372). Atieh Hashemi has received the research support. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

**Declarations**

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Ethical approval** This is an observational study. The school of pharmacy and nursing and midwifery Shahid Beheshti University of Medical Sciences’ Ethics Committee has confirmed that no ethical approval is required.

**References**

Arora K, Pedersen P (1995) Glucokinase of *Escherichia coli*: induction in response to the stress of overexpressing foreign proteins. Arch Biochem Biophys 319(2):574–578

Behravan A, Hashemi A (2021) Statistical optimization of culture conditions for expression of recombinant humanized anti-EpCAM single-chain antibody using response surface methodology. Res Pharm Sci 16:153–164. https://doi.org/10.4103/1735-5362.310522

Behravan A, Hashemi A, Marashi S-A (2022) A constraint-based modeling approach to reach an improved chemically defined minimal medium for recombinant antiEpEX-scFv production by *Escherichia coli*. Biochem Eng J 179:108339. https://doi.org/10.1016/j.bej.2022.108339

Bühning M, Friemel M, Leimkühler S (2017) Functional complementation studies reveal different interaction partners of *Escherichia coli* IscS and human NFS1. Biochemistry 56:4592–4605. https://doi.org/10.1012/acs.biochem.7b0627

Choi HS, Lee SY, Kim TY, Woo HM (2010) In silico identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76:3097–3105. https://doi.org/10.1128/AEM.00115-10

Christodoulou D, Link H, Fuhrer T et al (2018) Reserve flux capacity in the pentose phosphate pathway enables *Escherichia coli*’s rapid response to oxidative stress. Cell Syst 6:569-578.e7. https://doi.org/10.1126/Cell.2018.04.009

De Anda R, Lara AR, Hernández V et al (2006) Replacement of the glucose phosphotransferase transport system by galactose permease reduces acetate accumulation and improves process performance of *Escherichia coli* for recombinant protein production without impairment of growth rate. Metab Eng 8:281–290. https://doi.org/10.1016/j.ymben.2006.01.002

Eyyazi S, Farajnia S, Dastmalchi S et al (2018) Antibody based EpCAM targeted therapy of cancer, review and update. Curr Cancer Drug Targets 18:857–868. https://doi.org/10.2174/156809618666180102102311

Fernández-Cabezón L, Villaverde A (2020) Chapter 11—Advanced metabolic engineering strategies for the development of sustainable microbial processes. In: Rodrigues AG (ed) Metabolic engineering and bioengineering. Elsevier, Amsterdam, pp 225–246. https://doi.org/10.1016/bs.mee.2020.03.001

Ferrr-Miralles N, Villaverde A (2013) Bacterial cell factories for recombinant protein production; expanding the catalogue. Microb Cell Fact 12:113. https://doi.org/10.1186/1475-2859-12-113

Flores S, de Anda-Herrera R, Gosset G, Bolivar FG (2004) Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentosephosphate pathway. Biotechnol Bioeng 87:485–494. https://doi.org/10.1002/bit.201377

Fouladiha H, Marashi S-A, Torkashvand F et al (2020) A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. Bioprocess Biosyst Eng 43:1381–1389. https://doi.org/10.1007/s00449-020-02332-6
Gill RT, Valdes JJ, Bentley WE (2000) A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. Metab Eng 2:178–189. https://doi.org/10.1006/mben.2000.0148

Hernández-Montalvo V, Martínez A, Hernández-Chavez G et al (2003) Expression of galP and glk in a *Escherichia coli* PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. Biotechnol Bioeng 83:687–694. https://doi.org/10.1002/bit.10702

Ko Y-S, Kim JW, Lee JA et al (2020) Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. Chem Soc Rev 49:4615–4636. https://doi.org/10.1039/D0CS0155D

Li Z, Rinas U (2020) Recombinant protein production associated growth inhibition results mainly from transcription and not from translation. Microbiol Cell Fact 19:83. https://doi.org/10.1186/s12934-020-01343-y

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25:402–408. https://doi.org/10.1016/S1046-2023(01)00038-8

Nocon J, Steiger MG, Pfeffer M et al (2014) Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metab Eng 24:129–138. https://doi.org/10.1016/j.ymben.2013.12.003

Oh MK, Liao JC (2000) DNA Microarray detection of metabolic responses to protein overproduction in *Escherichia coli*. Metab Eng 2:201–209. https://doi.org/10.1006/mben.2000.0149

Orth JD, Conrad TM, Na J et al (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. Mol Syst Biol 7:535. https://doi.org/10.1038/msb.2011.65

Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci USA 104:7797–7802. https://doi.org/10.1073/pnas.0702609104

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.