Introduction

The methylotrophic yeast *Pichia pastoris* is a commonly used host for the production of recombinant proteins [1–4]. The most frequently used promoter for heterologous gene expression is derived from the AOX1 gene (alcohol oxidase, EC 1.1.3.13) [3]. The promoters pGAP (of the gene for glyceraldehyde-3-phosphate dehydrogenase, EC 1.2.1.12), pDHAS (of the gene for dihydroxyacetone synthase, EC 2.2.1.3) and pFLD1 (of the gene for formaldehyde dehydrogenase, EC 1.2.1.1) and others may be used as alternatives [5,6]. In order to give greater scope for optimizing gene expression in *P. pastoris* we set out to establish a bacteriophage T7 RNA polymerase (T7 RNAP) (EC 2.7.7.6) based gene expression strategy. T7 RNAP initiates transcription from a highly conserved 23-nucleotide promoter (pT7) and transcribes the template without involvement of any other cellular transcription factors [7,8]. T7 RNAP is widely used for the overproduction of foreign proteins in the prokaryote *Escherichia coli* [9] but was also reported to be a tool for gene expression in eukaryotes. For example, T7 RNAP encoded on a recombinant vaccinia virus successfully has been used for transient expression of proteins in mammalian cells by *in vitro* transcription with T7 RNAP, providing an economical alternative to chemical synthesis of siRNAs [11].

A novel *Pichia pastoris* expression vector (pEZT7) for the production of recombinant proteins employing prokaryotic bacteriophage T7 RNA polymerase (T7 RNAP) (EC 2.7.7.6) and the corresponding promoter pT7 was constructed. The gene for T7 RNAP was stably introduced into the *P. pastoris* chromosome 2 under control of the (endogenous) constitutive *P. pastoris* glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter (pGAP). The gene product T7 RNAP was engineered to contain a nuclear localization signal, which directed recombinant T7 RNAP to the *P. pastoris* nucleus. To promote translation of uncapped T7 RNAP derived transcripts, the internal ribosomal entry site from hepatitis C virus (HCV-IRES) was inserted directly upstream of the multiple cloning site of pEZT7. A *P. pastoris* autonomous replicating sequence (PARS1) was integrated into pEZT7 enabling propagation and recovery of plasmids from *P. pastoris*. Rapid amplification of 5’ complementary DNA ends (5’ RACE) experiments employing the test plasmid pEZT7-EGFP revealed that transcripts indeed initiated at pT7. HCV-IRES mediated translation of the latter mRNAs, however, was not observed. Surprisingly, HCV-IRES and the reverse complement of PARS1 (PARS1rc) were both found to display significant promoter activity as shown by 5’ RACE.
flasks were inoculated to an initial OD$_{600}$ of 0.15 and grown for 48 h at 250 rpm. Cells were harvested by centrifugation (5 min; 6000g), washed using lysis buffer (see section “protein extraction and Western blot analysis”) and stored at –80 °C.

**General DNA manipulations and construction of plasmids**

For standard molecular biological manipulations, well-described protocols were used [12]. E. coli strains Top10, Top10F or DH5α were employed as hosts for plasmid cloning experiments. Bacteria were grown in lysogeny broth (LB) [1]. Oligonucleotide primers were produced by Sigma Aldrich. For isolation of yeast chromosomal DNA the Epicentre® MasterPure™ Yeast DNA Purification Kit was used. For isolation of plasmid DNA from *P. pastoris* the ZymoPrep™ Yeast Plasmid Miniprep II Kit was utilized. The plasmids used for *P. pastoris* expression experiments are listed in Table S1 (Appendix A). The construction of these plasmids is described in “Supplementary materials and methods” (Appendix A).

**Yeast transformation and selection**

The transformation of *P. pastoris* was carried out essentially as described earlier [13]. *P. pastoris* strains were grown at 30 °C to an OD$_{600}$ of 1.5 in YPD. Cells were collected by centrifugation and washed twice in water. The cells were suspended in 1 M sorbitol and kept on ice until use. For integration of plasmid DNA into the genome of *P. pastoris* cells were transformed with linearized plasmids (5 μg) using a Bio-Rad Gene Pulser® II (1500 V, 50 μF, 200 Ω). For transformation using non-integrating PARS1-containing plasmids, circular DNA (300 ng) was used. Transformant strains were selected on YPDS (1% yeast extract, 2% peptone, 2% dextrose, 1 M sorbitol) containing zeocin™ (100 up to 500 μg/ml) or blasticidin (100 μg/ml).

**Rapid amplification of 5’ end RNA**

For analyzing pT7/T7 RNAP derived transcripts in X33T7 strains the SMARTer™ Amplification Kit (Clontech) was used. Total RNA was isolated using the RNeasy Mini Kit (Qiagen). For monitoring transcription initiation in *P. pastoris* X33 strains with Cap-dependent translation, the 5’ RACE method (GeneRacer kit, Life Technologies) was used. At least 15 different plasmid inserts derived from the individual reactions were analyzed by RNA sequencing (Eurofins MWG Operon).

**Protein extraction and Western blot analysis**

Frozen cell pellets were thawed on ice and suspended in lysis buffer (50 mM sodium phosphate, pH 7.4; 1 mM phenylmethylsulfonyl fluoride; 1 mM EDTA; 5% glycerol and 1 mM DTT). An equal volume of acid-washed glass beads (425–600 μm, product number G8772, Sigma Aldrich) was added. Samples were vigorously mixed (30 s; 2500 rpm) and subsequently placed on ice (30 s) (10 cycles). The samples were centrifuged (6000g) for 15 min at 4 °C. For mixing an MS2 Minishaker (IKA GmbH & Co. KG, Staufen, Germany) was used at maximum speed (2500 rpm). The protein concentration of cell lysates was estimated using the method of Bradford [14] and bovine serum albumin as a standard. The samples were analyzed by SDS-PAGE on 4–20% gradient polyacrylamide gels (Bio-Rad) using 30–50 μg of protein per lane. Proteins were transferred to nitrocellulose membranes (Fall GmbH) by electroblotting (Trans-Blot® SD, Bio-Rad). For the immunological detection of T7 RNAP, T7 RNAP monoclonal antibodies (product number 70566, Novagen) and goat anti-mouse IgG alkaline phosphatase (AP)-coupled secondary antibodies (product number 69266, Novagen) were used. For the immunological detection of EGFP, GFP monoclonal antibodies (product number MMS-118P, Covance, CA, USA) in combination with AP-coupled secondary antibodies (product number 69266, Novagen) were used.

**Fluorescence microscopy**

Fluorescence microscopy was carried out employing an Axiohot microscope (Carl Zeiss, Jena) equipped with a Hamamatsu Orca ER II CCD camera (Hamamatsu Photonics, Hamamatsu, Japan). The data were analyzed using Wasabi 1.4.

**Results**

**Construction of a Pichia pastoris strain producing a modified form of T7 RNAP**

The first step with regard to the establishment of a T7 RNAP/pT7 based expression system employing *P. pastoris* was the construction of a stable recombinant *P. pastoris* strain producing T7 RNAP. The latter strain was engineered to synthesize a modified version of T7 RNAP (NLS-T7 RNAP) containing the N-terminal nuclear localization signal (NLS) from Simian virus 40 T-antigen [15]. NLS was introduced in order to redirect the enzyme from the site of translation to the nucleus where transcription of pT7 controlled heterologous genes should occur. The gene for NLS-T7 RNAP was integrated into the GAP locus of chromosome 2 of *P. pastoris* X33 (to give *P. pastoris* X33T7) by homologous recombination using pGAPBA-NLS-T7 RNAP. In *P. pastoris* X33T7 the gene for NLS-T7 RNAP was under control of the constitutive GAP-promoter. Western-blot analysis of cell lysates of *P. pastoris* X33T7 using antibodies directed against T7 RNAP revealed that the enzyme indeed was synthesized (Fig. S1A). The nuclear localization of NLS-T7 RNAP was tested by analyzing a modified *P. pastoris* X33T7 strain (*P. pastoris* X33T7-EGFP), which synthesized a NLS-T7 RNAP fusion. Fluorescence microscopy of *P. pastoris* X33T7-EGFP suggested that most of the T7 RNAP-EGFP fusion protein was present in the nuclei of the recombinant strains (Fig. S1B). It is thus very likely that also in *P. pastoris* X33T7 (not containing the T7 RNAP-EGFP fusion) T7 RNAP localized to the nuclei. The novel strain *P. pastoris* X33T7 was employed as a host for the subsequent pT7/T7 RNAP based gene expression experiments.

**Construction of a test plasmid for evaluation of T7 RNAP/pT7 based expression in *P. pastoris* X33T7**

The vector pEZT7-EGFP, which was designed in order to test T7 RNAP/pT7 RNAP based gene expression in *P. pastoris*, is schematically shown in Fig. 1. A major challenge with regard to developing a T7 RNAP/pT7 based expression system in *P. pastoris* was to ensure translation of pT7 derived mRNAs. T7 RNAP is not known to produce mRNAs carrying a 5’ cap necessary in eukaryotic cells for translation initiation [16]. Thus, in order to drive translation of uncapped T7 RNAP/pT7 derived transcripts, a DNA fragment containing HCV-IRES was placed directly upstream in frame with the gene for EGFP, which was employed as a reporter. HCV-IRES was reported to allow assembly of a functional 80S translation complex on uncapped mRNAs upstream of the translation initiation codon [17]. Moreover, in order to ensure proper assembly of the ribosomal machinery and translation initiation at HCV-IRES the sequence coding for the 15 N-terminal amino acids of HCV polyprotein was included [18] which finally runs into the gene for EGFP. The latter gene was codon optimized for expression in

---

1 Abbreviations used: LB, lysogeny broth; NLS, nuclear localization signal.
Expression experiments using pEZT7-EGFP

The above described strain P. pastoris X33T7 was transformed with the test vector pEZT7-EGFP and analyzed with regard to EGFP production. Observation of the transformant cells by fluorescence microscopy readily distinguished the EGFP-containing cells from those which have been transformed with the control plasmid (pEZT7-EGFP<sup>-</sup>EGFP) not containing the EGFP gene (Fig. 2) (for all experiments at least three different transformant strains were analyzed). Western blot analyses using antibodies directed against EGFP supported these results (Fig. 2). Much to our surprise, however, fluorescence was also observed in pEZT7-EGFP transformed P. pastoris X33 strains (controls) not producing NLS-T7 RNAP. The Western blot data as well revealed that EGFP was produced in the control strains. Moreover, analysis of P. pastoris strains harboring a plasmid construct not containing the T7-promoter (pEZT7-EGFP<sup>-T7</sup>) revealed strong fluorescence (and strong Western blot signals) indicating that a T7 independent transcription had occurred (Fig. 2). Notably, the signal strength obtained by fluorescence microscopy and by Western blot analysis was similar in all strains indicating that the amount of EGFP produced in the P. pastoris X33 strains was similar to the EGFP amount produced employing P. pastoris X33T7. The above described findings suggested that transcription of the EGFP gene from pEZT7-EGFP was driven by genetic elements other than the T7 promoter, most likely by either HCV-RES [17] or PARs [PARs, origin for maintenance of extrachromosomal genetic elements in P. pastoris] [19].

It was suggested earlier that the DNA sequence corresponding to the HCV 5' UTR (HCV-RES) contained a functional promoter for mammalian expression systems [20] which directed the synthesis of mRNAs in HeLa cells. Similar conclusions were found for the IRES of the TIF4631 gene, which was tested for cap-independent translation in Saccharomyces cerevisiae [21–23]. In addition, it was reported that sequences surrounding autonomously replicating sequences (ARSs) in S. cerevisiae enhance the initiation of DNA replication by transcriptional activation [24]. We therefore hypothesized that the PARs sequence might have a similar activity.

In order to find an explanation for the observed T7 independent expression of EGFP a series of pEZT7-EGFP derivatives was tested and the data are summarized in Fig. 3. The plasmid pEZT7-EGFP<sup>PARsus</sup>-HCV<sub>0</sub> not containing HCV-RES, did not induce EGFP production in P. pastoris X33T7 and in P. pastoris X33. This experiment suggested that it was the HCV-RES element which was responsible for EGFP expression. The PARs sequence (PARs<sup>0</sup>) in the latter construct apparently was not able to drive expression of the reporter gene. However, when the PARs sequence was inverted (reverse-complement) (pEZT7-EGFP<sup>PARsus<sub>I</sub>/HCV<sub>0</sub></sup>) gene expression was observed. This was also true for P. pastoris strains harboring pEZT7-EGFP<sup>PARsus<sub>II</sub>/HCV<sub>0</sub></sup> where both, the inverted PARs element and HCV-RES,
were present. EGFP expression from HCV or PARS1 using pEZT7-EGFP<sup>PARSus/rc-HCV</sup> and pEZT7-EGFP<sup>PARSus/rc</sup> was stronger in <i>P. pastoris</i> X33 when compared to <i>P. pastoris</i> X33<sup>T7</sup>. A possible explanation for this finding is that T7 RNAP present in X33<sup>T7</sup> competed with RNA polymerase II and produced mRNAs which were uncapped and thus were not translated (see section “monitoring of PARS1 derived transcripts” below). When the PARS element was moved downstream with regard to the reporter gene (pEZT7-EGFP<sup>PARSigma-HCV</sup> and pEZT7-EGFP<sup>PARSigma/rc-HCV</sup>), no EGFP expression was found.

**Monitoring of T7 RNAP/pT7 derived transcripts**

Three independent transformant <i>P. pastoris</i> X33<sup>T7</sup> strains harboring pEZT7-EGFP were analyzed with regard to transcription start sites of the reporter gene EGFP employing a 5' RACE method (SMARTer<sup>™</sup> Amplification Kit). From these strains total RNA samples were isolated and processed as described in the Materials and Methods section. From each RNA sample, at least 15 different cloned cDNAs were characterized by DNA sequencing. In all strains, the transcription start of the EGFP mRNA could be located to the known transcription start of the bacteriophage T7 promoter (80% of the sequenced cDNAs) showing that recombinant EGFP, however, additional start sites were detected (Fig. S3B).

**Monitoring of HCV-IRES derived transcripts**

Three independent transformant <i>P. pastoris</i> X33 strains harboring pEZT7-EGFP<sup>ATT7</sup> and three independent transformant <i>P. pastoris</i> X33 strains harboring pEZ-EGFP<sup>PARSigma</sup> (see Figs. 2 and 3 for explanation of plasmid elements) were analyzed with regard to the EGFP transcriptional start sites employing the GeneRacer Kit (Life Technologies). From each RNA sample, at least 15 different cloned cDNAs were characterized by DNA sequencing. In all strains, the transcription start of the EGFP mRNA could be located to the 3'-end of HCV-IRES. Apparently, the main initiation site (80% of the sequenced cDNAs) was present in codon 12 of the HCV polyprotein, however, additional start sites were detected (Fig. S3B).

**Discussion**

The objective of this study was to establish a bacteriophage T7 promoter dependent expression system employing <i>P. pastoris</i> in order to provide an alternative to the commonly used promoters. T7 RNAP was reported to efficiently transcribe nuclear target genes in <i>S. cerevisiae</i> [25,26]. These transcripts, however, were not translated, probably because the 5' ends were not modified to contain a 7-methyl guanosine cap [25,27]. Although T7 transcripts were not capped in <i>S. cerevisiae</i> they were reported to be polyadenylated at a capdependent promoter II polyadenylation signal which leads to export to the cytoplasm [28]. Consequently, it appeared that the only difficulty in designing a pT7/T7 RNAP based expression system in <i>P. pastoris</i> would be to ensure translation of uncapped T7 transcripts. RNAs of several viruses are naturally uncapped yet are translated in eukaryotic cells [29]. These RNAs contain an IRES present in the 5' untranslated region allowing translation of viral polyproteins.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Plasmid elements</th>
<th>EGFP fluorescence in X33&lt;sup&gt;T7&lt;/sup&gt;</th>
<th>Western blot EGFP</th>
<th>EGFP fluorescence in X33</th>
<th>Western blot EGFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEZT7-EGFP&lt;sup&gt;PARSus-rc-HCV&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEZT7-EGFP&lt;sup&gt;PARSigma-HCV&lt;/sup&gt;</td>
<td></td>
<td>++</td>
<td></td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>pEZT7-EGFP&lt;sup&gt;PARSigma-rc-HCV&lt;/sup&gt;</td>
<td></td>
<td>+</td>
<td></td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>pEZT7-EGFP&lt;sup&gt;PARSigma/rc-HCV&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEZT7-EGFP&lt;sup&gt;PARSigma-rc-HCV&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

+++, strong fluorescence; −−−, absent fluorescence

Fig. 3. Expression experiments using variants of pEZT7-EGFP. Schematic representation of the reporter plasmid constructs used for transformation of the bacteriophage T7 RNA polymerase producing <i>Pichia pastoris</i> strain X33<sup>T7</sup> and <i>P. pastoris</i> X33 (wild-type = control). For simplicity the circular pEZT7 plasmids are shown as linear constructs and not all genetic elements are shown (see also Fig. 1). The relative amount of enhanced recombinant green fluorescent protein from <i>Anugera victoria</i> (codon optimized version EGFP<sub>opt</sub>) production was estimated by fluorescence microscopy (1000×). The plasmid elements are explained in Fig. 1. Western blot experiments were carried out as described in Fig. 2.

---
IRESes have widely been used in genetic engineering in order to direct translation of uncapped mRNAs. For our experiments an IRES derived from hepatitis C virus was employed [30,31]. Although our mapping experiments employing 5’ RACE unequivocally showed that T7 RNAP was able to generate mRNAs from pT7 present on our test plasmid pEZT7-EGFP, translation of the HCV-IRES containing transcripts was not observed. Thus, in our hands HCV-IRES was not sufficient in P. pastoris to favor ribosomal binding and to initiate translation. As an alternative we used an IRES from encephalomyocarditis virus (EMCV-IRES) which proved to be inactive as well (data not shown).

We introduced a variety of test vectors into the nucleus of P. pastoris and were surprised to find expression even when using control plasmids without T7 promoter and when using strains derived from hepatitis C virus was employed [30,31]. For our experiments an IRES has widely been used in genetic engineering in order to direct translation of uncapped mRNAs. For our experiments an IRES was employed [10,32]. We introduced a variety of test vectors into the nucleus of P. pastoris and were surprised to find expression even when using control plasmids without T7 promoter and when using strains derived from hepatitis C virus was employed [30,31]. Our transcription mapping experiments clearly show that the DNA sequence which produces HCV-IRES has a relatively strong promoter activity. To the best of our knowledge, the finding that PAR1s have promoter activity is novel. In summary our data suggest that (using the described genetic elements) it is not possible to establish an efficient pT7/T7 RNAP based expression system in the eukaryotic host P. pastoris.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.pep.2013.09.004.

References