**Supporting Information**

Amino acid supplementation for enhancing recombinant protein production in *E. coli*

**Jashwant Kumar, Ashish S. Chauhan, Rohan L. Shah, Jaya A. Gupta and Anurag S. Rathore**

Department of Chemical Engineering, Indian Institute of Technology, New Delhi, India

**Correspondence**

Anurag S. Rathore, Department of Chemical Engineering, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

**Email:** asrathore@biotechcmz.com

**This file contains**

**Supplementary Tables S1, S7, S8, S10**

**Supplementary Figures S2, S3, S4, S5, S6**

**Supplementary Method S9**

**Table S1: Chemically defined medium (CDM) used in the study**

|  |  |  |
| --- | --- | --- |
| **Nutrient** | **Batch (per L)** | **Fed-batch (per L)** |
| Glucose | 10 g | 150 g |
| Potassium dihydrogen phosphate | 13.3 g | - |
| di-Ammonium hydrogen phosphate | 4.0 g | - |
| Magnesium sulphate heptahydrate | 1.2 g | 20.0 g |
| Citric Acid | 1.7 g | - |
| EDTA | 8.0 mg | - |
| Thiamine HCl | 4.5 mg | - |
| Trace Stock (100X) | 5 ml | 10 ml |
| Kanamycin (100 mg mL-1) | 300 µl | - |



**(A)**



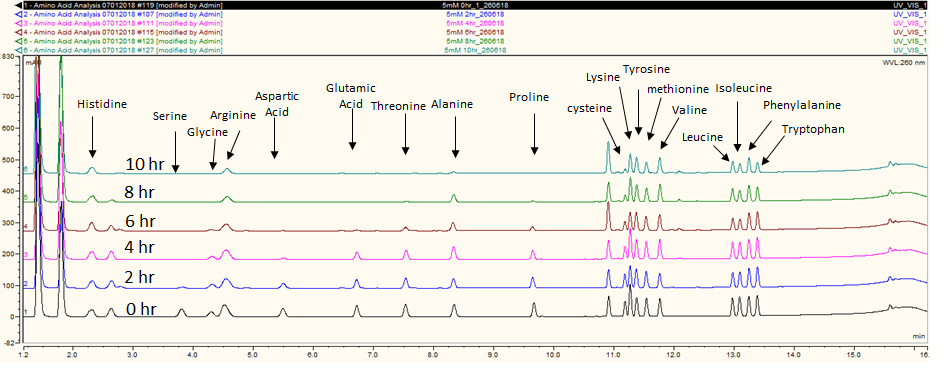
**(B)**

**Figure S2: Amino acid consumption profiles for initial optimization of amino acid concentration (A):** CDM + 2.5 mM of all 20 amino acids **(B):** CDM + 7.5 mM of all 20 amino acids Among 20 amino acids, 18 amino acids are detected by the method for amino acid analysis. Data are expressed as mean±SEM.

(A)

(**B)**

**Figure S3: Specific amino acid consumption profiles for initial optimization of amino acid study (A):** CDM + 2.5 mM of all 20 amino acids **(B):** CDM + 7.5 mM of all 20 amino acids. Data are expressed as mean±SEM.



**Figure S4: LC chromatogram of amino acid analysis:** Time based amino acid consumption profile of CDM + 5 mM amino acid is shown.

 (A)

 **(**B)

 (C)

(D)

**Figure S5: Amino acid consumption profiles for amino acid supplementation design (A)** Amino acid consumption profiles for SF2 where 5 mM of GP1 AA were supplemented prior to inoculation **(B):** Amino acid consumption profiles for SF3 where 5 mM of GP1 AA were supplemented prior to inoculation and 2.5 mM of GP2 prior to protein induction **(C)**Amino acid consumption profiles for SF4 where 2.5 mM of GP2 AA were supplemented prior to protein induction **(D)**Amino acid consumption profiles for SF5 where 5 mM of GP1 AA were supplemented prior to inoculation and 2.5 mM of GP1 and GP2 AA were supplemented prior to protein induction. Data are expressed as mean±SEM



Test

Control

A

B



C

## 

**Figure S6: Validation of best amino acid supplementation design strategy on recombinant *E. coli* strain expressing GCSF. Data are expressed as the mean±SEM of two independent experiments: (A) Glucose and cell density profile** No significant change in glucose uptake rate was observed. However, high cell density was observed in test culture **(B) Solubilized GCSF IB:** A significant increase in rGCSF concentration (g/L) was obtained in test culture as compared to control **(C) HPLC Chromatogram:** A RP-HPLC Chromatogram of GCSF of test (red) and control (blue) culture

We validated the best supplementation strategy of GP1 and GP2 amino acids with another recombinant *E. coli* strain expressing GCSF (Granulocyte-colony stimulating factor). Optimized GP1 and GP2 amino acid supplementation strategy ( i.e. 5 mM of GP1 at the start of culture and 2.5 mM of GP1 & 2.5 mM of GP2 before induction) were compared with the control culture.

The test culture data were compared with the control culture in terms of the cell density, glucose consumption and rGCSF concentration. There was no significant change in glucose uptake rate (Fig. S6(A) between control and test culture. However, a 35 % increase in cell density (Fig. S6(A)) was obtained in test culture as compared to culture at 10th hour with maximum cell density of 5.79±0.24 g/L. A significant increase in recombinant solubilized GCSF IB (Fig. S6(B)) was obtained in test culture with maximum concentration of 0.55±0.28 g/L as compared to control. A RP-HPLC chromatogram of solubilized GCSF IB of test (red) and control (blue) is shown in Figure S6(C). This confirms that the optimized supplementation strategy of GP1 and GP2 amino acids also works with other recombinant protein expressed in *E. coli*.

**Table S7: Log2 fold change in expression of stress related genes in test culture as compared to control**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gene** | **Gene Product** | **Function** | **Log2 Fold change** | | **References** |
| **Growth Phase** | **Protein Production Phase** |
| **Stringent like Response** | | | | |  |
| *relA* | ppGpp synthase I | Synthesis of alarmone ppGpp | -0.17 | -1.46 | (Wick & Egli, 2004) |
| *dksA* | RNA polymerase-binding transcription factor DksA | Inhibit rRNA promoter activity and amplify effect of ppGpp | 0.38 | -2.53 | (Paul et al., 2004) |
| *rsd* | Repressor of sigma D | Binds to RpoD/σ70 and negatively regulates σ70-mediated transcription | 0.58 | -2.53 | (Sharma & Chatterji, 2010) |
| *rpoS* | RNA polymerase sigma factor S (σ38) | Compete with σ70 for RNAP core and control stress response | -0.4 | -1.97 | (Lacour & Landini, 2004; Suh et al., 1999) |
|  |  |  |  |  |  |
| **Universal stress protein** | | | | |  |
| *uspA* | universal stress protein A | Accumulates during cell growth arrest and protect cell from DNA damage | 0.2 | -3.02 |  |
| *uspC* | universal stress protein C | -do- | 1.44 | -2.76 |  |
| *uspD* | universal stress protein D | -do- | 0.82 | -2.32 | (Liu et al., 2007) |
| *uspE* | universal stress protein E | -do- | -0.02 | -3.52 |  |
| *uspF* | universal stress protein F | -do- | -0.26 | -3.23 |  |
| *uspG* | universal stress protein G | -do- | 1.29 | -3.30 |  |
|  |  |  |  |  |  |
| **Biofilm formation** | | | | |  |
| *bssR* | Biofilm regulator BssR | Regulate biofilm formation through signal secretion | -0.89 | -4.45 | (Domka et al., 2006) |
| *bsmA* | Lipoprotein BsmA | Helps in microcolony formation and biofilm maturation | -0.30 | -4.12 | (Weber et al., 2010) |
| *yjfN* | Uncharacterized protein YjfN | Unknown function in biofilm formation | -4.54 | -3.85 | (Eletsky et al., 2014) |
| *yahO* | Uncharacterized protein YahO | Uncharacterized role in stress response and biofilm formation | 0.76 | -3.84 | (Eletsky et al., 2014) |
| *ydgH* | Protein YdgH | Unknown function in biofilm formation | 0.72 | -1.59 | (Eletsky et al., 2014) |
| *bssS* | Biofilm formation regulatory protein BssS | Regulate biofilm formation through signal secretion | 0.9 | -1.59 | (Domka et al., 2006) |
|  |  |  |  |  |  |
| **Ribosome Related** | | | | |  |
| *raiA* | Ribosome associated inhibitor A | Block translation initiation during stress condition | -2.05 | -4.99 | (Vila-Sanjurjo et al., 2004)(Starosta et al., 2014) |
| *hpf* | Ribosome hibernation promoting factor | Stabilization of 100S ribosome | 0.19 | -2.32 | (Yoshida & Wada, 2014)(Starosta et al., 2014) |
| *rmf* | Ribosome modulation factor | Dimerization of two 70S ribosomes into inactive 100S ribosome | -0.21 | -1.49 | (Yoshida & Wada, 2014)(Starosta et al., 2014) |
|  | | |  |

**Table S8: Log2 fold change in expression of genes associated with amino acid biosynthesis in control culture as compared to test**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gene** | **Gene Product** | **Function** | **Log2 Fold change** | | **References** |
| **Growth Phase** | **Protein Production Phase** |
| **Aromatic amino acid biosynthesis** | | | | |  |
| *trpD* | bifunctional glutamine amidotransferase/anthranilate phosphoribosyltransferase | Synthesizes Phosphoribosyl anthranilate (PRA) from anthranilate | 5.22 | 7.32 |  |
| *aroF* | DHAP synthase | Synthesizes DHAP from PEP and E4P | 0.44 | 7.24 |  |
| *trpA* | Tryptophan synthase alpha chain | Aldol cleavage of indole-3-glycerol phosphate to indole and glyceraldehyde-3-phosphate | 3.84 | 6.94 |  |
| *trpB* | Tryptophan synthase beta chain | Synthesizes L-tryptophan from indole and L-serine | 3.78 | 6.79 |  |
| *tyrA* | bifunctional chorismate mutase/prephenate dehydrogenase (T-protein) | Catalyses the rearrangement of chorismate to prephenate; converts prephenate to phenylpyruvate | 1.59 | 6.48 | (Wendisch, 2007) |
| *trpE* | Anthranilate synthase component 1 | Catalyses the synthesis of anthranilate | 4.38 | 6.38 |  |
| *trpC/trpF* | bifunctional indole-3-glycerol-phosphate synthase/ phosphoribosylanthranilate isomerase | Bifunctional enzyme that catalyzes two sequential steps of tryptophan biosynthetic pathway | 3.48 | 6.34 |  |
| *aroL* | Shikimate kinase 2 | Catalyses phosphorylation of the 3-hrdroxyl group of shikimic acid | -0.54 | 2.91 |  |
| *pheA* | bifunctional chorismate mutase/prephenate dehydrogenase (P-protein) | Catalyses conversion of chorismate to prephenate; converts prephenate to phenylpyruvate | 0.85 | 2.77 |  |
|  |  |  |  |  |  |

**Method S9: RNA-seq sample preparation and analysis**

Total RNA was isolated using RNeasy Mini Kit (Qiagen, USA). About 2.5µg of total RNA was taken for rRNA depletion using Ribo-Zero rRNA Removal Kit (Illumina Inc, CA, USA) Next, 50 ng of Qubit quantified ribo-depleted RNA was taken for fragmentation and priming. The fragmented and primed mRNA were further subjected to first strand synthesis in the presence of Actinomycin D (Gibco, life technologies, CA, USA) followed by second strand synthesis. The double stranded cDNA was purified using HighPrep PCR magnetic beads (Magbio Genomics Inc, USA). The purified cDNA was end-repaired, adenylated and ligated to Illumina multiplex barcode adapters as per NEBNext® Ultra™ Directional RNA Library Prep Kit protocol. RNA sequencing libraries were prepared with Illumina-compatible NEBNext® Ultra™ Directional RNA Library Prep Kit (New England BioLabs, MA, USA) where adapter ligated cDNA was purified using HighPrep beads and was subjected to 14 cycles of Indexing-PCR (37˚C for 15mins followed by denaturation at 98˚C for 30 sec), cycling (98˚C for 10sec, 65˚C for 75sec, and 65˚C for 5mins) to enrich the adapter-ligated fragments. The final PCR product (sequencing library) was purified with HighPrep beads, followed by library quality control check. The Illumina-compatible sequencing library was initially quantified by Qubit fluorimeter (Thermo Fisher Scientific, MA, USA) (Table S2) and its fragment size distribution was analyzed on Agilent Tapestation. After the libraries were constructed, paired end run was performed on Illumina’s HiSeq 2500 platform to obtain 2 x150 bp reads. The pre-processed high quality data was aligned to reference genome (GCF\_000022665.1 assembly was downloaded from NCBI) using Bowtie2. Bowtie with default parameter was run to identify the alignment percentage. HTSeqwas used to estimate and calculate gene abundance. Absolute counts for each genes were identified, which were used in differential expression calculations. DESeq was used to calculate the differentially expressed genes. Genes were categorized into up, down and neutrally regulated based on the log2fold change cutoff of 1 value.

**Table S10: Effect of amino acid supplementation on host cell protein expression**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Protein** | **Protein Entry Name** | **Area under the curve** | | | | **RSD% C7** | **RSD% T7** | **Log2FC C7/T7** |
| **C3** | **C7** | **T3** | **T7** |
| TrpA | TRPA\_Tryptophan synthase alpha chain | 5.21E+04 | 2.81E+05 | 0 | 0 | 23 | 0 | 0 |
| TrpB | TRPB\_Tryptophan synthase beta chain | 9.07E+04 | 1.46E+05 | 0 | 0 | 23 | 0 | 0 |
| TrpC | TrpCF\_Tryptophan biosynthesis protein | 5.94E+04 | 4.04E+05 | 0 | 0 | 12 | 0 | 0 |
| TrpGD | TRPGD\_Anthranilate synthase component II | 0 | 1.87E+05 | 0 | 0 | 16 | 0 | 0 |
| TrpE | TrpE\_Anthranilate synthase component 1 | 1.67E+05 | 0 | 0 | 0 | 0 | 0 | 0 |
| TrpCF | Tryptophan biosynthesis protein | 5.94E+04 | 1.18E+05 | 0 | 0 | 12 | 0 | 0 |
| PheT | Phenylalanine-tRNA ligase beta subunit | 0 | 2.11E+05 | 0 | 1.97E+05 | 38 | 17 | 0.098 |
| TnaA | Tryptophanase | 0 | 0 | 0 | 1.74E+06 | 0 | 20 | 0 |
| ThrA | Bifunctional aspartokinase / homoserine dehydrogenase 1 | 0 | 1.85E+05 | 0 | 0 | 26 | 0 | 0 |
| MetE | 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase | 6.13E+05 | 1.52E+06 | 0 | 2.25E+05 | 19 | 35 | 2.759 |
| MetH | Methionine synthase | 4.42E+04 | 1.33E+05 | 0 | 0 | 7 | 0 | 0 |
| GltB | Glutamate synthase | 0 | 4.17E+05 | 0 | 2.98E+05 | 13 | 31 | 0.487 |
| infB | Translation initiation factor IF3 | 0 | 1.17E+05 | 0 | 1.97E+05 | 7 | 33 | -0.750 |
| HtpG | Chaperone Protein | 0 | 1.29E+05 | 0 | 2.06E+05 | 9 | 24 | -0.680 |
| AspA | Aspartate ammonia-lyase | 0 | 0 | 3.25E+04 | 3.26E+05 | 0 | 31 | 0 |

**C3= Control 3rd h, C7=Control 7th h, T3= Test 3rd h, T7= Test 7th h**

**References**

Domka, J., Lee, J., & Wood, T. K. (2006). YliH (BssR) and YceP (BssS) regulate Escherichia coli K-12 biofilm formation by influencing cell signaling. *Applied and Environmental Microbiology*, *72*(4), 2449–2459. https://doi.org/10.1128/AEM.72.4.2449-2459.2006

Eletsky, A., Michalska, K., Houliston, S., Zhang, Q., Daily, M. D., Xu, X., Cui, H., Yee, A., Lemak, A., Wu, B., & others. (2014). Structural and functional characterization of DUF1471 domains of Salmonella proteins SrfN, YdgH/SssB, and YahO. *PloS One*, *9*(7), e101787. https://doi.org/10.1371/journal.pone.0101787

Lacour, S., & Landini, P. (2004). σS-dependent gene expression at the onset of stationary phase in Escherichia coli: function of σS-dependent genes and identification of their promoter sequences. *Journal of Bacteriology*, *186*(21), 7186–7195. https://doi.org/10.1128/JB.186.21.7186-7195.2004

Liu, W.-T., Karavolos, M. H., Bulmer, D. M., Allaoui, A., Hormaeche, R. D. C. E., Lee, J. J., & Khan, C. M. A. (2007). Role of the universal stress protein UspA of Salmonella in growth arrest, stress and virulence. *Microbial Pathogenesis*, *42*(1), 2–10. https://doi.org/10.1016/j.micpath.2006.09.002

Paul, B. J., Barker, M. M., Ross, W., Schneider, D. A., Webb, C., Foster, J. W., & Gourse, R. L. (2004). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell*, *118*(3), 311–322. https://doi.org/10.1016/j.cell.2004.07.009

Sharma, U. K., & Chatterji, D. (2010). Transcriptional switching in Escherichia coli during stress and starvation by modulation of σ70 activity. *FEMS Microbiology Reviews*, *34*(5), 646–657. https://doi.org/10.1111/j.1574-6976.2010.00223.x

Starosta, A. L., Lassak, J., Jung, K., & Wilson, D. N. (2014). The bacterial translation stress response. *FEMS Microbiology Reviews*, *38*(6), 1172–1201. https://doi.org/10.1111/1574-6976.12083

Suh, S.-J., Silo-Suh, L., Woods, D. E., Hassett, D. J., West, S. E. H., & Ohman, D. E. (1999). Effect of rpoS mutation on the stress response and expression of virulence factors in Pseudomonas aeruginosa. *Journal of Bacteriology*, *181*(13), 3890–3897. https://doi.org/10.1128/jb.181.13.3890-3897.1999

Vila-Sanjurjo, A., Schuwirth, B.-S., Hau, C. W., & Cate, J. H. D. (2004). Structural basis for the control of translation initiation during stress. *Nature Structural and Molecular Biology*, *11*(11), 1054. https://doi.org/10.1038/nsmb850

Weber, M. M., French, C. L., Barnes, M. B., Siegele, D. A., & McLean, R. J. C. (2010). A previously uncharacterized gene, yjfO (bsmA), influences Escherichia coli biofilm formation and stress response. *Microbiology*, *156*(1), 139–147. https://doi.org/10.1099/mic.0.031468-0

Wendisch, V. F. (2007). *Amino acid biosynthesis--pathways, regulation and metabolic engineering* (Vol. 5). Springer Science & Business Media.

Wick, L. M., & Egli, T. (2004). Molecular components of physiological stress responses in Escherichia coli. In *Physiological Stress Responses in Bioprocesses* (pp. 1–45). Springer. https://doi.org/10.1007/b93957

Yoshida, H., & Wada, A. (2014). The 100S ribosome: ribosomal hibernation induced by stress. *Wiley Interdisciplinary Reviews: RNA*, *5*(5), 723–732. https://doi.org/10.1002/wrna.1242