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.