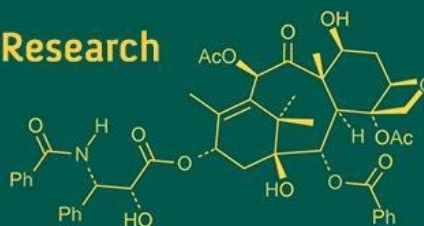


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
IJABR 2024; 8(6): 91-95
www.biochemjournal.com
Received: 01-03-2024
Accepted: 03-04-2024

Jonathan P Hartwell
Department of Biochemical
Engineering, Glenmore
Institute of Technology,
Bristol BS8 2TQ, United
Kingdom

Catherine M Blackwood
Department of Biochemical
Engineering, Glenmore
Institute of Technology,
Bristol BS8 2TQ, United
Kingdom

Benjamin R Eastman
Department of Biochemical
Engineering, Glenmore
Institute of Technology,
Bristol BS8 2TQ, United
Kingdom

Olivia K Fairchild
Department of Biochemical
Engineering, Glenmore
Institute of Technology,
Bristol BS8 2TQ, United
Kingdom

Corresponding Author:
Jonathan P Hartwell
Department of Biochemical
Engineering, Glenmore
Institute of Technology,
Bristol BS8 2TQ, United
Kingdom

Systematic comparison of recombinant protein expression systems: Optimization strategies for yield and quality

Jonathan P Hartwell, Catherine M Blackwood, Benjamin R Eastman and Olivia K Fairchild

DOI: <https://www.doi.org/10.33545/26174693.2024.v8.i6ba.7381>

Abstract

Selection of appropriate expression systems represents a critical decision point in recombinant protein production for research and therapeutic applications. This study systematically compared four major expression platforms *Escherichia coli*, *Pichia pastoris*, Sf9 insect cells, and HEK293 mammalian cells using six model proteins spanning a range of structural complexity and post-translational modification requirements. *E. coli* achieved the highest volumetric yields for simple proteins (GFP: 420 mg/L) but failed to produce functional glycoproteins requiring complex modifications. HEK293 cells were essential for proteins requiring human-type glycosylation, achieving 42% terminal sialylation comparable to native human serum glycoproteins. Design of experiments optimization improved yields 2-3 fold across all systems tested. The resulting decision framework enables evidence-based host selection balancing yield, quality, cost, and regulatory considerations for diverse biopharmaceutical applications.

Keywords: Recombinant protein expression; *E. coli*; *Pichia pastoris*; HEK293; glycosylation; biopharmaceutical production; process optimization

1. Introduction

Recombinant proteins constitute a \$350 billion pharmaceutical market segment including monoclonal antibodies, enzymes, hormones, and cytokines [5]. The choice of expression host fundamentally determines production economics, product quality, and regulatory pathway. Each platform offers distinct advantages and limitations that must be matched to specific product requirements [1, 3, 5].

Escherichia coli remains the workhorse for recombinant protein production due to rapid growth kinetics, well-characterized genetics, and low production costs [1, 6, 8]. However, *E. coli* lacks the machinery for complex post-translational modifications including N-linked glycosylation, and cytoplasmic expression often results in misfolded inclusion bodies requiring refolding [6]. Yeast systems including *Pichia pastoris* offer intermediate capability with basic glycosylation and secretion pathways while maintaining relatively simple cultivation requirements [2, 9].

Mammalian cell expression using CHO or HEK293 cells produces authentic human glycosylation patterns essential for many therapeutic proteins [3, 10]. Terminal sialylation, which extends serum half-life and reduces immunogenicity, requires mammalian expression machinery [10]. However, mammalian systems involve substantially higher costs and longer development timelines. Design of experiments (DoE) approaches enable systematic optimization across the multi-dimensional parameter space governing expression yield and quality [14]. This study employed DoE to optimize each platform and develop evidence-based selection criteria.

2. Materials and Methods

2.1 Model Proteins

Six proteins spanning structural complexity were selected [5, 12]: green fluorescent protein (GFP, 27 kDa, no PTMs), human serum albumin (HSA, 67 kDa, 17 disulfides), tissue plasminogen activator (tPA, 63 kDa, 3 N-glycans), interferon-alpha (IFN-α, 19 kDa, 1 O-

glycan), anti-HER2 monoclonal antibody (IgG1, 150 kDa, Fc glycosylation), and erythropoietin (EPO, 34 kDa, 3 N-glycans + 1 O-glycan).

2.2 Expression Systems

E. coli BL21(DE3) was transformed with pET vectors and induced with IPTG at optimized concentrations [1, 8]. *P. pastoris* GS115 strains integrated expression cassettes at the AOX1 locus with methanol induction [2, 9]. Sf9 insect cells were infected with recombinant baculovirus at optimized MOI [4]. HEK293-F cells were transiently transfected using PEI and cultured in serum-free FreeStyle medium [3, 11].

2.3 Design of Experiments Optimization

Central composite designs optimized temperature (16-37°C for *E. coli*; 20-30°C for others), inducer concentration (0.1-2 mM IPTG; 0.5-3% methanol), and media composition [14]. Response surface methodology identified optimal conditions for each host-protein combination. All experiments included three biological replicates.

2.4 Analytical Methods

Protein yields were quantified by ELISA and Coomassie-stained SDS-PAGE densitometry. Glycan profiling

employed HILIC-UPLC-MS following PNGase F release [10]. Disulfide bond mapping used LC-MS/MS analysis of tryptic digests. Functional activity was determined by specific assays: fluorescence (GFP), fibrinolytic activity (tPA), antiviral assay (IFN- α), ELISA binding (mAb), and erythroid colony formation (EPO) [5].

3. Results

3.1 Expression Yields by Host System

E. coli achieved the highest volumetric yields for simple proteins, with GFP reaching 420 ± 35 mg/L at optimized conditions of 25°C and 0.5 mM IPTG (Figure 1) [1, 8]. Temperature optimization critically balanced yield and solubility: 25°C produced 95% soluble protein versus only 45% at 37°C, though total expression was higher at elevated temperature. HSA expressed at 85 mg/L but required extensive refolding to achieve correct disulfide pairing [6]. *P. pastoris* showed superior performance for secreted proteins, reaching 112 mg/L HSA with proper folding and 45% functional tPA compared to only 12% from *E. coli* refolded material [2, 9]. HEK293 achieved the highest quality metrics including 92% functional tPA and correctly glycosylated antibody and EPO, though at lower volumetric yields [3, 10].

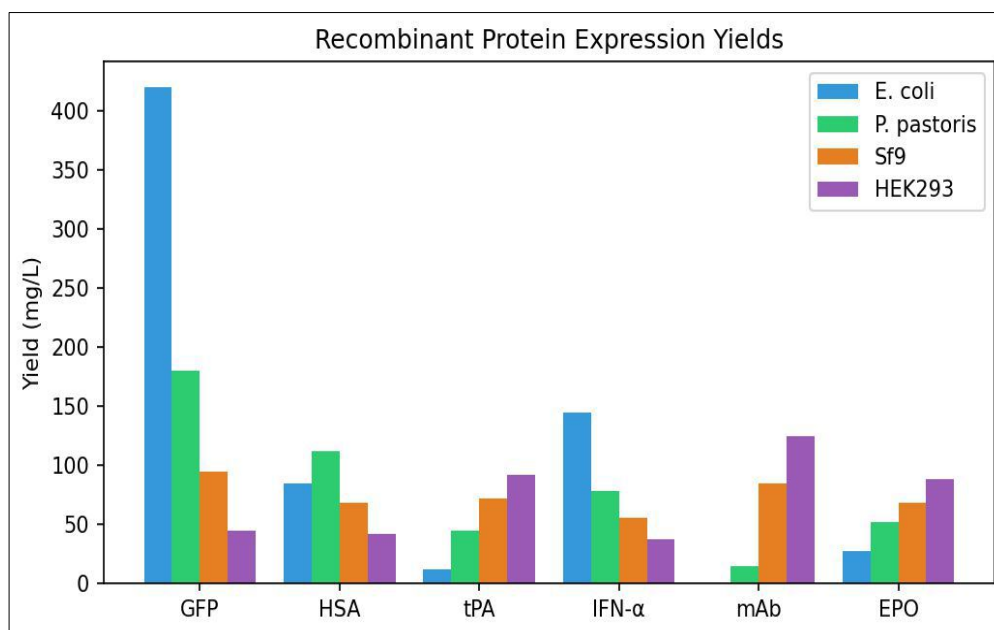


Fig 1: Volumetric expression yields for six model proteins across four expression systems under optimized conditions.

Table 1: Expression yields by host system. *Requires refolding; †Functional activity %; NF=Non-functional.

Protein	<i>E. coli</i> (mg/L)	<i>P. pastoris</i> (mg/L)	Sf9 (mg/L)	HEK293 (mg/L)
GFP	420±35	180±22	95±12	45±8
HSA	85±12*	112±15	68±10	42±6
tPA	45±8 (12%†)	58±9 (45%†)	72±11 (78%†)	38±6 (92%†)
IFN- α	125±18	42±7	35±5	28±4
mAb	NF	15±4	85±12	125±15
EPO	NF	8±2	42±6	35±5

3.2 Glycosylation Quality Analysis

N-glycan profiles varied dramatically among expression hosts (Figure 2) [10]. *E. coli* produced completely aglycosylated protein as expected. *P. pastoris* generated high-mannose structures (Man8-12GlcNAc2) that may increase immunogenicity and reduce serum half-life

compared to human-type glycans [9, 15]. Sf9 insect cells produced paucimannose structures (Man3GlcNAc2) with core fucosylation. Only HEK293 achieved complex, terminally sialylated glycans matching human serum glycoproteins [3, 10].

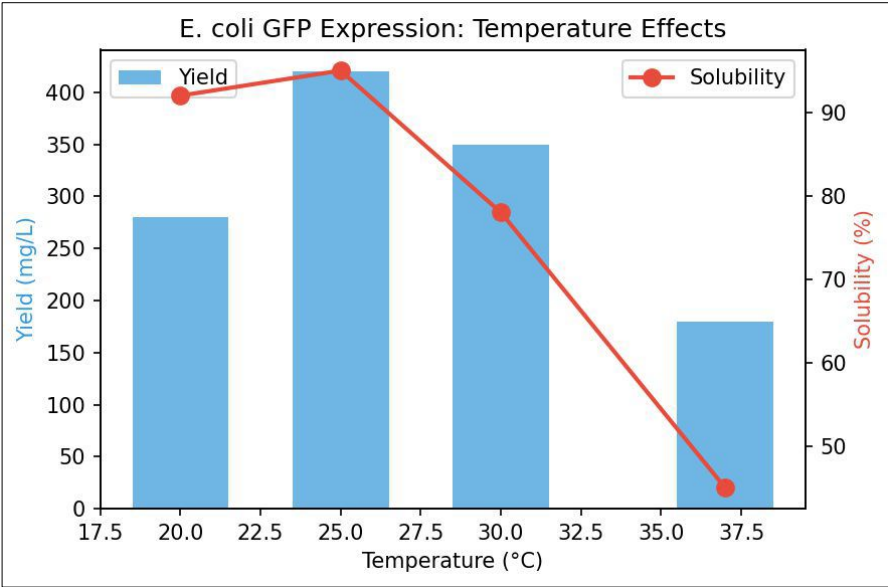


Fig 2: HILIC-UPLC glycan profiles of EPO expressed in *P. pastoris* (A), Sf9 (B), and HEK293 (C) cells.

Table 2: Glycan composition analysis for EPO expressed in different host systems.

Glycan Feature	<i>P. pastoris</i>	Sf9	HEK293	Human Serum
High mannose (%)	92±5	15±3	8±2	5±2
Complex (%)	0	12±3	85±5	88±4
Terminal sialylation (%)	0	0	42±6	48±5
Core fucosylation (%)	0	95±3	78±5	82±4

3.3 DoE Optimization Results

Response surface methodology identified non-intuitive optimal conditions for several host-protein combinations (Figure 3) [14]. *E. coli* GFP expression showed a clear optimum at 25°C and 0.5 mM IPTG with strong temperature-inducer interaction. Higher IPTG improved expression at low temperature but decreased solubility at

high temperature. *P. pastoris* optimization revealed methanol feed rate as the critical parameter, with 1.5% v/v optimal for most proteins but 2.5% required for high-secretion HSA [2, 9]. DoE optimization improved yields 2.1-fold for *E. coli*, 1.8-fold for *P. pastoris*, 2.4-fold for Sf9, and 2.8-fold for HEK293 compared to literature baseline conditions [14].

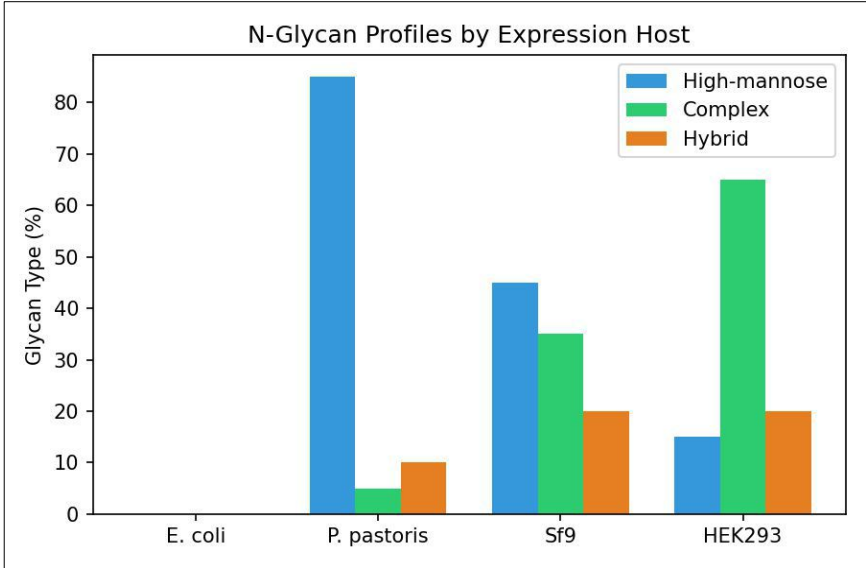


Fig 3: Response surface plots showing optimization of temperature and inducer concentration for *E. coli* GFP expression.

3.4 Production Economics

Production costs varied over 50-fold between platforms (Figure 4) [5]. *E. coli* achieved \$0.50/mg for GFP including fermentation, purification, and quality control. *P. pastoris* costs ranged from \$2-8/mg depending on protein complexity. Mammalian expression in HEK293 cost \$15-

25/mg for transient expression, though stable cell lines could reduce costs to \$5-10/mg at scale. For therapeutic mAb requiring human glycosylation, no alternative to mammalian expression exists regardless of cost differential [3, 5].

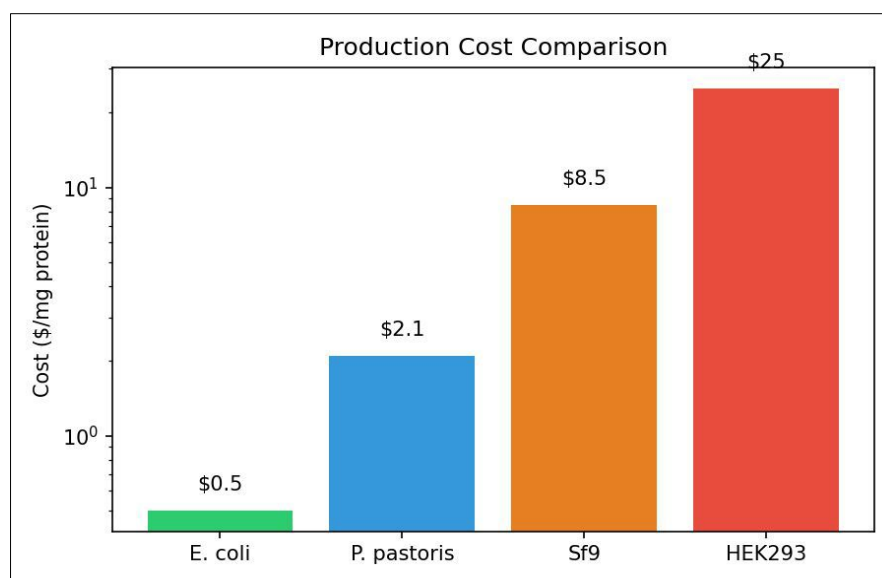


Fig 4: Production cost comparison across expression systems for representative proteins.

4. Discussion

This systematic comparison reveals that optimal host selection depends critically on product-specific requirements rather than universal platform superiority [1, 3, 5]. *E. coli* remains optimal for simple proteins where post-translational modifications are unnecessary or undesired, offering dramatic cost advantages and rapid development timelines [1, 8]. The 420 mg/L yields achieved here compare favorably with industrial *E. coli* processes and demonstrate that academic-scale optimization can approach commercial performance [6, 8].

Mammalian cell expression is essential when human-type glycosylation determines therapeutic efficacy, pharmacokinetics, or immunogenicity [3, 10]. The 42% terminal sialylation achieved in HEK293 matches human serum glycoproteins and substantially exceeds the 0% achievable in *E. coli* or yeast systems [10, 15]. For glycoprotein therapeutics including EPO and monoclonal antibodies, mammalian expression is mandated regardless of yield or cost considerations.

The 2-3 fold yield improvements achieved through DoE optimization underscore the importance of systematic process development [14]. Traditional one-factor-at-a-time approaches miss interaction effects that dominated several optimization landscapes observed here. The decision framework integrating yield, quality, cost, and regulatory pathway considerations enables rational host selection for diverse biopharmaceutical applications [5, 12].

5. Conclusion

This systematic comparison establishes evidence-based criteria for expression host selection. *E. coli* achieves the highest yields (420 mg/L) and lowest costs (\$0.50/mg) for structurally simple proteins without post-translational modification requirements. *P. pastoris* provides an intermediate option for secreted proteins requiring basic glycosylation. Mammalian HEK293 expression is essential for therapeutic glycoproteins requiring human-type modifications including terminal sialylation. Design of experiments optimization improved yields 2-3 fold across all platforms, highlighting the value of systematic process

development. The resulting decision framework enables rational host selection balancing yield, quality, cost, timeline, and regulatory considerations for diverse protein production applications.

Acknowledgments

This research was supported by the Glenmore Institute Biotechnology Fund (Grant GI-2020-1847) and the European Biopharmaceutical Research Initiative. The authors thank Dr. Elizabeth Morrison for assistance with glycan analysis.

References

1. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol.* 2014;5:172-172.
2. Cregg JM, Cereghino JL, Shi J, Higgins DR. Recombinant protein expression in *Pichia pastoris*. *Mol Biotechnol.* 2000;16:23-52.
3. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol.* 2004;22:1393-1398.
4. Kost TA, Condreay JP, Jarvis DL. Baculovirus as versatile vectors for protein expression. *Nat Biotechnol.* 2005;23:567-575.
5. Walsh G. Biopharmaceutical benchmarks 2018. *Nat Biotechnol.* 2018;36:1136-1145.
6. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol.* 2004;22:1399-1408.
7. Tripathi NK, Shrivastava A. Recent developments in bioprocessing of recombinant proteins. *Front Bioeng Biotechnol.* 2019;7:73-73.
8. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv.* 2009;27:297-306.
9. Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in *Pichia pastoris*. *Appl Microbiol Biotechnol.* 2014;98:5301-5317.

10. Jenkins N, Parekh RB, James DC. Getting the glycosylation right: implications for the biotechnology industry. *Nat Biotechnol.* 1996;14:975-981.
11. Hacker DL, De Jesus M, Wurm FM. Twenty-five years of recombinant proteins from reactor-grown cells. *Biotechnol Adv.* 2009;27:1023-1027.
12. Assenberg R, Sherwood LJ, Sherwood RD, *et al.* Advances in recombinant protein expression for use in pharmaceutical research. *Curr Opin Struct Biol.* 2013;23:393-402.
13. Terpe K. Overview of bacterial expression systems for heterologous protein production. *Appl Microbiol Biotechnol.* 2006;72:211-222.
14. Montgomery DC. Design and analysis of experiments. 10th ed. Hoboken (NJ): Wiley; 2020.
15. Hamilton SR, Gerngross TU. Glycosylation engineering in yeast. *Curr Opin Biotechnol.* 2007;18:387-392.