STAR Protocols

CellPress
OPEN ACCESS

Primer

A concise guide to choosing suitable gene expression systems for recombinant protein production

Anja Schütz,[1] Frank Bernhard,[2] Nick Berrow,[3] Johannes F. Buyel,[4] Frederico Ferreira-da-Silva,[5] Jurgen Haustraete,[6] Joop van den Heuvel,[7] Jan-Erik Hoffmann,[8,16] Ario de Marco,[9] Yoav Peleg,[10] Sabine Suppmann,[11,17] Tamar Unger,[12] Martine Vanhoucke,[13] Susanne Witt,[14] and Kim Remans[15,*]

[1]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Technology Platform for Protein Production & Characterization, Robert-Rössle-Str. 10, 13125 Berlin, Germany

[2]Institute of Biophysical Chemistry, Centre of Biomolecular Magnetic Resonance, Goethe-University of Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

[3]Protein Expression Core Facility, Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain

[4]Univeristy of Natural Resources and Life Sciences, Vienna (BOKU), Department of Biotechnology (DBT), Institute of Bioprocess Science and Engineering (IBSE), Muthgasse 18, 1190 Vienna, Austria

[5]Instituto de Biologia Molecular e Celular (IBMC) and Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto, Portugal

[6]VIB, Center for Inflammation Research & Ugent, Department of Biomedical Molecular Biology, Technologiepark-Zwijnaarde 71, 9052 Ghent, Belgium

[7]Helmholtz Centre for Infection Research (HZI), Department of Structure and Function of Proteins, Inhoffenstrasse 7, 38124 Braunschweig, Germany

[8]Protein Chemistry Facility, Max-Planck-Institute of Molecular Physiology, Otto-Hahn-Str. 11, 44227 Dortmund, Germany

[9]Laboratory of Environmental and Life Sciences, University of Nova Gorica, Vipavska Cesta 13, 5000 Nova Gorica, Slovenia

[10]Structural Proteomics Unit (SPU), Department of Life Sciences Core Facilities (LSCF), Weizmann Institute of Science, Rehovot 7610001, Israel

[11]Protein Expression and Purification Core Facility, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

[12]Structural Proteomics Unit (SPU), Department of Life Sciences Core Facilities (LSCF), Weizmann Institute of Science, Rehovot 7610001, Israel

[13]BCCM/GeneCorner Plasmid Collection, Department of Biomedical Molecular Biology, Ghent University, Technologiepark-Zwijnaarde 71, 9052 Gent, Belgium

[14]Centre for Structural Systems Biology (CSSB), University Medical Center Hamburg-Eppendorf (UKE), Notkestr. 85, 22607 Hamburg, Germany

[15]European Molecular Biology Laboratory (EMBL), Protein Expression and Purification Core Facility, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[16]Present address: ATTO-TEC GmbH, Martinshardt 7, 57074 Siegen, Germany

[17]Present address: Immunology, Infection and Pandemic Research IIP, Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Nonnenwald 2, 82377 Penzberg, Germany

*Correspondence: kim.remans@embl.de
https://doi.org/10.1016/j.xpro.2023.102572

SUMMARY

This overview guides both novices and experienced researchers facing challenging targets to select the most appropriate gene expression system for producing a particular protein. By answering four key questions, readers can determine the most suitable gene expression system following a decision scheme. This guide addresses the most commonly used and accessible systems and provides brief descriptions of the main gene expression systems' key characteristics to assist decision making. Additionally, information has been included for selected less frequently used "exotic" gene expression systems.

INTRODUCTION

The ready availability of biological resources and related genetic sequence data combined with advances in protein production systems have enabled many laboratories to begin production of their own proteins for use as biological reagents. This allows researchers to control the costs and the availability and quality of the proteins used in their experiments.[1] However, many researchers that are tasked with producing recombinant proteins in their respective laboratories have little or no previous experience with the gene expression systems available. This guide evaluates the key characteristics of the most commonly used gene expression systems in order to direct researchers wishing to begin protein production to the most appropriate system for their needs and resources. The evaluation of the main features of the systems are based on a survey (see supplemental information, "P4EU survey results") conducted among the members of the Protein Production and Purification Partnership in Europe (P4EU, https://p4eu.org), which is a network of professionals active in various protein production laboratories and platforms. We gathered and evaluated information from (mainly European) protein production centers represented by 60 experienced scientists. Their overall experience corresponds to the production of thousands of proteins belonging to many different classes.

The information on the different gene expression systems is presented in two ways:

1. A decision scheme that uses four key questions to help determine the most optimal gene expression system for a certain target protein. These questions are based on the biological characteristics of the protein of interest and direct the reader through key decision points, from which the different branches of the scheme can be followed to decide on the most appropriate gene expression system (Figure 1).
2. At-a-glance comparison of the key characteristics of the most commonly used gene expression systems, which includes features such as the ease of use, the speed, the capacity of each system for protein production, folding, (complex) assembly and secretion, and the estimated running costs. The results of these evaluations are summarized graphically in Figure 2.

The biological characteristics of the target protein and, to a lesser extent, the planned downstream applications will dictate the most appropriate gene expression system.[2] Therefore, it is important to collect information about the native localization of the protein of interest (intracellular, secreted, or membrane protein), the size/molecular weight, whether it is a single- or multi-domain protein, the number of disulfide bonds that are present, and post-translational modifications (e.g., glycosylation) and/or cofactors that might be required for correct folding and structural integrity. Some proteins that form part of multi-subunit complexes might not be stable on their own and hence require co-expression with their interaction partners. This type of information can be gathered by searching the scientific literature, consulting the Uniprot database (https://www.uniprot.org) and using bio-informatic tools such as ProtParam (https://web.expasy.org/protparam/) and AlphaFold structural predictions (https://alphafold.ebi.ac.uk).

Generally, the first choice for the production of prokaryotic target proteins is *E. coli* (Figure 1), although there are also other bacterial gene expression systems available (Table S1). For the production of eukaryotic target proteins, multiple factors play a role in the decision-making process. For simple eukaryotic target proteins that do not require post-translational modifications and that possess a limited amount of disulfide bonds, *E. coli* can be considered as an expression host as well (Figure 1). However, in many cases, eukaryotic gene expression systems such as yeast, insect cells, or mammalian cells might be more suitable. One of the main differences between these eukaryotic expression hosts lies in the type of glycosylation (N- and O-glycosylation) they can provide. Mammalian cells produce mainly complex type N-glycans, in which the glycan branches are modified with N-acetylglucosamine, galactose, fucose, and sialic acid.[3,4] In contrast, N-glycans from insect cells are generally not processed into terminally sialylated complex type structures and are instead modified into paucimannose or oligomannose structures.[5,6] Furthermore, the presence of

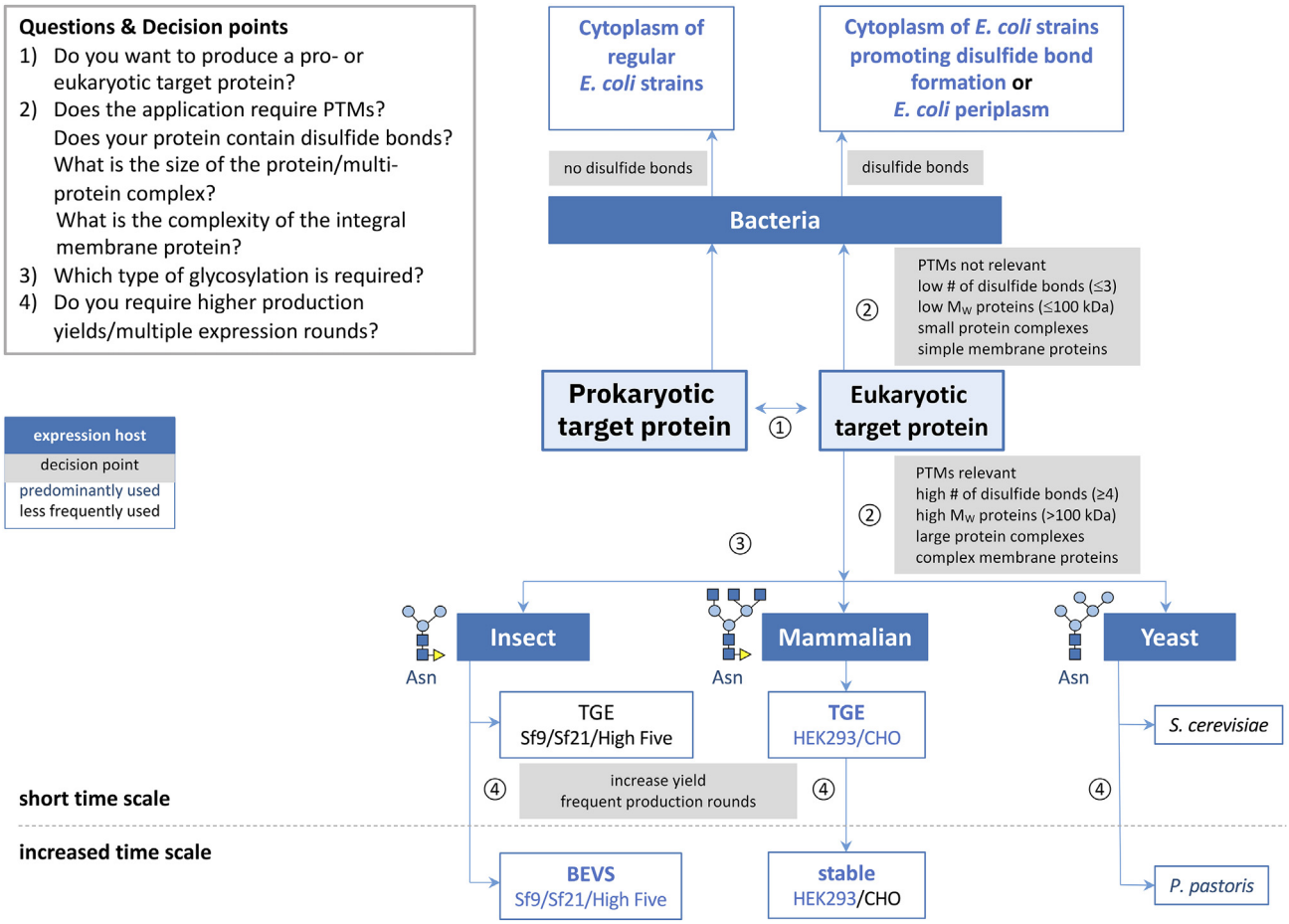
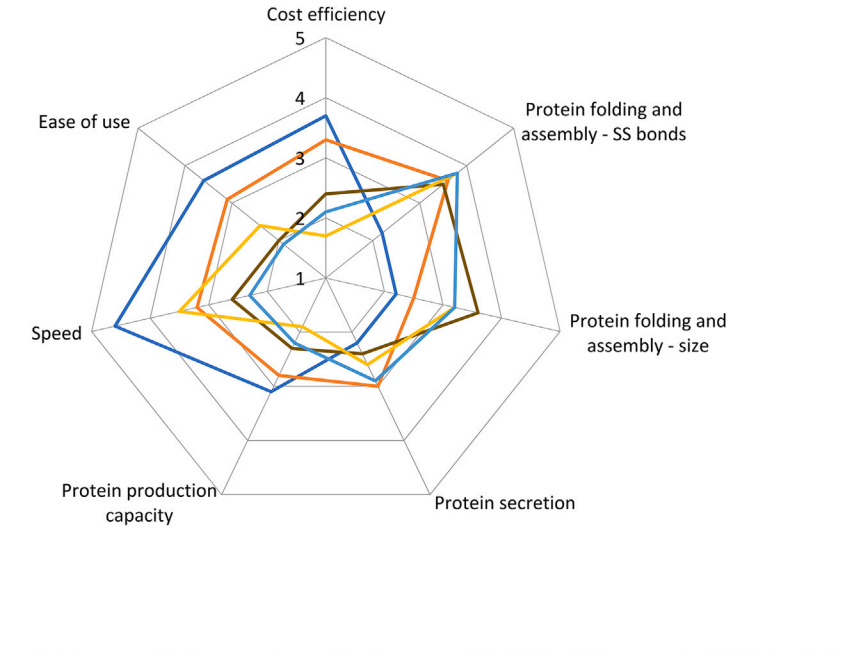


Figure 1. Decision scheme for gene expression system selection

There are four key decision points (circled numbers 1–4), each referring to specific questions shown in the inset text box. The blue arrows indicate the reading direction. The associated gray boxes describe the parameters to be considered at the various decision points. Expression hosts are presented in blue boxes, and predominantly and less frequently used systems are colored in blue and black, respectively. Decision points: (1) the initial decision point relates to the origin of the target protein to be produced, either being prokaryotic or eukaryotic in nature. Generally, prokaryotic proteins are produced in bacteria using different strains of *E. coli*. (2) For eukaryotic target proteins, however, multiple parameters have to be considered in the decision process. The production of such proteins in bacteria is only recommended for proteins that do not require post-translational modifications (PTMs; primarily glycosylation) for functional activity and/or stability, for proteins with up to 3 disulfide bonds, for proteins and protein complexes with a molecular weight (M_W) of up to 100 kDa, and for small integral membrane proteins (IMPs). Generally, for disulfide-containing proteins produced in bacteria, *E. coli* strains promoting cytoplasmic disulfide bond formation are used or proteins are secreted to the periplasm. On the contrary, the production of eukaryotic target proteins in eukaryotic systems is recommended for proteins requiring functional PTMs, for proteins with multiple (≥ 4) disulfide bonds, and large (>100 kDa) proteins/complexes and larger IMPs. (3) The decision as to which eukaryotic system (insect, mammalian, yeast) to use depends on the glycan type required for obtaining functional protein (see cartoon models for the different asparagine [Asn]-linked glycans). (4) If an increased protein yield and/or frequent production rounds are needed, the additional time investment (indicated by dashed line) for the generation of stable cell lines (mammalian systems) or baculovirus expression vector system (BEVS) compared to TGE (transient gene expression) is warranted.

core α(1,3)-linked fucose modifications, which are common in invertebrates but totally absent in mammals, can be immunogenic. Unicellular yeasts are capable of both N- and O-glycosylation,[7] but the pattern is quite different from mammalian cells. Yeast N-glycosylation is of the high/hyper-mannose type, which can cause antigenicity.[8] If the glycosylation type is important for the protein of interest and/or the intended downstream applications, this might be a critical factor to consider when choosing the optimal expression system.

When the target protein is a membrane-associated or integral membrane protein (IMP), the selection of a suitable gene expression system is essential. While it might be possible to produce small



Scoring:

Ease of use: (1) Possible with SOP + user training + 1 year hands-on experience; (2) possible with SOP + user training + short-term experience; (3) possible with SOP + user training; (4) possible with SOP + short-term experience; (5) possible with SOP only.
Speed: (1) >8 weeks; (2) 4–8 weeks; (3) 1–4 weeks; (4) 3–7 days; (5) 1–3 days.
Protein production capacity: (1) <1 mg/L; (2) 1–5 mg/L; (3) 5–20 mg/L; (4) 20–100 mg/L; (5) >100 mg/L.
Protein secretion: (1) <1 mg/L; (2) 1–5 mg/L; (3) 5–20 mg/L; (4) 20–100 mg/L; (5) >100 mg/L.
Protein folding and assembly—size: (1) <50 kDa; (2) 50–100 kDa; (3) 100–250 kDa; (4) 250–500 kDa; (5) >500 kDa.
Protein folding and assembly—SS bonds: (1) 1 SS bond; (2) 2 SS bonds; (3) 3–4 SS bonds; (4) 5–10 SS bonds; (5) >10 SS bonds.
Cost efficiency: (5) <50 €/L; (4) 50–100 €/L; (3) 100–500 €/L; (2) 500–1,000 €/L; (1) >1,000 €/L.

Figure 2. Comparative overview of the characteristics associated with the major gene expression systems

Currently, the most commonly used protein production systems are *E. coli*, yeast, mammalian cells, and insect cells. In mammalian cells, both TGE and stable cell lines are frequently used, whereas in insect cells baculovirus-mediated expression is the predominant method of choice. The main characteristics associated with protein production in these systems are ranked on a scale of 1–5, which allows for an easy comparison of the individual characteristics between the different gene expression systems. The results presented here are based on a survey, which was organized among the members of the Protein Production and Purification Partnership in Europe (P4EU) network. The scores are weighted averages calculated from the survey responses. The survey first queried about the different gene expression systems used in the participants' home laboratories. These data then formed the basis for deciding on the most commonly used gene expression systems in the community. Next, the participants were asked to score the individual characteristics associated with the gene expression systems they were familiar with based only on their own personal experiences (not on textbook knowledge). Sixty complete responses to the survey were received, which might seem like a small number of participants, but which in reality corresponds to a cumulative experience with thousands of different expression constructs. The characteristics that were assessed in the survey were: (i) ease of use, indicating how much experience/training is necessary to use a particular gene expression system; (ii) speed, which is the time required from plasmid DNA/expression construct to biomass (expressed protein) for processing; (iii) protein production capacity, which represents the average intracellular protein production capacity in mg/L of culture; (iv) protein secretion, which is the average range of secreted protein production capacity in mg/L of culture (secretion to the periplasm for *E. coli*, secretion to the extracellular milieu for yeast, mammalian, and insect cells); (v) protein folding and assembly related to the size of the protein(s) of interest, representing the ability to produce functional and correctly folded single-chain multi-domain proteins or multi-subunit protein complexes depending on their

Figure 2. *Continued*
respective maximum size; (vi) protein folding and assembly related to the number of disulfide (SS) bonds, indicating the ability to produce functional and correctly folded (secreted) proteins depending on their respective number of disulfide bonds; and (vii) cost efficiency, estimating the consumable costs (e.g., media, transfection reagents, disposable flasks, plasmid preparation, cell maintenance, virus production, cell counting, etc.) for a 1-L production. All criteria are scored in a positive way, meaning higher scores correspond to more beneficial outputs.

membrane proteins in *E. coli*, eukaryotic host organisms and cell lines are generally preferred for this more challenging class of target proteins.[9,10] Currently, the most commonly used gene expression systems for larger IMPs—such as, for example, GPCRs, ion channels, and transporters—are insect and mammalian cells.[11] Even though many complex membrane proteins can be produced successfully in insect cells,[12] it is useful to keep in mind that the lipidic membrane environments are not identical to those in mammalian cells. As insect cells are generally cultured at 27°C, the types of lipids required to maintain membrane fluidity are different from those in mammalian cells, which are mostly cultured at 37°C.[13]

In order to obtain milligram quantities of recombinant proteins, *in vivo* cell-based gene expression systems are the preferred way to go. However, if either a few micrograms of protein suffice for the downstream application or *in vivo* production is impossible due to toxicity, or if specific ligands or additives are required, then *in vitro* cell-free expression (CFE) might be a suitable alternative. As the proper set-up of CFE with homemade reagents generally requires specialist training and might not be so easily accessible, CFE is neither included in the decision scheme for gene expression system selection (Figure 1) nor in the key-characteristics comparison of gene expression systems (Figure 2). However, as it might be applicable for some specific projects, detailed information and appropriate references about CFE are provided in the section "cell-free expression."

Figures 1 and 2 focus on *E. coli*, yeast, insect cells, and mammalian cells, as these are commonly used, well-characterized, and easily accessible gene expression systems. Nevertheless, there are many other alternative gene expression systems available, which possess different features and might be suitable choices for specific target proteins as well. However, as these more "exotic" host organisms are generally less frequently used, we recommend seeking experts in these systems before attempting to set up such a system in-house. For example, plants and plant cells are able to fold and secrete more complex proteins and also possess the ability to direct the recombinantly produced proteins to different cellular compartments, which can be useful for, for example, toxic proteins (supplemental information, "protein production in plants"). Even though *E. coli* is by far the best-known prokaryotic gene expression system, other bacterial gene expression systems such as *Vibrio natriegens*, *Pseudomonas putida*, *Mycobacterium smegmatis*, and the Gram-positive bacteria *Lactococcus lactis* and *Bacillus subtilis* can be relevant options as well (Table S1). Furthermore, the eukaryotic expression hosts *Drosophila* S2 and the unicellular green algae *Chlamydomonas reinhardtii* represent other interesting alternative gene expression systems (Table S1).

The aim of this manuscript is to guide the reader to the most appropriate gene expression system by posing key questions regarding the characteristics of their proteins and matching them to the characteristics of the different available systems. Once an initial choice has been made regarding the most appropriate gene expression system(s), the reader can find more detailed descriptions in the specific sections of this primer. The different sections offer details about the individual systems, including key reviews and relevant references that can be consulted. Basic information, including the pros and cons of each system, is provided, as are ample references to relevant reading materials. As the availability of equipment might be an important factor as well, a more detailed overview of the instrumentation required for protein production in the respective gene expression systems can be found in Table S3. Additional information about the features of various commonly used expression strains/cell lines and vectors and how biological resources such as vectors, plasmids, and related host

strains can be acquired is provided in the supplemental information, "expression vectors and strains" and "biological resources."

E. COLI, ONE OF THE MOST COMMONLY USED GENE EXPRESSION SYSTEMS

E. coli is one of the most commonly used host organisms for protein production thanks to its ease of use, cost efficiency, speed, and minimal requirement in terms of equipment. *E. coli* is generally the first organism of choice for production of prokaryotic proteins, but many eukaryotic proteins can be produced successfully in *E. coli* as well. However, compared to eukaryotic systems, *E. coli* cannot provide most of the post-translational modifications (notably glycosylation) and often fails in folding complex proteins, such as those containing multiple disulfide bonds, eukaryotic membrane proteins, or large multi-domain assemblies and multi-subunit complexes.[14]

In *E. coli*, proteins can be produced intracellularly in the cytoplasm, directed into the periplasm, or secreted to the extracellular milieu. The cytoplasm is a reducing environment, whereas the periplasm is an oxidizing environment that allows the formation of disulfide bonds and also has lower proteolytic activity. However, directing produced proteins into the periplasm often results in a lower yield than cytosolic production and usually not all expressed protein will be secreted into the periplasm. To direct a recombinant protein to the periplasm, one needs to add a periplasmic signal sequence (such as *phoA*, *pelB*, *ompA*, *ompT*, *dsbA*, *torA*) to the N-terminus of the protein, which will be removed after crossing the inner membrane. Proteins can be secreted either post-translationally (Sec mechanism) or co-translationally (SPR mechanism).[15]

A large collection of *E. coli* expression vectors is widely available, either commercially or via institutional or non-profit plasmid repositories (see supplemental information, "biological resources"). Such expression vectors contain a set of genetic elements (e.g., promoter, terminator, origin of replication, antibiotic resistance cassette, etc.) that allow a regulated expression of the coding sequence of the protein(s) of interest (see supplemental information, "expression vectors and strains"). One of the most frequently used bacterial gene expression systems makes use of vectors in which the gene(s) of interest are placed under control of the strong T7 promoter, which requires the T7 RNA polymerase for transcription.

Although many different *E. coli* expression strains have been developed in the past decades, the most commonly used strains are based on *E. coli* BL21. The popular *E. coli* BL21(DE3)[16] strain and its derivatives contain a lambda prophage encoding the T7 RNA polymerase under control of the *lacUV5* promoter, allowing IPTG-regulated expression of gene(s) under control of the T7 promoter. Various *E. coli* expression strains also have specific characteristics (see Table S2), making them more suitable for specific subtypes of proteins. For example, some strains can be engineered to produce extra copies of rare tRNAs, which is very useful if the codon usage of the gene of interest is non-optimized for expression in *E. coli*.[17] Other strains are better equipped to deal with the expression of toxic proteins or are more suitable for the expression of disulfide bond-rich proteins in the cytoplasm.[18] The required plasmid-related host strains are also accessible on a non-profit (see supplemental information, "biological resources") or profit basis. When starting with the production of a new protein in *E. coli*, it is generally recommended to assess different strains and different expression conditions (e.g., different media,[19–21] growth and induction temperatures, time of induction, concentration of inducer, etc.). This type of approach is also amenable to automation and hence to high-throughput screening.[22]

Over the years, various approaches have been developed to alleviate some *E. coli* shortcomings regarding the production of more complex proteins. For example, a commonly used method is the addition of solubility-enhancing fusion tags to the protein of interest.[23] Often, slowing down the rate of gene expression by using low-copy plasmids and/or low induction temperatures improves solubility as well. Alternatively, co-expression of molecular chaperones can result in proper

folding in *E. coli*.[24] Auto-induction media[20] may also improve yields of soluble protein in *E. coli*. Another option is the engineering of protein sequences to increase their solubility in *E. coli*, for which easy-to-use and validated open-access algorithms are available.[25] In some cases, aggregation of the recombinantly produced proteins into insoluble inclusion bodies can also be exploited to purify relatively homogeneous target proteins and refold them.[26] However, it must be stressed that the refolding of proteins from inclusion bodies[27] requires time-consuming protocol optimization, and the yields are often low and the recovery of the native structure must be carefully verified.

Many useful general papers[28–32] and protocols to start approaching protein production in *E. coli* are available.

THE USE OF YEAST AS A PROTEIN PRODUCTION SYSTEM

Yeasts are single-cell eukaryotic host organisms which combine some of the advantages of prokaryotic and eukaryotic-based gene expression systems. They are amenable to high-density fermentation and possess the necessary cellular machinery to carry out certain post-translational modifications such as glycosylation, disulfide bond formation, and proteolytic processing.[33] Several yeasts are being used for protein production, including *Pichia pastoris* (syn. *Komagataella phaffii*), *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, and *Kluyveromyces lactis*.[33,34] Among these, the methylotrophic yeast *P. pastoris* has emerged in the past 20 years as one of the most popular yeast-based gene expression systems,[35,36] whereas *S. cerevisiae* is used as a major genetic tool.

In yeast, proteins can be produced intracellularly, or they can be secreted to the extracellular milieu, which requires the presence of an N-terminal signal peptide (e.g., α-mating factor or *Ost1*). *P. pastoris* is capable of both N- and O-linked glycosylation.[37] Glycosylation in yeast is rich in non-homogeneous hypermannosyl structures, which is different from the more complex mammalian glycan structures and can lead to antigenicity. Therefore, much effort has been put into developing *P. pastoris* strains capable of performing humanized N-glycosylation.[38]

P. pastoris is an easy-to-handle and relatively cheap gene expression system. Generating expression strains is more time consuming than for *E. coli*, but it can deliver very high recombinant protein yields and properly folded complex proteins without lipopolysaccharide contamination, which is highly beneficial for pharmaceutical and therapeutical proteins. *P. pastoris* expression vectors are generally integrated into the genome to create stable, high-expressing strains. Small-scale expression tests can be performed to screen for the highest-yielding clones. Commonly used strong promoters are the methanol-inducible *AOX1* promoter or the constitutively active *GAP* promoter. There's also a wide selection of *P. pastoris* expression vectors available (see Table S2 and supplemental information, "biological resources"), which can be wild-type strains used in combination with antibiotic selection or auxotrophic strains that allow complementation with specific marker genes present in the expression vectors.[36,39]

Due to the broad applicability of *P. pastoris* both in academic research labs and in industrial protein production setups, extensive efforts have been made to further improve protein yields and to optimize growth. New elements are being added to the *P. pastoris* expression toolkit continuously, such as the OPENPichia strains,[40] different promoters (e.g., *AOX1*, *UPP*, *PDF*),[41] signal peptides (α-mating factor, *Ost1*[39]), and optimized media with reduced protease activity and oxidation levels. Furthermore, novel high-cell-density fermentation methods are being developed.[36] Thanks to all these efforts in the field, there are currently already more than 70 licensed commercial products derived from *P. pastoris* available on the market (www.pichia.com).

P. pastoris can be used for the production of various types of (complex) proteins, but it's especially popular for the production of cytokines (IL3[42]), certain growth factors (GM-CSF[43]), and antibody derivatives without Fc fusion such as nanobodies,[44] bibodies, and tribodies.[45] Even though

S. cerevisiae is less popular for protein production purposes than *P. pastoris*, it is being used for the large-scale manufacturing of, for example, insulin, certain vaccines, and enzymes for industrial applications.[34,46]

For readers that are interested in using yeast for protein production, we recommend the following papers: Matsuzaki et al.,[44] De et al.,[47] Mastropietro et al.,[48] Rinnofer et al.,[49] and Higgins et al.[50] These are good starting papers to learn more about the technology in general and to obtain some initial protocols.

BACULOVIRUS-MEDIATED GENE EXPRESSION IN INSECT CELLS

Baculovirus-mediated gene expression in insect cells is one of the most widely used systems for heterologous protein production in academia and industry and has become a major technology for the manufacturing of membrane proteins, especially GPCRs and ion channels, multi-subunit protein complexes, secreted growth factors, virus-like particles (VLPs), and gene delivery vectors for mammalian cells (reviewed in Errey et al.,[11] Gupta et al.,[51] and Mahajan et al.[52]). A multitude of tools developed in the past four decades—extensively engineered and improved variants of the baculovirus *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV), commercially available insect cell lines (*Spodoptera frugiperda* cell lines Sf9 and Sf21, *Trichoplusia ni* cell lines High Five and Tnao38), and the manufacturing of serum-free media—have contributed to its success.

In insect cells, proteins can be produced intracellularly, or they can be secreted to the extracellular milieu, which requires the presence of an N-terminal signal peptide. In many cases, the insect cell peptidase can recognize mammalian signal sequences,[53,54] but it's possible to use native insect cell signal sequences (e.g., *gp67*, *HBM*, *SP1*, *SP2*) as well.[55–58] Even though insect cells are capable of N- and O-glycosylation, they lack complex type N-glycans, which is a limitation for the production of therapeutic proteins. However, different types of approaches are possible to obtain proteins with a more mammalian-like N-glycosylation from *lepidopteran* insect cell lines, which are generally based on the co-expression of various glycozymes.[59,60]

For protein production in insect cells, the gene(s) of interest are integrated into the baculovirus genome either by Tn7-mediated transposition within *E. coli* cells (DH10Bac, Thermo Fisher Scientific; DH10MultiBac and DH10EMBacY, Geneva Biotech) or by co-transfection of insect cells with a transfer vector and baculovirus DNA (*flash*BAC and derivatives, Oxford Expression Technologies; FlexiBac[61]). Although more time-consuming, transposition-based integration is easier to adopt for first-time users, as it allows users to control and monitor target gene insertion by antibiotic selection, blue-white screening, and bacmid PCR/sequencing. More advanced users may instead prefer the shorter co-transfection/homologous recombination protocol within insect cells. Due to the strong baculovirus-derived *polH* and *p10* promoters driving expression of the target protein(s), high yields can be achieved in infected insect cells, no matter which of the two integration tools has been applied, as shown in a benchmarking study conducted by 14 different expert laboratories.[62] Most standard procedures for baculovirus generation use amplification of the first transfection-based baculovirus passage P0 to generate P1, P2, or P3. However, due to the limited stability of baculovirus particles, shortened protocols using P0[63] or even working virus free (titerless infected-cells preservation and scale-up [TIPS][64]) has been introduced as well.

The baculovirus is unbeatable with regard to the size of the DNA cargo it can carry and transduce into insect or mammalian cells. As many as 17 subunits of a multiprotein complex have been successfully co-expressed in insect cells[65] and as many as 9 subunits have been transduced into HEK293 cells with BacMam baculovirus.[66] Different molecular cloning technologies—Golden Gate (GoldenBac),[67] Gibson assembly of PCR fragments (biGBac),[65] Cre-lox recombination (MultiBac)[68]—allow efficient multi-gene assembly in the baculovirus genome (see Table S2).

The main drawbacks regarding baculovirus-mediated expression in insect cells are the time required to go from DNA to target protein and the decay of the baculovirus over time. Therefore, transient plasmid-based gene expression methods have been developed as an alternative as well (see "transient gene expression in insect cells").

For readers that are interested in using baculovirus-mediated gene expression in insect cells, we recommend papers[68–72] as a good start to learn more about the technology in general and to obtain some initial protocols.

TRANSIENT GENE EXPRESSION IN INSECT CELLS

The use of baculoviral expression vectors (BEVs) to drive heterologous protein production in lepidopteran-derived insect cells is very well established. BEVs are a transient expression system lasting 3–4 days due to the disassembly of the insect cell secretion machinery, loss of cellular structures, and, finally, cell lysis. Plasmid-based transient gene expression (TGE), using chemical transfection of insect cells with expression plasmids, allows protein expression that is free of virus. The transfected cells remain viable and continue growing unhindered by a baculoviral infection process. However, plasmid-based TGE in insect cells is dependent on using strong endogenous insect cell promoters or immediate-early baculoviral promoters.

Since 2014, virus-free TGE in *Sf*9/*Sf*21 insect cells using plasmid-based vectors has been developed[73,74] to avoid the time-consuming generation of baculoviruses (the generation of high-titer baculoviral stocks may require more than 3 weeks[75]). The initial attempts to establish insect TGE resulted in low yields of produced protein in cells of *Sf* origin until the method was hugely improved by Beckmann et al.,[74,76,77] Shen et al.,[73] Mori et al.,[78] and Puente-Massaguer et al.[79,80] Replacement of *Sf*-derived cells with *Trichoplusia ni* (High Five) cells as the expression host and introduction of the strongest available RNA polymerase II-dependent immediate-early promoter (the pOpIE2 promoter from the *Orygia pseudotsugata* multicapsid nucleopolyhedrosis virus, OpMNPV) allowed the development of a rapid and simple virus-free gene expression system in High Five insect cells. Many other experimental parameters have since been optimized and TGE in High Five insect cells was established as a robust and efficient method to produce intra-cellular and secreted protein within one week.[80] Briefly, transient transfection in High Five cells is performed by the addition of, first, ultra-pure expression plasmids, harboring the gene of interest cloned between the insect-specific pOpIE2 promoter and an adequate terminator and, second, polyethylenimine (PEI40) as transfection agent to logarithmically growing High Five insect cells at high density. After a short 3–4 h incubation, the cells are diluted, and growth is continued for several days. The efficiency of each transfection may be followed by co-transfection of a GFP control vector (as 5% of the total plasmid DNA transfected). The transfected cells can be harvested and adequate amounts of correctly folded protein may be isolated from either cell biomass (for cytoplasmic proteins or IMPs) or the cell culture supernatant after removal of the cells (for secreted proteins) by standard affinity chromatographic techniques.

The main advantage of the TGE insect cell system is a simple scale up to several liters in affordable insect media, while the cells are cultivated in a 27°C incubator with shaker platform without the use of CO_2 (in contrast with mammalian cell growth requirements). The expression timeline is fast and requires only one week once the expression plasmid is available. The insect TGE also benefits from the homogeneous paucimannose type of glycosylation, which is ideal for structural analysis of secreted proteins.[81] Recently, its application for producing membrane proteins has been shown as well.[82]

Interested readers are recommended to check Shen et al.,[73] Bleckmann et al.,[74,77] Puente-Massaguer et al.,[79] and Shen et al.[83] as excellent papers to learn about the development of the technology and how to establish TGE in High Five insect cells.

PROTEIN PRODUCTION IN MAMMALIAN CELLS

Protein production in mammalian cells is particularly suited to larger or more complex eukaryotic proteins, as it can offer a cellular environment closely resembling the native one. Mammalian cells are a popular choice for the production of IMPs[84–86] and other (secreted) eukaryotic proteins requiring functional native-like post-translational modifications. Mammalian cell lines for protein production are generally derived from human embryonic kidney 293 (HEK293) or Chinese hamster ovary (CHO) cells (see Table S2). HEK293 cell lines are frequently used for research applications due to their ease of transfection, whereas CHO cells are often the system of choice for the production of bio-pharmaceutical proteins.

Mammalian cells can be grown as adherent cells or in suspension cultures. Adherent HEK293 have been used for almost 5 decades for transient transfections, as they are easy to culture and to maintain with high reproducibility, and high transfection efficiencies can be obtained with cheap reagents. The growth medium is inexpensive as well and can be prepared in house. However, for large-scale protein production, roller bottles may be necessary to avoid the need for manipulating a large number of culture plates. In contrast, HEK293-based suspension cultures, with simple passaging by dilution, present a more attractive alternative for obtaining production-level quantities of biomass. Popular suspension culture cell lines are, for example, HEK293-6E (293-EBNA1),[87] HEK293F, and Expi293F (see Table S2). The HEK293-6E cell line (transformed with Epstein-Barr virus nuclear antigen 1) combined with plasmids containing an oriP origin of replication allow the transfected expression plasmids to be replicated episomally, in turn leading to increased protein yields. Other suspension-adapted HEK293 derivatives include HEK293F and Expi293F, which are generally cultivated in a commercially available serum-free medium. The medium required for suspension cultures is much more expensive than for adherent cells though, and the composition is often proprietary. The high-density Expi293F commercial system combines both proprietary media and proprietary transfection reagents and may not be suitable for many academic research lab budgets.

Recombinant proteins can be produced transiently in mammalian cells by transfection with plasmid DNA or by transduction with baculoviruses (BacMam). The most widely used method for TGE is transfection with plasmid DNA, as it is fast and easy to adopt and affordable transfection reagents such as polyethylenimine (PEI) are readily available[88–91]. BacMam[92,93] is more time consuming, as it requires the generation of recombinant baculoviruses, but it can be efficient for difficult-to-transfect cell lines or when large DNA fragments need to be introduced—for example, for the expression of multi-component protein complexes (MultiBacMam).[94,95]

Stable mammalian cell pools can be generated by either non-targeted gene integration, using lentiviruses,[96,97] or transposase enzymes such as Sleeping Beauty, Frog Prince, Minos, or piggyBac.[98] PiggyBac transposase, isolated from cabbage looper moth *Trichoplusia ni*, and its hyperactive mutants can efficiently integrate up to 15 gene copies with a cargo capacity of 9–14 kb.[99] Stable pools of HEK293 and CHO cells generated with piggyBac transposase have been increasingly applied in protein production in the past 10 years for several reasons. Small amounts of plasmid DNA are needed for transfection, selection times are short (typically 11 days), the process is adaptable to many cell lines, the pools can produce high levels of protein, and the stable pools can be easily cryo-preserved. Transposase-based systems also allow the integration/expression of multiple genes, and it is possible to express cytotoxic proteins by using an inducible tetracycline promoter. Stable pools offer a lower-cost alternative to multiple rounds of TGE.[100–105]

For readers that are interested in using mammalian cells for protein production, Pieprzyk et al.,[85] Goehring et al.,[86] Baldi et al.,[91] Fornwald et al.,[92] Behiels and Elegheert,[97] and Suppmann[105] are recommended as good starting papers to learn more about the technology in general and to obtain some initial protocols.

CELL-FREE EXPRESSION

CFE is defined as the production of proteins using the components required for transcription and translation in a cell-free environment. CFE systems are based on lysates of *E. coli* or eukaryotic cells such as wheat germs or insect or tobacco cells.[106–108] Most CFE systems work with relatively crude cell lysates, although defined systems reconstituted from purified protein and RNA components are available as well.[109] The cell lysates are devoid of low-molecular substances and are complemented in CFE reactions by addition of amino acids, nucleotides, energy regeneration systems, and expression templates in the form of plasmid DNA, linear DNA, or mRNA.

The protein production efficiency of CFE strongly depends on the origin of the cell lysate as well as the reaction configuration. CFE systems based on *E. coli* or wheat germ lysates can reach protein synthesis levels of mg/mL reaction in two-compartment configurations, separating reaction mixtures from feeding mixtures that provide fresh low-molecular-weight precursors. Simpler one-pot batch configurations and CFE systems based on insect or mammalian cell lysates operate in the μg/mL production levels.

The advantages of CFE systems are their open, accessible nature and operation in low volumes. A wide range of ligands, stabilizers, and other additives, even those that are toxic or difficult to implement into cell-based expression systems, are tolerated. Tailored environments for the production of individual proteins can thus be created by co-expression of targets in the presence of cofactors, interaction partners, or ligands. CFE is of particular value for the production of membrane proteins as well. Insect and tobacco cell lysates retain microsomal fragments able to translocate and glycosylate synthesized membrane proteins. However, these modifications only work efficiently at low expression levels of a few μg/mL and may become readily overloaded.[110,111] Alternatively, membrane mimetics in the form of liposomes, nanodiscs, or even detergents can be supplied into CFE reactions to facilitate the instant co-translational solubilization of synthesized membrane proteins.[112] These strategies allow high-throughput applications[113] and are suitable to determine the functionality and even structures of membrane proteins by crystallization, NMR, or electron microscopy.[114–117]

Either commercial or individual in-house CFE systems may be used. Commercial systems are usually operated in one-pot batch configurations and the costs per milligram of product can become excessive. These systems may rather be considered if synthesis of a few micrograms of protein is sufficient. Protein synthesis is completed within a few hours, and no equipment other than pipets and a thermostat is required. For more frequent use and in order to profit from the full potential of CFE, in-house systems, ideally based on easy-to-prepare *E. coli* lysates and operated in two-compartment configurations, might be preferred. Necessary infrastructure would just be an adequately equipped biochemistry lab, whereas CFE protocol development might require some training and experience. The power and perspectives of this workflow were recently reviewed.[118]

Obtaining high-quality samples usually results from systematic screening to identify supporting additives, suitable template designs as well as optimal concentrations of additives, and critical basic reaction components. CFE is therefore not competitive for the production of standard protein samples that can be obtained in reasonable amounts from conventional cell-based systems. However, it could become a perfect choice for difficult targets such as membrane proteins, toxins, or the production of labeled protein samples for, for example, NMR studies.[119]

In summary, CFE can become a system of choice if either the entire platform, including cell lysate production, is available or if intended applications would require only low amounts of sample.

CONCLUSION

This gene expression system selection guide is based on the results of the consultation of more than 60 specialists in protein production and reflects the extensive practical experience of the authors. The decision scheme and the key characteristics comparison cover the currently most broadly

used, most widely available, and best understood gene expression systems. Unfortunately, there is no gene expression system which "fits" all, and, generally, the specific characteristics of the required protein and planned downstream application will determine which will be the most adequate gene expression system. The availability of local expertise and equipment should also be considered, as this may render a less commonly used gene expression system both accessible and viable/economical. Readers are encouraged to investigate potential gene expression systems more fully using the provided references before embarking on protein production in their own laboratories. Finally, this review is based on the authors' experience at the time of writing. As these gene expression systems continue to evolve, it is vital that readers regularly review their options for protein production systems. Today's "exotic" gene expression systems may become tomorrow's widely used gene expression systems for even more challenging protein targets.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xpro.2023.102572.

ACKNOWLEDGMENTS

The authors thank all members from the P4EU community who responded to the survey and contributed their expertise in the preparation of this manuscript. We acknowledge Elena Gubanova (MDC, Berlin, Germany) for administrative support in conducting the survey. We are very grateful to Dr. Christian Loew (EMBL Hamburg, Germany) and Dr. Eric Geertsma (MPI-CBG, Germany) for the discussions about membrane protein production. We thank Dr. Anna Arís and Dr. Elena Garcia-Fruitós (IRTA, Spain) for their insights about the use of *Lactococcus lactis* as a recombinant protein production system.

AUTHOR CONTRIBUTIONS

Conceptualization, A.S., N.B., F.F.d.S., J.v.d.H., Y.P., K.R.; survey, A.S., J.v.d.H.; writing & original draft, A.S., N.B., F.F.d.S., J.v.d.H., T.U.; figures, A.S., N.B., J.F.B., F.F.d.S., J.v.d.H., Y.P., S.S., T.U., S.W., K.R.; "*E. coli*, one of the most commonly used gene expression systems," A.S., F.F.d.S., J.E.H., A.d.M., S.W., K.R.; "the use of yeast as a protein production system," J.H., Y.P., K.R.; "baculovirus-mediated gene expression in insect cells," J.v.d.H., S.S., K.R.; "transient gene expression in insect cells," J.v.d.H., K.R.; "protein production in mammalian cells," S.S., T.U., K.R.; "cell-free expression," F.B., S.W., K.R.; supplemental information, "plants as an alternative protein production system," J.F.B.; supplemental information, "'exotic' gene expression systems" and Table S1, A.d.M., Y.P., S.W., K.R.; supplemental information, "expression vectors and strains/cell lines: how to choose them?" and Table S2, A.d.M.; supplemental information, "access to biological resources," M.V.; supplemental information, "equipment list" and Table S3, J.v.d.H.; review & editing, A.S., N.B., F.F.d.S., J.v.d.H., K.R.

DECLARATION OF INTERESTS

Kim Remans is a member of the STAR Protocols advisory board.

REFERENCES

1. De Marco, A., Berrow, N., Lebendiker, M., Garcia-Alai, M., Knauer, S.H., Lopez-Mendez, B., Matagne, A., Parret, A., Remans, K., Uebel, S., and Raynal, B. (2021). Quality control of protein reagents for the improvement of research data reproducibility. Nat. Commun. *12*, 2795. https://doi.org/10.1038/s41467-021-23167-z.

2. Remans, K., Lebendiker, M., Abreu, C., Maffei, M., Sellathurai, S., May, M.M., Vaněk, O., and De Marco, A. (2022). Protein purification strategies must consider downstream applications and individual biological characteristics. Microb. Cell Factories *21*, 52. https://doi.org/10.1186/s12934-022-01778-5.

3. Schjoldager, K.T., Narimatsu, Y., Joshi, H.J., and Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. Nat. Rev. Mol. Cell Biol. *21*, 729–749. https://doi.org/10.1038/s41580-020-00294-x.

4. Heffner, K.M., Wang, Q., Hizal, D.B., Can, Ö., and Betenbaugh, M.J. (2018). Glycoengineering of Mammalian Expression Systems on a Cellular Level. In Advances in Glycobiotechnology Advances in Biochemical Engineering/Biotechnology, E. Rapp and U. Reichl, eds. (Springer International Publishing), pp. 37–69. https://doi.org/10.1007/10_2017_57.

5. Tomiya, N., Narang, S., Lee, Y.C., and Betenbaugh, M.J. (2004). Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. Glycoconj. J. *21*, 343–360. https://doi.org/10.1023/B:GLYC.0000046275.28315.87.

6. Palomares, L.A., Srivastava, I.K., Ramírez, O.T., and Cox, M.M.J. (2018).

Glycobiotechnology of the Insect Cell-Baculovirus Expression System Technology. In Advances in Glycobiotechnology Advances in Biochemical Engineering/Biotechnology, E. Rapp and U. Reichl, eds. (Springer International Publishing), pp. 71–92. https://doi.org/10.1007/10_2018_61.

7. Delic, M., Valli, M., Graf, A.B., Pfeffer, M., Mattanovich, D., and Gasser, B. (2013). The secretory pathway: exploring yeast diversity. FEMS Microbiol. Rev. *37*, 872–914. https://doi.org/10.1111/1574-6976.12020.

8. De Wachter, C., Van Landuyt, L., and Callewaert, N. (2018). Engineering of Yeast Glycoprotein Expression. In Advances in Glycobiotechnology Advances in Biochemical Engineering/Biotechnology, E. Rapp and U. Reichl, eds. (Springer International Publishing), pp. 93–135. https://doi.org/10.1007/10_2018_69.

9. He, Y., Wang, K., and Yan, N. (2014). The recombinant expression systems for structure determination of eukaryotic membrane proteins. Protein Cell *5*, 658–672. https://doi.org/10.1007/s13238-014-0086-4.

10. I. Mus-Veteau, ed. (2022). Heterologous Expression of Membrane Proteins: Methods and Protocols (Springer US). https://doi.org/10.1007/978-1-0716-2368-8.

11. Errey, J.C., and Fiez-Vandal, C. (2020). Production of membrane proteins in industry: The example of GPCRs. Protein Expr. Purif. *169*, 105569. https://doi.org/10.1016/j.pep.2020.105569.

12. Hu, N.-J., Rada, H., Rahman, N., Nettleship, J.E., Bird, L., Iwata, S., Drew, D., Cameron, A.D., and Owens, R.J. (2015). GFP-Based Expression Screening of Membrane Proteins in Insect Cells Using the Baculovirus System. In Structural Proteomics Methods in Molecular Biology, R.J. Owens, ed. (Springer New York), pp. 197–209. https://doi.org/10.1007/978-1-4939-2230-7_11.

13. Marheineke, K., Grünewald, S., Christie, W., and Reilländer, H. (1998). Lipid composition of *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (Tn) insect cells used for baculovirus infection. FEBS Lett. *441*, 49–52. https://doi.org/10.1016/S0014-5793(98)01523-3.

14. Structural Genomics Consortium; Architecture et Fonction des Macromolécules Biologiques; Berkeley Structural Genomics Center; China Structural Genomics Consortium; Integrated Center for Structure and Function Innovation; Israel Structural Proteomics Center; Joint Center for Structural Genomics; Midwest Center for Structural Genomics; New York Structural GenomiX Research Center for Structural Genomics, and Northeast Structural Genomics Consortium, et al.. (2008). Protein production and purification. Nat. Methods *5*, 135–146. https://doi.org/10.1038/nmeth.f.202.

15. De Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microb. Cell Factories *8*, 26. https://doi.org/10.1186/1475-2859-8-26.

16. Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. *189*, 113–130. https://doi.org/10.1016/0022-2836(86)90385-2.

17. Eschenfeldt, W.H., Makowska-Grzyska, M., Stols, L., Donnelly, M.I., Jedrzejczak, R., and Joachimiak, A. (2013). New LIC vectors for production of proteins from genes containing rare codons. J. Struct. Funct. Genom. *14*, 135–144. https://doi.org/10.1007/s10969-013-9163-9.

18. Bertelsen, A.B., Hackney, C.M., Bayer, C.N., Kjelgaard, L.D., Rennig, M., Christensen, B., Sørensen, E.S., Safavi-Hemami, H., Wulff, T., Ellgaard, L., and Nørholm, M.H.H. (2021). DisCoTune: versatile auxiliary plasmids for the production of disulphide-containing proteins and peptides in the *E. coli* T7 system. Microb. Biotechnol. *14*, 2566–2580. https://doi.org/10.1111/1751-7915.13895.

19. Atlas, R.M., and Atlas, R.M. (2004). Handbook of Microbiological Media 0, ed. (CRC Press). https://doi.org/10.1201/9781420039726.

20. Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein Expr. Purif. *41*, 207–234. https://doi.org/10.1016/j.pep.2005.01.016.

21. Taylor, T., Denson, J.-P., and Esposito, D. (2017). Optimizing Expression and Solubility of Proteins in E. coli Using Modified Media and Induction Parameters. In Heterologous Gene Expression in E.coli Methods in Molecular Biology, N.A. Burgess-Brown, ed. (Springer New York), pp. 65–82. https://doi.org/10.1007/978-1-4939-6887-9_5.

22. Saez, N.J., and Vincentelli, R. (2014). High-Throughput Expression Screening and Purification of Recombinant Proteins in E. coli. In Structural Genomics Methods in Molecular Biology, Y.W. Chen, ed. (Humana Press), pp. 33–53. https://doi.org/10.1007/978-1-62703-691-7_3.

23. Dümmler, A., Lawrence, A.-M., and De Marco, A. (2005). Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. Microb. Cell Factories *4*, 34. https://doi.org/10.1186/1475-2859-4-34.

24. De Marco, A. (2007). Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in Escherichia coli. Nat. Protoc. *2*, 2632–2639. https://doi.org/10.1038/nprot.2007.400.

25. Peleg, Y., Vincentelli, R., Collins, B.M., Chen, K.-E., Livingstone, E.K., Weeratunga, S., Leneva, N., Guo, Q., Remans, K., Perez, K., et al. (2021). Community-Wide Experimental Evaluation of the PROSS Stability-Design Method. J. Mol. Biol. *433*, 166964. https://doi.org/10.1016/j.jmb.2021.166964.

26. Humer, D., and Spadiut, O. (2018). Wanted: more monitoring and control during inclusion body processing. World J. Microbiol. Biotechnol. *34*, 158. https://doi.org/10.1007/s11274-018-2541-5.

27. Singh, S.M., and Panda, A.K. (2005). Solubilization and refolding of bacterial inclusion body proteins. J. Biosci. Bioeng. *99*, 303–310. https://doi.org/10.1263/jbb.99.303.

28. De Marco, A. (2017). Acting on Folding Effectors to Improve Recombinant Protein Yields and Functional Quality. In Heterologous Gene Expression in E.coli Methods in Molecular Biology, N.A. Burgess-Brown, ed. (Springer New York), pp. 197–210. https://doi.org/10.1007/978-1-4939-6887-9_12.

29. Rosano, G.L., Morales, E.S., and Ceccarelli, E.A. (2019). New tools for recombinant protein production in *Escherichia coli* : A 5-year update. Protein Sci. *28*, 1412–1422. https://doi.org/10.1002/pro.3668.

30. Karyolaimos, A., and De Gier, J.-W. (2021). Strategies to Enhance Periplasmic Recombinant Protein Production Yields in Escherichia coli. Front. Bioeng. Biotechnol. *9*, 797334. https://doi.org/10.3389/fbioe.2021.797334.

31. Royes, J., Talbot, P., Le Bon, C., Moncoq, K., Uzan, M., Zito, F., and Miroux, B. (2022). Membrane Protein Production in Escherichia coli: Protocols and Rules. In Heterologous Expression of Membrane Proteins Methods in Molecular Biology, I. Mus-Veteau, ed. (Springer US)), pp. 19–39. https://doi.org/10.1007/978-1-0716-2368-8_2.

32. Rong, Y., Jensen, S.I., Lindorff-Larsen, K., and Nielsen, A.T. (2023). Folding of heterologous proteins in bacterial cell factories: Cellular mechanisms and engineering strategies. Biotechnol. Adv. *63*, 108079. https://doi.org/10.1016/j.biotechadv.2022.108079.

33. Patra, P., Das, M., Kundu, P., and Ghosh, A. (2021). Recent advances in systems and synthetic biology approaches for developing novel cell-factories in non-conventional yeasts. Biotechnol. Adv. *47*, 107695. https://doi.org/10.1016/j.biotechadv.2021.107695.

34. Baghban, R., Farajnia, S., Rajabibazl, M., Ghasemi, Y., Mafi, A., Hoseinpoor, R., Rahbarnia, L., and Aria, M. (2019). Yeast Expression Systems: Overview and Recent Advances. Mol. Biotechnol. *61*, 365–384. https://doi.org/10.1007/s12033-019-00164-8.

35. de Sá Magalhães, S., and Keshavarz-Moore, E. (2021). Pichia pastoris (Komagataella phaffii) as a Cost-Effective Tool for Vaccine Production for Low- and Middle-Income Countries (LMICs). Bioengineering *8*, 119. https://doi.org/10.3390/bioengineering8090119.

36. García-Ortega, X., Cámara, E., Ferrer, P., Albiol, J., Montesinos-Seguí, J.L., and Valero, F. (2019). Rational development of bioprocess engineering strategies for recombinant protein production in Pichia pastoris (Komagataella phaffii) using the methanol-free GAP promoter. Where do we stand? N. Biotech. *53*, 24–34. https://doi.org/10.1016/j.nbt.2019.06.002.

37. Liu, C., Gong, J.-S., Su, C., Li, H., Li, H., Rao, Z.-M., Xu, Z.-H., and Shi, J.-S. (2022). Pathway engineering facilitates efficient protein expression in Pichia pastoris. Appl. Microbiol. Biotechnol. *106*, 5893–5912. https://doi.org/10.1007/s00253-022-12139-y.

38. Laukens, B., Jacobs, P.P., Geysens, K., Martins, J., De Wachter, C., Ameloot, P., Morelle, W., Haustraete, J., Renauld, J.C., Samyn, B., et al. (2020). Off-target glycans encountered along the synthetic biology route toward humanized *N* -glycans in *Pichia*

pastoris. Biotechnol. Bioeng. *117*, 2479–2488. https://doi.org/10.1002/bit.27375.

39. Gao, J., Jiang, L., and Lian, J. (2021). Development of synthetic biology tools to engineer Pichia pastoris as a chassis for the production of natural products. Synth. Syst. Biotechnol. *6*, 110–119. https://doi.org/10.1016/j.synbio.2021.04.005.

40. Van Herpe, D., Vanluchene, R., Vandewalle, K., Vanmarcke, S., Wyseure, E., Van Moer, B., Eeckhaut, H., Fijalkowska, D., Grootaert, H., Lonigro, C., et al. (2022). OPENPichia: Building a Free-To-Operate *Komagataella phaffii* Protein Expression Toolkit. Mol. Biol. https://doi.org/10.1101/2022.12.13.519130.

41. Garrigós-Martínez, J., Vuoristo, K., Nieto-Taype, M.A., Tähtiharju, J., Uusitalo, J., Tukiainen, P., Schmid, C., Tolstorukov, I., Madden, K., Penttilä, M., et al. (2021). Bioprocess performance analysis of novel methanol-independent promoters for recombinant protein production with Pichia pastoris. Microb. Cell Factories *20*, 74. https://doi.org/10.1186/s12934-021-01564-9.

42. Dagar, V.K., and Khasa, Y.P. (2018). Combined effect of gene dosage and process optimization strategies on high-level production of recombinant human interleukin-3 (hIL-3) in Pichia pastoris fed-batch culture. Int. J. Biol. Macromol. *108*, 999–1009. https://doi.org/10.1016/j.ijbiomac.2017.11.008.

43. Jacobs, P.P., Inan, M., Festjens, N., Haustraete, J., Van Hecke, A., Contreras, R., Meagher, M.M., and Callewaert, N. (2010). Fed-batch fermentation of GM-CSF-producing glycoengineered Pichia pastoris under controlled specific growth rate. Microb. Cell Factories *9*, 93. https://doi.org/10.1186/1475-2859-9-93.

44. Matsuzaki, Y., Kajiwara, K., Aoki, W., and Ueda, M. (2022). Production of Single-Domain Antibodies in Pichia pastoris. In Single-Domain Antibodies Methods in Molecular Biology, G. Hussack and K.A. Henry, eds. (Springer US)), pp. 181–203. https://doi.org/10.1007/978-1-0716-2075-5_9.

45. Schoonooghe, S., Leoen, J., and Haustraete, J. (2012). Production of Antibody Derivatives in the Methylotrophic Yeast Pichia pastoris. In Antibody Engineering Methods in Molecular Biology, P. Chames, ed. (Humana Press), pp. 325–340. https://doi.org/10.1007/978-1-61779-974-7_19.

46. Wang, G., Huang, M., and Nielsen, J. (2017). Exploring the potential of Saccharomyces cerevisiae for biopharmaceutical protein production. Curr. Opin. Biotechnol. *48*, 77–84. https://doi.org/10.1016/j.copbio.2017.03.017.

47. De, S., Mattanovich, D., Ferrer, P., and Gasser, B. (2021). Established tools and emerging trends for the production of recombinant proteins and metabolites in *Pichia pastoris*. Essays Biochem. *65*, 293–307. https://doi.org/10.1042/EBC20200138.

48. Mastropietro, G., Aw, R., and Polizzi, K.M. (2021). Expression of proteins in Pichia pastoris. In Methods in Enzymology (Elsevier), pp. 53–80. https://doi.org/10.1016/bs.mie.2021.07.004.

49. Rinnofner, C., Felber, M., and Pichler, H. (2022). Strains and Molecular Tools for Recombinant Protein Production in Pichia pastoris. In Yeast Metabolic Engineering Methods in Molecular Biology, V. Mapelli and M. Bettiga, eds. (Springer US)), pp. 79–112. https://doi.org/10.1007/978-1-0716-2399-2_6.

50. Higgins, D.R., and Cregg, J.M. (1998). Introduction to Pichia pastoris. In Pichia Protocols Methods in Molecular Biology, D.R. Higgins and J.M. Cregg, eds. (Humana Press), pp. 1–15. https://doi.org/10.1385/0-89603-421-6:1.

51. Gupta, K., Tölzer, C., Sari-Ak, D., Fitzgerald, D.J., Schaffitzel, C., and Berger, I. (2019). MultiBac: Baculovirus-Mediated Multigene DNA Cargo Delivery in Insect and Mammalian Cells. Viruses *11*, 198. https://doi.org/10.3390/v11030198.

52. Mahajan, P., Ellis, K., Mukhopadhyay, S., Fernandez-Cid, A., Chi, G., Man, H., Dürr, K.L., and Burgess-Brown, N.A. (2021). Expression Screening of Human Integral Membrane Proteins Using BacMam. In Structural Genomics Methods in Molecular Biology, Y.W. Chen and C.-P.B. Yiu, eds. (Springer US)), pp. 95–115. https://doi.org/10.1007/978-1-0716-0892-0_6.

53. Smith, G.E., Ju, G., Ericson, B.L., Moschera, J., Lahm, H.W., Chizzonite, R., and Summers, M.D. (1985). Modification and secretion of human interleukin 2 produced in insect cells by a baculovirus expression vector. Proc. Natl. Acad. Sci. USA *82*, 8404–8408. https://doi.org/10.1073/pnas.82.24.8404.

54. Le, L.T.M., Nyengaard, J.R., Golas, M.M., and Sander, B. (2018). Vectors for Expression of Signal Peptide-Dependent Proteins in Baculovirus/Insect Cell Systems and Their Application to Expression and Purification of the High-Affinity Immunoglobulin Gamma Fc Receptor I in Complex with Its Gamma Chain. Mol. Biotechnol. *60*, 31–40. https://doi.org/10.1007/s12033-017-0041-8.

55. Wang, Y., Mao, Y., Xu, X., Tao, S., and Chen, H. (2015). Codon Usage in Signal Sequences Affects Protein Expression and Secretion Using Baculovirus/Insect Cell Expression System. PLoS One *10*, e0145887. https://doi.org/10.1371/journal.pone.0145887.

56. Soejima, Y., Lee, J., Nagata, Y., Mon, H., Iiyama, K., Kitano, H., Matsuyama, M., and Kusakabe, T. (2013). Comparison of signal peptides for efficient protein secretion in the baculovirus-silkworm system. Open Life Sci. *8*, 1–7. https://doi.org/10.2478/s11535-012-0112-6.

57. Ailor, E., and Betenbaugh, M.J. (1999). Modifying secretion and post-translational processing in insect cells. Curr. Opin. Biotechnol. *10*, 142–145. https://doi.org/10.1016/S0958-1669(99)80024-X.

58. Futatsumori-Sugai, M., and Tsumoto, K. (2010). Signal peptide design for improving recombinant protein secretion in the baculovirus expression vector system. Biochem. Biophys. Res. Commun. *391*, 931–935. https://doi.org/10.1016/j.bbrc.2009.11.167.

59. Palmberger, D., Wilson, I.B.H., Berger, I., Grabherr, R., and Rendic, D. (2012). SweetBac: A New Approach for the Production of Mammalianised Glycoproteins in Insect Cells. PLoS One *7*, e34226. https://doi.org/10.1371/journal.pone.0034226.

60. Maghodia, A.B., Geisler, C., and Jarvis, D.L. (2021). A New Bacmid for Customized Protein Glycosylation Pathway Engineering in the Baculovirus-Insect Cell System. ACS Chem. Biol. *16*, 1941–1950. https://doi.org/10.1021/acschembio.0c00974.

61. Lemaitre, R.P., Bogdanova, A., Borgonovo, B., Woodruff, J.B., and Drechsel, D.N. (2019). FlexiBAC: a versatile, open-source baculovirus vector system for protein expression, secretion, and proteolytic processing. BMC Biotechnol. *19*, 20. https://doi.org/10.1186/s12896-019-0512-z.

62. Stolt-Bergner, P., Benda, C., Bergbrede, T., Besir, H., Celie, P.H.N., Chang, C., Drechsel, D., Fischer, A., Geerlof, A., Giabbai, B., et al. (2018). Baculovirus-driven protein expression in insect cells: A benchmarking study. J. Struct. Biol. *203*, 71–80. https://doi.org/10.1016/j.jsb.2018.03.004.

63. Scholz, J., and Suppmann, S. (2021). A fast-track protocol for protein expression using the BEV system. Methods Enzymol *660*, 171–190. https://doi.org/10.1016/bs.mie.2021.06.015.

64. Wasilko, D.J., Lee, S.E., Stutzman-Engwall, K.J., Reitz, B.A., Emmons, T.L., Mathis, K.J., Bienkowski, M.J., Tomasselli, A.G., and Fischer, H.D. (2009). The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. Protein Expr. Purif. *65*, 122–132. https://doi.org/10.1016/j.pep.2009.01.002.

65. Weissmann, F., Petzold, G., VanderLinden, R., Huis In 'T Veld, P.J., Brown, N.G., Lampert, F., Westermann, S., Stark, H., Schulman, B.A., and Peters, J.-M. (2016). biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. Proc. Natl. Acad. Sci. USA *113*. https://doi.org/10.1073/pnas.1604935113.

66. Miller-Vedam, L.E., Bräuning, B., Popova, K.D., Schirle Oakdale, N.T., Bonnar, J.L., Prabu, J.R., Boydston, E.A., Sevillano, N., Shurtleff, M.J., Stroud, R.M., et al. (2020). Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients. eLife *9*, e62611. https://doi.org/10.7554/eLife.62611.

67. Neuhold, J., Radakovics, K., Lehner, A., Weissmann, F., Garcia, M.Q., Romero, M.C., Berrow, N.S., and Stolt-Bergner, P. (2020). GoldenBac: a simple, highly efficient, and widely applicable system for construction of multi-gene expression vectors for use with the baculovirus expression vector system. BMC Biotechnol. *20*, 26. https://doi.org/10.1186/s12896-020-00616-z.

68. Berger, I., Fitzgerald, D.J., and Richmond, T.J. (2004). Baculovirus expression system for heterologous multiprotein complexes. Nat. Biotechnol. *22*, 1583–1587. https://doi.org/10.1038/nbt1036.

69. Berger, I., and Poterszman, A. (2015). Baculovirus expression: old dog, new tricks. Bioengineered *6*, 316–322. https://doi.org/10.1080/21655979.2015.1104433.

70. Clem, R.J., and Passarelli, A.L. (2013). Baculoviruses: Sophisticated Pathogens of Insects. PLoS Pathog. *9*, e1003729. https://doi.org/10.1371/journal.ppat.1003729.

71. Scholz, J., and Suppmann, S. (2021). A fast-track protocol for protein expression using the BEV system. Methods Enzymol. *660*, 171–190. https://doi.org/10.1016/bs.mie.2021.06.015.

72. Weissmann, F., and Peters, J.-M. (2018). Expressing Multi-subunit Complexes Using biGBac. In Protein Complex Assembly Methods in Molecular Biology, J.A. Marsh, ed. (Springer New York), pp. 329–343. https://doi.org/10.1007/978-1-4939-7759-8_21.

73. Shen, X., Hacker, D.L., Baldi, L., and Wurm, F.M. (2014). Virus-free transient protein production in Sf9 cells. J. Biotechnol. *171*, 61–70. https://doi.org/10.1016/j.jbiotec.2013.11.018.

74. Bleckmann, M., Fritz, M.H.-Y., Bhuju, S., Jarek, M., Schürig, M., Geffers, R., Benes, V., Besir, H., and Van Den Heuvel, J. (2015). Genomic Analysis and Isolation of RNA Polymerase II Dependent Promoters from Spodoptera frugiperda. PLoS One *10*, e0132898. https://doi.org/10.1371/journal.pone.0132898.

75. Jarvis, D.L. (2014). Recombinant Protein Expression in Baculovirus-Infected Insect Cells. In Methods in Enzymology (Elsevier), pp. 149–163. https://doi.org/10.1016/B978-0-12-420070-8.00013-1.

76. Bleckmann, M., Schürig, M., Chen, F.-F., Yen, Z.-Z., Lindemann, N., Meyer, S., Spehr, J., and Van Den Heuvel, J. (2016). Identification of Essential Genetic Baculoviral Elements for Recombinant Protein Expression by Transactivation in Sf21 Insect Cells. PLoS One *11*, e0149424. https://doi.org/10.1371/journal.pone.0149424.

77. Bleckmann, M., Schürig, M., Endres, M., Samuels, A., Gebauer, D., Konisch, N., and Van Den Heuvel, J. (2019). Identifying parameters to improve the reproducibility of transient gene expression in High Five cells. PLoS One *14*, e0217878. https://doi.org/10.1371/journal.pone.0217878.

78. Mori, K., Hamada, H., Ogawa, T., Ohmuro-Matsuyama, Y., Katsuda, T., and Yamaji, H. (2017). Efficient production of antibody Fab fragment by transient gene expression in insect cells. J. Biosci. Bioeng. *124*, 221–226. https://doi.org/10.1016/j.jbiosc.2017.03.007.

79. Puente-Massaguer, E., Lecina, M., and Gòdia, F. (2018). Nanoscale characterization coupled to multi-parametric optimization of Hi5 cell transient gene expression. Appl. Microbiol. Biotechnol. *102*, 10495–10510. https://doi.org/10.1007/s00253-018-9423-5.

80. Puente-Massaguer, E., Strobl, F., Grabherr, R., Striedner, G., Lecina, M., and Gòdia, F. (2020). PEI-Mediated Transient Transfection of High Five Cells at Bioreactor Scale for HIV-1 VLP Production. Nanomaterials *10*, 1580. https://doi.org/10.3390/nano10081580.

81. Korn, J., Schäckermann, D., Kirmann, T., Bertoglio, F., Steinke, S., Heisig, J., Ruschig, M., Rojas, G., Langreder, N., Wenzel, E.V., et al. (2020). Baculovirus-free insect cell expression system for high yield antibody and antigen production. Sci. Rep. *10*, 21393. https://doi.org/10.1038/s41598-020-78425-9.

82. Kaipa, J.M., Krasnoselska, G., Owens, R.J., and van den Heuvel, J. (2023). Screening of Membrane Protein Production by Comparison of Transient Expression in Insect and Mammalian Cells. Biomolecules *13*, 817. https://doi.org/10.3390/biom13050817.

83. Shen, X., Pitol, A.K., Bachmann, V., Hacker, D.L., Baldi, L., and Wurm, F.M. (2015). A simple plasmid-based transient gene expression method using High Five cells. J. Biotechnol. *216*, 67–75. https://doi.org/10.1016/j.jbiotec.2015.10.007.

84. Kesidis, A., Depping, P., Lodé, A., Vaitsopoulou, A., Bill, R.M., Goddard, A.D., and Rothnie, A.J. (2020). Expression of eukaryotic membrane proteins in eukaryotic and prokaryotic hosts. Methods *180*, 3–18. https://doi.org/10.1016/j.ymeth.2020.06.006.

85. Pieprzyk, J., Pazicky, S., and Löw, C. (2018). Transient Expression of Recombinant Membrane-eGFP Fusion Proteins in HEK293 Cells. In Recombinant Protein Expression in Mammalian Cells Methods in Molecular Biology, D.L. Hacker, ed. (Springer New York), pp. 17–31. https://doi.org/10.1007/978-1-4939-8730-6_2.

86. Goehring, A., Lee, C.-H., Wang, K.H., Michel, J.C., Claxton, D.P., Baconguis, I., Althoff, T., Fischer, S., Garcia, K.C., and Gouaux, E. (2014). Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. Nat. Protoc. *9*, 2574–2585. https://doi.org/10.1038/nprot.2014.173.

87. Durocher, Y., Perret, S., and Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res. *30*, 9e–99e. https://doi.org/10.1093/nar/30.2.e9.

88. Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr. D Biol. Crystallogr. *62*, 1243–1250. https://doi.org/10.1107/S0907444906029799.

89. Delafosse, L., Xu, P., and Durocher, Y. (2016). Comparative study of polyethylenimines for transient gene expression in mammalian HEK293 and CHO cells. J. Biotechnol. *227*, 103–111. https://doi.org/10.1016/j.jbiotec.2016.04.028.

90. Kadlecova, Z., Nallet, S., Hacker, D.L., Baldi, L., Klok, H.-A., and Wurm, F.M. (2012). Poly(ethyleneimine)-mediated large-scale transient gene expression: Influence of molecular weight, polydispersity and *N*-propionyl groups. Macromol. Biosci. *12*, 628–636. https://doi.org/10.1002/mabi.201100404.

91. Baldi, L., Hacker, D.L., Meerschman, C., and Wurm, F.M. (2012). Large-Scale Transfection of Mammalian Cells. In Protein Expression in Mammalian Cells Methods in Molecular Biology, J.L. Hartley, ed. (Humana Press), pp. 13–26. https://doi.org/10.1007/978-1-61779-352-3_2.

92. Fornwald, J.A., Lu, Q., Boyce, F.M., and Ames, R.S. (2016). Gene Expression in Mammalian Cells Using BacMam, a Modified Baculovirus System. In Baculovirus and Insect Cell Expression Protocols Methods in Molecular Biology, D.W. Murhammer, ed. (Springer New York), pp. 95–116. https://doi.org/10.1007/978-1-4939-3043-2_5.

93. Barsoum, J., Brown, R., McKee, M., and Boyce, F.M. (1997). Efficient Transduction of Mammalian Cells by a Recombinant Baculovirus Having the Vesicular Stomatitis Virus G Glycoprotein. Hum. Gene Ther. *8*, 2011–2018. https://doi.org/10.1089/hum.1997.8.17-2011.

94. Mansouri, M., Bellon-Echeverria, I., Rizk, A., Ehsaei, Z., Cianciolo Cosentino, C., Silva, C.S., Xie, Y., Boyce, F.M., Davis, M.W., Neuhauss, S.C.F., et al. (2016). Highly efficient baculovirus-mediated multigene delivery in primary cells. Nat. Commun. *7*, 11529. https://doi.org/10.1038/ncomms11529.

95. Bellón-Echeverría, I., Carralot, J.-P., Del Rosario, A.A., Kueng, S., Mauser, H., Schmid, G., Thoma, R., and Berger, I. (2018). MultiBacMam Bimolecular Fluorescence Complementation (BiFC) tool-kit identifies new small-molecule inhibitors of the CDK5-p25 protein-protein interaction (PPI). Sci. Rep. *8*, 5083. https://doi.org/10.1038/s41598-018-23516-x.

96. Elegheert, J., Behiels, E., Bishop, B., Scott, S., Woolley, R.E., Griffiths, S.C., Byrne, E.F.X., Chang, V.T., Stuart, D.I., Jones, E.Y., et al. (2018). Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. Nat. Protoc. *13*, 2991–3017. https://doi.org/10.1038/s41596-018-0075-9.

97. Behiels, E., and Elegheert, J. (2021). High-Level Production of Recombinant Eukaryotic Proteins from Mammalian Cells Using Lentivirus. In Structural Proteomics Methods in Molecular Biology, R.J. Owens, ed. (Springer US), pp. 83–104. https://doi.org/10.1007/978-1-0716-1406-8_4.

98. Tschorn, N., Berg, K., and Stitz, J. (2020). Transposon vector-mediated stable gene transfer for the accelerated establishment of recombinant mammalian cell pools allowing for high-yield production of biologics. Biotechnol. Lett. *42*, 1103–1112. https://doi.org/10.1007/s10529-020-02889-y.

99. Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive *piggyBac* transposase for mammalian applications. Proc. Natl. Acad. Sci. USA *108*, 1531–1536. https://doi.org/10.1073/pnas.1008322108.

100. Li, Z., Michael, I.P., Zhou, D., Nagy, A., and Rini, J.M. (2013). Simple *piggyBac* transposon-based mammalian cell expression system for inducible protein production. Proc. Natl. Acad. Sci. USA *110*, 5004–5009. https://doi.org/10.1073/pnas.1218620110.

101. Balasubramanian, S., Matasci, M., Kadlecova, Z., Baldi, L., Hacker, D.L., and Wurm, F.M. (2015). Rapid recombinant protein production from piggyBac transposon-mediated stable

CHO cell pools. J. Biotechnol. *200*, 61–69. https://doi.org/10.1016/j.jbiotec.2015.03.001.

102. Caro, L.N., Li, Z., Balo, A.R., Van Eps, N., Rini, J.M., and Ernst, O.P. (2015). Rapid and Facile Recombinant Expression of Bovine Rhodopsin in HEK293S GnTI− Cells Using a PiggyBac Inducible System. In Methods in Enzymology (Elsevier), pp. 307–330. https://doi.org/10.1016/bs.mie.2015.01.005.

103. Hacker, D.L., and Balasubramanian, S. (2016). Recombinant protein production from stable mammalian cell lines and pools. Curr. Opin. Struct. Biol. *38*, 129–136. https://doi.org/10.1016/j.sbi.2016.06.005.

104. Michael, I.P., and Nagy, A. (2018). Inducible Protein Production in 293 Cells Using the piggyBac Transposon System. In Recombinant Protein Expression in Mammalian Cells Methods in Molecular Biology, D.L. Hacker, ed. (Springer New York), pp. 57–68. https://doi.org/10.1007/978-1-4939-8730-6_5.

105. Suppmann, S. (2021). Inducible protein expression in piggyBac transposase mediated stable HEK293 cell pools. In Methods in Enzymology (Elsevier), pp. 321–339. https://doi.org/10.1016/bs.mie.2021.06.016.

106. Silverman, A.D., Karim, A.S., and Jewett, M.C. (2020). Cell-free gene expression: an expanded repertoire of applications. Nat. Rev. Genet. *21*, 151–170. https://doi.org/10.1038/s41576-019-0186-3.

107. Foshag, D., Henrich, E., Hiller, E., Schäfer, M., Kerger, C., Burger-Kentischer, A., Diaz-Moreno, I., García-Mauriño, S.M., Dötsch, V., Rupp, S., and Bernhard, F. (2018). The E. coli S30 lysate proteome: A prototype for cell-free protein production. N. Biotech. *40*, 245–260. https://doi.org/10.1016/j.nbt.2017.09.005.

108. Kögler, L.M., Stichel, J., and Beck-Sickinger, A.G. (2019). Structural investigations of cell-free expressed G protein-coupled receptors. Biol. Chem. *401*, 97–116. https://doi.org/10.1515/hsz-2019-0292.

109. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. Nat. Biotechnol. *19*, 751–755. https://doi.org/10.1038/90802.

110. Zemella, A., Grossmann, S., Sachse, R., Sonnabend, A., Schaefer, M., and Kubick, S. (2017). Qualifying a eukaryotic cell-free system for fluorescence based GPCR analyses. Sci. Rep. *7*, 3740. https://doi.org/10.1038/s41598-017-03955-8.

111. Merk, H., Rues, R.-B., Gless, C., Beyer, K., Dong, F., Dötsch, V., Gerrits, M., and Bernhard, F. (2015). Biosynthesis of membrane dependent proteins in insect cell lysates: identification of limiting parameters for folding and processing. Biol. Chem. *396*, 1097–1107. https://doi.org/10.1515/hsz-2015-0105.

112. Hein, C., Henrich, E., Orbán, E., Dötsch, V., and Bernhard, F. (2014). Hydrophobic supplements in cell-free systems: Designing artificial environments for membrane proteins. Eng. Life Sci. *14*, 365–379. https://doi.org/10.1002/elsc.201300050.

113. Bruni, R., Laguerre, A., Kaminska, A.M., McSweeney, S., Hendrickson, W.A., and Liu, Q. (2022). High-throughput cell-free screening of eukaryotic membrane protein expression in lipidic mimetics. Protein Sci. *31*, 639–651. https://doi.org/10.1002/pro.4259.

114. Kuruma, Y., and Ueda, T. (2015). The PURE system for the cell-free synthesis of membrane proteins. Nat. Protoc. *10*, 1328–1344. https://doi.org/10.1038/nprot.2015.082.

115. Krug, U., Gloge, A., Schmidt, P., Becker-Baldus, J., Bernhard, F., Kaiser, A., Montag, C., Gauglitz, M., Vishnivetskiy, S.A., Gurevich, V.V., et al. (2020). The Conformational Equilibrium of the Neuropeptide Y2 Receptor in Bilayer Membranes. Angew. Chem. Int. Ed. *59*, 23854–23861. https://doi.org/10.1002/anie.202006075.

116. Reckel, S., Gottstein, D., Stehle, J., Löhr, F., Verhoefen, M.-K., Takeda, M., Silvers, R., Kainosho, M., Glaubitz, C., Wachtveitl, J., et al. (2011). Solution NMR Structure of Proteorhodopsin. Angew. Chem. Int. Ed. *50*, 11942–11946. https://doi.org/10.1002/anie.201105648.

117. Boland, C., Li, D., Shah, S.T.A., Haberstock, S., Dötsch, V., Bernhard, F., and Caffrey, M. (2014). Cell-free expression and in meso crystallisation of an integral membrane kinase for structure determination. Cell. Mol. Life Sci. *71*, 4895–4910. https://doi.org/10.1007/s00018-014-1655-7.

118. Garenne, D., Haines, M.C., Romantseva, E.F., Freemont, P., Strychalski, E.A., and Noireaux, V. (2021). Cell-free gene expression. Nat. Rev. Methods Primers *1*, 49. https://doi.org/10.1038/s43586-021-00046-x.

119. Takeda, M., and Kainosho, M. (2012). Cell-Free Protein Production for NMR Studies. In Protein NMR Techniques Methods in Molecular Biology, A. Shekhtman and D.S. Burz, eds. (Humana Press), pp. 71–84. https://doi.org/10.1007/978-1-61779-480-3_5.

STAR Protocols, Volume *4*

Supplemental information

A concise guide to choosing suitable gene expression systems for recombinant protein production

Anja Schütz, Frank Bernhard, Nick Berrow, Johannes F. Buyel, Frederico Ferreira-da-Silva, Jurgen Haustraete, Joop van den Heuvel, Jan-Erik Hoffmann, Ario de Marco, Yoav Peleg, Sabine Suppmann, Tamar Unger, Martine Vanhoucke, Susanne Witt, and Kim Remans

Supplemental information

Overview of the supplemental information files

I. Plants as an alternative protein production system

II. “Exotic” gene expression systems

III. Expression vectors and strains: how to choose them?

IV. Access to biological resources

V. Equipment list

VI. P4EU survey results

VII. Supplementary references

Supplementary File I: Plants as an alternative protein production system

Various plant species can be used for recombinant protein production through stable transformation, e.g. in transgenic maize or tobacco[1,2], or transient induction e.g. by infiltrating *Nicotiana benthamiana* plants with genetically modified viral vectors or *Rhizobium radiobacter* (formerly *Agrobacterium tumefaciens*) that has the natural ability to transfer DNA into plant cells. Expression in transgenic plants can be rapidly scaled up[3], but it is a labor-intensive, complex and lengthy process to obtain such plants. Therefore, rapid protein production will typically rely on transient gene expression that takes ~ 5-14 days from DNA sequence to milligram quantities of protein in intact plants[4], e.g. for activity studies. Such expression is easily carried out under non-sterile conditions and therefore adopted by many laboratories. Dedicated infrastructure such as greenhouses or phytotrons are necessary to ensure reproducibility of this approach[5]. Alternatively, plant cells from suspension cultures can be used for transient gene expression too[6]. For example, a semi-dry format called plant cell packs (PCPs), which is high-throughput compatible in 96-well plates with running costs of about 0.5 € per gene expression[7], achieves expression in 3-5 days from DNA to microgram quantities of protein. The necessary plant cell cultures are readily established using regular shake incubators and do not require dedicated equipment and reach cell wet masses of ~200–300 g L^{-1} within two weeks in case of *N. tabacum* bright yellow 2 cells[8].

Regardless of whether plants or plant cells are used, the recombinant protein production capabilities in terms of wet biomass are moderate, typically 10 to 500 mg kg^{-1} but levels up to 6000 mg kg^{-1} have been reported[9,10]. A major reason is the biosynthetically inactive vacuole that accounts for ~50% of the cell volume and mass depending on the cell type and culture conditions[11,12]. Note that for intact plants, 1 kg of wet plant biomass is approximately equivalent to 1 L of fermentation broth in terms of cell dry mass[13].

A substantial advantage of plants and plant cells is that they can effectively secrete and fold complex (human) proteins[14]. In fact, even toxic proteins such as abrin, ricin and viscumin as well as fusion proteins thereof can easily be produced, e.g. by targeting the proteins to compartments that separate them from potential molecular targets[15]. In general, producing target proteins in the different compartments of plant cells should be implemented in a regular screening approach to identify optimal conditions for protein accumulation and be guided by the properties and origin of the recombinant protein. For example, intracellular/cytosolic proteins will typically be targeted to the plant cytosol too, but targeting to the endoplasmic reticulum (ER) may improve accumulation, e.g. due to protection against proteases[16]. Targeting to the ER/secretory pathway will also facilitate disulfide bond formation and glycosylation due to the presence of oxidizing conditions and glycosyltransferases, respectively. Importantly, protein glycosylation is introduced properly and genetically modified host plants as well as plant cell lines exist that introduce human glycosylation[17].

In addition, the choice of the targeted cellular compartment can have implications on the subsequent downstream processing that is not described here. For example, some proteins targeted to the ER may require the presence of detergents to be recovered[18], whereas such additives can solubilize additional (membrane) proteins as well[7], which may complicate purification. In the purification context, plant cell cultures can be advantageous as they often do not contain chlorophylls and other pigments that need to be separated from a protein product.

As an additional feature, plants and especially plant cell cultures facilitate labeling of complex proteins with isotopes, e.g. for protein structure elucidation[19].

Supplementary File II: "Exotic" gene expression systems

"EXOTIC" EXPRESSION SYSTEMS			
Name	**Advantages**	**Limitations**	**References**
Lactococcus lactis (Gram-positive bacterium)	• Lipopolysaccharide (LPS)-free micro-organism • Fast growth rate (t_D= 30-60 min) • Secretion to the medium possible • Commercially available systems	• Low cloning efficiency • Codon optimisation of gene(s) of interest required • Frequent aggregation of heterologous proteins	20, 21, 22, 23, 24
Bacillus subtilis (Gram-positive bacterium)	• LPS-free micro-organism • Fast growth rate (t_D= 30 min) • Secretion to the medium possible • Broad codon usage • Commercially available systems • Important host for the production of industrially relevant proteins and chemicals	• Screening of various genetic elements (promoters, signal sequences, ribosome binding sites etc.) and strains can be required for optimising production titers • Protein production tools not as well characterised as for *E. coli*	25, 26, 27, 28, 29, 30
Vibrio natriegens (Gram-negative bacterium)	• Very fast growth rate (t_D= <20 min) • Growth to high cell densities • Compatible with many *E. coli* expression vectors • Commercially available systems (Vmax)	• Lower transformation efficiencies than *E. coli* • Commercially available media rather expensive • Cold sensitive • Natural resistance to kanamycin	31, 32, 33, 34, 35
Pseudomonas putida	• Important industrial metabolic engineering and synthetic biology chassis • High tolerance to xenobiotics • Variety of genetic tools available	• Most of the standard ORIs present in *E. coli* expression plasmids incapable of replication in *P. putida* • Well-characterised *E. coli* inducible promoter systems behave differently in *P. putida*	36, 37 ,38, 39, 40, 41
Mycobacterium smegmatis	• Used if expression of genes from different mycobacterial species in *E. coli* fails	• Slow growth (t_D= 3 h) • Expression process is lengthy • Sometimes low yields, no	42

	● Variety of genetic tools available ● Introduction of specific post-translational modifications (PTMs) for Mycobacteria ● Incorporation of ligands unique for mycobacteria	expression and insoluble expression ● Application requires adaptation of specific protocols and know-how in all stages	
Drosophila Schneider 2 (S2) cells	● Eukaryotic PTMs ● High secretion capacity ● Growth rate (t_D= 24 h) ● Growth to high cell densities ● Diverse growth conditions (serum-free and serum-containing media) ● Semi-adherent and suspension culturing possible ● Transient transfection or stable cell lines possible ● Commercially available systems	● Time required for establishing stable cell pools ● Limited cell growth at low cell densities	43, 44, 45, 46, 47
Chlamydomonas reinhardtii (unicellular green algae)	● Eukaryotic PTMs ● Growth rate (t_D= 7-14 h) ● Secretion to the medium possible ● Nuclear or chloroplast expression possible ● Correct folding and assembly of complex proteins ● Commercially available systems	● Codon optimisation of gene(s) of interest required ● Secretion/Glycosylation only possible upon nuclear production, but yields generally lower than for chloroplast production	48, 49, 50, 51, 52

Table S1: "Exotic" gene expression systems
This table presents an overview of some less commonly used protein production systems and their respective advantages and limitations. References for more in-depth information are provided as well for the readers that have an interest in these expression hosts organisms. In general, we recommend contacting experienced groups before attempting to set up some of these more "exotic" gene expression systems in-house.

Supplementary File III: Expression vectors and strains/cell lines: how to choose them?

The majority of expression vectors have reached a mature phase, meaning that the wide diversity in terms of characteristics present 30 years ago has now converged to relatively simplified backbones with certain individual features. Among the elements that can vary among vectors, it is important to underline the relevance of the origin of replication (ori), the promoter, the presence of purification and/or other fusion tags, protease cleavage sites, the presence of a signal sequence, the selection marker and the multiple cloning site[53,54].

The ori determines the vector copy number per cell and therefore contributes to establish the rate of accumulation of recombinant protein. Since the host cell folding machinery is limited, one option to slow down recombinant gene expression to favor correct folding is to use low copy number vectors. Another important element determining the expression rate is the promoter, which must be regulated to avoid "leakage" (basal expression of the gene(s) of interest in the absence of a specific inducer), as this can lead to cell toxicity. Promoters also vary in strength and hence differ in their efficiency in supporting RNA synthesis. Tags can be added to the N- or C-termini of the protein(s) of interest in order to simplify the affinity purification (His, Strep, Flag etc.[54,55]), but they can also be used to improve the stability of the recombinant protein (maltose binding protein, SUMO etc.[56]), to provide different functionalities (fluorescent proteins, enzymes) or to assist downstream derivatization and assembling (cysteine, SpyTag, recognition sequence for sortases, biotinylation sequence etc.). Protease cleavage sites (TEV, HRV 3C, thrombin etc.[57]) are often added to allow the removal of downstream tags, for instance when the protein will be used for X-ray crystallography. When proteins need to be secreted to the periplasm (*E. coli*) or to the extracellular milieu, signal sequences are required as well. Finally, expression plasmids generally contain (antibiotic) selection markers and a multiple cloning site, although the latter is less important than in the past given the increased use of sequence- and ligation-independent cloning methods.

Usually, there are several expression **strains and cell lines** available for a particular host organism, which might differ in their specific characteristics (e.g. expression levels, growth rate, folding capacity for certain types of proteins, glycosylation pattern etc.). Although information about commonly used strains and cell lines is available in literature, it's always useful to confer with experts before deciding which specific expression strains or cell lines to acquire. In general, setting up a new gene expression system and purchasing the right plasmid backbones and strains/cell lines will be greatly facilitated by discussions with experienced scientists. Table S2 describes some of the most commonly used expression strains/cell lines and vectors for the major gene expression systems (*E. coli*, yeast, insect and mammalian cells). However, this is by no means an exhaustive list of all available systems and more detailed information can be found in more focused (review) papers (appropriate references are mentioned in the individual sections of the main manuscript).

For protein production in ***E. coli***, the pET-based vectors are some of the most commonly used expression vectors. As the gene(s) of interest are placed under control of the T7 promoter in pET-based vectors, they must be used in combination with *E. coli* expression strains encoding the T7 RNA polymerase, such as *E. coli* BL21(DE3)[58] and its derivatives. For **yeast**, the pPICZ- and pPIC9-based expression vectors are popular choices for protein production in *Pichia pastoris*. For baculovirus-mediated gene expression in **insect cells**, pFastBac-derived plasmids are often used when transposition-based methods are utilised for the generation of bacmids in *E. coli*, whereas for example the FlexiBAC pOCC and *flash*BAC pOET vectors are suitable backbones when homologous recombination-based methods in insect cells are used. As **TGE in insect cells** is still an up-and-coming method, there are not so many different expression vectors available yet, but the pOpiE2 represents a good choice. For **TGE in mammalian cells** a large variety of expression vectors is available, with the pCDNA-, pCMV- and pHLsec-based plasmids being some of the most frequently used ones. The most

suitable vectors to generate **stable mammalian cell** lines for protein production depend very much on the chosen method for gene integration. The piggyBac plasmids for transposase-mediated gene integration represent a good example of a user-friendly and relatively quick method to establish stable mammalian pools (see section "protein production in mammalian cells" in the main manuscript)[59,60].

ESCHERICHIA COLI (E. coli)		
***E. coli* strains**	**Characteristics**	**Usage**
E. coli BL21	Deficient in *lon* and *ompT* proteases	Standard protein production strain
E. coli BL21(DE3)	T7 RNA polymerase gene under control of the *lacUV5* promotor	IPTG-inducible expression of genes under control of the T7 promotor
E. coli BL21(DE3) pLysS/pLysE	Extra plasmid that encodes T7 lysozyme, which represses the T7 RNA polymerase	Repression of basal expression for proteins causing toxicity issues
E. coli Origami2(DE3)	Mutations in *trx* and *gorB*, leading to a less reducing environment in the cytosol	Cytosolic production of proteins containing disulfide bonds
E. coli SHuffle T7 Express	Mutations in *trx* and *gorB*, leading to a less reducing environment in the cytosol; cytosolic expression of the DsbC isomerase	Cytosolic production of proteins containing disulfide bonds
E. coli Rosetta2(DE3)	Extra plasmid that encodes tRNAs for 7 rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG)	Expression of genes containing codons that are rare in *E. coli*
***E. coli* expression vectors**	**Characteristics**	**Usage**
pET-based plasmids	Strong bacteriophage T7 promoter	Protein production in host cells expressing the T7 RNA polymerase
pBAD-based plasmids	Arabinose inducible *ara*BAD promoter; tight regulation (repression) possible via glucose	Tightly regulatable and inducible expression of recombinant proteins
pGEX-based plasmids	*tac* promoter	Production of GST-fusion proteins
YEAST		
Yeast strains	**Characteristics**	**Usage**
Kogamataella pastoris KM71H	*aox1::Arg4, arg4* genotype	Selection of Zeocin-resistant strains with Mut^s phenotype
Kogamataella pastoris SMD1168H	*Pep4* genotype	Selection of Zeocin resistant strains with Mut^+ phenotype without Protease A activity
Kogamataella pastoris GS115	*his4* genotype	Auxotrophic selection of *HIS4*-containing vectors
Kogamataella pastoris X33	*Wild type* strain	Selection of Zeocin resistant strains

Yeast expression vectors	Characteristics	Usage
pPICZ-based plasmids	Enables direct selection of multiple integration events by increasing Zeocin resistance; integration in *AOX1* promoter region	Methanol-induced expression (Mut^+ phenotype); pPICZ-derivatives are used for the expression of intracellular proteins; pPICZα-derivatives are used for the expression of secreted proteins
pPIC9K	*HIS4* selection; enables direct selection of multiple integration events by increasing Geneticin (G418) resistance; integration in *AOX1* promoter region or gene replacement of *AOX1* by double cross-over	Methanol-induced expression (Mut^+ or Mut^S phenotype); used for the expression of secreted proteins
pPIC3.5K	*HIS4* selection; enables direct selection of multiple integration events by increasing Geneticin (G418) resistance; integration in *AOX1* promoter region or gene replacement of *AOX1* by double cross-over	Methanol-induced expression (Mut^+ or Mut^S phenotype); used for the expression of intracellular proteins
pGAPZ-based plasmids	Zeocin selection; integration in the *GAP* promoter region	Constitutive expression; pGAPZ-derivatives are used for the expression of intracellular proteins; pGAPZα-derivatives are used for the expression of secreted proteins
INSECT CELLS		
Insect cell lines	**Characteristics**	**Usage**
Spodoptera frugiperda (*Sf*9, *Sf*21)	Suspension cultivation at 27°C	BEVS
Trichoplusia ni (*Tni*5, High Five™)	Suspension cultivation at 27°C	BEVS, TGE
Vectors for baculovirus-mediated expression	**Characteristics**	**Usage**
pFastBac, pFastBac-Dual	Site-specific transposition into bacmid in *E. coli (*DH10Bac, DH10MultiBac, DH10EMBacY)	Single gene expression, co-expression of 2 genes
biGBac	PCR-based multi-gene assembly compatible with transposition-based integration	Single gene expression, multi-subunit protein complexes
MultiBac	Cre/Lox-based multi-gene assembly compatible with transposition-based integration	Single gene expression, multi-subunit protein complexes

MacroBac	Biobricks-type multi-gene assembly based on restriction/ligation or ligation-independent cloning; compatible with transposition-based integration	Single gene expression, multi-subunit protein complexes
GoldenBac	Restriction enzyme class II-based multi-gene assembly; compatible with transposition- and recombination-based integration	Multi-subunit protein complexes
FlexiBAC pOCC vectors	Recombination-based integration in insect cells	Linearized bacmid DNA
pOET transfer vectors	Recombination-based integration in insect cells	*Flash*BAC™-linearized bacmid DNA
Vectors for transient gene expression	**Characteristics**	**Usage**
pOpiE2-based plasmids	Strong constitutive immediate early promoter 2 (*Orgyia pseudotugata*)	PEI-mediated TGE
MAMMALIAN CELLS		
Mammalian cell lines	**Characteristics**	**Usage**
HEK293T	Growth in suspension; contains the SV40 T antigen in the genome	Plasmids with SV40 ori
HEK293F	Growth in suspension in serum-free medium	Large culture volumes
Expi293F	Growth in suspension in serum-free medium; high cell densities	High yields; lower culture volumes
MEXi-293E	Growth in suspension in serum-free medium; EBNA1 expression	Episomal replication of plasmids with oriP (e.g. pTT-derivatives)
HEK293-6E	Growth in suspension in serum-free medium; EBNA1 expression	Episomal replication of plasmids with oriP (e.g. pTT derivatives)
HEK293 GnTI- (ATCC CRL3022)	Growth in suspension; deficient in N-acetylglucosaminyltransferase I (GnTI) activity; lack of complex N-glycans	Protein crystallization
Expi293F™ GnTI- Cells	Growth in suspension; deficient in N-acetylglucosaminyltransferase I (GnTI) activity; lack of complex N-glycans	Protein crystallization
CHO DG44	Growth in suspension in serum-free medium; DHFR selection; long-term stability	High protein titers; ideal for GMP procedures
Expi CHO-S	Growth in suspension in serum-free medium; high cell densities	High protein titers; recombinant antibody production
NS0	Lymphoblast mouse myeloma; growth in suspension	Monoclonal antibody production

Vectors for transient gene expression	**Characteristics**	**Usage**
pCDNA derivatives, pCMV derivatives	CMV promoter; SV40 ori; Neomycin resistance	Constitutive expression
pHLSec	Secretion signal; C-terminal His-tag; chimeric intron	Secreted proteins
pTT derivatives	EBV oriP; improved CMV expression cassette	High levels of protein production
Vectors for generating stable pools	**Characteristics**	**Usage**
hyPBase (Sanger institute) Sleeping Beauty	Hyperactive PiggyBac transposase Sleeping Beauty transposase	Transposition based non-specific gene integration
Expression plasmid containing respective antibiotic selection marker and gene insertion flanked by transposition sites, e.g. PB-T-PAF / PB-RN	Hygromycin selection; Tet-on	Induced protein expression (Doxycycline)

Table S2: Overview of the most commonly used expression strains/cell lines and vectors for the major gene expression systems.

This table provides an overview of the most commonly used *E. coli*, yeast, insect and mammalian expression strains/cell lines and expression vectors for protein production. This is by no means an exhaustive list of all available systems, but rather a summary of easily accessible systems that are broadly used in protein production laboratories. More detailed information can be found in various focused references cited in the individual sections of the main manuscript as well as on the websites of the cited manufacturers of gene expression systems and of biological resource centers (see Suppl. File IV).

Supplementary File IV: Access to biological resources

High-quality biological resources and related information are key elements on which protein production systems are built. Easy access to valuable biological material is therefore essential in this regard, but it is often hampered by inefficient storage conditions, irreproducible quality, poor data registration, incorrect distribution modalities, scarce accessibility and, more often than expected, trivial mislabeling which results in handling material with characteristics different from those expected. Next to commercial companies and some institutional databanks, Biological Resource Centres (BRC) or culture collections in general, and those offering recombinant expression plasmids and host strains more specifically, meet the requirements to overcome the potential issues listed above.

BRCs have a longstanding experience in the preservation and distribution of bacterial, fungal and yeast strains, plasmids, DNA libraries and cell lines. They provide long-term storage of the biological material under quality-controlled conditions, applying the most appropriate storage methods and organizing a material back-up at another location. They subject the strains and genetic resources to stringent quality controls, guaranteeing the purity, viability and authenticity of the material. They process the related information according to internationally agreed norms and provide detailed open access data. By referring to the depositor, they increase the visibility of the scientist on one hand and of the related department/university on the other. Last but not least, they guarantee the rapid delivery of samples, respecting (inter)national legislation regarding packaging and shipping of biological material and carefully enforcing terms of use and any restrictions that may apply to the ordered samples.
For some of the BRCs, the activities are covered by a (internationally recognized) quality management system. Moreover, the BRCs dealing with recombinant (expression) plasmids have specific expertise to support researchers in their choice of suitable material.

The Belgian, ISO9001-certified BCCM/GeneCorner Plasmid Collection as well as the American plasmid repository Addgene both possess extensive plasmid collections and are often a good starting point for obtaining expression plasmids and plasmid vectors. BCCM/GeneCorner also offers quite some plasmid-related host strains. Furthermore, non-exhaustive lists of non-profit BRCs that distribute plasmids, vectors and/or production hosts are available on the websites of global or regional networks such as the World Federation for Culture Collections (WFCC, wfcc.info), the European Culture Collections' Organisation (ECCO, eccosite.org), the Microbial Resource Research Infrastructure (MIRRI, mirri.org), the Asian Network of Research Resource Centers (ANRRC, anrrc.info), the Asian Consortium for the Conservation and Sustainable Use of Microbial Resources (ACM, acm-mrc.asia), the United States Culture Collection Network (USCCN, usccn.org) and the Federación Latinoamericana de Colecciones de Cultivos (FELACC, felacc.cinvestav.mx). Some of these platforms offer single access points to an ever-growing number of high-quality, safe and legally fit-for-use biological material made available by its members and covering all types of microbial and genetic resources.

The responsibility to deposit microorganisms and genetic resources in public BRCs is shared by different key players, i.e. researchers, funding agencies and publishers[61]. Researchers can provide easy access to material by storing their biological resources in publicly available BRCs. In parallel, when applicable, it is necessary to deposit sequences to obtain an unambiguous reference to tag the biological material and refer to it in publications. Storing biological material in an internationally accessible public culture collection frees the researcher from the task of personally providing it to whom requests it and has a multiplier effect on further research related to that biological material[62]. A public deposit also contributes to transparency and reproducibility, and supports the principles of scientific integrity, open science and FAIR data (findability, accessibility, interoperability, and re-usability). Beyond the public deposit service, several BRCs also offer confidential (no public access) deposit possibilities, e.g. in the case of data related to intellectual property rights.

Supplementary File V: Equipment list

	E. coli	Yeast	Insect BEVs	Mammalian Transient	Mammalian Stable
Basic molecular biology laboratory equipment	x	x	x	x	x
Temperature-controlled shaker	x	x	x	x	x
Temperature-controlled shaker with CO_2 and humidity control				x	x
Laminar flow cabinet	(x)	(x)	x	x	x
Centrifuge for harvesting large scale cultures	x	x	x	x	x
High-pressure homogenizer or sonicator for cell lysis	x	x	x	x	x
Electroporation system and cuvettes		x			
Spectrophotometer	x	x			
Cell counter			(x)	(x)	(x)
Inverse cell culture microscope			x	x	x
Cell line storage at or below -150°C			x	x	x
Flow cytometry			(x)	(x)	(x)

Table S3: Equipment list for protein production experiments

This table provides an overview of the instrumentation that is commonly used for protein production. Basic molecular biology laboratory equipment (static incubators, gel electrophoresis set-up, Eppendorf and falcon tube centrifuges, power supply etc.) is required independent of the chosen host organism. As shown above, temperature-controlled shaking incubators are necessary for all gene expression systems as well, whereas mammalian cells require CO_2 and humidity control on top of temperature regulation. For working with insect and mammalian cells, a laminar flow cabinet is indispensable. For *E. coli* and yeast, it is possible to manipulate the cells on the bench (simply using a flame), although some laboratories prefer to work in laminar flow cabinets as well. For harvesting large scale expression cultures, specialized centrifuges that fit larger volumes are necessary. For analysis of protein production, cells need to be lysed. For cell lysis, high-pressure homogenizers or sonicators can be used, although insect and mammalian cells often break spontaneously after resuspension in buffer and/or a freeze-thaw cycle. The most efficient method to introduce foreign DNA in yeast is electroporation, which can also be used for other expression host organisms. For *E. coli*, chemically competent cells are a valid alternative to electrocompetent cells for introducing plasmid DNA. For insect and mammalian cells, both virus-based infections/transductions and plasmid DNA transfections can be used to introduce foreign DNA into the cells. To follow cell growth and measure the optical density for *E. coli* and yeast, standard spectrophotometers can be used. For insect and mammalian cells, specialized cell counters provide a convenient way to measure the cell density, but counting chambers in combination with a cell culture microscope are suitable as well. A cell culture microscope is essential when working with insect and mammalian cells to assess the state of the cells regularly (e.g. to check the shape, size, sources of contamination etc.). Storage of master banks of cell lines requires storage either in a freezer at or below -150°C or in the gas phase of a liquid nitrogen cryo-tank. Flow cytometry can be a useful technology to assess baculoviral titers, transfection efficiencies and expression levels when working with insect and mammalian cells.

Supplementary File VI: P4EU survey results

Questionnaire – Expression system selection for protein production

With this survey, we aim to collect experiences with the application of particular expression systems for protein production from different labs in the P4EU community. Please answer the questions based on **your personal PRACTICAL experience** (experiments performed in your lab) rather than textbook knowledge.

Most questions can be answered by a simple click. The survey will take about 15 min to accomplish.

Thanks for your participation and time!

Summary: Here, the results from the 60 fully answered surveys are shown.

Color scheme:

main
less frequently used
minor

Section 1 – Expression systems applied in your lab

1. Please estimate the number of target proteins you process per year.
range 3 - 1500 (∅ 112) (enter number) (mandatory to answer)

2. Please rank the frequency (% of expression experiments performed in your lab) of applying specific expression systems.

(mandatory to check one box each line)

Expression system	not used	less frequently used (<20%)	frequently used (20-50%)	very frequently used (50-75%)	most frequently used (≥75%)
E. coli	3 (5%)	10 (16,7%)	12 (20%)	17 (28,3%)	18 (30%)
Bacillus subtilis	58 (96,7%)	0 (0%)	1 (1,7%)	1 (1,7%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	46 (76,7%)	8 (13,3%)	2 (3,3%)	3 (5%)	1 (1,7%)
Saccharomyces cerevisiae	51 (85%)	5 (8,3%)	3 (5%)	1 (1,7%)	0 (0%)
HEK293 - transient gene expression (transfection agent-based)	16 (26,7%)	19 (31,7%)	14 (23,3%)	8 (13,3%)	3 (5%)
HEK293 - stable gene integration	27 (61,7%)	20 (33,3%)	1 (1,7%)	1 (1,7%)	1 (1,7%)
CHO - transient gene expression (transfection agent-based)	42 (70%)	13 (21,7%)	2 (3,3%)	3 (5%)	0 (0%)
CHO - stable gene integration	50 (83,3%)	8 (13,3%)	1 (1,7%)	0 (0%)	1 (1,7%)
BacMam transduced HEK293 cells	53 (88,3%)	5 (8,3%)	1 (1,7%)	0 (0%)	1 (1,7%)
BacMam transduced CHO cells	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced HEK293 cells	58 (96,7%)	2 (3,3%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced CHO cells	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Insect cells - BEVs	25 (41,7%)	8 (13,3%)	18 (30%)	5 (8,3%)	4 (6,7%)
Insect cells - transient gene expression	53 (88,3%)	6 (10%)	1 (1,7%)	0 (0%)	0 (0%)
Plants - transient gene expression	59 (98,3%)	0 (0%)	0 (0%)	1 (1,7%)	0 (0%)
Plants - stable gene integration	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60	0	0	0	0

	(100%)	(0%)	(0%)	(0%)	(0%)
Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cell-free expression (*in vitro*)	51 (85%)	8 (13,3%)	1 (1,7%)	0 (0%)	0 (0%)

Cut-off criteria:

- "≤ 80% not used" ⇒ main
- "80-90% not used" ⇒ less frequently used
- "91-100% not used" ⇒ minor

Main systems: *E. coli*, *Pichia pastoris*, HEK293-transient, HEK293-stable, CHO-transient and Insect cells-BEVs

Less frequently used systems: *Saccharomyces cerevisiae*, CHO-stable, HEK293-BacMam, Insect cells-transient and *in vitro* cell-free expression

Minor systems: *Bacillus subtilis*, *Lactococcus lactis*, CHO-BacMam, HEK293-Lentivirus, CHO-Lentivirus, Plants-transient, Plants-stable, *Algae*, *Leishmania tarentolae* and Filamentous fungi

3. Do you use any other expression system(s) not listed here? Please enter below and indicate the percentage of frequency used.
_____ (enter free text/numbers) (not mandatory to answer)

Answer	Count	Percentage
Answer	20	33.33%
No answer	40	66.67%

Alternative expression systems used occasionally in the P4EU community:

- *Mycobacterium smegmatis*
- Hybridoma cell lines
- *Vibrio natriegens*
- *Brevibacillus*

Section 2 – Ease of use

Depending on **your personal experience**, please rank the ease of use for various expression systems. Factors to consider are: SOP (protocol); user training, simple experience (> 5 projects hands-on), complex experience (more than 1 year hands-on).

1 = possible with SOP + user training + complex experience
2 = possible with SOP + user training + simple experience
3 = possible with SOP + user training
4 = possible with SOP + simple experience
5 = possible with SOP only

(mandatory to check one box each line)

Expression system	not used	1	2	3	4	5
E. coli	2 (3,3%)	4 (6,7%)	6 (10%)	10 (16,7%)	25 (41,7%)	13 (21,7%)
Bacillus subtilis	58 (96,7%)	0 (0%)	2 (3,3%)	0 (0%)	0 (0%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	44 (73,3%)	2 (3,3%)	4 (6,7%)	5 (8,3%)	5 (8,3%)	0 (0%)
Saccharomyces cerevisiae	50 (83,3%)	0 (0%)	2 (3,3%)	2 (3,3%)	5 (8,3%)	1 (1,7%)
HEK293 - transient gene expression (transfection agent-based)	13 (21,7%)	7 (11,7%)	**19** (31,7%)	19 (31,7%)	2 (3,3%)	0 (0%)
HEK293 - stable gene integration	33 (55%)	8 (13,3%)	17 (28,3%)	2 (3,3%)	0 (0%)	0 (0%)
CHO - transient gene expression (transfection agent-based)	40 (66,7%)	2 (3,3%)	8 (13,3%)	9 (15%)	1 (1,7%)	0 (0%)
CHO - stable gene integration	46 (76,7%)	4 (6,7%)	7 (11,7%)	2 (3,3%)	1 (1,7%)	0 (0%)
BacMam transduced HEK293 cells	52 (86,7%)	4 (6,7%)	1 (1,7%)	3 (5%)	0 (0%)	0 (0%)
BacMam transduced CHO cells	59 (98,3%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced HEK293 cells	55 (91,7%)	3 (5%)	2 (3,3%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced CHO cells	57 (95%)	2 (3,3%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Insect cells - BEVs	23 (38,3%)	11 (18,3%)	18 (30%)	6 (10%)	1 (1,7%)	1 (1,7%)
Insect cells - transient gene expression	49 (81,7%)	5 (8,3%)	3 (5%)	3 (5%)	0 (0%)	0 (0%)
Plants - transient gene expression	59 (98,3%)	0 (0%)	0 (0%)	0 (0%)	1 (1,7%)	0 (0%)
Plants - stable gene integration	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60	0	0	0	0	0

	(100%)	(0%)	(0%)	(0%)	(0%)	(0%)
filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
cell-free expression (in vitro)	52 (86,7%)	1 (1,7%)	2 (3,3%)	**2** (3,3%)	1 (1,7%)	2 (3,3%)

Conclusions:

- *E. coli* and *Saccharomyces cerevisiae* are the easiest to use systems (both possible with SOP + simple experience).
- BacMam, lentiviral transduction of mammalian cells and transient gene expression in insect cells are the most demanding systems (possible with SOP + user training + complex experience).

Section 3 – Speed

Based on **your practical experience**, please rank the speed of expression experiments from expression vector to biomass (produced protein) for various expression systems, assuming a 1 L scale.

(mandatory to check one box each line)

Expression system	not used	1-3 days	3-7 days	1-4 weeks	4-8 weeks	>8 weeks
E. coli	2 (3,3%)	37 (61,7%)	20 (33,3%)	1 (1,7%)	0 (0%)	0 (0%)
Bacillus subtilis	58 (96,7%)	1 (1,7%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	43 (71,7%)	0 (0%)	2 (3,3%)	12 (20%)	3 (5%)	0 (0%)
Saccharomyces cerevisiae	51 (85%)	0 (0%)	5 (8,3%)	4 (6,7%)	0 (0%)	0 (0%)
HEK293 - transient gene expression (transfection agent-based)	16 (26,7%)	1 (1,7%)	24 (40%)	18 (30%)	1 (1,7%)	0 (0%)
HEK293 - stable gene integration	35 (58,3%)	0 (0%)	1 (1,7%)	9 (15%)	11 (18,3%)	4 (6,7%)
CHO - transient gene expression (transfection agent-based)	40 (66,7%)	0 (0%)	7 (11,7%)	11 (18,3%)	2 (3,3%)	0 (0%)
CHO - stable gene integration	47 (78,3%)	0 (0%)	1 (1,7%)	3 (5%)	8 (13,3%)	1 (1,7%)
BacMam transduced HEK293 cells	53 (88,3%)	0 (0%)	1 (1,7%)	4 (6,7%)	1 (1,7%)	1 (1,7%)
BacMam transduced CHO cells	59 (98,3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1,7%)
Lentiviral transduced HEK293 cells	55 (91,7%)	0 (0%)	0 (0%)	2 (3,3%)	2 (3,3%)	1 (1,7%)
Lentiviral transduced CHO cells	57 (95%)	0 (0%)	0 (0%)	0 (0%)	2 (3,3%)	1 (1,7%)
Insect cells - BEVs	22 (36,7%)	0 (0%)	1 (1,7%)	22 (36,7%)	14 (23,3%)	1 (1,7%)
Insect cells - transient gene expression	52 (86,7%)	1 (1,7%)	2 (3,3%)	4 (6,7%)	1 (1,7%)	0 (0%)
Plants - transient gene expression	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Plants - stable gene integration	59 (98,3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1,7%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cell-free expression (*in vitro*)	52 (86,7%)	8 (13,3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Conclusions:

- *E. coli*, *Bacillus subtilis* and *in vitro* cell-free expression are the fastest systems
- BacMam and the generation of stable cell lines (mammalian and plants) are the systems that take the most time to go from expression vector to biomass

Section 4 – Intracellular protein production capacity

1. Please estimate the number of target proteins you produce intracellularly (targeted to cytoplasm) per year.

__**range 0 - 1200 (∅ 87)**_____ (enter number) (mandatory to answer)

2. Based on **your practical experiences**, please rank the average range of INTRACELLULAR protein expression for a particular expression system (in mg of protein per liter of culture).

(mandatory to check one box each line)

Expression system	not used	< 1 mg/L	1-5 mg/L	5-20 mg/L	20-100 mg/L	>100 mg/L
E. coli	4 (6,7%)	1 (1,7%)	11 (18,3%)	27 (45%)	15 (25%)	2 (3,3%)
Bacillus subtilis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	51 (85%)	0 (0%)	3 (5%)	3 (5%)	3 (5%)	0 (0%)
Saccharomyces cerevisiae	54 (90%)	1 (1,7%)	2 (3,3%)	2 (3,3%)	1 (1,7%)	0 (0%)
HEK293 - transient gene expression (transfection agent-based)	29 (48,3%)	10 (16,7%)	17 (28,3%)	3 (5%)	1 (1,7%)	0 (0%)
HEK293 - stable gene integration	43 (71,7%)	6 (10%)	6 (10%)	4 (6,7%)	1 (1,7%)	0 (0%)
CHO - transient gene expression (transfection agent-based)	52 (86,7%)	1 (1,7%)	6 (10%)	1 (1,7%)	0 (0%)	0 (0%)
CHO - stable gene integration	54 (90%)	0 (0%)	3 (5%)	2 (3,3%)	1 (1,7%)	0 (0%)
BacMam transduced HEK293 cells	54 (90%)	3 (5%)	3 (5%)	0 (0%)	0 (0%)	0 (0%)
BacMam transduced CHO cells	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced HEK293 cells	58 (96,7%)	1 (1,7%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced CHO cells	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Insect cells - BEVs	28 (46,7%)	1 (1,7%)	21 (35%)	8 (13,3%)	2 (3,3%)	0 (0%)
Insect cells - transient gene expression	54 (90%)	2 (3,3%)	4 (6,7%)	0 (0%)	0 (0%)	0 (0%)
Plants - transient gene expression	59 (98,3%)	0 (0%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)
Plants - stable gene integration	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
cell-free expression (*in vitro*)	52 (86,7%)	6 (10%)	1 (1,7%)	0 (0%)	0 (0%)	1 (1,7%)

Conclusions:

- Ranking according to usage: for intracellular protein production, *E. coli* is by far the most frequently applied system, followed by insect-BEVs and HEK293-transient.
- Ranking according to protein yield: for intracellular protein production, the best yields can be obtained by using *E. coli*, followed by yeast. Insect and mammalian cells also provide decent yields (1-5 mg/L).

Section 5 – Protein secretion capacity

1. Please estimate the number of target proteins you produce by secretion per year.
____**range 0 - 300 (∅ 27)**_____ (enter number) (mandatory to answer)

2. Based on **your practical experiences**, please rank the average range of SECRETED protein expression (in mg of protein per liter of culture) for a particular expression system (using standard laboratory strains).

(mandatory to check one box each line)

Expression system	not used	< 1 mg/L	1-5 mg/L	5-20 mg/L	20-100 mg/L	>100 mg/L
E. coli (secretion to periplasm)	33 (55%)	5 (8,3%)	13 (21,7%)	8 (13,3%)	0 (0%)	1 (1,7%)
E. coli (secretion to media)	51 (85%)	2 (3,3%)	4 (6,7%)	2 (3,3%)	0 (0%)	1 (1,7%)
Bacillus subtilis	58 (96,7%)	0 (0%)	0 (0%)	1 (1,7%)	0 (0%)	1 (1,7%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	47 (78,3%)	0 (0%)	3 (5%)	5 (8,3%)	3 (5%)	2 (3,3%)
Saccharomyces cerevisiae	54 (90%)	1 (1,7%)	3 (5%)	1 (1,7%)	1 (1,7%)	0 (0%)
HEK293 - transient gene expression (transfection agent-based)	22 (36,7%)	4 (6,7%)	16 (26,7%)	13 (21,7%)	5 (8,3%)	0 (0%)
HEK293 - stable gene integration	42 (70%)	1 (1,7%)	6 (10%)	7 (11,7%)	4 (6,7%)	0 (0%)
CHO - transient gene expression (transfection agent-based)	46 (76,7%)	2 (3,3%)	3 (5%)	5 (8,3%)	4 (6,7%)	0 (0%)
CHO - stable gene integration	49 (81,7%)	0 (0%)	3 (5%)	5 (8,3%)	3 (5%)	0 (0%)
BacMam transduced HEK293 cells	55 (91,7%)	3 (5%)	1 (1,7%)	1 (1,7%)	0 (0%)	0 (0%)
BacMam transduced CHO cells	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced HEK293 cells	57 (95%)	2 (3,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)
Lentiviral transduced CHO cells	59 (98,3%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Insect cells - BEVs	35 (58,3%)	3 (5%)	10 (16,7%)	12 (20%)	0 (0%)	0 (0%)
Insect cells - transient gene expression	55 (91,7%)	1 (1,7%)	3 (5%)	1 (1,7%)	0 (0%)	0 (0%)
Plants - transient gene expression	59 (98,3%)	0 (0%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)
Plants - stable gene integration	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Leishmania tarentolae (LEKSY)	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Conclusions:

- Ranking according to usage: for secreted proteins, transient gene expression in HEK293 cells is the most frequently applied system, followed by periplasmic expression in *E. coli* and insect-BEVs.
- Ranking according to protein yield: for secreted proteins, the best yields (5-20 mg/L) can be obtained by using insect-BEVs, stable HEK293 cell lines, transient gene expression in CHO cells and *Pichia pastoris*. Transient gene expression in HEK293 cells and periplasmic expression in *E. coli* also provide decent yields (1-5 mg/L) and are also recommended due to their ease of use and speed.

Section 6 – Membrane protein production capacity

1. Please estimate the number of INTEGRAL membrane proteins (targeted to the membrane) you produce per year.
___**range 0 - 20 (∅ 2)**_____ (enter number) (mandatory to answer)

2. Based on **your practical experiences**, please rank the average range of INTEGRAL membrane protein expression for a particular expression system (in mg of protein per liter of culture).

(mandatory to check one box each line)

Expression system	not used	< 1 mg/L	1-5 mg/L	5-10 mg/L	20-100 mg/L	>10 mg/L
E. coli	43 (71,7%)	5 (8,3%)	10 (16,7%)	2 (3,3%)	0 (0%)	0 (0%)
Bacillus subtilis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	57 (95%)	0 (0%)	1 (1,7%)	1 (1,7%)	1 (1,7%)	0 (0%)
Saccharomyces cerevisiae	56 (93,3%)	1 (1,7%)	3 (5%)	0 (0%)	0 (0%)	0 (0%)
HEK293 - transient gene expression (transfection agent-based)	43 (71,7%)	9 (15%)	8 (13,3%)	0 (0%)	0 (0%)	0 (0%)
HEK293 - stable gene integration	54 (90%)	3 (5%)	3 (5%)	0 (0%)	0 (0%)	0 (0%)
CHO - transient gene expression (transfection agent-based)	58 (96,7%)	0 (0%)	2 (3,3%)	0 (0%)	0 (0%)	0 (0%)
CHO - stable gene integration	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
BacMam transduced HEK293 cells	58 (96,7%)	0 (0%)	2 (3,3%)	0 (0%)	0 (0%)	0 (0%)
BacMam transduced CHO cells	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced HEK293 cells	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Lentiviral transduced CHO cells	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Insect cells - BEVs	49 (81,7%)	4 (6,7%)	7 (11,7%)	0 (0%)	0 (0%)	0 (0%)
Insect cells - transient gene expression	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cell-free expression (*in vitro*)	58	1	1	0	0	0

	(96,7%)	(1,7%)	(1,7%)	(0%)	(0%)	(0%)

Conclusions:

For the production of integral membrane proteins, *E. coli* and transient gene expression in HEK293 cells are the most frequently applied systems, followed by insect-BEVs, resulting in decent yields of 1-5 mg/L. Yeast and stable mammalian cell lines are less frequently applied, but yield the same amount of protein. Note that especially *Pichia pastoris* seems to be a good alternative choice in case a larger amount of protein is required.

Section 7 – Ability for correct folding and assembly of proteins – Size dependency

Based on **your practical experiences**, please rank the ability for a particular host organism to produce functional and correctly folded **single-chain-multidomain proteins and/or multisubunit protein complexes** depending on their respective maximum total size.

(mandatory to check one box each line)

Expression system	not used	< 50 kDa	50-100 kDa	100-250 kDa	250-500 kDa	> 500 kDa
E. coli	4 (6,7%)	10 (16,7%)	30 (50%)	13 (21,7%)	3 (5%)	0 (0%)
Bacillus subtilis	58 (96,7%)	0 (0%)	2 (3,3%)	0 (0%)	0 (0%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	47 (78,3%)	0 (0%)	10 (16,7%)	2 (3,3%)	0 (0%)	1 (1,7%)
Saccharomyces cerevisiae	53 (88,3%)	1 (1,7%)	2 (3,3%)	2 (3,3%)	2 (3,3%)	0 (0%)
HEK293 cells	16 (26,7%)	1 (1,7%)	9 (15%)	18 (30%)	13 (21,7%)	3 (5%)
CHO cells	40 (66,7%)	0 (0%)	3 (5%)	10 (16,7%)	5 (8,3%)	2 (3,3%)
Insect cells	22 (36,7%)	0 (0%)	7 (11,7%)	13 (21,7%)	8 (13,3%)	**10 (16,7%)**
Plants	59 (98,3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1,7%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cell-free expression (*in vitro*)	52 (86,7%)	1 (1,7%)	4 (6,7%)	2 (3,3%)	0 (0%)	1 (1,7%)

Conclusions:

Generally, *E. coli* and yeast systems can be used to produce proteins up to 100 kDa in size, although occasionally larger proteins up to 250 kDa are successfully produced as well. Mammalian and insect cells are more suitable for the production of larger proteins/complexes. Notably, insect cells seem to be the preferred system for very large proteins/complexes (>500 kDa).

Section 8 – Ability for correct folding and assembly of proteins – Disulfide-bond dependency

Based on **your practical experiences**, please rank the ability for a particular host organism to produce functional and correctly folded (secreted) proteins depending on their respective number of disulfide bonds.

(mandatory to check one box each line)

Expression system	not used	1 disulfide bond	2 disulfide bonds	3-4 disulfide bonds	5-10 disulfide bonds	>10 disulfide bonds
E. coli	18 (30%)	8 (13,3%)	19 (31,7)	12 (20%)	3 (5%)	0 (0%)
Bacillus subtilis	58 (96,7%)	0 (0%)	1 (1,7%)	**0** (0%)	1 (1,7%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	49 (81,7%)	0 (0%)	0 (0%)	6 (10%)	4 (6,7%)	1 (1,7%)
Saccharomyces cerevisiae	57 (95%)	0 (0%)	0 (0%)	2 (3,3%)	0 (0%)	1 (1,7%)
HEK293 cells	23 (38,3%)	1 (1,7%)	3 (5%)	12 (20%)	**9** (15%)	12 (20%)
CHO cells	41 (68,3%)	0 (0%)	0 (0%)	7 (11,7%)	8 (13,3%)	4 (6,7%)
Insect cells	30 (50%)	0 (0%)	6 (10%)	8 (13,3%)	10 (16,7%)	6 (10%)
Plants	59 (98,3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1,7%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cell-free expression (*in vitro*)	57 (95%)	0 (0%)	2 (3,3%)	1 (1,7%)	0 (0%)	0 (0%)

Conclusions:

E. coli is mostly used for proteins that contain up to 2 disulfide bonds, although proteins with a higher amount of disulfide bonds have been produced successfully in *E. coli* as well. Eukaryotic systems are clearly the preferred choice for proteins with a higher amount of disulfide bonds, with mammalian and insect cells being especially suitable for proteins with a high disulfide content.

Section 9 – Glycosylation properties

Please estimate for how many target proteins produced per year (in %) a particular glycosylation pattern is required.

(mandatory to answer)

Glycosylation pattern	not used	unknown	<20%	20-50%	50-75%	≥75%
Mannose-type (yeast)	47 (78,3%)	8 (13,3%)	2 (3,3%)	2 (3,3%)	1 (1,7%)	0 (0%)
Paucimannose-type (insect cell)	31 (51,7%)	13 (21,7%)	11 (18,3%)	2 (3,3%)	1 (1,7%)	2 (3,3%)
Complex glycosylation (CHO cells)	39 (65%)	8 (13,3%)	7 (11,7%)	3 (5%)	2 (3,3%)	1 (1,7%)
Complex human glycosylation (HEK293 cells)	21 (35%)	12 (20%)	12 (20%)	7 (11,7%)	5 (8,3%)	3 (5%)

Conclusions:

In the majority of the cases, the glycosylation pattern does not seem to be the determining factor to choose a specific protein expression host.

Section 10 – Running costs

In this section you will be asked to rank the running costs (Euro pricing for 1 liter production scale) for the various expression systems. Please consider costs for **consumables only** (media, transfection agent, disposable flasks, plasmid preparation, cell maintenance, virus production, cell counting, etc.).

(mandatory to check one box each line)

Expression system	not used	< 50 €/L	50-100 €/L	100-500 €/L	500-1000 €/L	> 1000 €/L
E. coli	2 (3,33%)	46 (76,7%)	7 (11,7%)	4 (6,7%)	1 (1,7%)	0 (0%)
Bacillus subtilis	58 (96,7%)	1 (1,7%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Lactococcus lactis	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pichia pastoris	46 (76,7%)	6 (10%)	5 (8,3%)	2 (3,3%)	1 (1,7%)	0 (0%)
Saccharomyces cerevisiae	51 (85%)	5 (8,3%)	3 (5%)	1 (1,7%)	0 (0%)	0 (0%)
HEK293 - transient gene expression (transfection agent-based)	16 (26,7%)	0 (0%)	8 (13,3%)	25 (41,7%)	4 (6,7%)	7 (11,7%)
HEK293 - stable gene integration	34 (56,7%)	2 (3,3%)	6 (10%)	13 (21,7%)	4 (6,7%)	1 (1,7%)
CHO - transient gene expression (transfection agent-based)	42 (70%)	0 (0%)	1 (1,7%)	11 (18,3%)	2 (3,3%)	4 (6,7%)
CHO - stable gene integration	47 (78,3%)	1 (1,7%)	3 (5%)	**4** (6,7%)	4 (6,7%)	1 (1,7%)
BacMam-transduced HEK293 cells	53 (88,3%)	0 (0%)	1 (1,7%)	5 (8,3%)	0 (0%)	1 (1,7%)
BacMam-transduced CHO cells	59 (98,3%)	0 (0%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)
Lentiviral transduced HEK293 cells	56 (93,3%)	0 (0%)	0 (0%)	4 (6,7%)	0 (0%)	0 (0%)
Lentiviral transduced CHO cells	58 (96,7%)	0 (0%)	0 (0%)	2 (3,3%)	0 (0%)	0 (0%)
Insect cells - BEVS	26 (43,3%)	2 (3,3%)	14 (23,3%)	16 (26,7%)	1 (1,7%)	1 (1,7%)
Insect cells - transient gene expression	54 (90%)	0 (0%)	4 (6,7%)	2 (3,3%)	0 (0%)	0 (0%)
Plants - transient gene expression	59 (98,3%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Plants - stable gene integration	59 (98,3%)	0 (0%)	1 (1,7%)	0 (0%)	0 (0%)	0 (0%)
Algae	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Leishmania tarentolae (LEKSY)	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Filamentous fungi	60 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cell-free expression (*in vitro*)	52 (86,7%)	0 (0%)	1 (1,7%)	1 (1,7%)	2 (3,3%)	4 (6,7%)

Conclusions:
Amongst the most frequently used systems, *E. coli* and yeast are clearly the most affordable ones (< 50 €/L). Next in ranking are insect cells (50-100 €/L), followed by the mammalian expression systems (100-500 €/L).

Section 11 – Please let us know your opinion

Based on **your practical experiences**, please rank the level of agreement with the following statements.
(1 = totally agree, 4 = totally disagree)

(mandatory to check one box each line)

Statement	not sure	1 totally agree	2	3	4 totally disagree
I would use a bacterial production host to produce a prokaryotic target protein.	0 (0%)	50 (83,3%)	7 (11,7%)	3 (5%)	0 (0%)
I would use a eukaryotic production host to produce a eukaryotic target protein.	2 (3,3%)	17 (28,3%)	24 (40%)	12 (20%)	5 (8,3%)
Regardless of the nature of an intracellular, single-chain target protein to be produced (prokaryotic, eukaryotic), I always would try *E. coli* as expression system first, unless PTMs (e.g. glycosylation) are known to be required for the planned downstream application or functional activity.	1 (1,7%)	24 (40%)	20 (33,3%)	4 (6,7%)	11 (18,3%)
If the task is to produce a human protein and native-like glycosylation is required for the downstream application (e.g. antibody generation), I would choose HEK cells as expression host.	8 (13,3%)	31 (51,7%)	11 (18,3%)	7 (11,7%)	3 (5%)
If the planned downstream application requires a larger amount (>5 mg) of an INTRACELLULARLY produced single-chain protein and *E. coli* attempts failed so far, I would rather choose insect than mammalian cells as expression host.	15 (25%)	22 (36,7%)	9 (15%)	6 (10%)	8 (13,3%)
Screening multiple expression constructs is key to success.	3 (5%)	23 (38,3%)	21 (35%)	11 (18,3%)	2 (3,3%)
Screening various expression hosts is key to success.	3 (5%)	16 (26,7%)	28 (46,7%)	8 (13,3%)	5 (8,3%)
I prefer to apply eukaryotic expression hosts to produce protein complexes.	8 (13,3%)	18 (30%)	22 (36,7%)	8 (13,3%)	4 (6,7%)

Supplementary references

1. Burnett, M.J.B., and Burnett, A.C. (2020). Therapeutic recombinant protein production in plants: Challenges and opportunities. Plants People Planet *2*, 121–132. 10.1002/ppp3.10073.

2. Spiegel, H., Stöger, E., Twyman, R.M., and Buyel, J.F. (2018). Current status and perspectives of the molecular farming landscape.

3. Buyel, J.F., Twyman, R.M., and Fischer, R. (2017). Very-large-scale production of antibodies in plants: The biologization of manufacturing. Biotechnology Advances *35*, 458–465. 10.1016/j.biotechadv.2017.03.011.

4. Shoji, Y., Farrance, C.E., Bautista, J., Bi, H., Musiychuk, K., Horsey, A., Park, H., Jaje, J., Green, B.J., Shamloul, M., et al. (2012). A plant-based system for rapid production of influenza vaccine antigens: Plant-based production of influenza vaccines. Influenza and Other Respiratory Viruses *6*, 204–210. 10.1111/j.1750-2659.2011.00295.x.

5. Huebbers, J.W., and Buyel, J.F. (2021). On the verge of the market – Plant factories for the automated and standardized production of biopharmaceuticals. Biotechnology Advances *46*, 107681. 10.1016/j.biotechadv.2020.107681.

6. Rademacher, T., Sack, M., Blessing, D., Fischer, R., Holland, T., and Buyel, J. (2019). Plant cell packs: a scalable platform for recombinant protein production and metabolic engineering. Plant Biotechnol J *17*, 1560–1566. 10.1111/pbi.13081.

7. Gengenbach, B.B., Opdensteinen, P., and Buyel, J.F. (2020). Robot Cookies – Plant Cell Packs as an Automated High-Throughput Screening Platform Based on Transient Expression. Front. Bioeng. Biotechnol. *8*, 393. 10.3389/fbioe.2020.00393.

8. Holland, T., Blessing, D., Hellwig, S., and Sack, M. (2013). The in-line measurement of plant cell biomass using radio frequency impedance spectroscopy as a component of process analytical technology. Biotechnology Journal, n/a-n/a. 10.1002/biot.201300125.

9. Zischewski, J., Sack, M., and Fischer, R. (2016). Overcoming low yields of plant-made antibodies by a protein engineering approach. Biotechnology Journal *11*, 107–116. 10.1002/biot.201500255.

10. Castilho, A., Windwarder, M., Gattinger, P., Mach, L., Strasser, R., Altmann, F., and Steinkellner, H. (2014). Proteolytic and *N* -Glycan Processing of Human *α* 1-Antitrypsin Expressed in *Nicotiana benthamiana*. Plant Physiology *166*, 1839–1851. 10.1104/pp.114.250720.

11. Schillberg, S., and Finnern, R. (2021). Plant molecular farming for the production of valuable proteins – Critical evaluation of achievements and future challenges. Journal of Plant Physiology *258–259*, 153359. 10.1016/j.jplph.2020.153359.

12. Opdensteinen, P., and Buyel, J.F. (2022). Reducing water uptake into BY-2 cells by systematically optimizing the cultivation parameters increases product yields achieved by transient expression in plant cell packs. Biotechnology Journal *17*, 2200134. 10.1002/biot.202200134.

13. Gengenbach, B.B., Keil, L.L., Opdensteinen, P., Müschen, C.R., Melmer, G., Lentzen, H., Bührmann, J., and Buyel, J.F. (2019). Comparison of microbial and transient expression (tobacco

plants and plant-cell packs) for the production and purification of the anticancer mistletoe lectin viscumin. Biotechnology and Bioengineering *116*, 2236–2249. 10.1002/bit.27076.

14. Buyel, J.F., Stöger, E., and Bortesi, L. (2021). Targeted genome editing of plants and plant cells for biomanufacturing. Transgenic Res *30*, 401–426. 10.1007/s11248-021-00236-z.

15. Knödler, M., and Buyel, J.F. (2021). Plant-made immunotoxin building blocks: A roadmap for producing therapeutic antibody-toxin fusions. Biotechnology Advances *47*, 107683. 10.1016/j.biotechadv.2020.107683.

16. Schiermeyer, A. (2020). Optimizing product quality in molecular farming. Current Opinion in Biotechnology *61*, 15–20. 10.1016/j.copbio.2019.08.012.

17. Schoberer, J., and Strasser, R. (2018). Plant glyco-biotechnology. Seminars in Cell & Developmental Biology *80*, 133–141. 10.1016/j.semcdb.2017.07.005.

18. Gengenbach, B.B., Müschen, C.R., and Buyel, J.F. (2018). Expression and purification of human phosphatase and actin regulator 1 (PHACTR1) in plant-based systems. Protein Expression and Purification *151*, 46–55. 10.1016/j.pep.2018.06.003.

19. Opdensteinen, P., Sperl, L.E., Mohamadi, M., Kündgen-Redding, N., Hagn, F., and Buyel, J.F. (2022). The transient expression of recombinant proteins in plant cell packs facilitates stable isotope labelling for NMR spectroscopy. Plant Biotechnology Journal *20*, 1928–1939. 10.1111/pbi.13873.

20. Song, A.A.-L., In, L.L.A., Lim, S.H.E., and Rahim, R.A. (2017). A review on Lactococcus lactis: from food to factory. Microb Cell Fact *16*, 55. 10.1186/s12934-017-0669-x.

21. Neef, J., Van Dijl, J.M., and Buist, G. (2021). Recombinant protein secretion by *Bacillus subtilis* and *Lactococcus lactis* : pathways, applications, and innovation potential. Essays in Biochemistry *65*, 187–195. 10.1042/EBC20200171.

22. Gifre-Renom, L., Cano-Garrido, O., Fàbregas, F., Roca-Pinilla, R., Seras-Franzoso, J., Ferrer-Miralles, N., Villaverde, A., Bach, À., Devant, M., Arís, A., et al. (2018). A new approach to obtain pure and active proteins from Lactococcus lactis protein aggregates. Sci Rep *8*, 13917. 10.1038/s41598-018-32213-8.

23. Frelet-Barrand, A. (2022). Lactococcus lactis, an Attractive Cell Factory for the Expression of Functional Membrane Proteins. Biomolecules *12*, 180. 10.3390/biom12020180.

24. Singh, S.K., Tiendrebeogo, R.W., Chourasia, B.K., Kana, I.H., Singh, S., and Theisen, M. (2018). Lactococcus lactis provides an efficient platform for production of disulfide-rich recombinant proteins from Plasmodium falciparum. Microb Cell Fact *17*, 55. 10.1186/s12934-018-0902-2.

25. Krüger, A., Welsch, N., Dürwald, A., Brundiek, H., Wardenga, R., Piascheck, H., Mengers, H.G., Krabbe, J., Beyer, S., Kabisch, J.F., et al. (2022). A host-vector toolbox for improved secretory protein overproduction in Bacillus subtilis. Appl Microbiol Biotechnol *106*, 5137–5151. 10.1007/s00253-022-12062-2.

26. Popp, P.F., Dotzler, M., Radeck, J., Bartels, J., and Mascher, T. (2017). The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with Bacillus subtilis. Sci Rep *7*, 15058. 10.1038/s41598-017-15107-z.

27. Su, Y., Liu, C., Fang, H., and Zhang, D. (2020). Bacillus subtilis: a universal cell factory for industry, agriculture, biomaterials and medicine. Microb Cell Fact *19*, 173. 10.1186/s12934-020-01436-8.

28. Nguyen, H.D., and Phan, T.T.P. (2022). A Protocol to Enhance Soluble Protein Expression in the Cytoplasm of Bacillus subtilis. In Insoluble Proteins Methods in Molecular Biology., E. Garcia Fruitós and A. Arís Giralt, eds. (Springer US), pp. 233–243. 10.1007/978-1-0716-1859-2_14.

29. Falkenberg, K.B., Mol, V., De La Maza Larrea, A.S., Pogrebnyakov, I., Nørholm, M.H.H., Nielsen, A.T., and Jensen, S.I. (2021). The ProUSER2.0 Toolbox: Genetic Parts and Highly Customizable Plasmids for Synthetic Biology in *Bacillus subtilis*. ACS Synth. Biol. *10*, 3278–3289. 10.1021/acssynbio.1c00130.

30. Yang, H., Qu, J., Zou, W., Shen, W., and Chen, X. (2021). An overview and future prospects of recombinant protein production in Bacillus subtilis. Appl Microbiol Biotechnol *105*, 6607–6626. 10.1007/s00253-021-11533-2.

31. Hoff, J., Daniel, B., Stukenberg, D., Thuronyi, B.W., Waldminghaus, T., and Fritz, G. (2020). *Vibrio natriegens* : an ultrafast-growing marine bacterium as emerging synthetic biology chassis. Environ Microbiol *22*, 4394–4408. 10.1111/1462-2920.15128.

32. Weinstock, M.T., Hesek, E.D., Wilson, C.M., and Gibson, D.G. (2016). Vibrio natriegens as a fast-growing host for molecular biology. Nat Methods *13*, 849–851. 10.1038/nmeth.3970.

33. Becker, W., Wimberger, F., and Zangger, K. (2019). *Vibrio natriegens* : An Alternative Expression System for the High-Yield Production of Isotopically Labeled Proteins. Biochemistry *58*, 2799–2803. 10.1021/acs.biochem.9b00403.

34. Tschirhart, T., Shukla, V., Kelly, E.E., Schultzhaus, Z., NewRingeisen, E., Erickson, J.S., Wang, Z., Garcia, W., Curl, E., Egbert, R.G., et al. (2019). Synthetic Biology Tools for the Fast-Growing Marine Bacterium *Vibrio natriegens*. ACS Synth. Biol. *8*, 2069–2079. 10.1021/acssynbio.9b00176.

35. Xu, J., Dong, F., Wu, M., Tao, R., Yang, J., Wu, M., Jiang, Y., Yang, S., and Yang, L. (2021). Vibrio natriegens as a pET-Compatible Expression Host Complementary to Escherichia coli. Front. Microbiol. *12*, 627181. 10.3389/fmicb.2021.627181.

36. Gauttam, R., Eng, T., Zhao, Z., Ul Ain Rana, Q., Simmons, B.A., Yoshikuni, Y., Mukhopadhyay, A., and Singer, S.W. (2023). Development of genetic tools for heterologous protein expression in a pentose-utilizing environmental isolate of *Pseudomonas putida*. Microbial Biotechnology *16*, 645–661. 10.1111/1751-7915.14205.

37. Liang, T., Sun, J., Ju, S., Su, S., Yang, L., and Wu, J. (2021). Construction of T7-Like Expression System in Pseudomonas putida KT2440 to Enhance the Heterologous Expression Level. Front. Chem. *9*, 664967. 10.3389/fchem.2021.664967.

38. Gauttam, R., Mukhopadhyay, A., and Singer, S.W. (2020). Construction of a novel dual-inducible duet-expression system for gene (over)expression in Pseudomonas putida. Plasmid *110*, 102514. 10.1016/j.plasmid.2020.102514.

39. Nikel, P.I., and De Lorenzo, V. (2018). Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. Metabolic Engineering *50*, 142–155. 10.1016/j.ymben.2018.05.005.

40. Martínez-García, E., Nikel, P.I., Aparicio, T., and De Lorenzo, V. (2014). Pseudomonas 2.0: genetic upgrading of P. putida KT2440 as an enhanced host for heterologous gene expression. Microb Cell Fact *13*, 159. 10.1186/s12934-014-0159-3.

41. Cook, T.B., Rand, J.M., Nurani, W., Courtney, D.K., Liu, S.A., and Pfleger, B.F. (2018). Genetic tools for reliable gene expression and recombineering in *Pseudomonas putida*. Journal of Industrial Microbiology and Biotechnology *45*, 517–527. 10.1007/s10295-017-2001-5.

42. Bashiri, G., and Baker, E.N. (2015). Production of recombinant proteins in *Mycobacterium smegmatis* for structural and functional studies: *Mycobacterium smegmatis* Expression System. Protein Science *24*, 1–10. 10.1002/pro.2584.

43. Coker, J.A., Katis, V.L., Fairhead, M., Schwenzer, A., Clemmensen, S.B., Frandsen, B.U., De Jongh, W.A., Gileadi, O., Burgess-Brown, N.A., Marsden, B.D., et al. (2022). FAS2FURIOUS: Moderate-Throughput Secreted Expression of Difficult Recombinant Proteins in Drosophila S2 Cells. Front. Bioeng. Biotechnol. *10*, 871933. 10.3389/fbioe.2022.871933.

44. Moraes, Â.M., Jorge, S.A.C., Astray, R.M., Suazo, C.A.T., Calderón Riquelme, C.E., Augusto, E.F.P., Tonso, A., Pamboukian, M.M., Piccoli, R.A.M., Barral, M.F., et al. (2012). Drosophila melanogaster S2 cells for expression of heterologous genes: From gene cloning to bioprocess development. Biotechnology Advances *30*, 613–628. 10.1016/j.biotechadv.2011.10.009.

45. Yamashita, A., Nango, E., and Ashikawa, Y. (2017). A large-scale expression strategy for multimeric extracellular protein complexes using Drosophila S2 cells and its application to the recombinant expression of heterodimeric ligand-binding domains of taste receptor: Protein Complex Expression by S2 Cells. Protein Science *26*, 2291–2301. 10.1002/pro.3271.

46. Brillet, K., Pereira, C.A., and Wagner, R. (2010). Expression of Membrane Proteins in Drosophila Melanogaster S2 Cells: Production and Analysis of a EGFP-Fused G Protein-Coupled Receptor as a Model. In Heterologous Expression of Membrane Proteins Methods in Molecular Biology., I. Mus-Veteau, ed. (Humana Press), pp. 119–133. 10.1007/978-1-60761-344-2_8.

47. Schetz, J.A., and Shankar, E.P.N. (2004). Protein Expression in the *Drosophila* Schneider 2 Cell System. CP Neuroscience *27*. 10.1002/0471142301.ns0416s27.

48. Rosales-Mendoza, S., Paz-Maldonado, L.M.T., and Soria-Guerra, R.E. (2012). Chlamydomonas reinhardtii as a viable platform for the production of recombinant proteins: current status and perspectives. Plant Cell Rep *31*, 479–494. 10.1007/s00299-011-1186-8.

49. Schroda, M. (2019). Good News for Nuclear Transgene Expression in Chlamydomonas. Cells *8*, 1534. 10.3390/cells8121534.

50. Rasala, B.A., Chao, S.-S., Pier, M., Barrera, D.J., and Mayfield, S.P. (2014). Enhanced Genetic Tools for Engineering Multigene Traits into Green Algae. PLoS ONE *9*, e94028. 10.1371/journal.pone.0094028.

51. Barrera, D., Gimpel, J., and Mayfield, S. (2014). Rapid Screening for the Robust Expression of Recombinant Proteins in Algal Plastids. In Chloroplast Biotechnology Methods in Molecular Biology., P. Maliga, ed. (Humana Press), pp. 391–399. 10.1007/978-1-62703-995-6_26.

52. Cutolo, E.A., Mandalà, G., Dall'Osto, L., and Bassi, R. (2022). Harnessing the Algal Chloroplast for Heterologous Protein Production. Microorganisms *10*, 743. 10.3390/microorganisms10040743.

53. Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Research *35*, e45–e45. 10.1093/nar/gkm047.

54. Scholz, J., Besir, H., Strasser, C., and Suppmann, S. (2013). A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. BMC Biotechnol *13*, 12. 10.1186/1472-6750-13-12.

55. Dammeyer, T., Timmis, K.N., and Tinnefeld, P. (2013). Broad host range vectors for expression of proteins with (Twin-) Strep-tag, His-tag and engineered, export optimized yellow fluorescent protein. Microb Cell Fact *12*, 49. 10.1186/1475-2859-12-49.

56. Ki, M.-R., and Pack, S.P. (2020). Fusion tags to enhance heterologous protein expression. Appl Microbiol Biotechnol *104*, 2411–2425. 10.1007/s00253-020-10402-8.

57. Waugh, D.S. (2011). An overview of enzymatic reagents for the removal of affinity tags. Protein Expression and Purification *80*, 283–293. 10.1016/j.pep.2011.08.005.

58. Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology *189*, 113–130. 10.1016/0022-2836(86)90385-2.

59. Li, Z., Michael, I.P., Zhou, D., Nagy, A., and Rini, J.M. (2013). Simple *piggyBac* transposon-based mammalian cell expression system for inducible protein production. Proc. Natl. Acad. Sci. U.S.A. *110*, 5004–5009. 10.1073/pnas.1218620110.

60. Suppmann, S. (2021). Inducible protein expression in piggyBac transposase mediated stable HEK293 cell pools. In Methods in Enzymology (Elsevier), pp. 321–339. 10.1016/bs.mie.2021.06.016.

61. Becker, P., Bosschaerts, M., Chaerle, P., Daniel, H.-M., Hellemans, A., Olbrechts, A., Rigouts, L., Wilmotte, A., and Hendrickx, M. (2019). Public Microbial Resource Centers: Key Hubs for Findable, Accessible, Interoperable, and Reusable (FAIR) Microorganisms and Genetic Materials. Appl Environ Microbiol *85*, e01444-19. 10.1128/AEM.01444-19.

62. Furman, J.L., and Stern, S. (2011). Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. American Economic Review *101*, 1933–1963. 10.1257/aer.101.5.1933.