Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Technologies to keep an eye on: alternative hosts for protein production in structural biology

Francisco J Fernández and M Cristina Vega

The conduct of many trials for the successful production of large quantities of pure proteins for structural biology and biotechnology applications, particularly of active, authentically processed enzymes, large eukaryotic multi-subunit complexes and membrane proteins, has spurred the development of recombinant expression systems. Beyond the well-established *Escherichia coli*, mammalian cell culture and baculovirus-infected insect cell expression systems, a plethora of alternative expression systems has been discovered, engineered, matured and deployed, resulting in crystal, nuclear magnetic resonance, and electron microscopy structures. In this review, we visit alternative expression hosts for structural biology ranging from bacteria and archaea to filamentous and unicellular yeasts and protozoa, with particular emphasis on their applicability to the structural determination of high-value, challenging proteins and complexes.

Address
Center for Biological Research, Spanish National Research Council (CIB-CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain

Corresponding author: Vega, M Cristina (cvega@cib.csic.es)

Current Opinion in Structural Biology 2013, **23**:365–373

This review comes from a themed issue on **New constructs and expression of proteins**

Edited by **Imre Berger** and **Lorenz M Mayr**

For a complete overview see the Issue and the Editorial

Available online 5th March 2013

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2013.02.002

Introduction

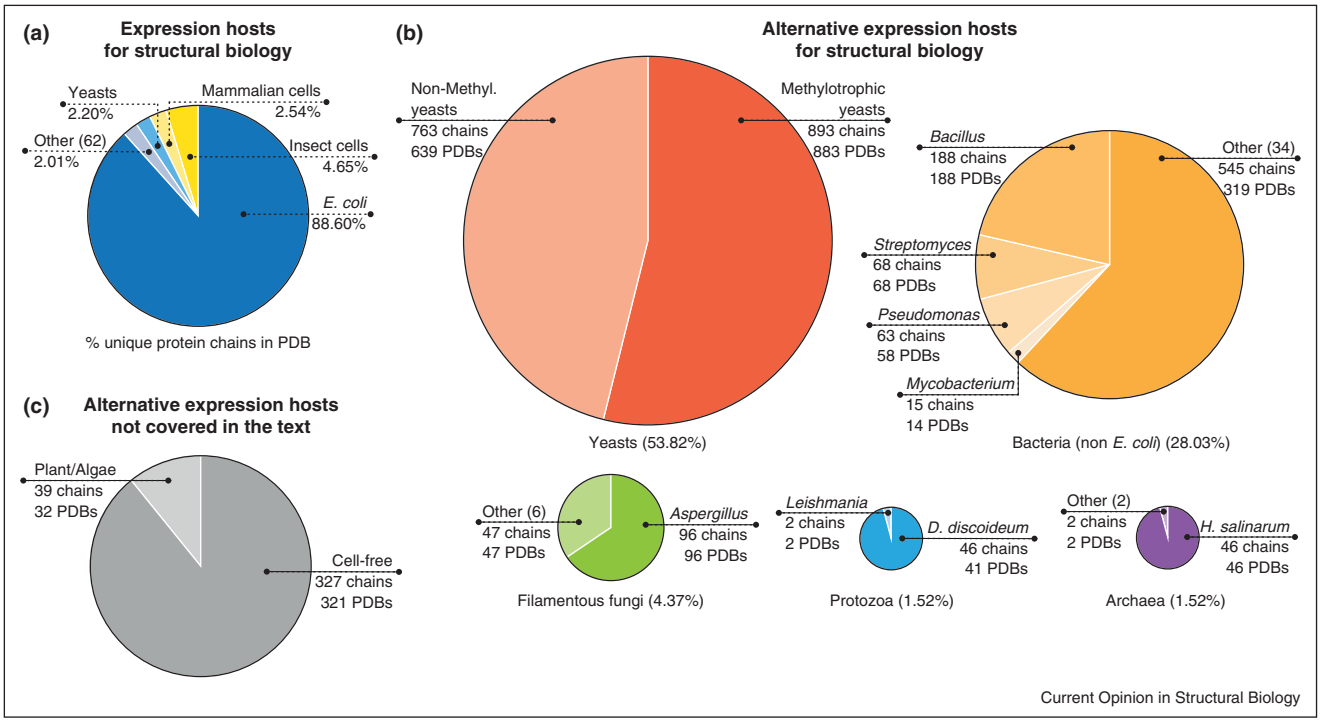
Recombinant protein production stands as an enabling technology in the life sciences and plays a crucial role in biotechnology and structural biology. At present, *Escherichia coli* ranks as the most prevalent microbial factory in structural biology as it has been used to produce more than 66 724 (88.6%) of all distinct protein chains in the Protein Data Bank (PDB) with a non-empty descriptor for 'Expression Host' (Figure 1a). Insect cells (including baculovirus-infected insect cells and *Drosophila* Schneider cells) rank second, accounting for 3499 distinct protein chains (4.65%), while mammalian cells occupy the third position with 1911 distinct protein chains (2.54%) (Figure 1a). The remaining 3102 protein chains have been expressed in 67 different expression hosts spanning the three domains of life with half of these entries recording a yeast expression system for their production (1658 protein chains or 2.20%) (Figure 1a). These dominant expression platforms share widespread use and a wealth of available resources (material and knowledge) and their performance has been demonstrated for many heterologous proteins and multi-subunit complexes (Figures 2 and 3). Hence, the cost associated with their adoption for a new protein production project is generally lower than trying less tested hosts. Nevertheless, research on unconventional expression hosts represents a valuable addition by shedding light on organisms with orthogonal genomic, metabolic, and recombinant expression properties (Table 1) that can outperform more established hosts for specific protein families — as specialty systems. This review aims to summarize state-of-the-art alternative expression hosts including various bacteria, archaea, filamentous fungi, yeasts, and protozoa, all of which have been successfully used in structural biology within the last decade (Table 1 and Figures 1 and 3); inevitably, several hosts will be left out from this selection: cell-free systems, algae and plants, and bacteria taxonomically redundant with those already included (Figure 1b and c). Except when noted otherwise, none of these expression hosts poses a threat to our health. Furthermore, we will touch briefly on what the future may hold for these unconventional hosts.

Alternative bacterial expression hosts

Pseudomonas

Pseudomonas spp. (typically *P. putida*, *P. aeruginosa*, *P. fluorescens* and *P. alcaligenes*) include Gram-negative rods with strict aerobic metabolism [1••] that grow at 30–42°C on inexpensive wastage products like corn molasses and can be transformed with plasmid DNA by chemical methods and by plasmid conjugation from *E. coli* [2]. *P. aeruginosa* is a moderate potential hazard for staff and the environment thereby requiring biosafety level 2 (S2) containment. Its relatively high G + C content (60.5–66.6% among sequenced species) makes *Pseudomonas* an ideal host for the expression of certain restriction endonucleases [3••] (Figure 2a) and naturally G + C rich gene products, for example, HsaD from *M. tuberculosis* [4]. Recombinant expression in *Pseudomonas* may be driven by temperature (42°C) using the temperature-sensitive P_L promoter as well as other promoters such as the IPTG-inducible tac promoter, the alkane hydroxylase gene promoter (alkBp) [5,6], the regulatory control region for naphthalene and phenanthrene degradation from *Comamonas testosteroni* (inducible by sodium salicylate) [7], and

Figure 1



Number of expression hosts used for recombinant protein production for structural biology, expressed either as percentage of unique protein chains in the PDB or as a count of unique chains and PDBs per host. Only PDB entries with a non-empty 'Expression Host' field are considered, amounting to 67 690 PDB entries containing 66 722 unique protein chains in total. Absolute counts and frequencies of expression host usage are depicted as pie charts. When several expression hosts are grouped together as Other, the total number of implied hosts is shown in parentheses. **(a)** *E. coli*, baculovirus-infected insect cells, and mammalian cells represent the dominant expression hosts in the PDB, together accounting for nearly 96% of all unique chains in protein structures. **(b)** Summarized statistics for alternative recombinant expression hosts, including yeasts and 'other' hosts in (a); hosts are grouped according to their taxonomic classification and, for yeasts, refined on the basis of their metabolic capacity to assimilate methanol (methylotrophic and non-methylotrophic yeasts). The area of each pie chart is in proportion to the contribution that each host set makes to the total number of protein chains expressed with alternative hosts; this frequency is shown underneath the pie charts. **(c)** As in (b), for alternative hosts not described in the text (plant/algae and cell-free systems).

the Pm/xylS expression system (Vectron Biosolutions AS) inducible by toluinic acid and other benzoic acids. Proteins expressed with this host include, among others, enzymes, receptors, fimbrial proteins, and heme-containing proteins. So far, 58 PDB entries report *Pseudomonas* as expression host of both homologous and heterologous proteins from other pathogenic bacteria and bacteriophages. Interestingly, non-pathogenic *Pseudomonas* species have been used to express and crystallize own membrane proteins and virulence factors.

Bacillus

B. subtilis and *B. megaterium* are Gram-positive soil bacteria increasingly used as hosts for intra-cellular and especially extracellular protein production that have grown over the past decade into attractive alternatives to *E. coli* because their genomes are fully sequenced [8,9], due to their GRAS (generally regarded as safe) status and the development of an extensive tool-kit of gene expression plasmids and production strains. Both stable episomal and integrative plasmids are available that can be efficiently delivered by protoplast transformation or trans-conjugation. Recombinant expression is typically driven from the xylose-inducible P_{xylA} [10] or from T7 RNA polymerase [11] promoters. Even though *B. subtilis* is currently more widely used in structural biology, the strong push to develop *B. megaterium* as the standard Gram-positive microbial factory for protein production, which includes a commercial version (MoBiTec) and ongoing efforts to characterize it at a systems biology level [12••], may change these numbers in the future. The 198 PDB entries reporting *Bacillus* as expression host include diverse glycoside hydrolases, lipases, proteases, as well as insecticidal peptides, virulence factors, for example, anthrax lethal factor and *Clostridium difficile* toxin A [13] (Figure 2b) and phage DNA polymerases.

Streptomyces

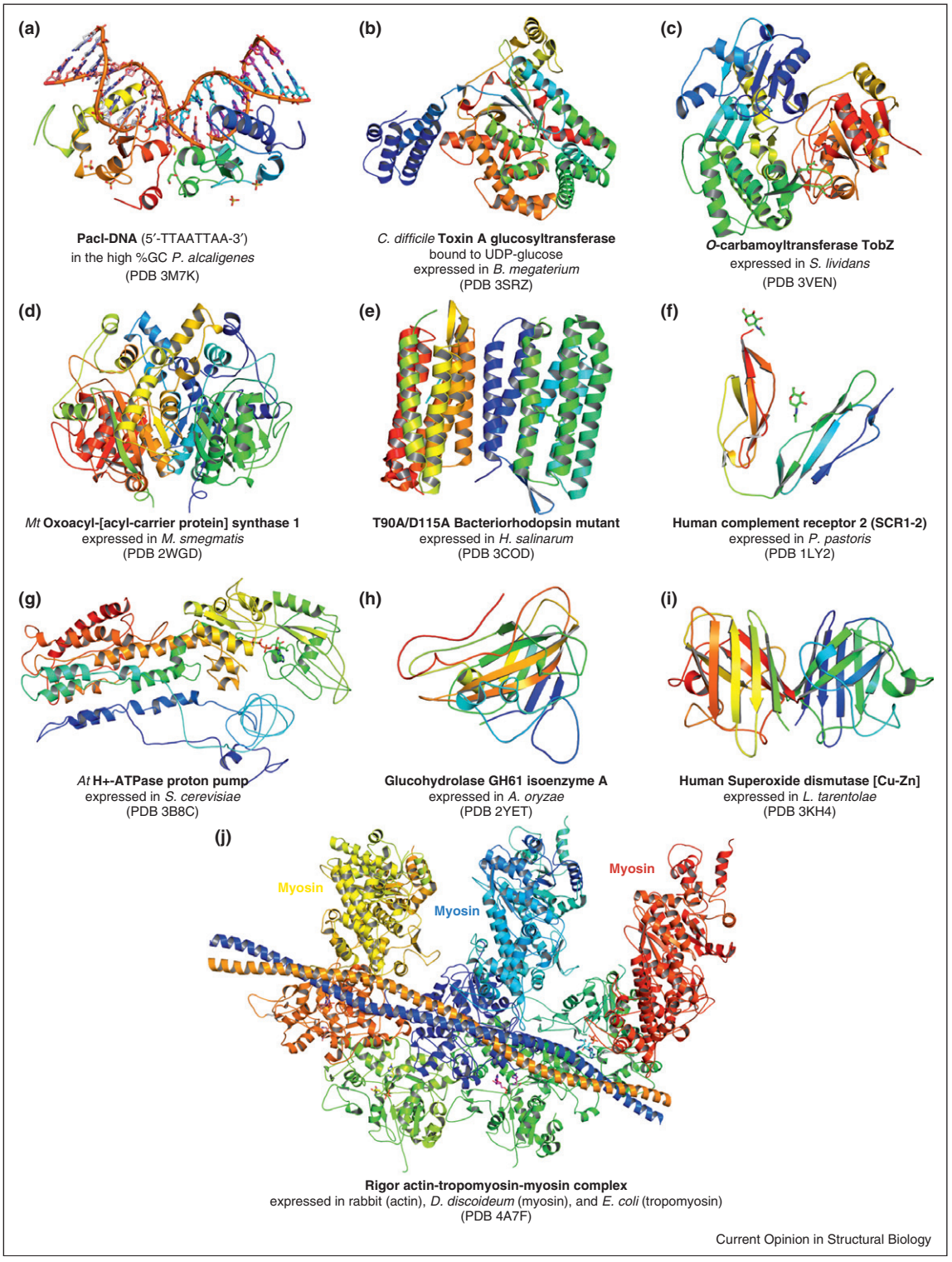
Several *Streptomyces* strains have been used for the production of proteins for 67 PDB entries, mostly *S. lividans* TK24 because of its exceptionally low level of extracellular proteases. These mycelial Actinobacteria are versatile

Table 1

	Expression host[a]		%GC [b]	Gene regulatory sequences (activation)[c]	t_D (h)[d]	Expression level[e]	Vectors[f]	Proteolytic processing[g]	Phosphorylation[h]	Glycosylation[i]	PTMs[j]	Learning curve[k]	Time (wk)[l]	Economic cost[m]	Recommended use[n]
Bacteria (simple, robust, lowest cost)	*Escherichia coli*		50%	*P*T7, *P*tac (IPTG) *P*BAD (arabinose)	0.3–0.5	High (inclusion bodies)	Episomal	No	Rare	No	No	Reference	1–2	Reference ($1000/protein)	Small proteins Single domains Antigens
	Other Enterobacteria	*Pseudomonas* spp.	60–67%	P_L (switch to 42°C) *P*tac (IPTG) *PalkBp* (sodium salicylate) *P*m/xylS (toluinic/benzoic ac.)	1.0–2.0	High	Episomal	No	Possible	No	No	<3 months	2–4	Low-moderate	Specific receptors Fimbrial proteins Heme proteins
	Firmicutes	*Bacillus* spp.	39–43%	*PxylA* (xylose) *P*T7 (IPTG)	0.6–1.0	High (secretion)	Episomal	No	Possible	No	No	<2 month	1–3	Low	Secreted enzymes Virulence factors
	Actinobacteria	*Streptomyces* spp.	72%	Ph. starvation promoter Erythromycin promoter Nitrilase promoter Hybrid T7 promoter	4.0–6.0	High (secretion)	Episomal	No	Possible	No	No	3–6 months	2–4	Low-moderate	Antibiotic peptides Antibiotic biosynthesis and drug-modifying enzymes
	Mycobacteria	*Mycobacterium* spp.	67%	Acetamidase (acetamide)	2.0–3.0	Low-high	Episomal	No	Possible	No	No	<3 months	2–4	Low-moderate	Mycobacterial antigens and cell-wall enzymes
Archaea (low cost, good for some membrane proteins)	Euryarcheota	*Halobacterium halobium*	65%	Strong constitutive promoter	4.0–8.0	High (membrane)	Episomal	Some	Possible	No	No	3–6 months	2–4	Low	Bacteriorhodopsin GPCRs
Eukarya (glycosylation, PTMs)	Unicellular fungi (yeast)	Methylotrophic (*P. pastoris*, *H. polymorpha*)	41%	*P*AOX1 (methanol) *P*FLD (acetamide) *P*GAP (constitutive)	1.5–3.0	Low-high (secretion)	Integrative	Some	Some	High	Some	<6 months	3–6	Low-moderate	Secreted proteins Complement factors Membrane proteins Peroxysome proteins
		Non-methylotrophic (*S. cerevisiae*, *K. lactis*)	36–39%	*P*GAL1/*P*GAL10 (galactose) Acetamidase (acetamide) Glycolytic (constitutive)	1.3–2.0	Low-high	Episomal/ integrative	Some	Some	High	Some	<3 months	2–4	Low	Kinases Nuclear proteins Membrane proteins Antibodies
	Filamentous fungi	*Aspergillus* spp. *Hypocrea/ Trichoderma* spp.	49–51%	IPTG inducible Glycolytic (constitutive) *PglaA* (xylose) *PalcA* (ethanol) Human estrogen (estrogen)	3.0–4.0	Low-high (secretion)	Episomal/ integrative	Some	Some	High	Some	<6 months	2–6	Low-moderate	Secreted enzymes Glycoenzymes
	Protozoa	*Leishmania terantolae*	58%	18S rRNA locus (constitutive) *P*T7/tetO (tetracycline)	6.0–13	Low-high	Integrative	Some	Extensive	Complex	High	<6 months	2–4	Moderate-high	General purpose Abundant PTMs
		Dictyostelium discoideum	23%	*P*Actin-15 (constitutive) *P*Discoidin I (constitutive) *PrnrB* (UV inducible)	4.0–12.0	Low-high	Episomal	Some	Extensive	Core	High	<6 months	2–4	Moderate-high	Glycoproteins Motor proteins
	BVS	Sf9, Sf21, Hi-5	35–41%	p10, polh (viral life cycle)	16–72	Low-high (cytosol)	Bacmid	Yes	Yes	Complex	Yes	6–12 months	3–5	Moderate-high	General purpose
	Mammalian cells	CHO, HEK293	41%	SV40, CMV (constitutive)	14–36	Low-moderate (secretion)	Episomal/ integrative	Yes	Yes	Complex	Yes	<12 months	3–8	High	General purpose

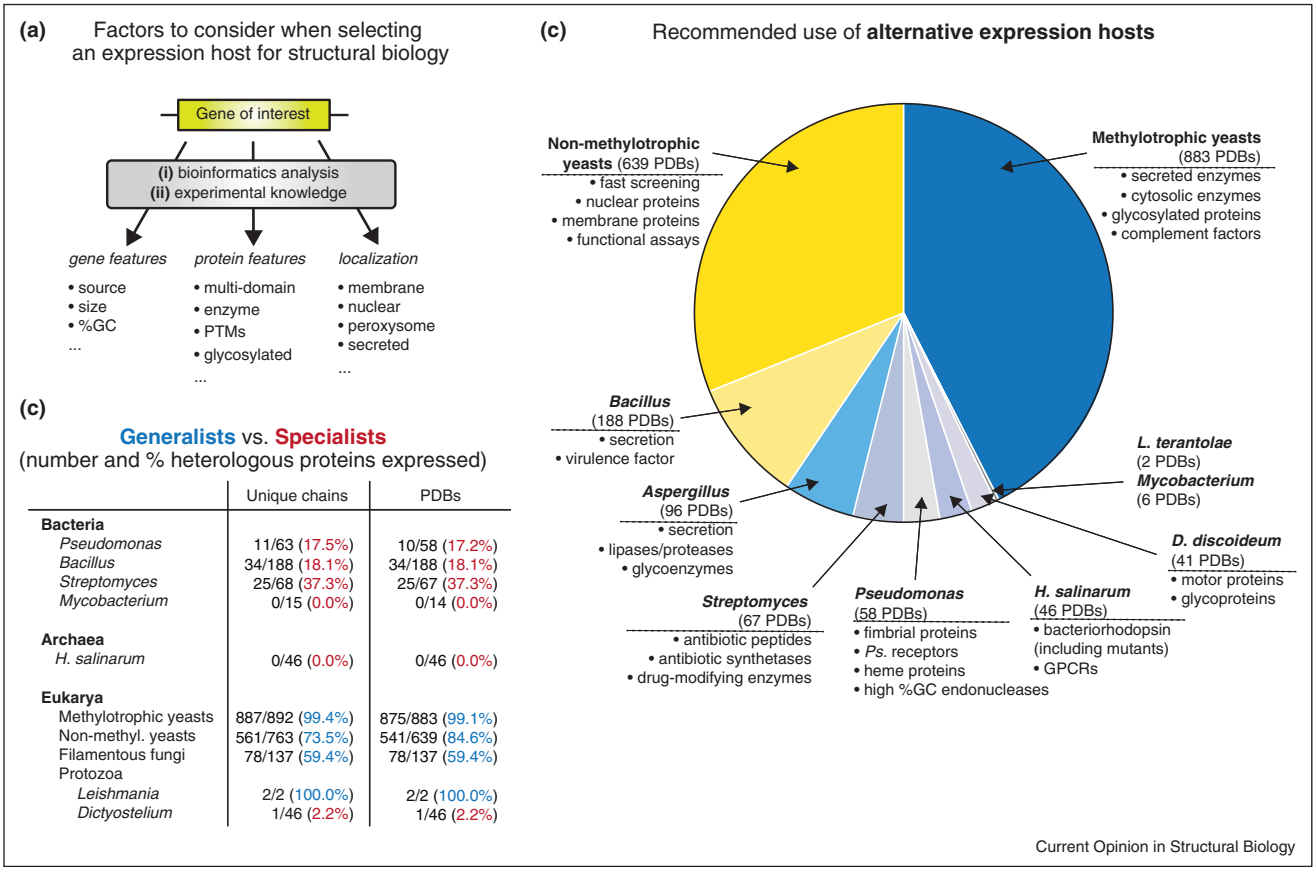
[a] Organisms exploited for recombinant expression of homologous or heterologous proteins (details of each host are expanded on the text), organized according to their taxonomy and, for yeasts, to their capacity to thrive in methanol.
[b] %GC: GC content (percentage) of complete genome. GC content may affect protein expression indirectly via codon bias and availability of tRNA pools.
[c] Selection of the most frequently used promoters and their activation (inducible or constitutive).
[d] t_D: Doubling time (h). Faster dividing expression hosts will generate more biomass per unit time; exact doubling times depend on temperature, media composition, and genetic background.
[e] Qualitative scale aimed at describing the potential for an expression host to achieve high-level expression levels; it is highly dependent on the type of protein being expressed. The rough scale translates into an expected yield in mg per liter culture: low, 0.1–5.0 mg/L; moderate: 5.0–50.0 mg/L; high: >50 mg/L. Further information is shown between parentheses.
[f] Typical vectors used to drive recombinant expression, either episomal (stably maintained extrachromosomally) or targeted for genomic integration (integrative).
[g] Capacity to perform mammalian-style proteolytic processing of the expressed recombinant protein.
[h] Capacity to perform mammalian-style phosphorylation of the expressed recombinant protein.
[i] Capacity to glycosylate recombinant proteins; high: abundant glycosylation of the high mannose type; core: mammalian-style core glycosylation; complex: full or nearly full mammalian-type glycosylation.
[j] Overall capacity for complex PTMs in comparison to mammalian cells.
[k] Learning curve tries to capture the approximate time it would take a scientist with previous exposure only to *E. coli* expression to become proficient (i.e. productive mostly independently) in the other expression system; no expert guidance is assumed since intensive next-door expert training or translational access programs to dedicated expert facilities would certainly accelerate learning these technologies.
[l] Time: time in weeks for an experienced user from start of project to the first large-scale expression culture, hence spanning cloning, transformation and selection, small-scale expression trials, and large-scale expression culture.
[m] Economic cost is a complex variable that depends on any particular target protein, dedicated person-months and length of the overall process (from clone to purified protein), gene synthesis (or not), target yield, fermentation scale, and the combined costs of media, antibiotics, inducers, and other reagents. Hence, the values shown are merely a rough guide that might oscillate (upwards and downwards). Low, comparable to the cost involved in the soluble expression of a single-domain protein in *E. coli* (estimated at $1000 from cloning to expressed protein); moderate, 2–5-fold higher overall cost; high, >5-fold higher cost.
[n] Recommended use: highlighted/recommended applications suggesting potential areas with the greatest promise for further development for each host.

Figure 2



Gallery of crystal structures where proteins were overexpressed using an unconventional platform. **(a)** Restriction endonuclease PacI complexed with its DNA recognition sequence expressed homologously in *Pseudomonas alcaligenes* (3MK7) [3••]. **(b)** Glucosyltransferase domain from *Clostridium difficile* toxin A bound to UDP-glucose expressed in *Bacillus megaterium* (3SRZ) [13]. **(c)** O-carbamoyltransferase TobZ from *Streptoalloteichus tenebrarius* expressed in *Streptomyces lividans* (3VEN) [14]. **(d)** KasA protein (3-oxoacyl-[acyl-carrier-protein] synthase) from *Mycobacterium tuberculosis* expressed in *Mycobacterium smegmatis* (2WGD) [17••]. **(e)** T90A/D115A mutant bacteriorhodopsin expressed homologously in *Halobacterium salinarum* (3COD) [22]. **(f)** Δ73 C-terminally truncated plasma membrane proton pump from *Arabidopsis thaliana* expressed in

Figure 3



Selection of a suitable alternative expression host for a protein expression project. **(a)** Features of the target gene and the encoded protein that bear on the choice of expression host can be collected from bioinformatics analysis and known experimental facts (own research or published). Some of those features may play crucial roles in tipping the balance towards an unconventional host, for example, if the target protein is secreted to the extracellular medium. **(b)** Number and percentage of heterologous proteins expressed with several alternative hosts, expressed as unique chains or as number of PDBs. A host used to express 50% or more of heterologous proteins is said to be a generalist, otherwise it is considered to be a specialist. **(c)** Pie chart showing the same alternative hosts as in (b), with section areas representing frequency of use in the PDB. To aid in choosing from them, guidelines are shown underneath each expression host regarding the recommended use for that host.

hosts for the production of secondary metabolites accounting for half of all commercialized antibiotics and are especially indicated for antibiotic peptides and enzymes involved in antibiotic biosynthesis and drug modification, for example, the O-carbamoyl transferase TobZ [14] (Figure 2c). Several strong promoters are available controlled by phosphate starvation, inducible by erythromycin, and the hyper-inducible nitrilase promoters [15], as well as hybrid promoters employed to drive expression from a T7 promoter to very high levels [16]. Hence, *Streptomyces* strains appear particularly well suited for the recombinant production of antibiotic and drug-modifying enzymes in addition to peptides and small-molecule antibiotics and drugs.

Mycobacterium

Two fully sequenced, non-pathogenic species for human, *M. smegmatis* and *M. vaccae*, have been successfully exploited for the production of *M. tuberculosis* proteins in work spearheaded by the *M. tuberculosis* structural genomics consortium that has resulted in 5 and 1 deposited PDB entries, respectively. Even if those entries represent only a small fraction of the >1000 structures concerning *M. tuberculosis* proteins, they include Mycobacterial antigens,

Saccharomyces cerevisiae (3B8C) [33•]. **(g)** Unliganded human CD21 (complement receptor 2, CR2) SCR domains 1-2 expressed in *Pichia pastoris* (1LY2) [39•]. **(h)** Glucohydrolase 61 (GH61) isozyme A (post-translationally modified by methylation and two disulfide bridges) from *Thermoascus aurantiacus* expressed in *Aspergillus oryzae* (2YET) [50•]. **(i)** Human Cu/Zn superoxide dismutase (SOD1) expressed in *Leishmania tarentolae* (3KH4) [53•]. **(j)** Actin–tropomyosin–myosin rigor complexes expressed in *Dictyostelium discoideum*, rabbit, and *E. coli* (4A7F) [60].

iron-superoxide dismutase, and cell-wall and specific lipid biosynthetic enzymes [17••] (Figure 2d) that may have required elaborate strategies for their expression in *E. coli*. Vectors for recombinant expression in *M. smegmatis* usually drive expression from the strong acetamidase promoter, which is induced by acetamide at mid-log phase and peaks at 24–36 hours [17••,18]. *M. smegmatis* and *M. vaccae* represent invaluable additions to the protein expression toolbox of researchers dealing with enzymes involved in *M. tuberculosis* specific pathways.

Archaeal expression hosts

Bacteriorhodopsins are model systems for light-harvesting membrane proteins and can be obtained in large amounts from the natural source owing to their strong constitutive expression. However, investigation of mutant bacteriorhodopsins cannot simply rely on the natural source or time-consuming and labor-intensive genome mutagenesis, therefore providing an incentive to developing recombinant systems to express bacteriorhodopsins. One such system uses the extreme halophilic *Euryarcheaota Halobacterium salinarum* (syn. *halobium*) [19–21], which has been harnessed in 46 PDB entries, including several mutant structures [22] (Figure 2e); another extreme halophilic archaeon, *Natromonas pharaonis*, has been used at least once (PDB 1JGJ). Shuttle *E. coli–Halobacterium* vectors have been constructed that contain the pHH9 *H. salinarum* replication origin, *E. coli* tetracycline resistance gene (Tet^r) and the *H. volcanii* novobiocin resistance gene (Nov^r) [23]. In addition, human G-protein coupled beta2-adrenergic receptor has also been successfully expressed in *H. salinarum* [24••], thus providing proof-of-concept of the potential usefulness of these archaeal hosts for the production of mammalian membrane proteins in general and GPCRs in particular.

Yeasts expression systems

These single-celled fungi have been extensively researched as hosts for heterologous protein production, including non-methylotrophic genera as *Saccharomyces*, *Schizosaccharomyces*, *Kluyveromyces* [25], and *Yarrowia* [26], and methylotrophic genera as *Pichia pastoris* (syn. *Komagataella phaffii*) [27,28] and *Ogataea* (*Hansenula*) polymorpha (syn. *P. angusta*) [29,30,31•]. Attractive features include well-known genetics and metabolism, easy to perform gene knockout and gene replacement, availability of strong promoters (constitutive as well as tightly controlled and inducible), and fast growth to very high cell densities on inexpensive media (in microtiter plate, shake flask and bioreactors). Being lower eukaryotes, yeasts can perform many post-translational modifications as higher eukaryotes and in particular possess a secretory pathway capable of correct protein processing and high-mannose glycosylation. With serious efforts well underway to engineer humanized glycosylation in several yeasts platforms, available tools for whole genome analysis and metabolic flux analysis in yeast render them promising tools for present and future biotechnology.

Non-methylotrophic yeasts

763 distinct protein chains in 639 PDB entries list a non-methylotrophic yeast as the recombinant host for expression. Of those, *Saccharomyces cerevisiae* is the most frequently used (741) followed by *Yarrowia* (11), *Kluyveromyces lactis* (8) and *Schizosaccharomyces pombe* (3). Nearly 50% of them are human soluble proteins and protein complexes [32•], but there are also membrane proteins [33•] (Figure 2f) and even viral capsids completely reconstituted in yeast.

Methylotrophic yeasts

These yeasts can feed on methanol as sole carbon and energy source and represent the yeast-based expression host with the highest adoption rate; especially because of the high cell densities they reach in mineral minimum medium and the strength of the alcohol oxidase 1 (AOX1) and methanol and nitrogen-regulated FLD [34] promoters or the constitutive GAP promoter [35]. Heterologous production of fungal, plant, mammalian and human proteins is well-documented, especially by secretion to the extracellular medium. Highlight examples of this strategy are the full-length versions of innate immunity proteins and protein complexes, for example, complement control proteins and complement receptors, including the human decay-accelerating factor (DAF, CD55) [36], and human complement receptors 1 (CR1) [37,38] and 2 (CR2, CD21) [39•] (Figure 2g). To date, *P. pastoris* has been employed in about 880 PDB entries since first reported in the 1994 NMR structure of tick anticoagulant peptide [40].

Filamentous fungi

Aspergillus

In contrast to yeasts and the other microbial factories described here, *Aspergillus* is a multicellular organism, a filamentous fungus. Several *Aspergillus* species, chiefly *A. niger* and *A. oryzae*, are workhorses in many biotechnology applications [41,42••] and have also been used in structural biology. Their amenable genetics and strain optimization features, an unmatched capacity for secreting homologous and heterologous enzymes, as well as detailed knowledge of its industrial-scale fermentation, make *Aspergillus* an attractive expression host. Random mutagenesis screenings aimed at abolishing the proteolytic secretoma have also succeeded in isolating mutants with 98% less proteolytic activity than the wild type, apparently by the gene disruption of a fungal-specific master regulator of protease gene expression [43]. Although harder to transform than yeasts, methods for the efficient transformation of *Aspergillus* have been developed that include *Agrobacterium tumefaciens*-mediated DNA transfer [44] and transformation of protoplasts after enzymatic cell wall degradation [45]. Both

constitutive and inducible promoters are available for *A. niger* expression of foreign proteins, with inducible promoters being preferred because of the beneficial effect on protein yield of decoupling biomass production and protein expression (e.g. in the production of toxic proteins). Two inducible promoters are commonly used, the *A. niger* glucoamylase gene promoter P*glaA* repressed by xylose and highly induced when maltose or starch is used as single carbon source [46] and the *A. nidulans* alcohol dehydrogenase gene promoter P*alc*A, induced by ethanol and derivatives [47]. Recently, promising human estrogen inducible promoters ensuring both tight regulation and a linear inducer response have been developed [48]. Glycosylation in *Aspergillus* is of the high mannose type and both N-linked glycosylation and O-linked glycosylation are efficiently catalyzed. The *Aspergillus* toolbox was until recently lacking in expression vectors encoding fusion tags for purification [49]. To date, 130 NMR and X-ray crystal structures of fungal proteins expressed in *A. niger* and *A. oryzae* have been reported, for example, [50•] (Figure 2h), demonstrating the feasibility of using filamentous fungi for protein production for structural biology applications, which require very homogeneous, pure and abundant protein. In fact, some of the above issues have been addressed and solved, the greatest success having been reached using synthetic gene sequences which were codon optimized for *A. niger* [51].

Protozoa

Leishmania tarentolae

Several characteristics of *L. tarentolae* make it a convenient microorganism for the production of authentically processed eukaryotic proteins: this trypanosomatid protozoan infects lizards and is non-human pathogenic; therefore it is innocuous and only requires biosafety level 1 laboratories; and it is a robust organism that grows optimally at 26°C and can be kept under sterile conditions outside the hood using a flame. Constitutive or inducible vectors for intracellular or secretion expression are commercially available with optimized splicing signals for expression of heterologous proteins [52], which can be either stably integrated into the genome (typically into the strongly transcribed 18S rRNA locus) or maintained episomally. *L. tarentolae* cells, transformed by electroporation, can grow in adherent culture or be cultivated to higher densities in shaking cultures (5×10^8 cells/ml) in inexpensive, bacteriological media without serum. Examples of expressed proteins include human erythropoietin and interferon-gamma. To date, two PDB entries report having used *L. tarantolae* as expression host in two different crystal forms of human Cu/Zn superoxide dismutase [53•] (Figure 2i).

Dictyostelium discoideum

The social amoeba *D. discoideum* is recommended as a non-mammalian model organism for functional analysis of sequenced genes by the National Institutes of Health (Bethesda, MD, USA) since it provides an intermediate level of complexity between yeasts and plants and animals [54]. More than 30 genes orthologous to human disease genes have been found in *D. discoideum* 34 Mb genome, which shows a higher degree of gene conservation with higher eukaryotes than with fungi [55]. With an A + T content of >75%, the expression of certain genes has been enhanced by gene optimization (adjusting for a similar codon preference) and using authentic signal sequences [56,57]. Foreign DNA is efficiently delivered to *D. discoideum* by electroporation and transformants can be cultured xenically on a bacterial food source or axenically on simple salt medium, with a doubling time of 4–12 hours that exceeds the growth rate of mammalian cells. *D. discoideum* performs mammalian-style N-linked and O-linked glycosylation but cannot accomplish complex terminal glycosylation since it lacks the ability to add galactose, N-acetyl galactosamine, or sialic acids [58,59]. In fact, several therapeutic glycoproteins have been successfully expressed in *D. discoideum* including rotavirus VP7, muscarinic receptor m2 and m3, antithrombin III, gonadotropins, or erythropoietin. The most remarkable applications of *D. discoideum* in structural biology concern the production of wild-type, chimeric, and mutant motor proteins, for example, the actin–tropomyosin–myosin rigor complexes [60] (Figure 2j) and dynein [61].

Conclusion

Success in recombinant protein expression for structural biology involves making crucial choices about construct design, coexpression partners, culture parameters, and induction regime, among many others — including switching to unconventional expression hosts that may be particularly well-suited to the target gene for various reasons (Figure 3). These lesser known organisms, which can accomplish expression of heterologous proteins (as general purpose hosts or 'generalists') as well as revealing themselves most valuable when over-producing their own proteins or specific classes of macromolecules (as 'specialist' hosts) (Figure 3b), include Gram-negative and Gram-positive bacteria, extreme halophilic archaea, and eukaryotic fungi (yeast and filamentous fungi) and protozoa. As demonstrated in more than 3100 distinct PDB protein chains, these hosts can efficiently generate high yields of proteins endowed with correct fold, activity, and post-translational modifications, and possess sufficient quality for structural and biophysical analysis. Growing access to those hosts via commercial or academic ready-to-use toolkits should facilitate their comparative screening and the solid experimental assessment of their relative merits for the production of different classes of proteins and protein complexes.

Acknowledgements

This work was supported by the EC project ComplexINC (Framework Programme 7 (FP7) under grant agreement no. 279039) and Comunidad de Madrid project Complemento-CM (S2010/BMD-2316).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, Schweizer HP: **A Tn7-based broad-range bacterial cloning and expression system**. *Nat Methods* 2005, **2**:443-448.

The paper shows that constructed mini-T7 vectors can be targeted for integration into the bacterial chromosome with the aid of a helper vector encoding the site-specific transposition pathway and used to drive protein expression in diverse Gram-negative bacteria of biomedical and environmental importance.

2. De Laurentis W, Leang K, Hahn K, Podemski B, Adam A, Kroschwald S, Carter LG, van Pee KH, Naismith JH: **Preliminary crystallographic characterization of PrnB, the second enzyme in the pyrrolnitrin biosynthetic pathway**. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2006, **62**:1134-1137.

3. •• Shen BW, Heiter DF, Chan SH, Wang H, Xu SY, Morgan RD, Wilson GG, Stoddard BL: **Unusual target site disruption by the rare-cutting HNH restriction endonuclease PacI**. *Structure* 2010, **18**:734-743.

In this article the high %G + C Gram-negative bacterium *P. aeruginosa* is exploited to recombinantly express the restriction endonuclase PacI, which cleaves inside an exclusively A/T site, without requiring the co-expression of the associated methylase.

4. Lack N, Lowe ED, Liu J, Eltis LD, Noble ME, Sim E, Westwood IM: **Structure of HsaD, a steroid-degrading hydrolase, from *Mycobacterium tuberculosis***. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2008, **64**:2-7.

5. Kok M, Oldenhuis R, van der Linden MP, Raatjes P, Kingma J, van Lelyveld PH, Witholt B: **The *Pseudomonas oleovorans* alkane hydroxylase gene. Sequence and expression**. *J Biol Chem* 1989, **264**:5435-5441.

6. Panke S, Meyer A, Huber CM, Witholt B, Wubbolts MG: **An alkane-responsive expression system for the production of fine chemicals**. *Appl Environ Microbiol* 1999, **65**:2324-2332.

7. Husken LE, Beeftink R, de Bont JA, Wery J: **High-rate 3-methylcatechol production in *Pseudomonas putida* strains by means of a novel expression system**. *Appl Microbiol Biotechnol* 2001, **55**:571-577.

8. Eppinger M, Bunk B, Johns MA, Edirisinghe JN, Kutumbaka KK, Koenig SS, Creasy HH, Rosovitz MJ, Riley DR, Daugherty S *et al.*: **Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319**. *J Bacteriol* 2011, **193**:4199-4213.

9. Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Medigue C *et al.*: **From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later**. *Microbiology* 2009, **155**:1758-1775.

10. Rygus T, Hillen W: **Inducible high-level expression of heterologous genes in *Bacillus megaterium* using the regulatory elements of the xylose-utilization operon**. *Appl Microbiol Biotechnol* 1991, **35**:594-599.

11. Gamer M, Frode D, Biedendieck R, Stammen S, Jahn D: **A T7 RNA polymerase-dependent gene expression system for *Bacillus megaterium***. *Appl Microbiol Biotechnol* 2009, **82**:1195-1203.

12. •• Biedendieck R, Borgmeier C, Bunk B, Stammen S, Scherling C, Meinhardt F, Wittmann C, Jahn D: **Systems biology of recombinant protein production using *Bacillus megaterium***. *Methods Enzymol* 2011, **500**:165-195.

Concise yet fairly complete description of the *B. megaterium* expression system with a strong emphasis on how to use the toolkit in systems biology applications.

13. Pruitt RN, Chumbler NM, Rutherford SA, Farrow MA, Friedman DB, Spiller B, Lacy DB: **Structural determinants of *Clostridium* difficile toxin A glucosyltransferase activity**. *J Biol Chem* 2012, **287**:8013-8020.

14. Parthier C, Gorlich S, Jaenecke F, Breithaupt C, Brauer U, Fandrich U, Clausnitzer D, Wehmeier UF, Bottcher C, Scheel D *et al.*: **The O-carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction**. *Angew Chem Int Ed Engl* 2012, **51**:4046-4052.

15. Herai S, Hashimoto Y, Higashibata H, Maseda H, Ikeda H, Omura S, Kobayashi M: **Hyper-inducible expression system for streptomycetes**. *Proc Natl Acad Sci U S A* 2004, **101**:14031-14035.

16. Lussier FX, Denis F, Shareck F: **Adaptation of the highly productive T7 expression system to *Streptomyces lividans***. *Appl Environ Microbiol* 2010, **76**:967-970.

17. •• Luckner SR, Machutta CA, Tonge PJ, Kisker C: **Crystal structures of *Mycobacterium tuberculosis* KasA show mode of action within cell wall biosynthesis and its inhibition by thiolactomycin**. *Structure* 2009, **17**:1004-1013.

This article shows the use of *Mycobacterium smegmatis* as a recombinant expression host for KasA, an enzyme that participates in the biosynthesis of the highly specialized cell-wall of *M. tuberculosis*.

18. Changsen C, Franzblau SG, Palittapongarnpim P: **Improved green fluorescent protein reporter gene-based microplate screening for antituberculosis compounds by utilizing an acetamidase promoter**. *Antimicrob Agents Chemother* 2003, **47**:3682-3687.

19. Ni BF, Chang M, Duschl A, Lanyi J, Needleman R: **An efficient system for the synthesis of bacteriorhodopsin in *Halobacterium halobium***. *Gene* 1990, **90**:169-172.

20. Krebs MP, Hauss T, Heyn MP, RajBhandary UL, Khorana HG: **Expression of the bacterioopsin gene in *Halobacterium halobium* using a multicopy plasmid**. *Proc Natl Acad Sci U S A* 1991, **88**:859-863.

21. Heymann JA, Havelka WA, Oesterhelt D: **Homologous overexpression of a light-driven anion pump in an archaebacterium**. *Mol Microbiol* 1993, **7**:623-630.

22. Joh NH, Min A, Faham S, Whitelegge JP, Yang D, Woods VL, Bowie JU: **Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins**. *Nature* 2008, **453**:1266-1270.

23. Cline SW, Schalkwyk LC, Doolittle WF: **Transformation of the archaebacterium *Halobacterium volcanii* with genomic DNA**. *J Bacteriol* 1989, **171**:4987-4991.

24. •• Jaakola VP, Rehn M, Moeller M, Alexiev U, Goldman A, Turner GJ: **G-protein-coupled receptor domain overexpression in *Halobacterium salinarum*: long-range transmembrane interactions in heptahelical membrane proteins**. *Proteins* 2005, **60**:412-423.

In addition to naturally overexpressing the model membrane GPCR bacteriorhodopsin, *H. salinarum* can be harnessed for the production of heterologous mammalian and human GPCRs by fusing the third intracellular loop of bacteriorhodopsin into the third intracellular loop of the target GPCR molecule.

25. van Ooyen AJ, Dekker P, Huang M, Olsthoorn MM, Jacobs DI, Colussi PA, Taron CH: **Heterologous protein production in the yeast Kluyveromyces lactis**. *FEMS Yeast Res* 2006, **6**:381-392.

26. Madzak C, Gaillardin C, Beckerich JM: **Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review**. *J Biotechnol* 2004, **109**:63-81.

27. Cregg JM, Cereghino JL, Shi J, Higgins DR: **Recombinant protein expression in *Pichia pastoris***. *Mol Biotechnol* 2000, **16**:23-52.

28. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM: **Heterologous protein production using the *Pichia pastoris* expression system**. *Yeast* 2005, **22**:249-270.

29. Sudbery PE, Gleeson MA, Veale RA, Ledeboer AM, Zoetmulder MC: ***Hansenula polymorpha* as a novel yeast system for the expression of heterologous genes**. *Biochem Soc Trans* 1988, **16**:1081-1083.

30. van Dijk R, Faber KN, Kiel JA, Veenhuis M, van der Klei I: **The methylotrophic yeast *Hansenula polymorpha*: a versatile cell factory**. *Enzyme Microb Technol* 2000, **26**:793-800.

31. Gellissen G, Kunze G, Gaillardin C, Cregg JM, Berardi E, Veenhuis M, van der Klei I: **New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adeninivorans* and *Yarrowia lipolyticainosa* is exploited to recombinantly express the— a comparison**. *FEMS Yeast Res* 2005, **5**:1079-1096.

In this mini-review the authors compared two methylotrophic yeasts and two non-methylotrophic yeasts with respect to protein expression capacity, illustrating the versatility of yeast-based expression platforms as microbial cell factories.

32. Son SY, Ma J, Kondou Y, Yoshimura M, Yamashita E, Tsukihara T: **Structure of human monoamine oxidase A at 2.2-A resolution: the control of opening the entry for substrates/inhibitors**. *Proc Natl Acad Sci U S A* 2008, **105**:5739-5744.

In this article the authors describe the crystal structure of a human mitochondrial outer membrane-associated protein expressed in *S. cerevisiae*, extending previous studies using the rat enzyme.

33. Pedersen BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, Nissen P: **Crystal structure of the plasma membrane proton pump**. *Nature* 2007, **450**:1111-1114.

The first crystal structure of a plasma membrane P-type proton pump was obtained from purified material heterologously expressed in *S. cerevisiae*.

34. Shen S, Sulter G, Jeffries TW, Cregg JM: **A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris**. *Gene* 1998, **216**:93-102.

35. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM: **Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter**. *Gene* 1997, **186**:37-44.

36. Uhrinova S, Lin F, Ball G, Bromek K, Uhrin D, Medof ME, Barlow PN: **Solution structure of a functionally active fragment of decay-accelerating factor**. *Proc Natl Acad Sci U S A* 2003, **100**:4718-4723.

37. O'Leary JM, Bromek K, Black GM, Uhrinova S, Schmitz C, Wang X, Krych M, Atkinson JP, Uhrin D, Barlow PN: **Backbone dynamics of complement control protein (CCP) modules reveals mobility in binding surfaces**. *Protein Sci* 2004, **13**:1238-1250.

38. Smith BO, Mallin RL, Krych-Goldberg M, Wang X, Hauhart RE, Bromek K, Uhrin D, Atkinson JP, Barlow PN: **Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor**. *Cell* 2002, **108**:769-780.

39. Prota AE, Sage DR, Stehle T, Fingeroth JD: **The crystal structure of human CD21: implications for Epstein-Barr virus and C3d binding**. *Proc Natl Acad Sci U S A* 2002, **99**:10641-10646.

P. pastoris was used to express the two N-terminal short consensus repeats (SCR1-2) of human complement receptor 2 (CR2, CD21), the region that recognizes the Epstein-Barr virus and the C3 proteolytic fragment C3d.

40. Lim-Wilby MS, Hallenga K, de Maeyer M, Lasters I, Vlasuk GP, Brunck TK: **NMR structure determination of tick anticoagulant peptide (TAP)**. *Protein Sci* 1995, **4**:178-186.

41. Fleissner A, Dersch P: **Expression and export: recombinant protein production systems for *Aspergillus***. *Appl Microbiol Biotechnol* 2010, **87**:1255-1270.

42. Lubertozzi D, Keasling JD: **Developing *Aspergillus* as a host for heterologous expression**. *Biotechnol Adv* 2009, **27**:53-75.

In this review paper the *Aspergillus* expression systems is presented as a platform for manifold applications, from recombinant protein production to the production of fungal secondary metabolites and mycoremediation.

43. Punt PJ, Schuren FH, Lehmbeck J, Christensen T, Hjort C, van den Hondel CA: **Characterization of the *Aspergillus niger* prtT, a unique regulator of extracellular protease encoding genes**. *Fungal Genet Biol* 2008, **45**:1591-1599.

44. Michielse CB, Hooykaas PJ, van den Hondel CA, Ram AF: **Agrobacterium-mediated transformation of the filamentous fungus *Aspergillus awamori***. *Nat Protoc* 2008, **3**:1671-1678.

45. de Bekker C, Wiebenga A, Aguilar G, Wosten HA: **An enzyme cocktail for efficient protoplast formation in *Aspergillus niger***. *J Microbiol Methods* 2009, **76**:305-306.

46. Fowler T, Berka RM, Ward M: **Regulation of the glaA gene of *Aspergillus niger***. *Curr Genet* 1990, **18**:537-545.

47. Nikolaev I, Mathieu M, van de Vondervoort P, Visser J, Felenbok B: **Heterologous expression of the *Aspergillus nidulans* alcR-alcA system in *Aspergillus niger***. *Fungal Genet Biol* 2002, **37**:89-97.

48. Pachlinger R, Mitterbauer R, Adam G, Strauss J: **Metabolically independent and accurately adjustable *Aspergillus* sp. expression system**. *Appl Environ Microbiol* 2005, **71**:672-678.

49. Roth AH, Dersch P: **A novel expression system for intracellular production and purification of recombinant affinity-tagged proteins in *Aspergillus niger***. *Appl Microbiol Biotechnol* 2010, **86**:659-670.

50. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JC, Johansen KS, Krogh KB, Jorgensen CI, Tovborg M, Anthonsen A *et al.*: **Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components**. *Proc Natl Acad Sci U S A* 2011, **108**:15079-15084.

Finding factors that enhance cellulose degradation is an ongoing challenge in biotechnology. In this paper *A. oryzae* is used as recombinant host for the expression of a copper-dependent oxidase that favors cellulose degradation by GH61, showing that *Aspergillus* is capable of correctly charging the enzyme's active site with copper and catalyzing the methylation of an active site's histidine residue.

51. Koda A, Bogaki T, Minetoki T, Hirotsune M: **High expression of a synthetic gene encoding potato alpha-glucan phosphorylase in *Aspergillus niger***. *J Biosci Bioeng* 2005, **100**:531-537.

52. Breitling R, Klingner S, Callewaert N, Pietrucha R, Geyer A, Ehrlich G, Hartung R, Muller A, Contreras R, Beverley SM *et al.*: **Non-pathogenic trypanosomatid protozoa as a platform for protein research and production**. *Protein Exp Purif* 2002, **25**:209-218.

53. Gazdag EM, Cirstea IC, Breitling R, Lukes J, Blankenfeldt W, Alexandrov K: **Purification and crystallization of human Cu/Zn superoxide dismutase recombinantly produced in the protozoan *Leishmania tarentolae***. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2010, **66**:871-877.

Presentation of *L. tarentoale* as a viable expression host for authentic, high-quality eukaryotic proteins in recombinant form for structural biology, using a model human protein.

54. Kreppel L, Kimmel AR: **Genomic database resources for *Dictyostelium discoideum***. *Nucleic Acids Res* 2002, **30**:84-86.

55. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q *et al.*: **The genome of the social amoeba *Dictyostelium discoideum***. *Nature* 2005, **435**:43-57.

56. Warrick HM, Spudich JA: **Codon preference in *Dictyostelium discoideum***. *Nucleic Acids Res* 1988, **16**:6617-6635.

57. Reymond CD, Beghdadi-Rais C, Roggero M, Duarte EA, Desponds C, Bernard M, Groux D, Matile H, Bron C, Corradin G *et al.*: **Anchoring of an immunogenic *Plasmodium falciparum* circumsporozoite protein on the surface of *Dictyostelium discoideum***. *J Biol Chem* 1995, **270**:12941-12947.

58. Williams KL, Emslie KR, Slade MB: **Recombinant glycoprotein production in the slime mould *Dictyostelium discoideum***. *Curr Opin Biotechnol* 1995, **6**:538-542.

59. Asgari S, Arun S, Slade MB, Marshall J, Williams KL, Wheldrake JF: **Expression of growth factors in *Dictyostelium discoideum***. *J Mol Microbiol Biotechnol* 2001, **3**:491-497.

60. Behrmann E, Muller M, Penczek PA, Mannherz HG, Manstein DJ, Raunser S: **Structure of the rigor actin–tropomyosin–myosin complex**. *Cell* 2012, **150**:327-338.

61. Kon T, Oyama T, Shimo-Kon R, Imamula K, Shima T, Sutoh K, Kurisu G: **The 2.8 A crystal structure of the dynein motor domain**. *Nature* 2012, **484**:345-350.