New tools for heterologous protein production in the yeast Yarrowia lipolytica

Catherine Madzak

▶ **To cite this version:**

Catherine Madzak. New tools for heterologous protein production in the yeast Yarrowia lipolytica. Recent Research Developments in Microbiology, volume 7, Research Signpost, 2003. hal-02831211

HAL Id: hal-02831211
https://hal.inrae.fr/hal-02831211

Submitted on 7 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Research Signpost
37/661 (2), Fort P.O., Trivandrum-695 023, Kerala, India

Recent Res. Devel. Microbiology, 7(2003): 453-479 ISBN: 81-271-0008-0

29 New tools for heterologous protein production in the yeast *Yarrowia lipolytica*

Catherine Madzak
Laboratoire de Microbiologie et Génétique Moléculaire (LMGM)
INRA / CNRS / INA P-G, Centre de Biotechnologie Agro-Industrielle
(CBAI), BP01, 78850 Thiverval-Grignon, France

1. Abstract

The production of heterologous proteins is a developing area, in which yeasts offer a number of advantages as host systems. Among them, the "non-conventional" yeast Yarrowia lipolytica appears as one of the most attractive, in terms of secretion efficiency and performance regularity. This review presents the main characteristics of Y. lipolytica, and the genetic and molecular tools available in this yeast. It focuses on the recent development of new tools: (i) a recombinant promoter, hp4d, able to drive a strong expression in virtually any medium, (ii) an "auto-cloning" process leading to producing strains devoid of bacterial DNA, (iii) an efficient amplification system using a defective selection marker,and (iv) a "zeta-targeting" process providing a non-homologous

Correspondence/Reprint request: Dr. Catherine Madzak, Laboratoire de Microbiologie et Génétique Moléculaire (LMGM) INRA / CNRS / INA P-G, Centre de Biotechnologie Agro-Industrielle (CBAI), BP01, 78850 Thiverval-Grignon, France
E-mail: madzak@grignon.inra.fr

method of transformation for Y. lipolytica. On the basis of these new tools, our laboratory designed an integrated system for expression/secretion in Y. lipolytica, comprising different kinds of shuttle vectors, and the corresponding recipient strains. This system can fit various purposes, from the small-scale laboratory production of a protein for its analysis, to its high-scale industrial production. The recent examples of heterologous protein production in Y. lipolytica, using the newly developed expression/secretion system, are reviewed and discussed.

Introduction

Heterologous protein production consists in using an easy-to-handle organism (the host) to produce a protein of interest from another (generally more complex) organism. Since several decades, this research area has arisen high interest, for both academic and commercial purposes. Recombinant proteins are implicated in a variety of biomedical applications (i. e. interferon, insulin, interleukins or viruses antigens) or of industrial processes (i. e. lipases for oleochemistry, laccases for bleaching, or chymosin for cheese-making). Historically, the first host system used for the industrial production of heterologous proteins has been the well-studied eubacterium *Escherichia coli.* The choice of this model organism allowed to beneficiate from the high amount of information and know-how accumulated on its genetics and biochemistry. But prokaryotic hosts, albeit easy to handle, present some drawbacks: unability to perform post-translational processing, folding problems and agglomeration of the proteins into insoluble complexes, etc. Thus, as more complex proteins were needed to be expressed, the demand for new host systems increased. The development of transformation systems for many new organisms increased greatly the range of interesting hosts. Among the host systems now currently used are prokaryotes (*E. coli*), several yeasts (i. e. *Saccharomyces cerevisiae* and *Pichia pastoris*), moss protoplasts, insect cells (baculoviral expression system), mammalian cells (i. e. episomal adenoviral and integrated retroviral expression systems), and even transgenic animals or plants (i. e. production in goat milk or in tobacco leaves).

In the developing area of heterologous production, yeasts offer a number of advantages as host systems, especially for complex proteins. As unicellular organisms, they present two advantages: ease of manipulation and capacity of growth. But, in contrast to bacteria, they also possess an eukaryotic subcellular organisation able to perform the post-translational processing of complex (i. e. mammalian) proteins. Yeasts secrete proteins via a multi-component apparatus, allowing their proteolytic maturation, a correct folding and disulfide bonds formation, their N- and O-linked glycosylation, and other post-translational modifications. As a "yeast host" alternative to *E. coli*, *Saccharomyces cerevisiae* was the first historical choice, mainly because of the large amount of knowledge accumulated on this model organism. *S. cerevisiae* also presented a safety guaranty, considering its millenial use in baking and brewing, and, more recently, the GRAS (Generally Regarded As Safe) classification of industrial processes involving this yeast. However, as a host, *S. cerevisiae* was found to be limited in several aspects: a low product yield, a poor plasmid stability, some difficulties in scaling-up production, the hyperglycosylation of the recombinant proteins, and a low secretion capacity. Since several years, heterologous production using *S. cerevisiae* has however much progressed, especially by the design of improved strains overexpressing

chaperones and foldases. Meanwhile, some other laboratories choosed to develop alternative yeast systems.

The non-*Saccharomyces* yeasts tested for heterologous production comprise the well-studied fission yeast *Schizosaccharomyces pombe*, and a heterogenous group, termed the "non-conventional" yeasts. This group include the methylotrophs *Pichia pastoris* and *Hansenula polymorpha*, the dairy yeast *Kluyveromyces lactis*, the amylolytic yeast *Schwanniomyces occidentalis* and the alkane-utilizer *Yarrowia lipolytica*. The development of these alternative host systems was greatly accelerated by the application of the technical know-how established on *S. cerevisiae*. These new hosts exhibit better production yields and secretion efficiencies than *S. cerevisiae*, especially with high molecular weight proteins (reviewed in [1, 2]). They also generally produce less-hyperglycosylated recombinant proteins. Since several years, the successful *P. pastoris* expression system, developed by Invitrogen (USA), allowed the efficient production of more than 400 heterologous proteins (reviewed in [3, 4, 5]). However, this system still sometimes encounters some problems (lower secretion level, hyperglycosylation [5]), leaving place to further improvements or to competitors.

In order to allow a reliable comparison between different yeast host systems, Muller *et al.* [6] tested *S. cerevisiae*, *H. polymorpha*, *K. lactis*, *S. pombe*, and *Y. lipolytica* for their capacity to secrete active forms of six fungal genes. All the alternative yeasts tested were found to be more efficient than *S. cerevisiae*. However, their relative performances varied significantly with the different heterologous proteins. Although not always the best choice for a given protein, *Y. lipolytica* however exhibited the most reproducible results for the whole series of proteins. Therefore, this study defined *Y. lipolytica* as the most attractive alternative host for heterologous production.

The non-conventional yeast *Yarrowia lipolytica*

Historical survey and main characteristics

The hemiascomycetous yeast *Y. lipolytica* has been formely known as *Candida*, *Endomycopsis* or *Saccharomycopsis lipolytica*. It was first classified in the *Candida* genus through ignorance of any sexual state. The perfect form of this yeast was identified later on, in the late 1960s [7]. *Y. lipolytica* is the only known species in its genus. It is heterothallic, with two mating types A and B. In the early 1940s, this yeast was noticed for its uncommon characteristics: it was isolated mostly from lipid- or protein-containing substrates (dairy products or delicatessen), and rarely from sugar-containing substrates. Indeed, *Y. lipolytica* can metabolize only a few sugars, and mainly glucose (but not sucrose, for wild-type strains). In contrast, It can assimilate alcohols, acetate and hydrophobic substrates, such as alkanes, fatty acids and oils (reviewed in [8]). In the late 1940s, *Y. lipolytica* cultures were shown to exhibit high extracellular protease and lipase activities. As we will see, the corresponding secreted proteins (especially the alkaline extracellular protease: AEP) have been of great importance, since then, in the development of the *Y. lipolytica* technology. In the mid 1960s, *Y. lipolytica* has arisen a strong industrial interest: it was able to grow on n-paraffins, which were then a cheap and abundant substrate. This yeast was applied to the production of single-cell proteins, an important objective in these years. *Y. lipolytica* was also noticed for its capacity to produce high amounts of organic acids (i. e. 2-ketoglutaric and citric acids [9]). Thus, the large-scale utilisation of *Y. lipolytica*, for citric acid or single-cell proteins

production, allowed the accumulation of a large amount of data on its cultivation under industrial conditions. In the 1970s and 1980s, several research groups, from academic or industrial laboratories, devoted themselves to the genetics, biology and biochemistry of *Y. lipolytica* (reviewed in [10]). Thus, an important amount of data began to accumulate, and an efficient integrative transformation system became available in the mid 1980s [11, 12].

The first attempts to express heterologous proteins in *Y. lipolytica* were successful, with an efficient secretion in the culture medium (see [13] for an early review). It was observed that an efficient expression of a heterologous gene in *Y. lipolytica*, like in other yeasts, was obtained only when it was inserted between homologous promoter and terminator sequences [14]. The processing pathway of the abundantly secreted AEP was described in details [15], and *Y. lipolytica* was found to possess a signal recognition particle ressembling those of higher eucaryotes [16]. These results arouse a new interest for this yeast, which became a model organism for studies of protein secretion.

No natural episomal DNA has ever been detected in *Y. lipolytica*. For a long time, it has also been impossible to construct replicative vectors. This problem was finally solved by the discovery of centromeric ARSs [17]. Indeed, replicative and centromeric functions cannot be separated in *Y. lipolytica*, and the replicative vectors containing the isolated ARSs behave like mini-chromosomes. They can be maintained at only one to three copies per cell [18, 19]. Other important progresses in *Yarrowia* research were made in the 1990s. Since then, as we will see, they have inspired the design of new tools for heterologous production: the development of a powerful gene amplification system [20], the identification of a retrotransposon [21], and the characterization of strong promoters [22, 23].

Other aspects of *Y. lipolytica* biology are now being explored, such as the metabolism of alkanes and fatty acids, the peroxisome biogenesis, and the control of the dimorphic transition. Indeed, *Y. lipolytica* is a dimorphic yeast, able to form either yeast cells or hyphae and pseudohyphae, depending on growth conditions (aeration, carbon and nitrogen sources, pH, etc). Thus, in many fields, *Y. lipolytica* constitute an interesting model organism. It also appears to be a very peculiar yeast. Many observations show that it diverges greatly from other ascomycetous yeasts: its high GC content, the unusual structure of its rDNA genes, its peculiar codon bias, the low level of similarity of its genes with their counterparts from other yeasts (typically 50-60% at amino acid level), etc. On evolutionary trees (based on sequences of genes encoding "house-keeping" functions: glycolyse, rDNA), *Y. lipolytica* appears on an isolated branch, clearly separated from *S. pombe* on one hand, and from the group of other ascomycetous yeasts on the other [24, 25]. The sequencing of *Y. lipolytica* genome, now in progress, will undoubtly clarify this point. Partial sequencing results, obtained during the "Génolevures I" project, have been made available recently [26].

Y. lipolytica present also the interesting characteristic to be considered as non-pathogenic. No pathology has ever been linked to this yeast, despite its frequent presence in food. It does not endure temperatures higher than 32 to 34°C, and, in contrast to several yeasts from *Candida* genus, it was shown to be inocuous even following intravenous injection [27]. Moreover, several industrial processes implicating *Y. lipolytica* have been classified as GRAS by the FDA (Food and Drug Administration).

Secretion of proteins

Since the 1940s, *Y. lipolytica* has been distinguished for its capacity to secrete naturally several proteins into the culture medium. These extracellular proteins include two proteases (the AEP in neutralkaline conditions, and the AXP in acidic conditions), several lipases and phosphatases, a RNase and an esterase (reviewed in [8]). Under convenient inducing conditions, *Y. lipolytica* is able to secrete very large amounts of AEP: 1 to 2 g per litre [28, 29]. Indeed, this protein became a model in secretion research, and its gene (*XPR2*) a useful molecular tool. Thus, the *XPR2* promoter was of great historical importance in the development of heterologous production in *Y. lipolytica.*

At the early steps of its secretion, a nascent protein can follow two pathways: the co-translational or the post-translational translocation from the cytoplasm to the endoplasmic reticulum. The co-translational pathway was shown to be predominant in *Y. lipolytica* [30], as in higher eukaryotes (i. e. in mammalian cells). This characteristic constitutes an important advantage of *Y. lipolytica* over *S. cerevisiae*, in which the post-translational pathway is predominant, for the efficient production of complex heterologous proteins.

Its natural secretion capacities, together with some peculiar characteristics ressembling those of higher eukaryotes (signal recognition particle, co-translational pathway) render *Y. lipolytica* particularly attractive for heterologous production. Twelve years ago, Buckholz and Gleeson [1] had listed a total of five heterologous proteins that had been produced using *Y. lipolytica.* This number is now higher than forty (to be reviewed elsewhere), which shows the recent development of this yeast as an alternative host organism.

Tools for heterologous production in *Y. Lipolytica*

Host strains

The genetic background of the recipient strain seems to be an important parameter governing heterologous production, as examplified by the work of De Baetselier *et al.* [31] on *S. cerevisiae*. However, despite its potential importance, the role of the genetic background is often neglected, since variations do not correlate with known genetic markers ("black box" effect). Among *Y. lipolytica* laboratory strains, some were selected more or less empirically for their capacity to secrete homologous proteins. They became the best candidates for assays of heterologous proteins production. One of the most frequently used recipient strain, Po1d [20], presents several interesting features: (i) its genetic background, enabling a high level of secretion, (ii) the deletion of the AEP, a potent threat for secreted heterologous proteins, and (iii) its production of recombinant invertase, allowing the utilization of sucrose as a new carbon source [32]. This last point is particularly relevant for industrial applications, since invertase-positive *Y. lipolytica* strains can grow efficiently on molasses, a cheap and abundant substrate [33]). As we will see, our laboratory recently developed a series of derivatives of Po1d strain that were further improved for heterologous protein production.

Expression vectors and transformation of *Y. lipolytica*

Shuttle vectors used for heterologous production in *Y. lipolytica* typically contain a bacterial moiety (for isolation and amplification in *E. coli*) and, for expression *per se*, a selection marker, the expression cassette and, possibly, additional elements for

maintenance into yeast cells. In the case of classical integrative vectors, these elements are sequences homologous to the genomic DNA from the recipient strain (this homology can be part of the selection marker gene). In the case of replicative vectors, a centromeric ARS is present. The shuttle vectors can be introduced into the host strains using either the lithium acetate method [34] in the case of integrative vectors, or electroporation [35] in the case of replicative ones.

The unusual properties of *Y. lipolytica* ARS elements, due to the colocalization of centromeric and replicative functions, render them not very attractive for vector design. Thus, in *Y. lipolytica*, replicative vectors allow only a limited amplification of the copy number [18, 19], and of the correlated gene expression [36, 19]. In addition, they require the maintenance of a selective pressure, which may not be always compatible with efficient production conditions and/or industrial management. Because of these drawbacks, the use of integrative vectors remains the best choice for heterologous expression in *Y. lipolytica*.

The integration of exogenous DNA into *Y. lipolytica* genome occurs almost exclusively by homologous recombination. The integration of a plasmid was shown to be greatly stimulated by linearization within the homology region, resulting in very high transformation frequencies (up to 10^6 transformants per μg of DNA [34]). When the homologous region is large enough (a few hundred bp), a single complete copy of the vector is integrated at the chosen site in more than 80% of the cases (Barth and Gaillardin, 1996), while the remaining events include multiple tandem integrations, gene conversions, and out-of-site integrations. The replacement of a chromosomal sequence using integration by double crossover can also be obtained, with a tenfold less efficiency [37]. In *Y. lipolytica*, integrated vectors exhibit a very high stability: Hamsa and Chattoo [38] showed that they were retained without any rearrangement after hundred generations under non-selective conditions.

Homologous multiple integrations

In addition to their advantages in term of easy handling and stability, the integrative vectors offer the possibility of multiple integrations, with correlatedly increased gene expression. The first strategy used to increase the copy number in *Y. lipolytica* was to target integration into the ribosomal DNA (rDNA) cluster, and to select for multiple integrants using a defective marker [39, 20]. The number of potential target integration sites (rDNA units) was estimated to be at least two hundred copies per genome [40]. In our laboratory, Le Dall *et al.* [20], selected for multiple integrations in the rDNA using the defective *ura3d4* allele as a selection marker (see hereafter). The *XPR2* gene was used as a reporter gene, in order to quantify expression from different integrants. Transformants carrying up to sixty integrated copies of the vector were obtained. These multiple integrations were found mostly as tandem repeats, and also as dispersed copies, on one or two chromosomes. These high-copy-number integrants were fairly stable under non-inducing conditions. However, when the expression of the *XPR2* gene was induced, deamplification occurred during the cultivation, and the copy number eventually stabilized around ten. The amount of secreted AEP activity increased linearly with gene dosage, up to around ten copies, but a higher overproduction seemed to be deleterious. The stability of the high-copy-number integrants is likely to depend on the nature of the gene inserted into the rDNA-based vector.

Surprisingly, it was afterwards observed that, even with a single integration site, multiple integrations can be selected for. Vectors carrying the defective *ura3d4* allele were integrated either at the multiple rDNA or Ylt1 retrotransposon sites, or at the single *XPR2* locus. Multicopy transformants were obtained in all cases, regardless of the target used for integration [41]. This unexpected result indicates the determining role of the defective selection marker in the process. As observed previously [20], multiple integrations occurred almost always in tandemly-repeated copies at one or two sites, and the copy numbers tended to stabilize around ten copies during cultivation, by an amplification/deamplification process [41]. This copy number probably reflects the optimum auxotrophy complementation driven by the *ura3d4* allele.

Selection markers

Although several attempts have been made, no efficient dominant selection system was developed in *Y. lipolytica* (reviewed in [37]). The use of auxotrophy markers remains the best choice. The most commonly used are *LEU2* and *URA3*, especially since non-leaky non-reverting *leu2* and *ura3* recipient strains were constructed by the way of internal deletions (i. e. in the Po1d strain).

In order to select for multiple vector integrations, several defective versions of the *URA3* marker were designed in our laboratory [20]. Sequential deletions were performed in the *URA3* promoter, and the corresponding constructions tested for their capacity to transform *Y. lipolytica*. The *ura3d4* allele, retaining only 6 bp upstream from the ATG sequence, was no more able to confer a Ura$^+$ phenotype as a single copy, but was able to promote the amplification of the vector copy number in a multiple integration experiment [20]. Surprisingly, another allele, *ura3d1*, retaining 41 bp upstream from the ATG sequence, was still able to confer a Ura$^+$ phenotype as a single copy: it constitutes a minimal *URA3* selection marker.

Promoters

Heterologous proteins production requires the use of strong promoters. Historically, the promoter from the *XPR2* gene has arisen a strong interest since the corresponding protein, AEP, was abundantly expressed. The *XPR2* promoter (p*XPR2*) has been extensively studied [23, 19], and used for heterologous production [42, 32, 36, 38, 6]. However, its complex regulation hindered its industrial use. This strong promoter is active only at pH above 6, on media lacking preferred carbon and nitrogen sources, and its full induction requires high levels of peptones in the culture medium [43]. Thus, the search for new promoters, more compatible with industrial production, was undertaken, and is still important in *Y. lipolytica* research.

Muller *et al.* [6] described two strong constitutive promoters, isolated from *TEF* and *RPS7* genes. They were intended to be used for isolation of enzyme genes by expression cloning, and not for heterologous production *per se*. Actually, fully constitutive promoters are not recommanded for that purpose since the early expression of the heterologous protein can be detrimentous to the culture growth. For this reason, several laboratories searched for inducible promoters. Dominguez *et al.* [2] described the bi-directional metallothionein promoter, but the use of metallic salts as inducer would impair its industrial use.

Secretion signals

In order to secrete a heterologous protein in the culture medium, a secretion signal is required to target it to the secretion pathway. Both homologous and heterologous (native) secretion signals have been used successfully in *Y. lipolytica*. Among the native secretion signals that had proved efficient for the secretion of the cognate heterologous protein, were mainly signals from several fungal enzymes. Thus, good results were observed using native signals from *Aspergillus aculeatus*, *Humicola insolens* and *Thermomyces lanuginosus*, [6], from *Trichoderma reesei* [44], from *Alternaria alternata*, [45], and from *Trametes versicolor* [46]. However, a native secretion signal from another filamentous fungus, *Pycnoporus cinnabarinus*, did not gave satisfying results [47]. Filamentous fungi remain relatively close to *Y. lipolytica* in terms of phylogeny. But native secretion signals from more remote origin can also be used in *Y. lipolytica*, as shown by an example from a higher plant: the efficient secretion of the rice α-amylase using its native signal [48]. Despite these examples, the use of *Y. lipolytica* secretion signals is in general considered as more reliable, especially for mammalian proteins.

The *XPR2* prepro region has been by far the most widely used signal sequence. This sequence was shown to target the early steps of protein secretion to the co-translational pathway of translocation [49, 50]. The *XPR2* pro region itself is required for the transit of the AEP: it acts as an internal chaperone, and allow the mature protein to adopt a conformation compatible with its secretion [51, 52]. Its use in heterologous production allows the conservation of the initiator ATG codon environment. This is intended to "adapt" the chimaeric protein to the co-translational translocation. The removal of the pro region is thought to occur in the late Golgi apparatus (reviewed in [53]). The *XPR2* pre region alone (with or without the following dipeptide stretch) was also shown to be sufficient to drive efficient heterologous secretion [14, 54].

New tools for a yarrowia expression/secretion system

New inducible promoters

In its search for new inducible promoters for heterologous production, our laboratory focused on *Y. lipolytica* capacity to grow on hydrophobic substrates. Several promoters of key enzymes from this pathway were isolated and studied [55]. The promoters from *ICL1*, *POT1*, and *POX2* genes were found to be the strongest inducible promoters available for *Y. lipolytica*. All of them are highly inducible by fatty acids and alkanes, the last two are totally repressed by glucose and glycerol, and the first one is also strongly inducible by ethanol and acetate, but is not completely repressed by glucose and glycerol. However, despite their high efficiency, the industrial use of these new promoters could encounter some problems. For *POT1* and *POX2* promoters, the peculiar nature of the inducers (fatty acids and alkanes) may not be always compatible with efficient heterologous production or protein purification. For *ICL1* promoter, a complete repression is not possible, which limits its interest for the production of toxic proteins. Thus, in *Y. lipolytica*, the "perfect" inducible promoter remains to be found.

A new recombinant promoter: Hp4d

During the functional dissection of the p*XPR2* performed in our laboratory [23, 56], one upstream activating sequence (the UAS1) was found to be poorly affected by

environmental conditions. This element was used to design a hybrid promoter, composed of four tandem UAS1 copies inserted upstream from a minimal *LEU2* promoter (reduced to its TATA box). In contrast to the native p*XPR2*, this recombinant promoter, termed hp4d, is almost independent from environmental conditions (pH, carbon and nitrogen sources, peptones). Hp4d is able to drive a strong expression in virtually any medium [19]. However, this recombinant promoter is not, properly speaking, constitutive: it retained some unidentified elements that drive a growth-phase dependent gene expression. Heterologous gene expression driven by hp4d was found to occur essentially at the beginning of the stationary phase [19, 57]. This characteristic can be particularly interesting for heterologous production: it allows a partial dissociation of the growth and expression phases, which can be highly desirable in order to avoid toxicity problems.

New secretion signals

The cloning and characterization of the *Y. lipolytica* extracellular lipase gene *LIP2*, performed in our laboratory, revealed the presence of a prepro region [58]. This *LIP2* prepro region has an important potential interest as an alternative secretion signal for heterologous production in *Y. lipolytica.* An hybrid signal, consisting in a recombination between *XPR2* and *LIP2* prepro regions, has also recently been used to secrete efficiently several heterologous proteins ([57, 59, 60] and C. Gysler, personal communication). Some interesting production results have also been obtained using a modified *XPR2* prepro sequence, suggesting that this secretion signal could be improved by mutagenesis (N. Libessart, personal communication).

Zeta sequences and non-homologous integration in *Y. lipolytica*

The first retrotransposon described in *Y. lipolytica*, Ylt1 [21], belongs to the Ty3-gypsy group. Ylt1 is characterized by its very large (714 bp) long terminal repeats (LTRs) termed "zeta". This retrotransposon was found in the wild American isolate YB423, and in all derivative strains (inbreeding programs reviewed in [37]). In contrast, Ylt1 was not detected in the wild German isolate H222, nor in the wild French isolate W29 and its derivative Po1d [26, 41]. In Ylt1-carrying strains, this retrotransposon was shown to be present in at least thirty-five copies per genome, in a dispersed manner. In addition, the zeta LTRs were also found as solo elements, with thirty to sixty copies per genome, depending on the strain [21, 41]. Thus, the zeta sequence provided at least hundred potential targeting sites per genome of Ylt1-carrying strain. This element constitute an interesting alternative to the rDNA target sites, for the homologous integration of multicopy vectors. Zeta-targeting multicopy vectors have been used successfully to overexpress *Y. lipolytica* extracellular lipase [58] and to express heterologous genes [41]. The lipase-overexpressing transformants, carrying around ten copies of the vector integrated at zeta loci, were found to be stable after two weeks of exponential growth (namely 120 generations), under both non-inducing or lipase-inducing conditions [58].

Surprisingly, the zeta-targeting vectors were also found able to integrate into the genome of *Y. lipolytica* strains devoid of Ylt1 retrotransposon, such as Po1d strain [61, 58, 41]. Interestingly, these events were not "ordinary" non-homologous integrations, since the comparison of multicopy vectors with or without zeta sequences showed that the former exhibit a higher transformation efficiency, and a higher copy number [61, 58].

Thus, the zeta-targeting vectors provide a non-homologous process of transformation for Ylt1-free *Y. lipolytica* strains. This method can be applied particularly to the Po1d strain and its improved derivatives (see hereafter).

Auto-cloning vectors

The shuttle vectors used for heterologous production contain a bacterial moiety, consisting basically in a bacterial replication origin and a marker gene for selection in *E. coli* (resistance to ampicillin, tetracyclin or kanamycin). In classical yeast transformation experiments, the shuttle vectors are linearized, and integrate entirely into the genome of the recipient strain. This process results in the presence of bacterial DNA into the producing yeast strain. This is a drawback for commercial applications, since current European regulations classify this kind of constructions as genetically modified organisms (OGMs). To alleviate this problem, our laboratory developed "auto-cloning" expression vectors, from which the bacterial moiety could be removed prior to transformation, leading to producing strains devoid of bacterial DNA [61, 58]. The bacterial moiety, on one hand, and the rest of the vector ("yeast cassette"), on the other, can be separated using a restriction digestion followed by an agarose gel electrophoresis. The purified "yeast cassette" alone is used for the transformation of the recipient strain. The term "auto-cloning" is to mean that the *Y. lipolytica* recipient strain is transformed only with DNA from the same species (plus the heterologous gene of interest).

This new vector design can be combined with the principle of zeta-targeting: our laboratory developed auto-cloning expression vectors in which the "yeast cassette" is bordered by zeta sequences, which promote its non-homologous integration in the recipient strain. Additionally, the use of a defective selection marker can allow the amplification of the "yeast cassette" copy number. Such auto-cloning multicopy vectors have already been used for the production of homologous [58] or heterologous proteins [57, 59].

An integrated yarrowia expression/secretion system

On the basis of the new tools described hereabove, our laboratory designed an integrated system for expression/secretion in *Y. lipolytica*, comprising different kinds of shuttle vectors, and the corresponding recipient strains. This system was intended to fit various purposes, from the small-scale laboratory production of a protein for its analysis, to its high-scale industrial production. All proposed expression/secretion vectors are of the integrative type, and contain the recombinant hp4d promoter. As discussed hereabove, this promoter exhibits interesting characteristics: high efficiency, versatility for the culture conditions, and growth-phase-dependent expression, allowing a partial dissociation of growth and production phases. As we will see, the positive results already obtained using hp4d promoter, with various heterologous proteins, show that its non-inducibility should not be a problem, except perhaps with really toxic proteins.

We designed three series of expression/secretion shuttle vectors: (i) monocopy vectors, (ii) auto-cloning vectors, and (iii) multicopy auto-cloning vectors. They are presented in the Table 1, and their maps are shown in the Figures 1 to 3. The characteristics of each type of vector will be described in the following chapters. In each series are included expression vectors, without any secretion signal (possibly allowing the use of a native one), and expression/secretion vectors, carrying various *Y. lipolytica*

Table 1. Vectors for expression/secretion in *Y. lipolytica* using the hp4d promoter

Vector	Secretion signal	Cloning sites: upstream / downstream	Marker	Reference
pBR-based mono-integrative vectors				
pINA1269	none	*Pml*I (blunt) / *Bam*HI, *Kpn*I	*LEU2*	[19]
pINA1296	*XPR2* pre	*Sfi*I (in pre) / *Hin*dIII, *Xba*I, *Kpn*I	*LEU2*	[19]
pINA1267	*XPR2* prepro	*Sfi*I (in pro) / *Xma*I (*Sma*I), *Xba*I, *Kpn*I	*LEU2*	[19]
monocopy auto-cloning vectors, with a fully functional selection marker gene				
pINA1311	none	*Pml*I (blunt) / *Bam*HI, *Kpn*I, *Avr*II	*ura3d1*	[57]
pINA1312	none	*Pml*I (blunt) / *Bam*HI, *Kpn*I	*ura3d1*	[57]
pINA1313	*LIP2* prepro	*Xmn*I (in pro) / *Bam*HI, *Kpn*I, *Avr*II	*ura3d1*	[57]
pINA1317	*XPR2* pre	*Sfi*I (in pre) / *Avr*II, *Bam*HI, *Kpn*I	*ura3d1*	[57]
pINA1314	*XPR2* prepro	*Sfi*I (in pro) / *Kpn*I	*ura3d1*	[57]
multicopy auto-cloning vectors, with a defective selection marker gene (copy number amplification)				
pINA1291	none	*Pml*I (blunt) / *Bam*HI, *Kpn*I, *Avr*II	*ura3d4*	[57]
pINA1292	none	*Pml*I (blunt) / *Bam*HI, *Kpn*I	*ura3d4*	[57]
pINA1293	*LIP2* prepro	*Xmn*I (in pro) / *Bam*HI, *Kpn*I, *Avr*II	*ura3d4*	[57]
pINA1297	*XPR2* pre	*Sfi*I (in pre) / *Avr*II, *Bam*HI, *Kpn*I	*ura3d4*	[57]
pINA1294	*XPR2* prepro	*Sfi*I (in pro) / *Kpn*I	*ura3d4*	[57]

Plasmids pINA1269, 1296 and 1267 are pBR322-based monocopy integrative vectors carrying the *LEU2* selection marker gene. They can be targeted to the docking platform (integrated pBR322 sequence) of Po1e or Po1g recipient strains, after *Not*I digestion (linearization in the pBR region). The other plasmids are autocloning plasmids, from which a "yeast cassette" devoid of bacterial sequences can be liberated by *Not*I digestion prior to transformation. They are either monocopy or multicopy integrative vectors, respectively carrying the fully functional *ura3d1* allele, or the defective *ura3d4* allele, from the *URA3* selection marker gene. The "yeast cassette" can integrate at random in ylT1-free Po1d, Po1f or Po1h recipient strains.

secretion signals. All vectors carry a slightly modified version of the promoter or secretion signal, introducing a new unique cloning site (upstream site) for inserting the gene of interest. Downstream cloning sites are also available, in order to allow a directed insertion. In general, the gene of interest is amplified in a PCR experiment, using oligonucleotides providing the desired restriction sites. The upstream oligonucleotide must also reconstitute the correct sequence of the promoter or secretion signal, just upstream from that of the gene of interest. Thus, the restriction and insertion will result

in a perfect fusion between the promoter or secretion signal and the gene of interest. This vector design was intended to ensure the production of a protein of interest exactly identical in sequence to the native one (provided a correct maturation, when a secretion signal is present), since no nucleotide is added nor modified for the purpose of construction. This point appeared important to us, for both academic and industrial applications.

The choice of the type of vector to be used will depend on each purpose, since they differ in terms of handling and efficiency. Briefly, the easy-to-handle monocopy vectors are well adapted to the small-scale laboratory production of a protein for its analysis, and the multicopy auto-cloning vectors (requiring more technical investment) to its high-scale industrial production.

New expression/secretion vectors (I): Monocopy vectors

A general scheme of these shuttle vectors is presented in the Figure 1. The bacterial moeity is derived from the plasmid pBR322, and comprises a bacterial origin of replication (ori) and the *bla* gene, conferring resistance to ampicilline (Amp^R). The "yeast cassette" comprises the *LEU2* selection marker gene and the expression cassette, composed of (i) the hp4d promoter, (ii) an optional secretion signal, (iii) a "multiple cloning site" (MCS) and (iv) the *XPR2* terminator. Three vectors are available: pINA1269, containing no secretion signal, pINA1296, carrying the *XPR2* pre sequence as a secretion signal, and pINA1267, carrying the *XPR2* prepro sequence. Each of these vectors is designed to integrate, by homology, as a monocopy into the *Y. lipolytica* genome. However, the accidental tandem integration of two copies was sometimes observed [19]. These pBR322-based vectors can be targeted to a docking platform, composed of an integrated pBR322 sequence, in the genome of Po1e or Po1g recipient strains. This targeting is achieved by linearization of the vectors in the pBR region (i. e. at the *Not*I unique restriction site).

The use of this kind of mono-integrative vectors presents several advantages: (i) the transformation efficiency is very high, (ii) the fact that one copy is integrated at a precise known location allows a comparison between experiments, (iii) it also avoids the interference of the integration with cellular fonctions, and (iv), as the transformants are expected to behave similarly, the number to be analyzed can be greatly reduced. A very high transformation frequency can be obtained in *Y. lipolytica*, with mono-integrative vectors: routinely, using frozen competent yeast cells, 10^3 to 10^4 transformants per μg of DNA are obtained, and efficiencies up to 10^5 to 10^6 transformants per μg of DNA can be attained when using freshly-grown competent yeast cells [34].

New expression/secretion vectors (II): Auto-cloning vectors

A general scheme of these shuttle vectors is presented in the Figures 2 and 3. The bacterial moeity is derived from the plasmid pHSS6, and comprises a bacterial origin of replication (ori) and a gene conferring resistance to kanamycin (Kan^R). The "yeast cassette" is bordered by zeta sequences. It comprises the *URA3* selection marker gene and the expression cassette, composed of (i) the hp4d promoter, (ii) an optional secretion signal, (iii) a "multiple cloning site" (MCS) and (iv) a terminator. These auto-cloning vectors are either monocopy or multicopy integrative vectors, provided they respectively

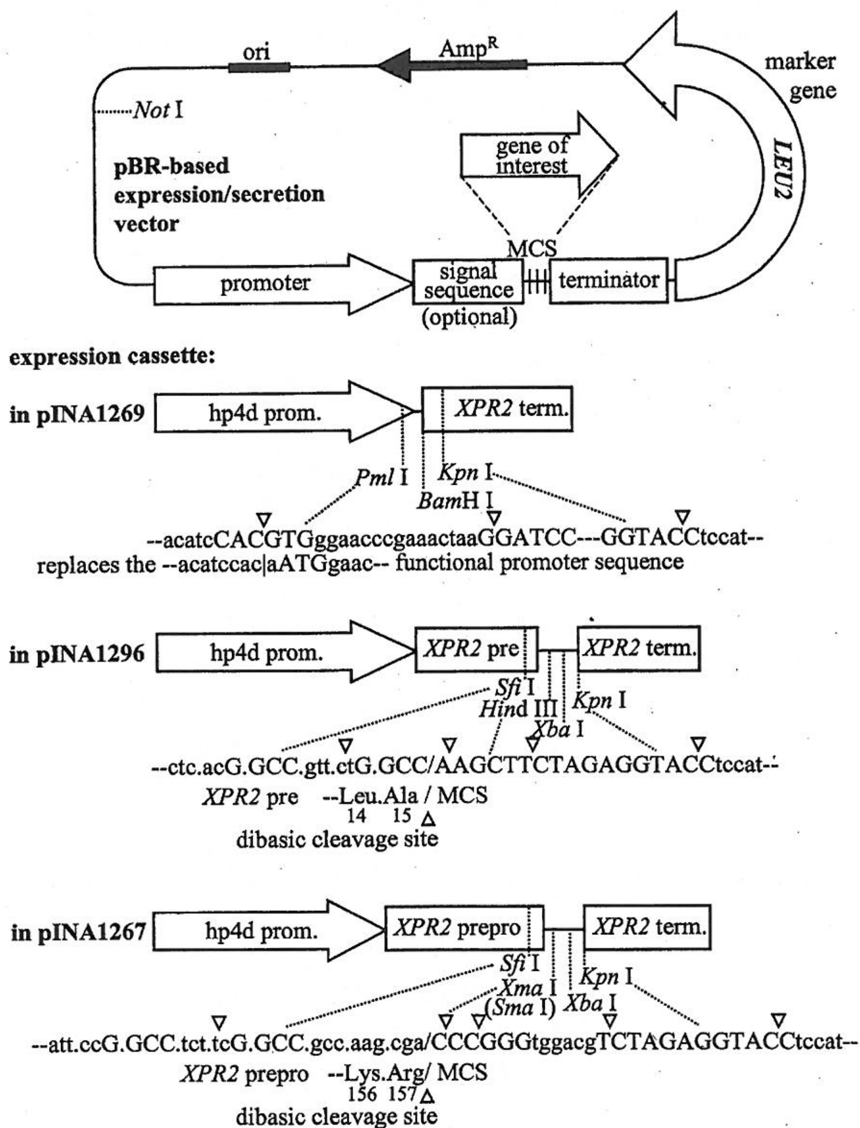

Figure 1. General scheme of pBR-based monocopy vectors. Cf commentary below Figure 3.

carry the fully functional *ura3d1* allele, or the defective *ura3d4* allele, from the *URA3* selection marker gene. Five monocopy auto-cloning vectors are available. Three of them carry the *XPR2* terminator (see Figure 2): pINA1312, containing no secretion signal, pINA1317, carrying the *XPR2* pre sequence as a secretion signal, and pINA1314, carrying the *XPR2* prepro sequence. The two others carry the *LIP2* terminator (see Figure 3): pINA1311, containing no secretion signal, and pINA1313, carrying the *LIP2*

prepro sequence. For each of these autocloning vectors, a "yeast cassette" devoid of bacterial sequences can be liberated by *Not*I digestion prior to transformation. This "yeast cassette", bordered by zeta sequences, is designed to integrate non-homologously into the genome of Ylt1-free *Y. lipolytica* strains, such as Po1d, Po1f and Po1h.

The use of these auto-cloning vectors presents several advantages, especially for industrial applications: (i) the resulting producing strains are devoid of bacterial DNA,

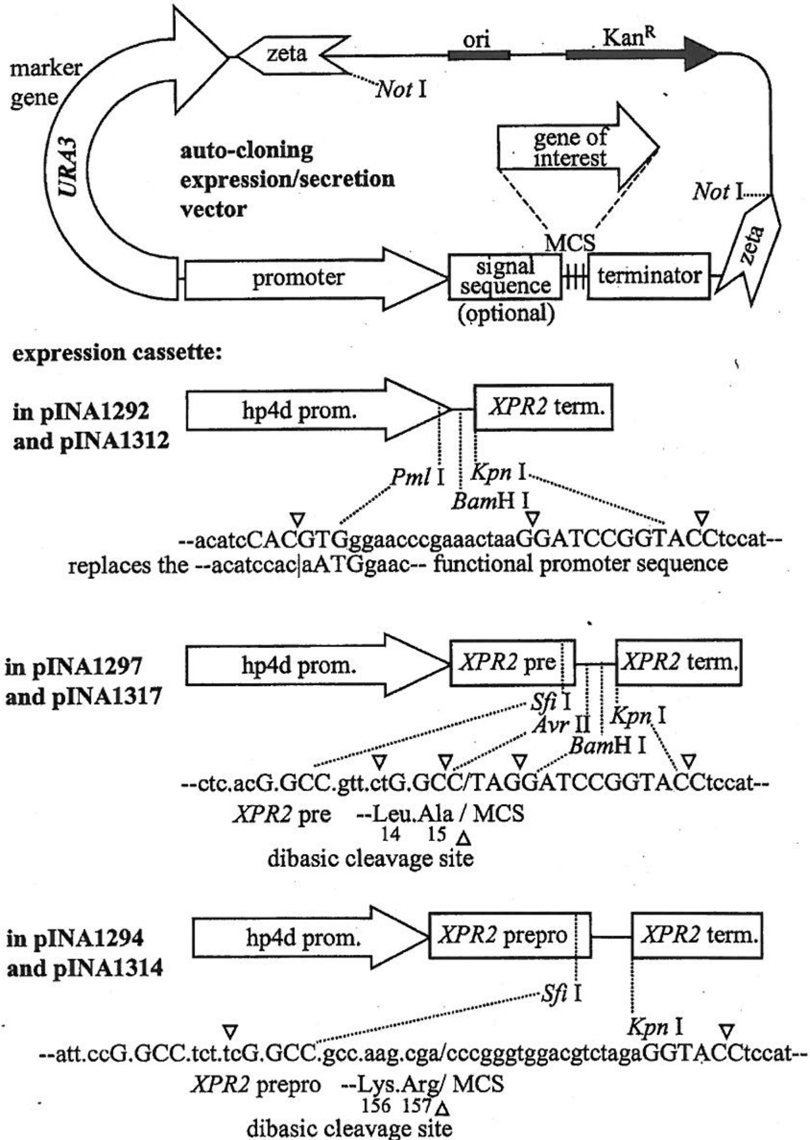

Figure 2. General scheme of auto-cloning vectors (1/2). Cf commentary below Figure 3.

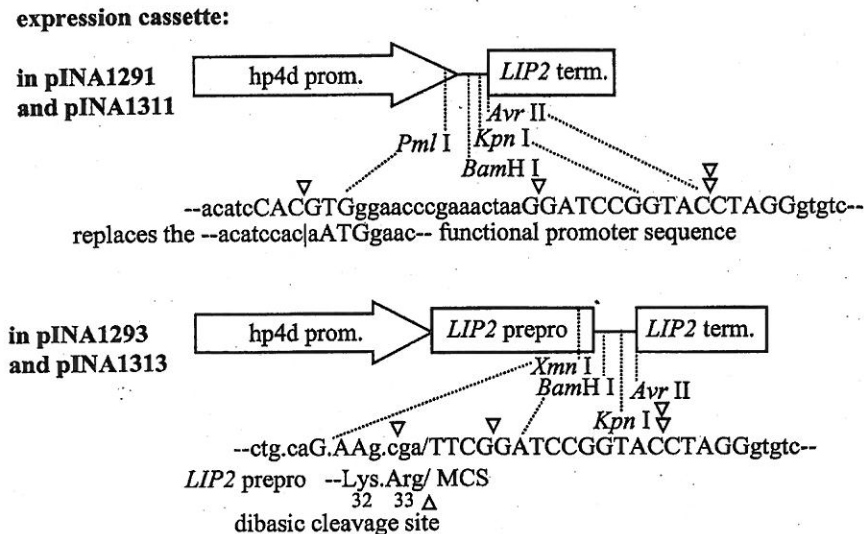

Figure 3. General scheme of auto-cloning vectors (2/2). For each vector type, a general vector map is shown, with the bacterial moiety represented as a thin line with black elements, and the "yeast cassette" represented as empty boxes and arrows. "MCS" indicates a "multiple cloning site" for the insertion of the gene of interest. For each vector, the features of the "expression cassette" are shown below, with the detailed sequence of the MCS (restriction sites indicated in high-casse letters, with cutting sites marked by arrowheads). In the expression vectors, devoid of secretion signal, the downstream extremity of the hp4d promoter was modified to create a blunt-ended *Pml*I upstream cloning site. Consequently, the blunt-ended upstream extremity of the inserted fragment (to be ligated to *Pml*I upstream cloning site) must exhibit the aATG sequence, in order to reconstitute a correct in-frame fusion of the hp4d promoter with the gene sequence. For the expression/secretion vectors, the amino acid sequence of the dibasic cleavage site, at the end of the secretion signal, is indicated. The secretion signals were modified to create *Sfi*I or *Xmn*I upstream cloning sites. Consequently, the *Sfi*I- or *Xmn*I-compatible upstream extremity of the inserted fragment must reconstitute the end of the signal sequence (up to the dibasic cleavage site), in order to obtain a correct in-frame translational fusion of the secretion signal sequence with that of the mature protein.

and (ii) the non-homologous transformation process of Ylt1-free *Y. lipolytica* strains by zeta-based vectors is independent from pre-existing patents. However, the efficiency of transformation of a Ylt1-free strain by a zeta-based vector is lower than with a classical integrative vector, by at least one order of magnitude (unpublished data). The fact that the "yeast cassette" integrates at random into the genome of the recipient strains can also bring interferences with their cellular fonctions: a higher number of transformants needs to be analyzed.

New expression/secretion vectors (III): Multicopy auto-cloning vectors

A general scheme of these shuttle vectors is presented in the Figures 2 and 3. They present all the characteristics described in the previous chapter, except that the use of the

defective *ura3d4* selection marker also allows to select for multiple integration events of the "yeast cassette". Five multicopy auto-cloning vectors are available, which are the exact counterparts of the monocopy ones, except for *ura3d4* replacing *ura3d1*. Three of them carry the *XPR2* terminator (see Figure 2): pINA1292, containing no secretion signal, pINA1297, carrying the *XPR2* pre sequence as a secretion signal, and pINA1294, carrying the *XPR2* prepro sequence. The two others carry the *LIP2* terminator (see Figure 3): pINA1291, containing no secretion signal, and pINA1293, carrying the *LIP2* prepro sequence. These multicopy auto-cloning vectors present the same advantages than the monocopy ones, plus the copy number amplification, which render them particularly interesting for industrial applications. Their use is however less straightforward than that of classical integrative vectors.

In contrast to monocopy pBR-based vectors, which integrate at a precisely identified genomic site, the multicopy vectors generate integrants which can greatly differ from each other in terms of heterologous production, depending on the copy number. These differences between transformants are still greater when using non-homologous integration, since some integration sites can impair cell growth or production. It is therefore desirable, in this case, to obtain and test a much higher number of transformants. As we have seen, the efficiency of transformation of a Ylt1-free strain by a zeta-based vector is lower than with a classical integrative vector, and the use of a defective selection marker still decreases the transformation frequency by several orders of magnitude. The efficiency of transformation of a zeta-based vector carrying a defective selection marker in a Ylt1-free strain is less than 10 transformants per µg [41]. Thus, with multicopy auto-cloning vectors, it is more technically difficult to obtain enough transformants to found at least one of them interesting in terms of production. However, despite this low transformation efficiency, the random transformation of Ylt1-free *Y. lipolytica* strains by zeta-based vectors remains an interesting method, since the non-homologously integrated copies appeared to be highly dispersed in the genome, leading to a better stability of the high-copy-number integrants [61].

New host strains

Our laboratory developed a series of derivatives of Po1d strain, that were specially adapted to the production of heterologous proteins ([19] and unpublished data). These new strains retain the interesting features of Po1d (genetic background enabling a high level of secretion, deletion of the AEP, and metabolization of sucrose) and are further adapted for heterologous production using the newly developed vectors. They are described in the Figure 4, and exhibit the following characteristics: (i) Po1f, Po1g, Po1h and Po1t were further deleted for the acid extracellular protease (AXP), eliminating all sources of secreted proteasic activity; (ii) Po1e and Po1g were fitted with an integrated pBR322 docking platform, allowing the easy further integration of pBR-based vectors; (iii) Po1f and Po1h can integrate non-homologously a "yeast cassette" from an auto-cloning vector, allowing the obtention of producing strains devoid of bacterial DNA; (iv) Po1g and Po1h retained only one auxotrophy, allowing their transformants to be prototrophs for a easier handling; (v) Po1t was rendered prototroph, in order to provide a negative control strain in production assays.

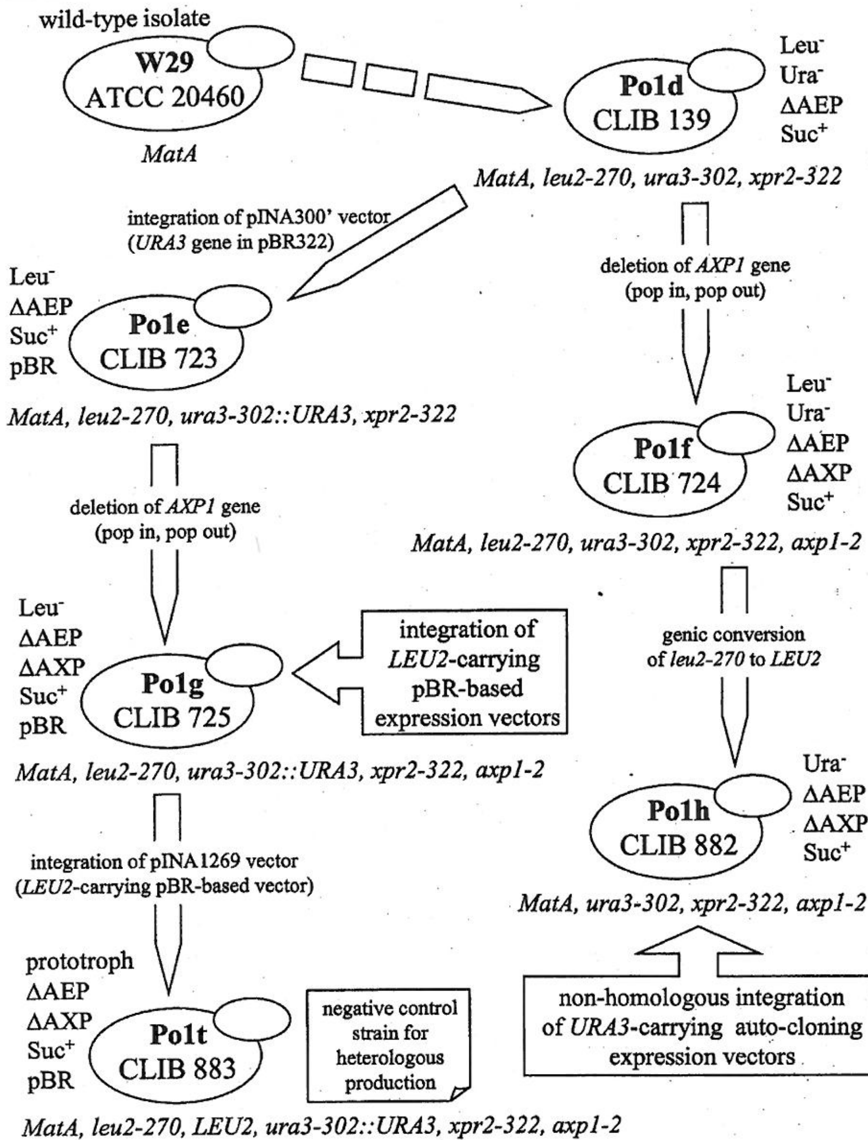

Figure 4. Genealogy of the newly developed *Y. lipolytica* host strains. For each strain, the phenotype, the genotype, and a library reference number is indicated (CLIB: Collection de Levures d'Intérêt Biotechnologique,CBAI, INAPG, 78850 Thiverval-Grignon, France).

Examples of heterologous production

This chapter presents and comments recent examples of heterologous production involving the new tools developed in our laboratory. It focuses on the use of the recombinant hp4d promoter. The Table 2 presents the 18 heterologous proteins produced

until now, in *Y. lipolytica*, using this promoter. In order to make this review as exhaustive as possible, unpublished data was included, in the form of personal communications and abstract reports from recent congresses. The works resumed in the Table 2 all used either the vectors from our integrated expression/secretion system (cf Table 1) and their close relatives or derivatives, or the vector p0 and its derivatives [54], which structure is similar to that of the pBR-based monocopy vectors (cf notes from Table 2). Except for the production of β-galactosidase [19], the *Y. lipolytica* strains used for these works were all from, or derivatives from, the "Po1" series (see Figure 1). The strains Po1f and Po1g [19] have been the most frequently used. With the exception of the two first examples (intracellular productions of Cre recombinase and of β-galactosidase [62, 19]), for which replicative vectors were used, all the other examples concern integrative vectors. A qualitative (+/-) or quantitative (international units of activity, or weight/volume concentration) estimation of the amount of heterologous protein producted is indicated. It is to note that the production was quantified by measuring heterologous activity in the culture supernatant: it thus corresponds to an amount of secreted active heterologous protein.

Hp4d: An efficient and versatile promoter

Although only recently made available [19], the recombinant hp4d promoter has already been abundantly used, since the eighteen proteins cited in the Table 2 represent almost half the total number of proteins ever produced in *Y. lipolytica* (slightly more than forty, with examples dating back from 1987 [64, 42] - to be reviewed elsewhere). The overall results appear clearly positive: all the proteins cited in the Table 2 were expressed, and secreted when a secretion signal was present, under an active form, although with various levels of efficiency. The examples cover a large range of phylogenic origins (from viruses and eubacteria to plants and mammals) and protein sizes (from 12 to 116 kDa). Some concern classical reporter proteins (β-galactosidase, bovine prochymosin) when the others implicate proteins which production is known to present difficulties, such as single-chain antibodies (scFv and VHH) and metalloproteins (laccases).

Several previous works, and especially the comparative study of Muller *et al.* [6] had already established that *Y. lipolytica* was a reliable and versatile host for heterologous production, combining performance and regularity. The more recent works using the new hp4d promoter and expression/secretion system reinforce this conclusion. In most of the examples from the Table 2, heterologous proteins were produced at the mg per litre level (often 5 to 20 mg per litre), in shaked flask. Several examples also show that the scale-up to batch or fed-batch cultivation can allow a ten-fold increase in production [19, 57, 60]. Additionally, a ten-fold increase in production was generally observed, when using multicopy vectors [57, 60]. Thus, the increase obtained between a monocopy transformant, grown in batch, and a multicopy transformant, grown in fed-batch, can be of two orders of magnitude [57].

For two heterologous proteins, human α-foetoprotein and β2-microglobulin, the hp4d-driven production was low (K. Uchida, personal communication), but it was not determined if the problem was a weak expression or a poor secretion. It is to note that the codon usage in *Y. lipolytica* is rather different from that of *S cerevisiae*, and more similar to that of the *Aspergillus* genus ([13] and unpublished data). This could be a

Table 2. Hp4d-directed heterologous protein expression/secretion in *Y. lipolytica*

Organism Protein	Expression vector Secretion signal	Vector type [a] Production [b]	Reference
Virus			
Bacteriophage P1 Cre recombinase (41 kDa)	pRRQ1 (pINA1269 + *ARS68*) none	replicative shaked flask: [intracellular: +]	[62] [c]
Eubacteria			
Escherichia coli β-galactosidase (116 kDa)	related to pINA1269 none	replicative or mono-integrative shaked flask: [intracellular: 420 U]	[19] [d]
E. coli amylolytic enzyme (85 kDa)	derivative of pINA1267 modified *XPR2* prepro	mono-integrative shaked flask: 1 g/l	N Libessart, Roquette, France: personal communication
Fungi			
Arxula adeninivorans glucoamylase (90 kDa)	p0 [e] + secretion signal *XPR2* pre or *XPR2* prepro	mono-integrative shaked flask: (+)	[54]
Aspergillus oryzae leucine aminopeptidase (90 kDa)	derivative of pINA1267 hybrid *LIP2/XPR2* prepro	mono-integrative batch: 320 U/l	[57]
	derivative of pINA1293 hybrid *LIP2/XPR2* prepro	multi-integrative fed-batch: 28000 U/l	[57]
Aspergillus aculeatus galactanase I (44 kDa)	p0 [e] + secretion signal *XPR2* pre	mono-integrative shaked flask: 3 mg/l	D Swennen, LMGM, INRA, France: personal communication

Table 2 continued

Pycnoporus cinnabarinus laccase I (54 kDa)	p0 [e] + secretion signal native	mono-integrative shaked flask: (+/-)	[47]
	pINA1296 *XPR2* pre	mono-integrative shaked flask: 10 mg/l	[47] [f]
	pINA1267 *XPR2* prepro	mono-integrative shaked flask: 10 mg/l	[47] [f]
Trametes versicolor laccase IIIb (58 kDa)	pINA1269 native	mono-integrative shaked flask: (+)	[46]
Plants			
Zea mays cytokinin oxidase I (55 kDa)	pINA1267 *XPR2* prepro	mono-integrative shaked flask: 5 mg/l	[63]
Theobroma cacao aspartic proteinase II (62 kDa)	derivative of pINA1293 hybrid *LIP2/ XPR2* prepro	multi-integrative shaked flask: (+)	[59]
Mammals (non human)			
Bos taurus prochymosin (40 kDa)	related to pINA1267 *XPR2* prepro	mono-integrative shaked flask: 20 mg/l batch: 160 mg/l	[19]
Mus musculus interleukin 6 (20 kDa)	pINA1267 *XPR2* prepro	mono-integrative shaked flask: 15 mg/l	B-C Sang, Pharmingen, USA: personal communication
Lama glama anti-ACE VHH antibody (30 kDa)	pINA1296 *XPR2* pre	mono-integrative shaked flask: (+)	M Chartier, LISM, CNRS, France: personal communication
Human (Homo sapiens sapiens)			
α-foetoprotein (74 kDa)	pINA1267 *XPR2* prepro	mono-integrative shaked flask: 250 μg/l	K Uchida, Oriental Yeast, Japan: personal communication [g]

Table 2 continued

β2-microglobulin (12 kDa)	pINA1267 *XPR2* prepro	mono-integrative shaked flask: 5 μg/l	K Uchida, Oriental Yeast, Japan: personal communication
soluble CD14 variants (48 kDa)	derivative of pINA1293 hybrid *LIP2*/ *XPR2* prepro	multi-integrative fed-batch: 500 mg/l	[60] and C Gysler: personal communication
anti-Ras single-chain antibody scFv (30 kDa)	p0 [(e)] + secretion signal *XPR2* pre or *XPR2* prepro	mono-integrative shaked flask: 20 mg/l	[54] [(h)]
anti-estradiol scFv (30 kDa)	pINA1296 *XPR2* pre	mono-integrative shaked flask: (+)	M Chartier, LISM, CNRS, France: personal communication

The vectors cited are described in the Table 1, or hereafter. The recipient strains used (except the JM23SB strain used for intracellular β-galactosidase production [19]) were all from, or derivatives from, the "Po1" series (see Figure 1). The most frequently used were Po1g for the *LEU2*-selected monocopy vectors, and Po1f for the *URA3*-selected monocopy and multicopy vectors.

(a) The expression/secretion vectors can be either replicative or integrative vectors. Centromeric replicative vectors can maintain 1-3 copies per cell [18, 19]. The integrative vectors can be either monocopy (mono-integrative), or multicopy (multi-integrative) when using promoter-defective marker genes (10-13 copies per cell [20, 61, 41, 57]).

(b) Yeast cultures were performed in shaked flasks or in bioreactor (batch or fed-batch cultivation). A qualitative (+/-) or quantitative (international units of activity, or weight/volume concentration) estimation of the production is indicated.When unspecified, it was measured as heterologous activity in the culture supernatant, and thus corresponds to secreted active heterologous protein. When no secretion signal is present, the intracellular production is indicated between square brackets.

(c) The intracellular production of Cre recombinase was sufficient to mediate the site-specific excision of a DNA region flanked by a pair of loxP sites, in a transposon tagging mutagenesis experiment.

(d) The intracellular production of β-galactosidase directed by the recombinant hp4d promoter, in various media, was similar to that driven by the strong native *XPR2* promoter under inducing conditions.

(e) p0 is a Bluescript-based vector carrying *URA3* as a selection marker, and an expression cassette composed of hp4d, a polylinker and the *XPR2* terminator [54].

(f) The production of *P. cinnabarinus* laccase was compared in *Y. lipolytica* and in *Pichia pastoris*: it was roughly threefold higher in the former. The size of the secreted *Yarrowia*-produced laccase (90 kDa), reflecting its glycosylation status, was also closest to the native one (81 kDa) than the size of the *Pichia*-produced laccase (110 kDa).

(g) The production of human α-foetoprotein was compared in *Y. lipolytica* and in *P. pastoris*: it was roughly twofold higher in the former.

(h) The production of anti-Ras scFv was compared in *Y. lipolytica* and in *Kluyveromyces lactis*: it was twofold higher in the former. This single-chain antibody production level is similar to what can be obtained in *S. cerevisiae* only using chaperones and foldases overexpression.

problem for the correct expression of some heterologous genes, which can be solved by directed mutagenesis allowing to fit *Y. lipolytica* peculiar codon usage.

The hp4d promoter was able to drive very high production levels with some heterologous proteins. Using monocopy vectors, 160 mg per litre of batch culture were obtained for bovine prochymosin [19], and 1 g per litre of shaked flask culture for a bacterial amylolytic enzyme (N. Libessart, personal communication). Using multicopy vectors, 28000 international units per litre of fed-batch culture were obtained for a fungal leucine aminopeptidase [57], and 500 mg per litre for the human sCD14 ([60] and C. Gysler, personal communication).

In the work on β-galactosidase, the authors compared the performances of hp4d with those of the strong *pXPR2*: on various media, the recombinant promoter was able to drive activities similar to that attained by native *pXPR2* only under very peculiar inducing conditions [19]. In the same work, the authors also compared the performances of integrative and replicative (+ *ARS68*) versions of the same expression vector: the hp4d-driven expression levels were similar in both cases [19]. The mean copy number for the replicative vector being one per plasmid-bearing cell, this result indicates that the properties of the hp4d promoter were intrinsic, and did not depend on the integration site of the pBR-based integrative vector, namely the pBR322 docking platform.

The fact that hp4d is able to drive a strong expression in virtually any medium proved very useful in some particular cases, for which the efficient secretion of the heterologous protein was strongly dependent on the culture conditions. This was especially the case for the production of the *P. cinnabarinus* laccase: this protein was efficiently secreted only when the growth rate of the culture was low, and when the cells remained essentially in the yeast form, with no hyphae formation [47]. The versatility of the hp4d promoter allowed, in this peculiar case, to test various media in order to determine the more adapted to an efficient production. Such an adaptation would not have been possible with an inducible promoter requiring a precise medium composition.

This and other results obtained in our laboratory show that each heterologous protein seems to remain a particular case, and that an optimization of the medium composition may be necessary to obtain a satisfying production ([47]; L. Otterbein, N. Houba-Hérin and C. Mougin, personal communications). This point is clearly in favor of the use of versatile promoters, such as hp4d, especially for the production of more complex heterologous proteins. The scale-up to batch or fed-batch cultivation can also require an adaptation of the medium and culture conditions. Although extensive data on *Y. lipolytica* behaviour in large bioreactors was accumulated in the past, it concerned the production of citric acid or single cell proteins, and was not adapted to heterologous proteins production. Recently, a partial optimization of some standard conditions for bioreactor cultivation of *Y. lipolytica* strains from the "Po1" series, for the production of homologous or heterologous proteins, was performed [57].

Secretion and maturation

In most of the examples from the Table 2, the secretion signal used was either the *XPR2* prepro or the *XPR2* pre region. In some cases, the efficiency of the secretion driven by these signals was quantified by testing for remains of intracellular heterologous activity. This was performed for bovine prochymosin, and for maize cytokinin oxidase: for both, intracellular activity remained undetectable, when high

levels were observed in the culture supernatant ([19, 63]; N. Houba-Hérin and C. Pethe, personal communication). This shows that the secretion efficiency was very near to 100%, as previously observed when using the p*XPR2*, namely when these secretion signals were in their native environment [14]. This very high secretion efficiency does not always imply, however, a correct maturation (namely, the correct removal, at the amino acid level, of the secretion signal), as we will see hereafter.

Some of the examples from the Table 2 compared the secretion driven by different signals. Interestingly, the proteins studied were known to be difficult to produce under an active form in classical expression systems. These were a single-chain antibody (the anti-Ras scFv), containing disulfide bridges, and a fungal laccase (the *P. cinnabarinus* laccase I), a metalloprotein with disulfide bridges and four associated coppers. Swennen *et al.* [54] compared the efficiencies of *XPR2* pre and prepro regions for the production of the scFv, and found them similar. However, the use of the *XPR2* prepro lead to a slightly incomplete maturation: a fraction of the secreted scFv exhibited a higher molecular weight, corresponding to an incomplete processing of the pro region [54]. The same signals were also used for the production of the laccase: they exhibited similar efficiencies, and were much more efficient than the native signal peptide ([47] and L. Otterbein, personal communication). However, interestingly, another fungal laccase, from *T. versicolor*, was efficiently secreted using its native signal peptide ([46] and C. Mougin, personal communication).

In some of the examples from the Table 2, the amino-terminal sequence of the processed secreted heterologous protein has been determined. This was the case for the maize cytokinin oxidase ([63] and N. Houba-Hérin, personal communication), the anti-Ras scFv [54] and the *P. cinnabarinus* laccase ([47] and L. Otterbein, personal communication). In all these examples, the processing of the pre or the prepro regions was found to have been correct at the amino acid level. However, some cases of abnormal processing of the *XPR2* prepro region had previously been reported, in the secretion of rice α-amylase or human EGF [48, 65]. It was then hypothesized that the pro processing enzyme could became unable to access the pro cleavage site in some fusion proteins, leading to an aberrant upstream processing.

It appears, from the examples described above, and from previous works, that (i) the *XPR2* pre and prepro sequences seem to behave as efficiently when downstream from the hp4d promoter than in their native environment, (ii) the pro sequence is not necessary for secretion of heterologous proteins, (iii) the pre sequence alone can drive a similarly efficient secretion, (iv) the pro sequence can became, in some peculiar cases, a source of incomplete or imprecise processing.

Recently, interesting results were obtained using a modified *XPR2* prepro sequence, or a hybrid *LIP2*/*XPR2* prepro sequence, as can be seen from the examples of the bacterial amylolytic enzyme, the fungal leucine aminopeptidase, and the human sCD14 ([57]; N. Libessart, personal communication; [60] and C. Gysler, personal communication). The properties of these alternative secretion signals are still under investigation.

Glycosylation

The glycosylation of the secreted proteins is a major post-translational modification. In particular, many proteins of therapeutic interest are glycosylated, and the exact feature

of these additions affects their activity and their immunogenicity. Unfortunately, there are many differences in the glycosylation pattern between yeasts and mammalian cells (reviewed in [3]). In all eukaryotes, N-linked glycosylation begins in the endoplasmic reticulum with the addition of a complex core oligosaccharide residue to the asparagine in the consensus sequence Asn-X-Ser/Thr (X for any amino acid). This core is then slightly reduced, and the glycosylation patterns of higher and lower eukaryotes differ hereafter [66]. The mammalian Golgi apparatus performs a series of reactions that generate three types of oligosaccharide residues: (i) high-mannose type, (ii) complex type (mixture of different sugars), and (iii) hybrid type (combination of both previous types). In contrast, the lower eukaryotes such as yeasts perform only the addition of mannose outer chains. However, these chains vary in lenght for endogenous and heterologous proteins, and can be much longer than in mammals: in *S. cerevisiae*, they are typically of fifty to hundred-fifty mannose residues. This hyperglycosylation can impair heterologous proteins activity, and is more particularly a problem for the pharmaceutical industry. Namely, some hyperglycosylated therapeutical proteins exhibit a rapid cleareance and are very antigenic. Yeasts other than *S. cerevisiae* present in general less problematic glycosylation patterns: *P. pastoris* and *H. polymorpha* produce glycoproteins with chains of eight to forteen mannose residues [67, 68]. Although glycosylation patterns have not yet been extensively studied in *Y. lipolytica*, a secreted heterologous protein was found to contain only short oligosaccharide chains, of about eight to ten mannose residues (A. Franke, personal communication). Thus, it appears that the glycosylation pattern of some non-conventional yeasts, and especially of *Y. lipolytica*, is closer to the mammalian high-mannose type of glycosylation than to that of *S. cerevisiae*.

In some of the examples from the Table 2, the overall glycosylation pattern of the secreted heterologous proteins was checked using Western blotting and endoglycosidase treatment. No heavy hyperglycosylation was reported, but a moderate overglycosylation was sometimes observed, as in the case of the *P. cinnabarinus* laccase. This overglycosylation was in the range of ten additional kDa, and did not impaired the heterologous activity ([47] and L. Otterbein, personal communication). It was however lesser than that previously observed when producing this same laccase in *P. pastoris* (see the following chapter). Interestingly, in the case of the bovine prochymosin, two potential glycosylation sites, which were present in the protein but not glycosylated naturally, were also not glycosylated in *Y. lipolytica* [19]. The precise feature of the glycosylation in *Y. lipolytica* is currently under investigation. Its knowledge could possibly lead to a "humanization" of the pattern of glycosylation in this yeast, which would greatly benefit to the quality of the secreted heterologous proteins.

Comparison with other expression systems

As discussed in the introduction, *Y. lipolytica* compared favorably with other expression systems, in particular with *S. cerevisiae* and with alternative yeasts. Some examples from the Table 2, for which a precise comparison of *Y. lipolytica* with other expression systems has been made, reinforce this conclusion. Namely: (i) the production of human α-foetoprotein was twofold higher in *Y. lipolytica* than in *P. pastoris* (K. Uchida, personal communication), (ii) the production of anti-Ras scFv was twofold higher than in *K. lactis* [54], and (iii) the production of *P. cinnabarinus* laccase was

roughly threefold higher than in *P. pastoris* ([47] and L. Otterbein, personal communication). In this later case, the overglycosylation of the recombinant laccase was also less pronounced in *Y. lipolytica*: its molecular weight (90 kDa [47]) was closer to that of the naturally glycosylated enzyme (81 kDa) than that observed in *P. pastoris* (110 kDa [69]).

Conclusion

Together with previous works, the recent examples of heterologous production, using newly developed tools, examplify the high potential of *Y. lipolytica* as an expression system. Its secretory apparatus present some peculiar characteristics: (i) importance of the co-translational pathway, (ii) high efficiency, and (iii) low overglycosylation. With these features, the secretion in *Y. lipolytica* appear closer to that of mammalian cells than to those of many yeasts. These characteristics render *Y. lipolytica* particularly interesting as an host, especially for the production of complex proteins. With the recent development of new tools, easy-to-handle expression/secretion vectors and recipient strains are now available, including an efficient amplification system, and offering the potential for scaled-up production. Research on *Y. lipolytica* can now focus on further improvements, such as the identification and correction of some bottlenecks in its secretion pathway, or the "humanization" of its pattern of glycosylation.

Acknowledgements

I wish to thank all colleagues and collaborators who accepted to communicate unpublished data: Drs Martine Chartier, Christof Gysler, Nicole Houba-Hérin, Claude Jolivalt, Nathalie Libessart, Christian Mougin, Ludovic Otterbein, Claude Pethe, Bi-Ching Sang, Dominique Swennen and Kohji Uchida. I am also indebted to Dr Jean-Marie Beckerich for his support, and for the critical reading of the manuscript.

References

1. Buckholz, R.G., and Gleeson, M.A.G. 1991, Bio/Technology, 9, 1067.
2. Dominguez, A., Ferminan, E., Sanchez, M., Gonzalez, F.J., Perez-Campo, F.M., Garcia, S., Herrero, A.B., San Vicente, A., Cabello, J., Prado, M., Iglesias, F.J., Choupina, A., Burguillo, F.J., Fernandez-Lago, L., and Lopez, M.C. 1998, Internatl. Microbiol., 1, 131.
3. Cereghino, J.L., and Cregg, J.M. 2000, FEMS Microbiol. Rev., 24, 45.
4. Cregg, J.M., Cereghino, J.L., Shi, J., and Higgins, D.R. 2000, Mol. Biotechnol., 16, 23.
5. Cereghino, G.P.L., Cereghino, J.L., Ilgen, C., and Cregg, J.M. 2002, Curr. Opinion in Biotechnol., 13, 329.
6. Muller, S., Sandal, T., Kamp-Hansen, P., and Dalboge, H. 1998, Yeast, 14, 1267.
7. Wickerham, L.J., Kurtzman, C.P., and Herman, A.I. 1970, Science 167, 1141.
8. Barth, G., and Gaillardin, C. 1997, FEMS Microbiol. Rev., 19, 219.
9. Tsugawa, R., Nakase, T., Koyabashi, T., Yamashita, K., and Okumura, S. 1969, Agr. Biol. Chem. (Tokyo), 33, 929.
10. Heslot H. 1990, Adv. Biochem. Eng. Biotechnol., 43, 43.
11. Davidow, L.S., Apostolakos, D., O'Donnell , M.M., Proctor, A.R., Ogrydziak, D.M., Wing, R.A., Stasko, I., and De Zeeuw, J.R. 1985, Curr. Genet., 10, 39.
12. Gaillardin, C., Ribet, A.M., and Heslot H. 1985, Curr. Genet., 10, 49.
13. Gaillardin, C., and Heslot H. 1988, J. Basic Microbiol., 28, 161.

14. Franke, A.E., Kaczmarek, F.S., Eisenhard, M.E., Geoghegan, K.F., Danley, D.E., De Zeeuw, J.R., O'Donnell, M.M., Gollaher, M.G., and Davidow, L.S. 1988, Developments in Industrial Microbiology, vol. 29, G. Pierce (Ed.), Elsevier, Amsterdam, 43.
15. Matoba, S., Fukuyama, J., Wing, R.A., and Ogrydziak, D.M. 1988, Mol. Cell. Biol., 8, 4904.
16. Poritz, M.A., Siegel, V., Hansen, W., and Walter, P. 1988, Proc. Natl. Acad. Sci., 85, 4315.
17. Fournier, P., Guyaneux, L., Chasles, M., and Gaillardin, C. 1991, Yeast, 7, 25.
18. Vernis, L., Abbas, A., Chasles, M., Gaillardin, C., Brun, C., Huberman, J.A., and Fournier, P. 1997, Mol. Cell Biol., 17, 1995.
19. Madzak, C., Treton, B., and Blanchin-Roland, S. 2000, J. Molecular Microbiol. Biotechnol., 2, 207.
20. Le Dall, M.-T., Nicaud, J.-M., and Gaillardin, C. 1994, Curr. Genet., 26, 38.
21. Schmid-Berger, N., Schmid, B., and Barth, G. 1994, J. Bacteriol., 176, 2477.
22. Barth, G., and Scheuber, T. 1993, Mol. Gen. Genet., 241, 422.
23. Blanchin-Roland, S., Cordero Otero, R., and Gaillardin, C. 1994, Mol. Cell. Biol., 14, 327.
24. Barns, S.M., Lane, D.J., Sogin, M.L., Bibeau, C., and Weisburg, W.G. 1991, J. Bacteriol., 173, 2250.
25. Okuma, M., Hwang, C.W., Masuda, Y., Nishida, H., Sugiyama, J., Ohta, A., and Takagi, M. 1993, Biosci. Biotechnol. Biochem., 57, 1793.
26. Casaregola, S., Neuvéglise, C., Lépingle, A., Bon, E., Feynerol, C., Artiguenave, F., Winckler, P., and Gaillardin, C. 2000, FEBS Lett., 487, 95.
27. Holzschu, D.L., Chandler, F.W., Ajello, L., and Ahearn, D.G. 1979, Sabouraudia, 17, 71.
28. Tobe, S., Takami, T., Ikeda, S., and Mitsugi, K. 1976, Agric. Biol. Chem., 40, 1087.
29. Ogrydziak, D.M., and Scharf, S.J. 1982, J. Gen. Microbiol., 128, 1225.
30. Boisramé, A., Kabani, M., Beckerich, J.-M., Hartmann, E., and Gaillardin, C. 1998, J. Biol. Chem., 273 (47), 30903.
31. De Baetselier, A., Vasavada, A., Dohet, P., Ha-Thi, V., De Beukelaer, M., Erpicum, T., De Clerck, L., Hanotier, J., and Rosemberg, S. 1991, Bio/Technology, 9, 559.
32. Nicaud, J.-M., Fabre, E., and Gaillardin, C. 1989, Curr. Genet., 16, 253.
33. Wojtatowicz, M., Rymowicz, W., Robak, M., Zarowska, B., and Nicaud, J-M. 1997, Pol. J. Food Nutr. Sci., 6/47 (4), 49.
34. Xuan, J.W., Fournier, P., and Gaillardin, C. 1988, Curr. Genet., 14, 15.
35. Fournier, P., Abbas, A., Chasles, M., Kudla, B., Ogrydziak, D.M., Yaver, D., Xuan, J.W., Peito, A., Ribet, A.-M., Feynerol, C., He, F., and Gaillardin, C. 1993, Proc. Natl. Acad. Sc., 90, 4912.
36. Nicaud, J.-M., Fournier, P., La Bonnardiere, C., Chasles, M., and Gaillardin, C. 1991, J. Biotechnol., 19, 259.
37. Barth, G., and Gaillardin, C. 1996, Nonconventional Yeasts in Biotechnology: a Handbook, K. Wolf (Ed.), Springer-Verlag, Heidelberg, 313.
38. Hamsa, P.V., and Chattoo, B.B. 1994, Gene, 143, 165.
39. James, L.C., and Strick, C.A. 1993, US Patent Application US08/117.375.
40. Casaregola, S., Feynerol, C., Diez, M., Fournier, P., and Gaillardin, C. 1997, Chromosoma, 106, 380.
41. Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G., and Nicaud, J.-M. 2001, Yeast, 18, 97.
42. Davidow, L.S., Franke, A.E., and De Zeeuw, J.R. 1987, European Patent Application EP86307839.
43. Ogrydziak, D.M., Demain, A.L., and Tannenbaum, S.R. 1977, Biochim. Biophys. Acta, 497, 525.
44. Park, C.S., Chang, C.C., and Ryu, D.D.Y. 2000, Appl. Biochem. Biotechnol., 87 (1), 1.
45. Morin, M., and Dominguez, A. 2002, Abstracts report book from Third *Yarrowia lipolytica* International Meeting, Dresden University of Technology, Dresden, Germany.

46. Jolivalt, C., Briozzo, P., Bertrand, T., Madzak, C., Caminade, E., and Mougin, C. 2002, Abstracts report book from Engineering Enzymes for Biocatalysis by Using Directed Evolution, European Federation of Biotechnology, Pasteur Institute, Paris, France.
47. Madzak, C., Otterbein, L., Moukha, S., Gaillardin, C., and Beckerich, J.M. 2002, Abstracts report book from Second International Conference on Recombinant Protein Production, European Federation of Biotechnology, Cernobbio, Como, Italy.
48. Park, C.S., Chang, C.C., Kim, J-Y., Ogrydziak, D.M., and Ryu, D.D.Y. 1997, J. Biol. Chem., 272, 6876.
49. He, F., Beckerich, J.-M., and Gaillardin, C. 1992, J. Biol. Chem., 267 (3), 1932.
50. Yaver, D.S., Matoba, S., and Ogrydziak, D.M. 1992, J. Cell Biol., 116, 605.
51. Fabre, E., Nicaud, J.-M., Lopez, M.C., and Gaillardin, C. 1991, J. Biol. Chem., 266, 3782.
52. Fabre, E., Tharaud, C., and Gaillardin, C. 1992, J. Biol. Chem., 267, 15049.
53. Beckerich, J.-M., Boisramé-Baudevin, A., and Gaillardin, C. 1998, Internatl. Microbiol., 1, 123.
54. Swennen, D., Paul, M.-F., Vernis, L., Beckerich, J.-M., Fournier, A., and Gaillardin, C. 2002, Microbiol., 148, 41.
55. Juretzek, T., Wang, H.-J., Nicaud, J.-M., Mauersberger, S., and Barth, G. 2000, Biotechnol. Bioprocess. Eng., 5, 320.
56. Madzak, C., Blanchin-Roland, S., Cordero Otero, R., and Gaillardin, C. 1999, Microbiol., 145, 75.
57. Nicaud, J.-M., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P., and Gaillardin, C. 2002, FEMS Yeast Res., 2, 371.
58. Pignède, G., Wang, H., Fudalej, F., Seman, M., Gaillardin, C., and Nicaud, J-M. 2000, Appl. Environ. Microbiol., 66 (8), 3283.
59. Laloi, M., Mac Carthy, J., Morandi, O., Gysler, C., and Bucheli, P. 2002, Planta, 215, 754.
60. Gysler, C., Duboc, P., Vidal, K., Nicaud, J.-M., Affolter, M., Pampel, L., and Niederberger, P. 2002, Abstracts report book from Third *Yarrowia lipolytica* International Meeting, Dresden University of Technology, Dresden, Germany.
61. Nicaud, J.-M., Gaillardin, C., Seman, M., and Pignède, G. 1998, French Patent Application PCT/FR99/02079.
62. Richard, M., Quijano, R.R., Bezzate, S., Bordon-Pallier, F., and Gaillardin, C. 2001, J. Bacteriol., 183 (10), 3098.
63. Madzak, C., Houba-Hérin, N., Pethe, C., Laloue, M., Gaillardin, C., and Beckerich, J.-M. 2001, Yeast, 18 (S1), 297.
64. Gaillardin, C., and Ribet A.-M. 1987, Curr. Genet., 11, 369.
65. Hamsa, P.V., Kachroo, P., and Chattoo, B.B. 1998, Curr. Genet., 33, 231.
66. Goochee, C.F., Gramer, M.J., Andersen, D.C., Bahr, J.B., and Rasmussen, J.R. 1991, Biotechnol. (NY), 9, 1347.
67. Grinna, L.S., and Tschopp, J.F. 1989, Yeast, 5, 107.
68. Gellisen, G., and Hollenberg, C.P. 1997, Gene, 190, 87.
69. Otterbein, L., Record, E., Longhi, S., Asther, M., and Moukha, S. 2000, Eur. J. Biochem., 267, 1619.