



Overview of Bacterial and Yeast Systems for Protein Expression

**Maheswara Reddy Mallu^{1*}, Siva Reddy Golamari¹,
Sree Rama Chandra Karthik Kotikalapudi¹ and Renuka Vemparala¹**

¹Department of Biotechnology, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India.

Authors' contributions

This work was carried out in collaboration among all authors. Author MRM designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors SRG and SRCKK managed the analyses of the study. Author RK managed literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JPRI/2021/v33i32A31722

Editor(s):

(1) Dr. Prem K. Ramasamy, Brandeis University, USA.

Reviewers:

(1) Qingbao Ding, The Wistar Institute, United States.

(2) Dongming Li, Fuzhou Medical College of Nanchang University, China.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/69665>

Review Article

Received 09 April 2021

Accepted 14 June 2021

Published 18 June 2021

ABSTRACT

Over the past decade the variety of hosts and vector systems for recombinant protein expression has increased dramatically. Researchers now select from among mammalian, insect, yeast, and prokaryotic hosts, and the number of vectors available for use in these organisms continues to grow. With the increased availability of cDNAs and protein coding sequencing information, it is certain that these and other, yet to be developed systems will be important in the future. Despite the development of eukaryotic systems, *E. coli* remains the most widely used host for recombinant protein expression. Optimization of recombinant protein expression in prokaryotic and eukaryotic host systems has been carried out by varying simple parameters such as expression vectors, host strains, media composition, and growth temperature. Recombinant gene expression in eukaryotic systems is often the only viable route to the large-scale production of authentic, post translationally modified proteins. It is becoming increasingly easy to find a suitable system to overexpress virtually any gene product, provided that it is properly engineered into an appropriate expression vector.

Keywords: *Prokaryotic systems; eukaryotic systems; recombinant; heterologous; protein expression; cloning hosts.*

*Corresponding author: E-mail: mahesh_bt@kluniversity.in;

1. INTRODUCTION

Protein quantity is an important consideration, since substantial time and effort are required to achieve higher quantities and is often easily obtained from prokaryotic or eukaryotic expression systems. Protein targets represent the majority of expressed proteins used in classical pharmaceutical drug discovery, which involves the configuration of a high-throughput screen of a chemical or natural product library in order to find selective antagonists or agonists of the protein's biological activity. In contrast to reagent proteins, therapeutic protein agents are the most demanding in terms of resource. Therapeutic proteins have intrinsic biological properties like medical drugs. The ultimate objective for expression of a therapeutic protein is the production of clinical-grade protein approaching or exceeding gram per liter quantities. For most expression systems this is not readily achievable. Other than bacterial and yeast expression, the most robust system for producing these levels is the Chinese hamster ovary (CHO) system. Due to the lack of proper post-translational modifications (e.g., glycosylation) in bacteria and yeast, CHO cell expression is often the only choice to achieve sufficient expression.

1.1 *Escherichia coli*

E.coli has rapid growth rate, ease of high cell density fermentation, affordability and the availability of excellent genetic tools, due to these characteristics of *E.coli* it have been extensively used for protein over expression [1]. *E. coli* is comprehensively used in laboratories which resulted in technologies to target protein over expression to different intracellular compartments (Fig. 1). This is beneficial because these compartments have various environments which might enable folding of particular proteins of interest [2]. The optimization of recombinant protein expression in *E.coli* was carried out widely by trial and error by altering the simple parameters like expression vectors, host strains, media composition, and growth temperature [3]. During the past years, pervasive studies have shown that replacing codons with a heterologous gene with synonymous ones used superiorly on the expression of host (codon optimization), and manipulating the nucleotide sequence of the translational initiation region could have a profound effect on recombinant protein yields [4]. In order to perform molecular epigenetic techniques successfully, it is important to have a

full understanding of the properties and of the various *Escherichia coli* host strains that are commonly used for the propagation and manipulation of recombinant DNA [5]. *E.coli* is an enteric rod-shaped Gram-negative bacterium with a circular genome of 4.6Mb [6]. It was initially chosen as a model system because of its ability to grow on chemically defined media and its rapid growth rate. In rich media, during the log phase of its growth, *E.coli* doubles every 20-30min; hence, with overnight incubation period, single selected organisms will double enough times to yield a colony on an agar plate 0r 1-2 billion cells per millilitre of the liquid media. The ease of its transformability and genetic manipulation has subsequently solidified the role of *E.coli* as the host of choice for the propagation, manipulation, and characterization of recombinant DNA [7]. In the past 60 years, intensive research was performed on *E.coli*, as a result, now a lot more is known about these bacilli than any other organisms on earth. A diverse variety of *E.coli* mutants were isolated and characterized. Almost all strains currently used in recombinant DNA experiments are derived from a single strain: *E.coli* K-12, which is isolated from the feces of a patient with diphtheria in 1922 [8]. Genetic state of the DNA in an organism can be indicated by Genotype. It is associated with an observed behaviour known as Phenotype [9]. Genotypes of *E.coli* strains are described in conformity with a standard nomenclature and genes are given in three-letter, lowercase, italicized names that are often mnemonics. The methylotrophic yeast *Pichia pastoris* is widely used as a host system for production of recombinant protein [10]. It is also commonly used as model organism for basic research of peroxisome and secretory organelles biosynthesis. Additionally, it has come into focus for the glycol-proteins production with human-like N-glycan structures, also for several metabolites and recombinant proteins [11]. Recently, *P. pastoris* has been reclassified into a new genus, namely *Komagataella* and divided into three species *K. pastoris*, *K. phaffii*, and *K. pseudopastoris* [12]. The strains GS115 and X-33 have been made available by Invitrogen and belong to the species *K. phaffii*. Apart from that, other strains belonging to either *K. pastoris* or *K. phaffii* are freely used by researchers. Moreover, in accordance with published literature, all strains are further named *P. pastoris*, standing for the entire genus *Komagataella*. At present, the genomes of two *P. pastoris* strains (DSMZ 70382 and GS 115) have been fully sequenced [13]. There from, two genome browsers were set up.

Until then, most data on genetic and physiological background for strain X-33 and process design relied on analogies to other, well studied yeasts like *Saccharomyces cerevisiae*. Accordingly, *P. pastoris* gene names follow mainly the format established for *S. Cerevisiae* [14].

1.2 *Pichia pastoris*

Pichia pastoris is alike *Saccharomyces cerevisiae* in terms of general growth conditions and handling [15]. Knowledge of basic microbiological and sterile techniques is essential for attempting to grow and manipulate any microorganism and also familiarity with basic molecular biology and protein chemistry is required. The number of functionally annotated genes (9.4 Mb; 5,450 ORFs) can be comparable to other yeasts, majority of metabolic enzymes are present in single copies and the count of actually secreted proteins is low, making secretory production of heterologous proteins attractive. For selection on neomycin and large-scale growth, a wild type of *Pichia* strain, X-33 is used [16]. It usually grows in YPD and in minimal media. For liquid cultures, plates, and slants the

growth temperature of *Pichia pastoris* is 28–30°C [17]. Protein expression will be damaged and Cell death can also be caused if there is growth above 32°C during induction [18-20].

Other major facts are, doubling time of exponential phase Mut+ or MutS *Pichia* in YPD is ~2 hours, Mut+ and MutS strains do not vary in growth rates unless grown on methanol, Doubling time of log phase Mut+ *Pichia* in methanol medium (MM) is 4–6 hours [21]. Doubling time of log phase MutS *Pichia* in MM is ~18 hours. OD600 = ~5 × 10⁷ cells/ml. YPD medium and YPD agar slants can be used to store cells for weeks to months. Each strain is to be streaked for single colonies on YPD. One colony must be transferred to YPD stab and grown for 2 days at 30°C. At 4°C, cells can be stored on YPD for several weeks. In order to store the cells for months to years, they should be stored at –80°C. A single colony of each strain is to be cultured overnight in YPD. Cells need to be harvested and suspended in YPD containing 15% glycerol at a final OD₆₀₀ of 50–100 approximately 2.5 × 10⁹–5.0 × 10⁹ cells/ml). In liquid nitrogen or a dry ice/ethanol bath cells are frozen and then stored at –80°C.

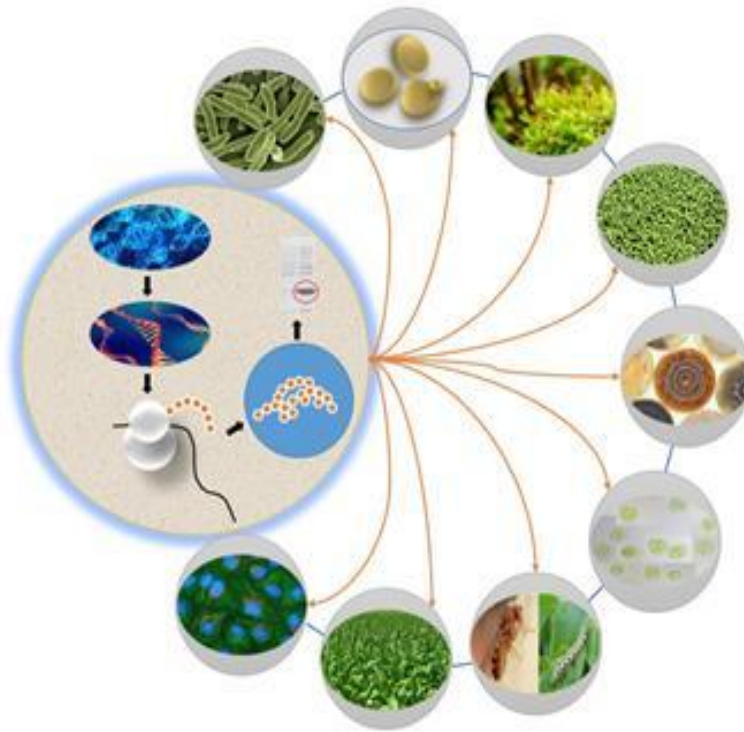


Fig. 1. Heterologous protein expression and production platforms using bacterial host systems

Source: frontiersin.org

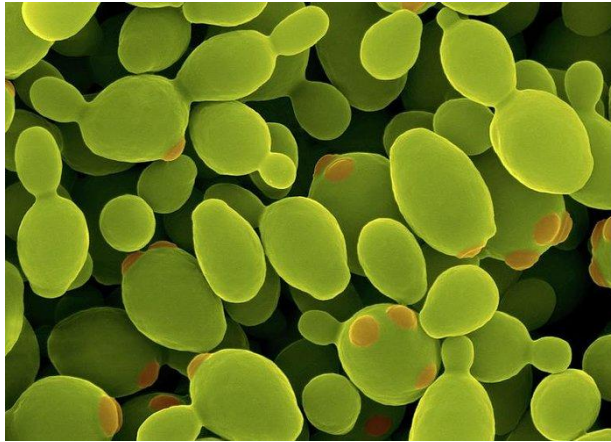


Fig. 2. *Pichia pastoris*
source: Igem.org

1.2.1 Recombinant protein production in *Pichia pastoris*

The use of *P.pastoris* as a cellular host for recombinant protein production unwaveringly increases [22]. It is easily manipulated, genetically and also cultured, can reach upto high cell densities (> 130 g l-1 dry cell weight) on methanol and glucose [23]. Equally salient, as eukaryote, it provides the potential for producing soluble, well folded proteins, which underwent post-translational modifications, such as glycosylation (O- and N-linked; less over-glycosylating than *S. cerevisiae*), disulfide bridge formation and processing of signal sequences [24]. For intracellular expression, the amino-terminal methionine residue is cleaved off, unlike proteins expressed in *E. coli*, or specific amino acid residues are likely to be phosphorylated, generating phospho-proteins without limitations and bottlenecks obtained by the secretory pathway and the protein can also be acetylated.

Transformation of the haploid, homothallic *P. Pastoris* host with recombinant DNA is mediated either by integrative plasmids or by autonomously replicating plasmids [25]. Homology of the introduced DNA with chromosomal locus is required for directed integration or replacement. Multiple integrations are commonly obtained on purpose. Only restriction at unique site, homologous to the *P. Pastoris* genome is required for integration [26]. Genetically stable transformants with high transcription rates are result of transformation mediated by electroporation. Most vectors are hybrids between bacterial and yeast sequences, possessing an origin of replication for *E. coli* and

selection markers. In addition, plasmids also contain a multiple cloning site (MCS). Both auxotrophic markers, like functional histidine dehydrogenase gene, which can be used in histidine dehydrogenase defective GS115, as well as dominant markers are present [27].

Additional benefits of the *P. Pastoris* system are strong constitutive and inducible promoter systems. A major issue of recombinant protein production is the transcription efficiency; therefore, the choice of the promoter is vital. The number of available promoters is limited and however, mainly contains the methanol inducible alcohol oxidase 1 (AOX1) promoter and the constitutive glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter [28]. Ellis and co-workers were the first to isolate the *P. pastoris* alcohol oxidase and two other genes regulated by methanol [29]. Alcohol oxidase is the vital enzyme in the methanol utilization pathway; specific for methylotrophic yeasts. It is encoded by two genes, *AOX1* and *AOX2*, and structurally and functionally characterized as well as reviewed several times. The AOX promoters are tightly regulated by a carbon source-dependent repression/induction mechanism; showing full repression during growth on glucose or glycerol excess conditions, and maximal induction during growth on methanol. In contradiction, the glyceraldehyde-3-phosphate dehydrogenase promoter (PGAP) is constitutively expressed, even though its strength differs depending on the carbon source used for cell growth. This offers an attractive alternative to PAOX1 on glucose, especially if induction by methanol may be inappropriate or inconvenient; simultaneously increasing cell viability. Furthermore, the activity

of PAOX1 in methanol-grown shake-flask cultures and slightly lower on methanol is not as strong as PGAP in glucose-grown shake-flask cultures [30].

2. CONCLUSION

All genes must be expressed to exhibit their biological activities. How genes are expressed and regulated is a central question in molecular biology and our knowledge in this area has been expanding enormously in recent years. The complexity of gene regulation is compounded by the fact that gene activities reach every corner of biology. Transcription is universally the first step toward expressing a gene. It is a highly regulated process. Understanding the molecular mechanisms of transcription regulation is of fundamental importance. For protein-coding genes, post-transcriptional steps, including pre-mRNA processing, mRNA transport and translation, can also play important roles in regulating gene expression.

CONSENT

Not applicable.

ETHICAL APPROVAL

Not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

- Makrides SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological reviews*. 1996;60(3):512-38.
- Vemuri PK, Anjum SM, Medida LS, Sai PH, Gummavajjala M, Panigrahi BK, Dash S, Senapati S. Studies on Physiological Effects and Plasmid Stability of JM109 Host System. *Journal of Pharmaceutical Research International*. 2020 ;13:102-8.
- Hayat SM, Farahani N, Golichenari B, Sahebkar A. Recombinant protein expression in *Escherichia coli* (*E. coli*): what we need to know. *Current pharmaceutical design*. 2018;24(6):718-25.
- Kante RK, Vemula S, Somavarapu S, Mallu MR, Gowd BB, Ronda SR. Optimized upstream and downstream process conditions for the improved production of recombinant human asparaginase (rhASP) from *Escherichia coli* and its characterization. *Biologicals*. 2018;56:45-53.
- Vemuri PK. Effect of plasmid presence and stability on growth of *Escherichia coli* DH5 α with use of drugs, chemicals, and radiation. *Asian Journal of Pharmaceutics (AJP): Free full text articles from Asian J Pharm*. 2019;13(3).
- Gupta S, Pandey S. The Structural and Functional Analysis of *Escherichia coli* Genome. *Food Molecular Microbiology*. 2019;14:1936.
- Monica Adapala, Vandana Kamjula, Renuka Vemparala and Praveen Kumar Vemuri. Recombinant heterologous protein expression in bacterial and yeast systems. *Int. J. of Adv. Res.* 2021;07-13. ISSN 2320-5407
- Bachmann BJ. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriological reviews*. 1972;36(4):525.
- Soorya L, Kolevzon A, Zweifach J, Lim T, Dobry Y, Schwartz L, Frank Y, Wang AT, Cai G, Parkhomenko E, Halpern D. Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. *Molecular autism*. 2013;4(1):1-7.
- Weidner M, Taupp M, Hallam SJ. Expression of recombinant proteins in the methylotrophic yeast *Pichia pastoris*. *JOVE (Journal of Visualized Experiments)*. 2010;25(36):e1862.
- Kumar Vemuri P, Veeravalli S. Expression, purification and characterization of human recombinant galectin 3 in *Pichia pastoris*. *Iranian Journal of Biotechnology*. 2014;12(2):3-8.
- Mallu MR, Vemula S, Ronda SR. Production, purification and characterization of recombinant human antithrombin III by *Saccharomyces cerevisiae*. *Electronic Journal of Biotechnology*. 2016;19(4):81-9.
- Love KR, Shah KA, Whittaker CA, Wu J, Bartlett MC, Ma D, Leeson RL, Priest M, Borowsky J, Young SK, Love JC. Comparative genomics and transcriptomics of *Pichia pastoris*. *BMC genomics*. 2016;17(1):1-7.
- Mallu MR, Vemula S, Ronda SR. Efficient single step chromatographic purification of recombinant human antithrombin (rhAT)

- from *Saccharomyces cerevisiae*. 3 Biotech. 2016;6(1):112.
15. Morton CL, Potter PM. Comparison of *Escherichia coli*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Spodoptera frugiperda*, and COS7 cells for recombinant gene expression. *Molecular biotechnology*. 2000;16(3):193-202.
 16. Roulet M, Välikkilä M, Chanut-Delalande H, Hämäläinen ER, Kessler E, Ala-Kokko L, Männikkö M, Bonod-Bidaud C, Ruggiero F. The Collagen V Homotrimer [$\alpha 1$ (V)]₃ Production Is Unexpectedly Favored over the Heterotrimer [$\alpha 1$ (V)]₂ $\alpha 2$ (V) in Recombinant Expression Systems. *Journal of Biomedicine and Biotechnology*;2010.
 17. Nurdiani D, Utami N, Putro EW, Kusharyoto W. Stability analysis of a *Pichia pastoris* recombinant clone expressing human insulin precursor. *IOP Conference Series: Earth and Environmental Science*. IOP Publishing. 2020 ;572(1):012008.
 18. Karbalaei M, Rezaee SA, Farsiani H. *Pichia pastoris*: A highly successful expression system for optimal synthesis of heterologous proteins. *Journal of cellular physiology*. 2020;235(9):5867-81.
 19. Kumar VP, Suryanarayana V, Sudhakar P, Rao KR. Cloning, Expression and purification of recombinant human Galectin-3 by using pPUZZLE shuttle vector in bacterial strain (DE3) host system. *Adv. Biotech*. 2013;13:9-12.
 20. Vemuri PK. Low Intensity Microwave Radiation on E. Coli Host System for Its Compatibility and Transformation Efficiency. *Low Intensity Microwave Radiation on E. Coli Host System for Its Compatibility and Transformation Efficiency*. *Biosciences Biotechnology Research Asia*. 2016;6(1):401-3.
 21. Singh A, Narang A. The Mut⁺ strain of *Komagataella phaffii* (*Pichia pastoris*) expresses P AOX1 5 and 10 times faster than Mut^s and Mut⁻ strains: evidence that formaldehyde or/and formate are true inducers of P AOX1. *Applied Microbiology and Biotechnology*. 2020;104(18):7801-14.
 22. Damasceno LM, Huang CJ, Batt CA. Protein secretion in *Pichia pastoris* and advances in protein production. *Applied microbiology and biotechnology*. 2012;93(1):31-9.
 23. Daly R, Hearn MT. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition: An Interdisciplinary Journal*. 2005;18(2):119-38.
 24. Heiss S. Establishment and characterization of a novel secretion system in the methylotrophic yeast *Pichia pastoris* (Doctoral dissertation, uniwienn).
 25. Cregg JM, Barringer KJ, Hessler AY, Madden KR. *Pichia pastoris* as a host system for transformations. *Molecular and cellular biology*. 1985;5(12):3376-85.
 26. Vogl T, Gebbie L, Palfreyman RW, Speight R. Effect of plasmid design and type of integration event on recombinant protein expression in *Pichia pastoris*. *Applied and environmental microbiology*. 2018;84(6).
 27. Shukla AK. *Membrane Proteins—Production and Functional Characterization*. Academic Press; 2015.
 28. Wang J, Wang X, Shi L, Qi F, Zhang P, Zhang Y, Zhou X, Song Z, Cai M. Methanol-independent protein expression by AOX1 promoter with trans-acting elements engineering and glucose-glycerol-shift induction in *Pichia pastoris*. *Scientific reports*. 2017;7(1):1-2.
 29. Koutz P, Davis GR, Stillman C, Barringer K, Cregg J, Thill G. Structural comparison of the *Pichia pastoris* alcohol oxidase genes. *Yeast*. 1989;5(3):167-77.
 30. Ardila-Leal LD, Poutou-Piñales RA, Morales-Álvarez ED, Rivera-Hoyos CM, Pedroza-Rodríguez AM, Quevedo-Hidalgo BE, Pérez-Flórez A. Methanol addition after glucose depletion improves rPOXA 1B production under the pGap in *P. pastoris* X33: breaking the habit. *SN Applied Sciences*. 2021;3(1):1-1.

© 2021 Mallu et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
 The peer review history for this paper can be accessed here:
<http://www.sdiarticle4.com/review-history/69665>