# Recombinant Protein Production in *Escherichia coli*

## A Retrospective and Prospective Review

**Peter Alexander, Christopher Dale, Sunil Desai, and Dan Rudolph**

*Escherichia coli* launched the biopharmaceutical manufacturing industry in 1982 with the approval of recombinant human insulin (Humulin), licensed by Genentech to Eli Lilly & Company. The long history of this microorganism's use in fundamental genetic research led to a serendipitous convergence of methodologies for DNA extraction and manipulation, bacterial transformation, and induction of heterologous protein expression. Once *E. coli* was established as a familiar host organism to regulatory authorities, further approvals followed steadily in 1985–2007 (Figure 1). The sole challenge to its leadership position came from *Saccharomyces cerevisiae* as the only credible (yeast-based) alternative microbial expression system.

In total, 38 products expressed in *E. coli* have been approved by the FDA compared with 11 in *S. cerevisiae* and 55 in mammalian cell culture. Microbial systems account for 30–40% of the total biopharmaceutical industry output. The segment may be further subdivided to account for the distribution of microbial expression systems and mode of expression (Figures 2 and 3). Analyses of the modes of expression in *E. coli* have shown a fundamental shift in product form and cellular localization over the past two decades. Expression strategies involving accumulation of insoluble protein as intracellular (cytoplasmic) inclusion bodies have been replaced with strategies for soluble protein expression localized to the cytoplasm or translocated into the periplasmic space. Accumulation of periplasmic inclusion bodies and passive secretion into the fermentation medium have been reported in rare cases.

The strategy for expression of recombinant proteins in *E. coli* is deceptively simple: The bacteria are grown in batch or fed-batch mode until sufficient cell density is achieved (usually to exponential or the early stage exponential–stationary transition), then controlled expression of the target protein is induced with a bias to either soluble or insoluble form and localization as described above. Maintaining specific productivities coupled with high cell densities maximizes volumetric productivity on a unit fermentation volume basis. Understanding of the metabolic and engineering aspects of fed-batch cultures predates the biopharmaceutical industry, and little has changed since 1987 except for some advances in sensor technology and automation control (software and hardware). The evolution of E. coli as a production platform has therefore focused on molecular biology and drug design (Figure 4).

### DOMINANCE OF THE T7 EXPRESSION SYSTEMS

Within the field of *E. coli* expression, one particular system has achieved and maintained a leadership position. The pET T7 (plasmid for expression by T7 RNA polymerase) expression system was first introduced by Studier and Moffet in 1986 (**1**) and later further developed as a fundamental expression tool for the broad scientific community (**2**). As listed in the box on this page, the key elements of this system confer significant advances in control of expression if compared with transcriptional mechanisms that rely solely on host RNA polymerase.

In the T7 system, a target-protein–coding sequence is under the control of a promoter exclusively recognized by the T7 RNA polymerase, so control of expression is mediated at the level of transcription initiation. T7 RNA polymerase is only expressed when the transacting T7 RNA polymerase gene is induced by addition of lactose or a lactose analog such as IPTG.

**The Details of *E. coli* Genetic Engineering:** The preeminent position of pET plasmids warrants their detailed discussion. All such plasmids are based on pBR322. The T7 promoter is cloned into the plasmid with a unique cloning site, a transcription terminator, and these expression vectors contain translation initiation signals. *****Chris: Did that edit change the meaning at all? —CAS***** Specifically, the ϕ10 promoter was inserted into the BamHI site of plasmid BR322 (in the opposite direction from the tet promoter) for transcription vectors. The ϕ10 promoter with the *s*10 translation-initiation region from the gene *10* protein of λ was inserted into the BamHI site of pBR322 (in the opposite direction from the tet promoter) for expression vectors, with an NdeI or NcoI site available for cloning. The Tϕ transcription termination signal was inserted downstream of the BamHI cloning site of some vectors. The native *bla* gene for ampicillin resistance was thus oriented such that it could be expressed from the T7 promoter along with a target gene (the *tetR* gene was interrupted by the T7 promoter).

The pET3 series incorporated the above elements. The pET6 and pET7 removed the 29-nucleotide leader RNA after the ϕ10 promoter and in front of the BamHI cloning site. The *bla* gene was replaced with the gene for kanamycin resistance in the pET-9 series, with the gene inserted in the opposite direction from the T7 promoter. Series pET10 and pET11 introduced the T7*lac* promoter, in which the *lac* operator sequence was introduced into the pET7 vector immediately downstream of the start site of the T7 ϕ10 promoter.

For additional control, the *lacI* element was added to those vectors to tighten regulation at both the *lacUV5* promoter for expression of T7 RNA polymerase and the target product on the plasmid. Cloning sites were developed to allow for introduction of a coding sequence in frame with the first 11 amino acids of the gene *10* protein and an arginine (based on three different orientations of the BamHI cloning site). Alternatively, DNA coding sequences could be cloned directly into the NdeI/NcoI site to eliminate fusion with the gene *10* protein that was also possible with these higher numbered pET vectors.

Culture growth inhibition due to low-level expression of toxic products was recognized as a possible consequence of low-level basal expression of T7 RNA polymerase due to high levels of activity of the T7 RNA polymerase. The primary method for mitigating this problem was to introduce pLysS and pLysE vectors to expression strains for constitutive expression of T7 lysozyme, which inhibits T7 RNA polymerase and thus minimizes basal T7 RNA polymerase activity. Alternatively, metabolic control with glucose could lower basal activity of T7 RNA polymerase in uninduced cells.

Some potential improvements to the pET system were envisioned in 1990, including placing the T7 RNA polymerase under the control of a promoter that is more tightly repressed than *lacUV5* and use of an antisense promoter to reduce the basal expression of toxic genes. The *E. coli* RNA polymerase promoter in antisense direction can stabilize toxic target plasmids, but the level of product expression may depend on that gene and the strength of the antisense promoter.

After 1990, the availability of fusion constructs for many purposes increased dramatically (**#**). Purification tags such as the His-tag, the GST-tag, KSI (for small proteins), the Strep-tag, and the T7-tag allow for simple affinity-based purification of products from crude cell lysates. Secretion tags such as the *pelB* signal sequence and the Dsb-tag allow for periplasmic localization (the *pelB* signal is processed away in transit to the periplasmic space and can leave a native protein sequence without a leading methionine, a strategy that may require licensing from Xoma for commercial use) (**3**). The KSI sequence can add stability to small proteins and peptides that might otherwise be degraded by proteases. The Dsb-tag, *pelB* leader sequence, and Trx-tag could improve soluble expression. The GST-tag, HSV-tag, S-tag, Step-tag, and T7-tag allow for Western blot analysis using tag-specific antibodies. The PKA (protein kinase A) tag allows in vivo phosphorylation of expressed proteins.

## Host *E. coli* Strains

Two strains of *E. coli* were described by Studier et al (**4**): HMS174, which was described primarily as more of a cloning strain for its *recA*⁻ properties, and BL21, which was the preferred protein-expression strain because of *lon* and *ompT* protease deficiencies. The T7 RNA polymerase gene is transferred to the expression strains either by infection of a culture with bacteriophage CE6 or integration of a T7 gene 1 RNA polymerase into the host genome. Bacteriophage CE6 is a λ derivative that carries the T7 polymerase gene under the control of the phage pL and pI promoters. CE6 infection can produce such high levels of target product RNA expression that normal phage development cannot occur.

Integration of the T7 RNA polymerase into the E. coli genome was accomplished through bacteriophage DE3, another derivative of λ phage. Its primary features were imparted by insertion of the T7 *gene 1* RNA polymerase coding region into the int gene of λ phage with the *lacUV5* promoter (*lac* operator) and *lacI* gene. Inactivation of the *int* gene causes the DE3 bacteriophage to require a helper for integration into or excision from a chromosome. In practice, the frequency of shedding is very low, although a lytic cycle can be induced under certain conditions of physiological stress. Use of *recA*⁻ host strains can eliminate the possibility of lytic events. Combining the *lacUV5* promoter with the *lac* operator and *lacI* gene product allows for tight regulation of T7 RNA polymerase. The specificity of the T7 promoter, in turn, allows target protein expression to be initiated only in the presence of T7 RNA polymerase.

A wide range of well-characterized *E. coli* strains are now available for heterologous protein expression, and have originated from K or B parent strains. BL21(DE3) is the most commonly used and available through many sources (Novagen/EMD, Promega, Edge BioSystems, Sigma-Aldrich, Genechoice, and others). This strain has gained wide popularity due to its growth characteristics, target protein expression levels, and protease deficiencies (*lon* and *ompT*). BLR is a *recA*⁻ mutant of the BL21 strain that decreases the likelihood of excision of the DE3 under stress.

Orgami strains of *E. coli* are either K-12 or B strains with mutations in their theoredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which improves the formation of disulfuide bonds in the cytoplasm. The Rosetta series includes B or K-12 strains that supply tRNAs for 6–7 codons that are otherwise rare in *E. coli* and can increase translation efficiency. The Rosetta-gami offering combines the *gor/trx* knockouts and the expression of rare tRNAs. Tuner strains are BL21 derivatives with a *lacZYA* deletion, in which a *lacY* (*lac* permease) gene deletion allows for uniform entry of IPTG into all cells for a concentration-dependant supply of IPTG (and thus tunable regulation of product expression).

An estimate of the current distribution of *E. coli* host strains in preclinical and clinical development

(Figure 5) suggests that B strains predominate due to the laboratory consumer market penetration of BL21 and BL21(DE3). It is not unusual, however, to switch from B- to K-derived strains during the (early stage) evolution of a drug product. Certainly HMS 174 and W3110 may exhibit product-specific advantages with respect to processing and productivity/yield.

The technology dominance of T7-based systems and associated licensing requirements for commercial production have led to a number of alternative *E. coli* expression systems from several technology developers. Xoma pioneered the exploitation of the arabinose-inducible operon for control of protein expression. Coupled with use of the *pelB* leader sequence from *Erwinia carotovora,* it conferred the ability to have recombinant proteins expressed in soluble form and translocated efficiently to the periplasmic space. This expression platform has been targeted to the manufacture of antibody-derived molecules (scFVs).

An alternative to conventional modes of induction (chemical, thermal, and nutrient starvation) has recently received widespread attention. Autoinduction, also referred to as hyper-expression, has been recently studied in a T7-promoter–based expression system by Studier and other researchers (**4, 5**). Auto-induction in *E. coli* cultures expressing recombinant proteins under T7 promoter control occurs when these cultures are allowed to grow to saturation in a complex medium. They express a product of interest without addition of any external inducing agent (such as IPTG or lactose).

Based on numerous media experiments, Studier observed that the autoinduction happens because of the trace amounts of lactose in the growth medium. To reduce the basal expression of a T7 promoter, a lac operator sequence was placed downstream of the promoter (**6**). It was also observed that lower rates of aeration enhanced autoinduction, whereas presence of glucose in the medium blocks the phenomenon by preventing uptake of residual lactose. By contrast, glycerol-containing fermentation media did not interfere with autoinduction. Studier also evaluated the effects of amino acids on autoinduction, finding serine to have the most negative impact (but it was limited to the log phase alone).

Autoinduction could be used for expressing a wide range of proteins in *E. coli*. It is particularly convenient for high-throughput studies in which cultures must be induced at similar biomass levels, although individual cultures typically reach saturation at slightly different times. Autoinduction methods are more easily performed under such conditions because induction is intrinsically triggered. The autoinduction methodology also could be appropriate for large-scale manufacture of recombinant proteins under moderate cell density conditions.

## Competing Systems

*E. coli* is beginning to face serious competition from alternative bacterial expression hosts. Although expression systems in Gram-positive bacteria (*Bacillus* species) and other Gram-negative hosts (for example *Salmonella* species) have long been in existence, it is only within the past two years that a credible system for recombinant protein biopharmaceutical production has emerged. Dowpharma launched the Pfenex expression system based on *Pseudomonas fluorescens* bacteria for high-level production of soluble proteins under simple fermentation conditions.

The principal difference between the launch of Pfenex and those of other microbial (bacterial, yeast, and fungal) expression systems is the level and depth of its commercial support. Dowpharma provided a concerted technical marketing effort in conjunction with a business model that included in-house development and system support. Whether this will translate to widespread acceptance for biologics manufacture and ultimately commercialization of Pfenex-expressed protein drugs remains to be determined. Nevertheless, the concept that *E. coli* has no rational monopoly on expression of heterologous proteins is gaining some recognition.

## A Future in Recombinant Protein Production

The global biopharmaceutical market leadership held by mammalian cell culture and threats from alternative microbial expression systems raises the question of whether *E. coli* can sustain a dominant position in the future. Several factors will offset erosion of this technology's market share based primarily on inertia (reluctance to change): overwhelming acceptance of *E. coli* in academic and research institute laboratories, changes in drug design, and the growing ability of *E. coli* to emulate eukaryotic posttranslational modifications.

The leading role of (whole) antibodies as therapeutic agents will be increasingly challenged by antibody-derived molecules and the emergence of active and highly stable antibody components ("nanobodies"). Efforts to produce both heavy- and light-chain antibody components within microbial hosts (bacteria and yeasts), are gaining momentum, along with assembly and export to the periplasmic space or extracellular secretion. Antibody-derived molecules and nanobodies are ideally suited to microbial expression systems for soluble protein expression.

An exciting development within microbial expression systems is the ability to perform controlled glycosylation of recombinant proteins. Initial successes have been demonstrated within yeast systems, most notably human-like glycoforms in Pichia pastoris (**7, 8**) and more recently emerging strategies to augment bacterial systems with glycosylation machinery. With the exception of Campylobacter jejuni, bacteria do not naturally glycosylate the proteins they express. However both posttranslational and cotranslational approaches to bacterial glycosylation are under active research. Aebi and co-workers at ETH Zurich have exploited a novel N-linked glycosylation pathway from *C. jejuni* in *E. coli* to confer the ability to conjugate simple monosaccharides to asparagine residues within expressed heterologous proteins. An alternative approach to glycosylation in *E. coli* is uptake of N-acetylglucosamine (GlcNAc) or N-acetylgalactosamine (GalNAc) derivatives of amino acids and incorporation during protein synthesis.

Increased efficacy of next-generation therapeutics coupled with high expression titers in microbial systems could present a longer-term shift of manufacturing capacity within the biopharmaceutical industry. Clearly, mammalian manufacturing systems will remain the principal core technology for marketed drugs. Indeed the market share captured by mammalian-cell-culture–derived products may be expected to increase over the short term as antibodies currently under clinical evaluation achieve commercialization. Nevertheless, microbial expression platforms may capture a greater proportion of an enlarged market as the current segregation of products becomes less defined because of changes in drug design and enabling new technologies in *E. coli*.

The entire product life cycle — from research to commercialization — still sustains a healthy interest in exploitation of *E. coli* systems. Based solely on technology use by companies in preclinical and early stage clinical development, the continued dominance of *E. coli* may be expected to be sustained for 5–10 years without major technological enhancements. However, the longer-term future for this tenured servant of biotechnology will depend on acceptance/uptake of new technologies coupled with the approval and commercialization of custom-designed protein therapeutics.

## References

**1** Studier FW, Moffatt BA. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J. Mol Biol.* 189(1) 1986: 113–30.

**2** Studier FW, et al. Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol.* 185, 1990: 60–89.

**#** pET System Manual, TB055, 10th edition, Rev. B. Novagen: San Diego, CA, April 2003.

**3** *Microbial Expression.* Xoma (US) LLC: Berkeley, CA, 2007; www.xoma.com/wt/page/microbial_expression (accessed 7 September 2007).

**4** Studier WF. Protein Production By Auto-Induction in High-Density Shaking Cultures. *Protein Expr. Purif.* 41(1) 2005: 207–234.

**5** Blommel PG, et al. Factorial Design Improvement of Auto-Induction Medium. *Biotechnol. Prog.* 23(3) 2007: 85-98.

**6** Dubbendorff JW, Studier FW. Controlling Basal Expression in an Inducible T7 Expression System By Blocking the Target T7 Promoter with lac Repressor. *J. Mol. Biol.* 219, 1991: 45–59.

**7** Hamilton SR, et al. Production of Complex Human Glycoproteins in Yeast. *Science* 301, 2003: 1244–1246.

**8** Gerngross, TU. Advances in the Production of Human Therapeutic Proteins in Yeasts and Filamentous Fungi. *Nature Biotechnol.* 22 (11) 2004: 1409–1414.



***Peter Alexander,*** *corresponding author* ***Christopher Dale, Sunil Desai,*** *and* ***Dan Rudolph*** *are with Lonza Biopharma Services, 97 South Street, Hopkinton, MA 01748, USA; chris.dale@lonza.com; www.lonza.com.*


## Key Elements of the T7 Expression System

Two-step control process for tight regulation of gene expression (induction of T7 RNA polymerase required for expression of target protein)

Bacteriophage T7 RNA polymerase encoded within E. coli chromosome

Specificity of the T7 promoter

Absence of T7 promoter sequence in *E. coli* genomic or plasmid DNA sequences

Speed of transcriptional elongation by T7 RNA polymerase

Efficiency and rarity of T7 termination sequence

**Figure 2:** Distribution of microbial expression systems encountered in outsourced projects

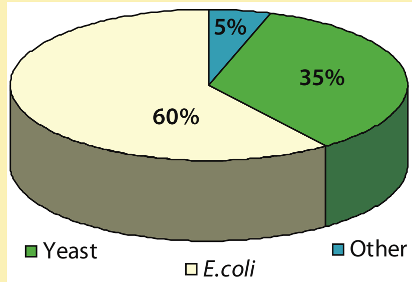
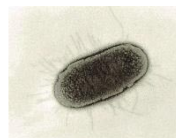



**Figure 3:** Distribution of expression mode, cellular localization, and induction system in *E. coli*–based outsourced projects

**Expression System:**
Cytosolic insoluble (IB) 35%
Cytosolic soluble 40%
Periplasmic soluble 20%
Periplasmic insoluble (IB) ~rare
Extracellular soluble <5%



**Induction System:**
IPTG 80%
Arabinose 10%
Phosphate depletion 5%
Other 5% ~thermal/constitutive

**Figure 1:** Analysis of FDA-approved recombinant protein biopharmaceuticals (1982–2007)

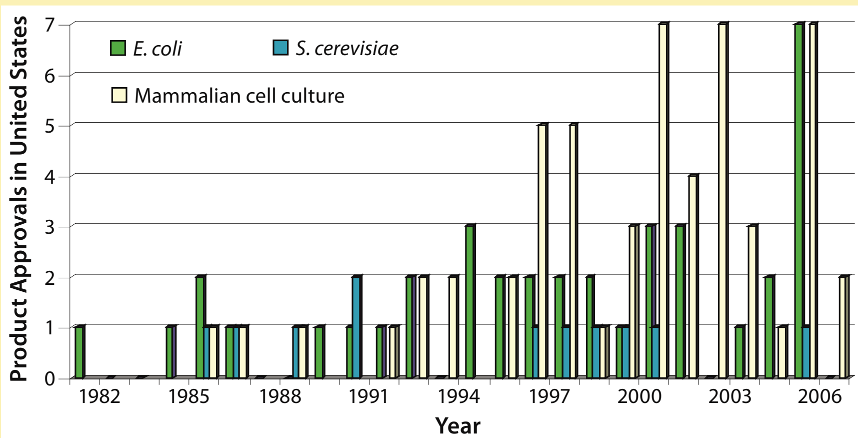


**Figure 4:** Analysis of FDA-approved recombinant protein biopharmaceuticals (1982–2007)

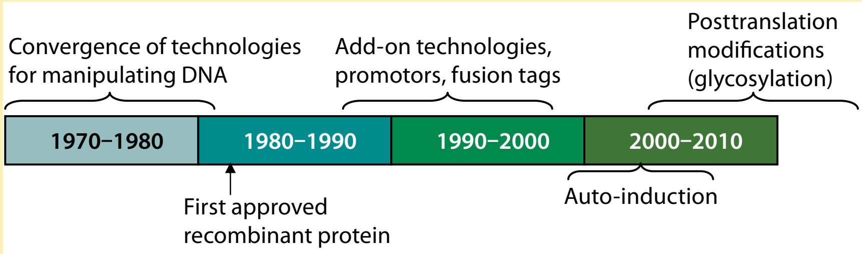



**Figure 5:** Distribution of *E .coli* host strains for production of recombinant proteins (data based on outsourced projects)

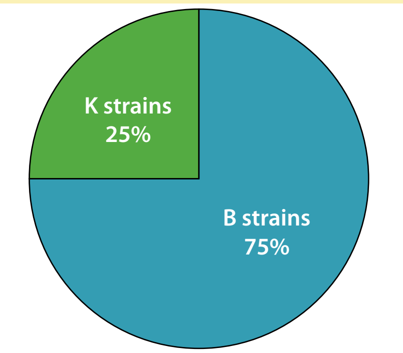
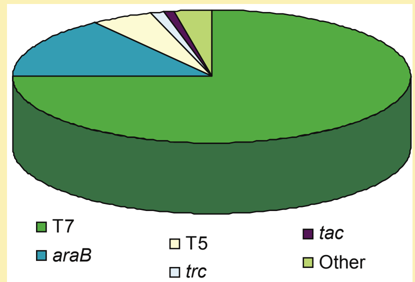



**Figure 6:** Frequency of promoter usage in *E. coli* systems (based on a survey of outsourced projects)




## Current Trends

Below are listed current trends in the production of recombinant protein biopharmaceuticals with *E. coli*.

Shift from inclusion body systems to soluble protein production (cytoplasm/ periplasmic space)

Significant improvements in expression levels

Bolt-on technologies for enhancing the effectiveness of *E. coli* expression (diversity of promoters, induction mechanisms, protease deficient strains, difficult-to-express proteins)

Engineering of *E. coli* strains to confer an ability to glycosylate proteins in a controlled and predictable manner

Dominant platform for protein production and reemergence of *E. coli* for pDNA manufacture