The methylotrophic yeast *Hansenula polymorpha*: a versatile cell factory

Ralf van Dijk, Klaas Nico Faber, Jan A.K.W. Kiel, Marten Veenhuis, Ida van der Klei*

Eukaryotic Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Postbus 14, 9750 AA Haren, The Netherlands

Received 24 March 1999; received in revised form 19 July 1999; accepted 10 August 1999

Abstract

The development of heterologous overexpression systems for soluble proteins has greatly advanced the study of the structure/function relationships of these proteins and their biotechnological and pharmaceutical applications. In this paper we present an overview on several aspects of the use of the methylotrophic yeast *Hansenula polymorpha* as a host for heterologous gene expression. *H. polymorpha* has been successfully exploited as a cell factory for the large-scale production of such components. Stable, engineered strains can be obtained by site-directed integration of expression cassettes into the genome, for which various constitutive and inducible promoters are available to control the expression of the foreign genes. New developments have now opened the way to additional applications of *H. polymorpha*, which are unprecedented for other organisms. Most importantly, it may be the organism of choice for reliable, large-scale production of heterologous membrane proteins, using inducible intracellular membranes and targeting sequences to specifically insert these proteins stably into these membranes. Furthermore, the use of *H. polymorpha* offers the possibility to accumulate the produced components into specific compartments, namely peroxisomes. These organelles are massively induced during growth of the organism on methanol and may occupy up to 80% of the cell volume. Accumulation inside peroxisomes prevents undesired modifications (e.g. proteolytic processing or glycosylation) and is also in particular advantageous when proteins are produced which are toxic or harmful for the host. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Cell factory; Heterologous proteins; Genetic tools; *Hansenula polymorpha*

1. Introduction

Heterologous production of biologically active proteins has become an important tool in fundamental and applied research. At present a variety of hosts (ranging from bacteria to higher eukaryotes) have been explored for their capacity to produce heterologous proteins. In general, microorganisms are favored, because they have short generation times, are easy to grow and are readily accessible to genetic manipulations. Yeasts are often preferred for the production of plant or animal proteins, because in these organisms protein modifications typical for eukaryotes normally occur. These modifications (e.g. glycosylation, acylation, phosphorylation, formation of disulfide bonds; [1]) are often essential for the function and/or stability of the protein.

Initially, *Saccharomyces cerevisiae* has been the yeast species of choice for foreign protein production for obvious reasons; its genetics were well developed and the organism is generally regarded as safe. Despite successful applications specific disadvantages in the use of *S. cerevisiae* have also been encountered. These include among others the oft-observed instability of the engineered strains, undesired hyperglycosylation and relatively low yields, due to the lack of strong promoters.

The methylotrophic yeast *Hansenula polymorpha* has been recognized as an attractive alternative. Originally, methylotrophic yeasts have been isolated for the production of single-cell protein at the expense of cheap carbon sources, like methanol. Studies on the physiology, biochemistry and ultrastructure of these yeasts revealed that methanol is oxidized by a hydrogen-peroxide producing alcohol oxidase (AOX), which is localized in peroxisomes. In these organelles also catalase, which decomposes the hydrogen-peroxide produced, and dihydroxyacetone synthase...
(DHAS), which catalyzes the first step in the assimilatory pathway, are present. Other enzymes involved in methanol assimilation (enzymes of the xylulose-5-phosphate pathway [2]) or dissimilation [(formaldehyde dehydrogenase (FdDH) and formate dehydrogenase (FMD)] are cytosolic [3].

A striking feature of methanol-grown yeasts is the excessive proliferation of peroxisomes. In methanol-limited continuous cultures these organelles can occupy up to 80% of the total cell volume [3]. In such cells AOX and DHAS may constitute over 60% of the total cell protein, which illustrates that the genes encoding these proteins are controlled by very strong promoters. Also, the expression of the genes encoding FdDH and FMD are under control of strong, inducible promoters [4,5].

H. polymorpha has some specific advantages over other methylotrophic yeasts (Candida boidinii, Pichia pastoris), being more thermotolerant and capable to grow at higher rates on simple, defined media. The relatively high optimal growth temperature for H. polymorpha (37–43°C vs. 30°C for C. boidinii, P. pastoris, and S. cerevisiae) may be favorable for the production of mammalian (including human) proteins and furthermore has the advantage that it allows a better cooling management and reduces the risk of contaminations in large scale fermentations.

In this paper recent relevant developments on the use of H. polymorpha as a reliable, efficient cell factory are discussed.

2. Tools for introducing (heterologous) genes in H. polymorpha

2.1. Transformation procedures

The molecular techniques for the manipulation of H. polymorpha have now been advanced to a level similar to those available for S. cerevisiae. Efficient and reliable transformation procedures for H. polymorpha by using electroporation have been developed by Faber et al. [6] (transformation frequency of $1.7 \times 10^6$ /μg DNA). Optimal results are obtained by using batch cells from the midexponential growth phase that are pretreated with the reducing agent DTT to weaken the cell wall. Upon addition of plasmid DNA, an electrical pulse is applied that allows entry of the DNA molecules. The electroporation procedure is now the method of choice for efficient H. polymorpha transformations.

2.2. Expression vectors

An essential tool for the selection of transformants is the availability of suitable genetic markers. Various auxotrophic H. polymorpha mutants (leu, ura, trp, and ade) have been generated and vectors have been constructed that carry homologous (LEU1.1, URA3, TRP3, and ADE11) or heterologous (LEU2, URA3 from S. cerevisiae, LEU2 from Candida albicans) genes that functionally complement these mutants [7–9]. Also, genes that confer resistance against G418 [10], phleomycin [11] or Zeocin [12] have successfully been used as dominant selection markers in H. polymorpha.

Several H. polymorpha autonomously replicating sequences (HARS) and other DNA-fragments that promote autonomous replication of plasmids in H. polymorpha have been isolated and are currently being used in various vectors [8,13–18]. The DNA fragment that contains the S. cerevisiae LEU2 gene as the selectable marker has also been shown to promote autonomous replication in H. polymorpha [18].

Stable transformants, which do not have to be grown under selective growth conditions, are obtained upon integration of plasmid DNA into the genome. For H. polymorpha both legitimate and illegitimate integrations have been reported.

Targeted integration can be obtained both by single or double cross-over recombination, resulting in additive integration or gene replacement, respectively. Integration is routinely achieved by using linearized plasmids containing DNA sequences from homologous genes, which are specifically targeted to the corresponding sites in the genome [19]. Most commonly used are the genes encoding alcohol oxidase (AOX) or amine oxidase (AMO).

An elegant disruption-replacement method for targeted single copy integration has recently been described [20]. First, the TRP3 gene of an H. polymorpha leu1.2 strain was disrupted by insertion of the S. cerevisiae LEU2 gene. The genotype of the resulting strain is trp3Δ LEU2. The trp3Δ allele in this strain and the tightly linked AOX gene can subsequently be replaced by a DNA fragment containing part of the TRP3 gene and expression cassettes containing P_AOX. The proper integrants, which are again TRP3 leu2, can easily be selected. An attractive advantage of this procedure is that no vector sequences are introduced into the genome of the resulting strains. Moreover, the resulting strain is again leu1.2, which offers the possibility to introduce additional genes in the same strain using LEU2 as selectable marker.

Single- or multicopy integrants can easily be selected by using an H. polymorpha leu1.1 strain and a vector containing S. cerevisiae LEU2 as selectable marker. Because multicopy integrants form large colonies on selective media, whereas single-copy integrants develop relatively small ones. The latter is due to the fact that the S. cerevisiae gene does not fully complement the leucine auxotrophy in H. polymorpha. A similar strategy can be followed by using an H. polymorpha ura3 strain and the S. cerevisiae URA3 gene. These strategies have enabled the generation of strains containing more than 100 copies of the vector per genome equivalent [13,21].

Another strategy [22] to select strains with multiple integrations is the use of resistance markers and several consecutive selection rounds by using increasing concentrations of the antibiotic.
2.3. Inducible and constitutive promoters

Like other methylotrophic yeasts, *H. polymorpha* possesses several strong promoters, which are induced during growth on methanol. The promoters of the alcohol oxidase (PAOX), dihydroxyacetone synthase (PDHAS), and formate dehydrogenase (PFMD) genes are most commonly used to control heterologous gene expression [4,5,23,24].

The P_AOX promoter is fully repressed by glucose or ethanol and strongly induced by methanol. When glucose or ethanol concentrations are low (e.g. in late exponential batch cultures or carbon-limited continuous cultures), P_AOX is derepressed, resulting in low AOX expression levels. Derepression also occurs when cells grow in batch cultures containing glycerol or xylitol as sole carbon source. In glycerol or xylitol grown cells the AO levels can reach up to 30% to 70% of the maximal values. The PDHAS and PFMD are also strongly induced during growth on methanol and the kinetics of derepression resemble that of P_AOX. These strong regulated promoters are, therefore, tools of choice to control high gene expression levels in *H. polymorpha*

In constructed strains that are not able to grow on methanol (e.g. upon gene replacement at the AOX-locus) induction of the P_AOX, PDHAS, and PFMD can be readily achieved by using continuous cultures and mixed substrates (e.g. glucose/methanol or glucose/choline mixtures). In the latter case the actual inducer is formaldehyde that is released from choline during the metabolism of this nitrogen source [25].

As an independent, inducible promoter the amine oxidase promoter (P_AMO) can be used. Amine oxidase (AMO) is a peroxisomal enzyme involved in oxidation of primary amines (methylamine, ethylamine), which can be utilized as sole source of nitrogen by *H. polymorpha*. P_AMO is induced by these amines and fully repressed by ammonium [26]. Compared to the P_AOX the P_AMO is much weaker (up to 20% of P_AOX).

Also the promoters of three nitrate assimilation genes, YNT1, YNI1, and YNR1, which are induced by nitrate but strongly repressed by ammonium may serve as additional tools for controlling heterologous gene expression [27]. Furthermore, the recently identified promoter of the *H. polymorpha* PHO1 gene encoding a repressible acid phosphatase [28] may be suitable for the control of heterologous genes that have to be expressed at relatively moderate levels.

By using the above promoters various attractive induction strategies can now be designed. For instance, in case of the production of harmful proteins, first biomass can be generated under repressing conditions, followed by an alteration in the composition of the growth medium, thereby inducing the expression of the heterologous gene.

Strong constitutive promoters have been identified in *H. polymorpha* as well. These include the promoters of the genes encoding a plasma membrane ATPase (PMAI) [24] or encoding the translation elongation factor-1α (TEF1/TEF2; [29] and J.A.K.W. Kiel, unpublished results).

In the course of our studies on the biogenesis of peroxisomes, 10 *H. polymorpha* PEX genes have been cloned, which all encode proteins essential for peroxisome biogenesis. The promoters of the PEX genes are relatively weak. Moreover, most of them are constitutively expressed or only slightly induced upon a shift of cells to peroxisome-inducing growth conditions [30–32]. As a consequence these promoters are not very suitable to control heterologous gene expression.

2.4. Sorting of heterologous proteins to specific subcellular locations

By using the above tools, it is possible to introduce foreign genes in *H. polymorpha* and to carefully control their expression. *H. polymorpha* has however an additional attractive advantage that is unprecedented for other cell factories. This includes the possibility to select specific subcellular compartments as optimal targets for specific proteins. Targeting signals have now been identified for various cell compartments of *H. polymorpha* (mitochondrion, peroxisome, endoplasmic reticulum, secretory pathway, vacuole). Such signals can easily be added to any protein, simply by constructing chimeric genes encoding fusion proteins that contain a (homologous or heterologous) targeting signal. If necessary, cleavage of the protein and the signal can be achieved in vivo or in vitro by using highly specific endopeptidases (e.g. the TEV protease [33] or factor Xa [34]) provided that the appropriate cleavage site is introduced between the protein and the signal sequence.

Below some examples are given of the advantage of the sort the of heterologous proteins to specific target organelles.

2.5. Accumulation in the peroxisomal matrix

Soluble proteins can be specifically targeted to the peroxisomal matrix. These organelles have the capacity to accumulate proteins to very high concentrations. It should be stressed that massive peroxisome proliferation can also be induced under conditions where the organelles are not essential for growth (e.g. in carbon-limited chemostat cultures by using mixed substrates; see above). This implies that the storage capacity of the organelles can in principle be used without affecting the viability of the cell. The advantage of storage in the peroxisomal matrix is the absence of protein modifying enzymes in this cell compartment (e.g. mediating phosphorylation, glycosylation, or proteolytic processing), which may give rise to undesired modifications upon production in the cytosol or during passage in the endoplasmic reticulum. The absence of proteolytic activity in the peroxisomal matrix of *H. polymorpha* is highly advantageous in case of heterologous proteins that are relatively unstable and/or are sensitive to proteolytic degradation.

Accumulation in peroxisomes may also be the method of choice for producing components that are toxic for the host;
compartmentalization in peroxisomes will prevent that such components can exert their toxic activity.

Targeting of heterologous proteins to the peroxisomal matrix can be mediated by addition of one of the two known peroxisomal targeting signals, PTS1 or PTS2 [35,36]. The PTS1 consists of only three amino acids and is present at the extreme carboxyterminus of many peroxisomal matrix proteins. The first identified PTS1, -SKL-COOH), was found in firefly luciferase. Later studies revealed that several amino acid substitutions are allowed in this sequence [37]. In H. polymorpha the enzymes AOX (-ARF), DHAS (-NKL), catalase (-SKI), and the peroxisomal matrix protein Pex8p (-AKL; see Fig. 1) contain a PTS1, that have been shown to be functional in targeting reporter proteins to peroxisomes of H. polymorpha.

The successful use of a PTS1 to accumulate large amounts of a heterologous protein in H. polymorpha peroxisomes has been demonstrated [30,38]. The PTS1 (SKL-COOH) was added to the carboxyterminus of a fusion protein, which consisted of human insulin-like growth factor II and a carrier protein. Overproduction of this protein, which was properly targeted to the peroxisomes, resulted in levels of >20% of the total cell protein [34].

The PTS2 is present at the N-terminus of peroxisomal matrix proteins and characterized by the consensus sequence (R/K)(L/V/I)X5(H/Q)(L/A) [35,36]. In H. polymorpha this signal has been found in AMO (RLX5QA) and Pex8p (KLX5QL). Import of PTS1 proteins is efficient under all growth conditions tested so far. However, for efficient import of PTS2 proteins the cells have to be grown in media containing primary amines as sole nitrogen source (i.e. conditions that induce expression of the AMO gene), possibly because only under these conditions the genes encoding essential components involved in PTS2 protein import are sufficiently expressed [18].

2.6. Intracellular membranes

Of special interest is the establishment of expression systems for overproduction of heterologous membrane proteins. Large-scale production of foreign membrane proteins is troublesome in all systems used to date. On the other hand, the production of membrane proteins (e.g. human hormone receptors) is of major importance in fundamental studies for obtaining sufficient amounts of proteins for functional studies or for three-dimensional structure determinations. Membrane proteins form a large and important class of proteins. Analysis of the available eukaryotic genomes shows that at least 30% of the open reading frames encode membrane proteins. Moreover, membrane protein malfunctioning is the cause of many inherited diseases in man. Hence, the availability of a reliable system to overproduce these proteins is of great medical interest, e.g. for rational drug design. H. polymorpha can fill in the existing gap in membrane protein production systems. The organism has the advantage that excessive amounts of peroxisomal membranes can be produced that largely lack homologous proteins. These membranes are induced during incubation of cells on oleic acid-containing media (H. polymorpha is not capable to grow on oleic acid) [39], but do not form organelles and thus have no matrix contents. Therefore, these membranes are ideal targets to accumulate foreign membrane proteins. Because these membranes are peroxisomal in nature, sorting can be mediated by a specific targeting signal of a homologous H. polymorpha peroxisomal membrane protein. Efficient sorting of specific human membrane proteins that were biologically active has already been achieved by using the N-terminus of Pex3p [32] as sorting signal.

2.7. Secretion

Secretion of an (over)produced heterologous protein is often favorable, because it facilitates the recovery of the protein from the culture fluid. Also in this respect H. polymorpha offers a distinct advantage, because this yeast species does not secrete significant amounts of proteins. Therefore, the secreted heterologous protein can be recovered in a relatively pure state from the culture medium (>90% of total extracellular protein). Another important aspect of protein secretion is that in the secretory pathway of eukaryotes several protein modifications take place, which are often essential for the activation of proteins (proteolytic processing, formation of the correct disulphide bonds, glycosylation etc.).

To establish secretion it is necessary to fuse the protein of interest to an N-terminal presequence (also called leader
sequence) that directs the protein to the secretory pathway. Leader sequences have a tripartite structure, consisting of a positively charged region at the extreme N-terminus, a central hydrophobic stretch of 7 to 15 amino acids, followed by a polar region (3–7 amino acids) that contains the recognition site for a specific ER-located leader peptidase. The cleavage site is characterized by small side chains at position –1 and –3 relative to the residue where the presequence is processed. Some secreted proteins contain, next to the presequence, a proregion. These preprosequences are processed twice. The presequence is cleaved by the ER-leader peptidase, whereas the second one is cleaved in the late Golgi by a functional equivalent of \( S. \text{cerevisiae} \) Kex2 [40,41]. Two basic residues (KR or RR) precede the second cleavage site.

So far only one homologous \( H. \text{polymorpha} \) secretion signal is known, namely the presequence of repressible acid phosphatase (PHO1). This presequence is a good candidate to mediate secretion of heterologous proteins in \( H. \text{polymorpha} \) [28], but has not been tested yet.

Of the various heterologous secretion signals available to drive protein secretion in \( H. \text{polymorpha} \), the preprosequence of \( S. \text{cerevisiae} \) mating factor 1α is mostly used [42,43,44]. As illustrated in Table 1, this signal invariably resulted in efficient secretion in \( H. \text{polymorpha} \).

Other heterologous prepro-sequences, which have successfully been used in \( H. \text{polymorpha} \) are derived from \( \text{Schwanniomyces occidentalis} \) glucoamylase [5,45], \( \text{Carcinus maenas} \) hyperglycemic hormone [45], \( \text{Aspergillus niger} \) glucoamylase [20], \( \text{Kluyveromyces lactis} \) killer toxin [20], and \( \text{Kluyveromyces marxianus} \) inulinase [46] (see Table 1).

Expression of \( S. \text{cerevisiae} \) SUC2 [47], which codes for invertase including its authentic leader sequence, resulted in efficient secretion of Sc-invertase protein by \( H. \text{polymorpha} \) [48]. Also upon fusion of the Sc-invertase presequence to Ct-α-galactosidase (see Table 1), the heterologous protein was successfully secreted. However, when the Sc-invertase leader sequence was fused to human tyroid peroxidase this signal was not effective (Table 1). Experiments in which different leader sequences were compared with respect to their capacity to direct a protein to the secretory pathway of \( H. \text{polymorpha} \), suggested that the search for the most efficient leader sequence is often a matter of trial and error (Table 1).

Several examples have been described of proteins (ranging from fungal to human ones) that carried the authentic, endogenous presequence and were properly directed to the secretory pathway of \( H. \text{polymorpha} \) and subsequently secreted (see Table 1).

Although the secretory pathway is highly conserved in eukaryotes, not all heterologous targeting sequences function efficiently in \( H. \text{polymorpha} \).

Proper translation of the presequence could be hampered due to the presence of rare codons for \( H. \text{polymorpha} \) [20] resulting in a decrease in translation rate or even in a translation block. Adjustments in the coding regions can solve such problems. Incorrect or incomplete processing of the pre(pro)sequences can also result in reduced secretion efficiency. Examples in \( H. \text{polymorpha} \) are the secretion of human urinary plasmalogen activator (slow processing of the presequence) or leech hirudin (incorrect processing resulting in one amino acid extension) [20,45]. These problems are generally related to insufficient recognition of the cleavage site(s) by the processing peptidases. A rational design of pre(pro)sequences and processing sites based on empirical rules as postulated by von Heijne [49], has proven to be effective for \( S. \text{occidentalis} \) preprosequences [45], and could be applied for \( H. \text{polymorpha} \) as well.

Proper glycosylation is another important aspect of the production of functional proteins in heterologous hosts. The
majority of human therapeutic proteins are glycoproteins, but improper glycosylation of heterologous proteins may cause an immunogenic response. Hence, it is of utmost importance that the glycosylation patterns resemble those present in the authentic proteins.

In *S. cerevisiae*, hyperglycosylation has often been observed (i.e. heterogenous glycosylation by addition of more than 40 mannose residues to the glycosylation core). In general, the overall length of the mannose chains is lower in *H. polymorpha* compared to *S. cerevisiae* and only few cases of hyperglycosylation have been reported in *H. polymorpha* [50,51]. Hence, *H. polymorpha* may be a much better candidate to produce human therapeutic proteins compared to *S. cerevisiae*.

3. Conclusion

At present a range of heterologous expression systems is available to cover the need for various proteins to be used for fundamental structure/function analysis and for biotechnological and pharmaceutical purposes. Among these systems, *H. polymorpha* is now recognized as a very suitable one. The first recombinant products produced in *H. polymorpha* have now passed the clinical examinations and are being launched on the market (e.g. a novel Hepatitis B vaccine) [52,53] and others will soon follow. Also, the use of *H. polymorpha* as a biocatalyst [54] will be extended in the near future in particular when the option to anchor factors involved in the activation of complex heterologous proteins at the periplasmic space or outer cell wall will be realized. This strategy allows the creation of regeneratable systems in continuous cultures.

However, the real challenge for the next coming years is to further develop the high potential of the organism for the production of various membrane proteins (including human). The need for such a system is obvious and would strongly stimulate the functional and structural analyses of this important class of proteins.

Acknowledgments

R.v.D. and I.J.v.d.K. are supported by a grant of the Netherlands Technology Foundation (STW), which is subsidized by the Netherlands Organization for Scientific Research (NWO). K.N.F. is supported by a PULS grant from NWO.

References


