



**HAL**  
open science

## Real time PCR based determination of gene copy numbers in *Pichia pastoris*

Sandra Abad, Kerstin Kitz, Ulrike Schreiner, Astrid Hoermann, Franz Hartner, Anton Glieder

► **To cite this version:**

Sandra Abad, Kerstin Kitz, Ulrike Schreiner, Astrid Hoermann, Franz Hartner, et al.. Real time PCR based determination of gene copy numbers in *Pichia pastoris*. *Biotechnology Journal*, 2010, 5 (4), pp.413. 10.1002/biot.200900233 . hal-00552336

**HAL Id: hal-00552336**

**<https://hal.science/hal-00552336>**

Submitted on 6 Jan 2011

**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.



## Real time PCR based determination of gene copy numbers in *Pichia pastoris*



Journal:	<i>Biotechnology Journal</i>
Manuscript ID:	biot.200900233.R2
Wiley - Manuscript type:	Technical Report
Date Submitted by the Author:	23-Jan-2010
Complete List of Authors:	Abad, Sandra; Research Centre Applied Biocatalysis, Molekulare Biotechnologie Kitz, Kerstin; Research Centre Applied Biocatalysis, Molekulare Biotechnologie Schreiner, Ulrike; Research Centre Applied Biocatalysis, Molekulare Biotechnologie Hoermann, Astrid; Graz University of Technology, Institute for Molecular Biotechnology Hartner, Franz; Graz University of Technology, Institute for Molecular Biotechnology Glieder, Anton; Research Centre Applied Biocatalysis, Molekulare Biotechnologie
Primary Keywords:	White/Industrial Biotechnology
Secondary Keywords:	Genetic Engineering
Keywords:	real-time PCR, <i>Pichia pastoris</i> , CaIB



1  
2 Technical Report  
3  
4  
5  
6  
7  
8  
9

10 **Real time PCR based determination of gene copy numbers in *Pichia***  
11 ***pastoris***  
12  
13  
14  
15  
16  
17

18 Sandra Abad <sup>1,\*</sup>, Kerstin Kitz <sup>1,\*</sup>, Astrid Hörmann <sup>2</sup>, Ulrike Schreiner<sup>1</sup>, Franz S. Hartner  
19  
20 <sup>1</sup>, Anton Glieder <sup>1,2,§</sup>.  
21  
22  
23  
24

25 <sup>1</sup>Applied Biocatalysis Research Centre, Petersgasse 14, 8010 Graz, Austria  
26

27 <sup>2</sup>Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14,  
28  
29 8010 Graz, Austria  
30

31  
32 \*equally contributing authors  
33

34 §Corresponding author  
35  
36  
37  
38

39 **Keywords:** real-time PCR, *Pichia pastoris*, CalB  
40  
41  
42  
43

44 **Correspondence:** Dr. Anton Glieder  
45

46 Petersgasse 14/II, 8010 Graz, Austria  
47

48 email: [anton.glieder@a-b.at](mailto:anton.glieder@a-b.at)  
49

50  
51 tel: +43 316 8734074  
52

53 fax: +43 316 8739302  
54  
55  
56  
57  
58  
59  
60

**ABSTRACT**

*Pichia pastoris* is a preferred host for heterologous protein production. Expression cassettes are usually integrated into the genome of this methylotrophic yeast. This manuscript describes a method for fast and reliable gene copy number determinations for *P. pastoris* expression strains. We believe that gene copy number determinations are important for all researchers working with *P. pastoris* and also many other research groups using similar gene integration techniques for the transformation of other yeasts. The described method uses Real Time PCR to quantify the integrated expression cassettes. Similar methods were employed previously for other host systems such as animal and plant cells but no such method comparing different detection methods and describing details for yeast analysis by quantitative PCR is known to us especially not for methylotrophic yeasts such as *P. pastoris*. Neglecting gene copy numbers can easily lead to false interpretations of experimental results from codon optimization or promoter studies and coexpression of helper proteins as demonstrated in an application example, which is also described in this manuscript.

## INTRODUCTION

The methylotrophic yeast *Pichia pastoris* has emerged to a frequently used protein expression host. High product titers have been obtained with the standard expression system [1, 2] employing the methanol inducible *AOX1* promoter ( $P_{AOX1}$ ). Nevertheless, recent studies reported even higher expression rates when the target gene was under the control of deletion variants of the *AOX1* promoter [3] or in some cases also with the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter ( $P_{GAP}$ ) [4]. Depending on the target gene, high copy number expression strains do not necessarily lead to optimal expression rates. For example, for some secreted proteins rising numbers of expression cassettes resulted in a decreased expression due to other bottlenecks than transcript levels [5, 16]. Other approaches such as co-expression of helper proteins e.g. chaperones or proteins which initiate the unfolded protein response have shown reasonable expression improvements [6-9]

Although many studies describe improved expression due to different promoters variants, codon optimization or co-expression of helper proteins, hardly any information of copy numbers of the expression cassettes is given. Consequently, false interpretation of experimental results is possible, leaving the question whether improvements either correlated with a certain promoter, co-expressed chaperone or an accompanying copy number effect. Inan *et al.* pointed out the importance of the copy number determination for *P. pastoris* expression strains, reporting the correlation between increased protein secretion and the copy number of the co-expressed chaperone protein disulfide isomerase from *P. pastoris*, *PpPDI*. Copy numbers were determined by southern blot analysis, which was so far the method of choice for this question. However, it is a time-consuming and laborious method, and requires large amounts of genomic DNA. In addition, restriction site loss during

1  
2 integration is possible, complicating the molecular analysis. To overcome these  
3  
4 drawbacks and also in order to deal with increasing numbers of expression strains  
5  
6 from high throughput experiments, an additional reliable and faster method for copy  
7  
8 number determination of integrated expression cassettes was needed.  
9

10  
11 Quantitative PCR emerged as an important and widely used analytical tool (reviewed  
12  
13 in [7]). But, there are only few studies reporting on real time PCR based copy number  
14  
15 determination in human-, animal- and plants cells [10, 11]. No detailed method for  
16  
17 yeast could be found, probably also since for many yeasts plasmid systems are still  
18  
19 quite common. In general, assay design, template preparation and analytical  
20  
21 methods are essential for accurate quantitative gene amplification [12, 13] and were  
22  
23 especially adapted for *P. pastoris* in this study as a fast and simple alternative to  
24  
25 southern blotting.  
26  
27

28  
29 Two settled detection methods, namely TaqMan and SYBR Green were compared,  
30  
31 as well as approaches for absolute and relative quantification. For a simple  
32  
33 visualization of our results and for a correlation between copy numbers and  
34  
35 expression, green fluorescent protein (GFP) fusion with a Zeocin resistance marker  
36  
37 has been chosen as a simple intracellular reporter system [14]. Furthermore the  
38  
39 importance of expression cassettes quantitation in *P. pastoris* was emphasized in an  
40  
41 industrially relevant example. Real time PCR was already used for copy number  
42  
43 determination in *P. pastoris* in our group and independently also by others before [3,  
44  
45 15, 16]. A demonstration of the importance of copy number determination as a  
46  
47 routinely performed experiment for molecular strain characterization and a first  
48  
49 detailed protocol especially adapted for *P. pastoris* and including the comparison of  
50  
51 two different detection methods, were the main goals of this work.  
52  
53  
54  
55  
56  
57  
58  
59  
60

## MATERIALS AND METHODS

## Chemicals and materials

Sterile water was purchased from Fresenius Kabi Austria (Graz, Austria). Unless otherwise stated, all chemicals were purchased from Carl Roth GmbH (Karlsruhe, Germany), Becton, Dickinson and Company (Franklin Lakes, NJ, USA) and Sigma-Aldrich (St Louis, MO, USA). DNA modifying enzymes were supplied by Fermentas GmbH (Burlington, Ontario, Canada). All PCRs were performed with Phusion HF according to the producer's manual (Finnzymes Oy, Espoo, Finland).

## GFP reporter expression cassette

$P_{AOX1}$  and the *AOX1* terminator were amplified from vector pPICZ-B (Invitrogen, Carlsbad, USA), and cycle-3-GFP-Zeocin from pTracer<sup>TM</sup>-CMV2 [14, 17]. All PCR products were purified by agarose gel electrophoresis prior to overlap extension PCR [18].

The obtained PCR product displayed 2.4 kb and was purified by agarose gel electrophoresis and gel extraction using Wizard<sup>®</sup> SV Gel and PCR Clean-Up system (Promega, Madison, WI, USA). pAOX was constructed by digesting the  $P_{AOX1}$ -GFP-Zeo-*AOX1*TT fragment with *KpnI* and *NotI* and ligation with *KpnI/NotI*-digested pBlueScript<sup>®</sup> SK<sup>-</sup> vector (Stratagene, La Jolla, CA, USA).

The *P. pastoris*  $P_{AOX1d6^*}$  promoter variant was excised from vector pPICZ-GFP-d6\* [3] employing *BglII/EcoRI* restriction sites into the vector pAOX thereby generating vector pAOXd6\*. The correct sequence was verified by sequencing (AGOWA Genomics, Berlin, Germany).

## *P. pastoris* transformation

*BglII* and *NotI* linearized pAOX or pAOX $\Delta$ 6\* was transformed into electro-competent *P. pastoris* cells. Electro-transformation was performed as described in [19]. After a regeneration phase, aliquots were spread on MSM-Zeo agar plates (1.5% Bacto<sup>TM</sup>

1  
2 Agar, 1.34% Difco™ Yeast Nitrogen Base w/o Amino Acids, 4·10<sup>-5</sup>% d-Biotin, 2%  
3  
4 Sorbitol, 1% methanol, 100 µg/mL Zeocin™) and incubated for 3-4 days at 30 °C.  
5  
6

### 7 ***Pichia pastoris* microscale cultivation and GFP reporter expression**

8  
9 *P. pastoris* cultivation for GFP expressing strains was performed as described in [20].  
10  
11 GFP intensity was detected with SPECTRA MAX Gemini XS plate reader (Molecular  
12  
13 Devices Corp., Sunnyvale, CA, USA) as described in [3].  
14  
15

### 16 **Generation of a *Pichia pastoris* genomic DNA-library**

17  
18 Genomic DNA (gDNA) from *P. pastoris* X-33 was isolated employing the Easy-  
19  
20 DNA™-Kit (Invitrogen). 10µg from the DNA was partially digested with *Sau3AI*. Gel  
21  
22 purified fragments (6-9 kb) were ligated into *XhoI* digested pGAPZ plasmid  
23  
24 (Invitrogen) which was partially filled with Klenow fragment. Ligation was transformed  
25  
26 into *E.coli* Top 10 F' (Invitrogen). Transformation was plated into LB-Zeocin (25  
27  
28 mg/mL). Transformants were pooled and cultivated for plasmid isolation with Wizard®  
29  
30 Plus SV (Promega) according to manufacturer's protocol. The library contained  
31  
32 10,500 clones and the average insert size of 14 independent transformants was  
33  
34 estimated to 4.5 kb. The probability to find any given unique sequence in this library  
35  
36 is 99%.  
37  
38  
39  
40  
41

### 42 ***CalB* reporter strain construction**

43  
44 Zeocin resistance cassette from pPICZ-B (Invitrogen) was replaced with the KanMX6  
45  
46 cassette from PFA6a-kanMX6 [21], which confers resistance against Kanamycin in  
47  
48 *E. coli* and Geneticin (G418) in yeasts. The resulting plasmid was named pPICK.  
49  
50 Codon optimized *Candida antartica* lipase B (*CalB*, GenBank: ACIO6118) was  
51  
52 assembled via oePCR. The resulting gene was *EcoRI/NotI* digested, gel purified and  
53  
54 cloned into an equally digested pPICK-plasmid. *E.coli* Top 10 F' (Invitrogen) was  
55  
56 transformed with this construct and plated onto LB-Kanamycin (50 mg/mL). The  
57  
58 construct was analyzed by sequencing, and linearized with *BglII* for *P. pastoris*  
59  
60



1  
2 KM71H (Invitrogen) transformation as described before. After transformation the cell  
3  
4 suspensions were plated on YPD-G418-plates (300 mg/L) and incubated at 28 °C for  
5  
6 2 days. Transformants were cultivated in deep-well plates as described before, and  
7  
8 screened for *CalB* activity.  
9

### 10 11 **Transformation of X-33 gDNA library into *CalB* reporter strain**

12  
13 The *P. pastoris CalB*-H5 (multicopy *CalB* strain) was transformed with 3 µg of *Bgl*II  
14  
15 linearized genomic DNA (gDNA) library. Cells were plated in YPD-Zeocin-plates (50  
16  
17 mg/L) and incubated at 28 °C for 2 days. Using a Genetix QPixII robot (Genetix  
18  
19 Limited, Hampshire, UK) the transformants were picked into 384-plates filled with 50  
20  
21 µL BMD1%. Incubation was carried out at 28 °C for 5 days. Glycerol-stocks were  
22  
23 prepared by addition of 30 µL of glycerol/BMD1% to a final concentration of 15%  
24  
25 glycerol. The 384-plates were covered with foil seal (Silverseal Aluminium-tape,  
26  
27 Greiner, Frickenhausen, Germany) and stored at -80 °C. Transformants were  
28  
29 replicated into deep-well plates and cultivated as described before with one minimal  
30  
31 modification, cell harvest after 60 h of methanol induction.  
32  
33  
34  
35

### 36 37 ***CalB* activity assay**

38  
39 *P. pastoris* strains were cultivated in deep-well plates as described before. Esterase  
40  
41 activity determination of *CalB* was performed according to [22] with minimal  
42  
43 modifications. Briefly, 20 µL of the supernatant or a dilution of it was pipetted into a  
44  
45 micro-titer plate well. The *CalB* esterase activity was measured following color  
46  
47 development (due to *p*-nitro phenolate formation) at 405 nm for 5min at RT after  
48  
49 addition of 180 µL of the assay solution. The assay solution consisted of 300 mM  
50  
51 Tris/HCl pH7, 1% ethanol and 4 mM of *p*-nitrophenyl butyrate. One unit was defined  
52  
53 as the formation of 1 µmol *para* nitrophenol per minute ( $\epsilon = 17,700 \text{ M}^{-1}\text{cm}^{-1}$ )  
54  
55  
56  
57  
58

### 59 60 **Recovery of the library-insert from the *CalB*-clones**

1  
2 gDNA from selected strains was isolated employing the Easy-DNA™-Kit from  
3 Invitrogen. 500 ng gDNA were digested with *Bgl*II, after purification with Wizard® SV  
4 Gel and PCR Clean-Up system (Promega), self-ligation with T4-ligase was incubated  
5 at 16°C over night. 2 µL of the ligation mixture were transformed into chemical  
6 competent One Shot® Top10 (Invitrogen). Resulting transformants were sent for  
7 sequencing (AGOWA genomics).  
8  
9

### 16 Isolation of genomic DNA for RT-PCR

17  
18 The isolation procedure was performed as described by Hoffman and Winston [23].  
19 The DNA concentration was measured at 260 nm. DNA quality was checked by gel  
20 electrophoresis and photometrically (SPECTRAMax PLUS plate reader, Molecular  
21 Devices Corp.). Samples displaying 260 nm/280 nm ratios higher than 1.8 were used  
22 for further analysis. Alternatively gDNA was isolated employing the E.N.Z.A.® Yeast  
23 DNA Kit (Omega Bio-Tek, Norcross, Georgia, USA) and the Easy-DNA™-Kit  
24 (Invitrogen) as described by the suppliers protocols. gDNA isolation was also  
25 performed with Y-PER Yeast Protein Extraction Reagent (Thermo Fisher Scientific  
26 Inc., Rockford, Illinois, USA). Briefly, the cells of a 10-15 mL an overnight culture were  
27 lysed with 300 µL Y-PER (30 min, room temperature, mixed on an Eppendorf shaker  
28 with 600 rpm). Lysed cells were spun down at 13,000 rpm, 10 min at room  
29 temperature. The supernatant then was subjected to a chloroform/phenol/isopropanol  
30 extraction followed by ethanol precipitation. Finally, the DNA pellet was resuspended  
31 with 50 µL H<sub>2</sub>O, dest.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

### 51 Primer sets for real time PCR

52  
53 The primer design was performed using *Primer Express v2.0* software (Applied  
54 Biosystems, Foster City, CA, USA). The primers had similar melting temperatures  
55 ( $T_m$ ) (58 – 60°C) and the  $T_m$  of the probe was 10°C higher. The amplicon size  
56 settings were defined between 80 and 150 bp. The nucleotide sequences are shown  
57  
58  
59  
60

1  
2 in Table 1. Primers were purchased from Invitrogen and probes were purchased from  
3  
4 MWG-Biotech (Ebersberg, Germany).

5  
6 The real time PCR assays were optimized with respect to the primer concentration by  
7  
8 titration experiments. Thereby, combinations of different concentrations (100 nM –  
9  
10 300 nM) of forward and reverse primers for *GFP* and *ARG4* genes, respectively,  
11  
12 were tested, using the same template amount per real time PCR reaction (see  
13  
14 below). The primer combinations with lowest Ct values and no occurrence of by-  
15  
16 products in the non-template control (NTC) reactions were considered as optimal.  
17  
18

### 19 20 **Real time PCR using SYBR Green and TaqMan probes**

21  
22 Real-time PCR amplification was performed using an ABI 7300 instrument with  
23  
24 Sequence Detection Software SDS v1.2 (Applied Biosystems). Real-time PCR  
25  
26 mixtures were prepared using the TaqMan<sup>®</sup> Universal PCR Master Mix or Power  
27  
28 SYBR<sup>®</sup> Green Master Mix (both Applied Biosystems), respectively.  
29  
30

31  
32 For TaqMan probe based determinations, a single reaction contained 1x TaqMan<sup>®</sup>  
33  
34 Universal PCR Master Mix, 200 nM of each primer and 125 nM of the corresponding  
35  
36 FAM/TAMRA labeled probe. For SYBR Green examinations, 200 nM of each primer  
37  
38 were added to the 1x Power SYBR<sup>®</sup> Green Master Mix. In both cases, reactions were  
39  
40 performed in triplicate for the method establishment and in duplicate for faster routine  
41  
42 strain analysis. 2 ng of genomic DNA were added in a reaction volume of 20  $\mu$ L.  
43  
44

45  
46 For each unknown strain, reactions with GFP and ARG4 primer sets were performed  
47  
48 in separate tubes. Standard curves and NTCs for both genes were recorded in every  
49  
50 plate. The reactions were done in MicroAmp<sup>®</sup> Optical 96-Well Reaction Plates sealed  
51  
52 with MicroAmp<sup>®</sup> Optical Adhesive Covers (Applied Biosystems).  
53  
54

55  
56 The thermal profile initiates with a 10 min step at 95°C followed by 40 cycles of  
57  
58 15 sec at 95°C and 1 min at 60°C. The data collection of the fluorescence signal was  
59  
60 performed at the end of the elongation step. These settings were the same for both

1  
2 detection strategies. In case of the SYBR Green samples, the amplification period  
3  
4 was followed by a melting curve analysis with a temperature gradient of 0.1 °C/s from  
5  
6 70 to 95 °C to exclude amplification of unspecific products.  
7  
8

### 9 **Establishment of standard curves for copy number determination**

10  
11 The GFP-Zeo strain X-33 pAOX D2 was first identified as single copy strain via two  
12  
13 colony PCRs. The first primer pair served as control and binds in the coding region of  
14  
15 the GFP gene, the second pair was set on the origin of replication region (forward) of  
16  
17 the integrated linearized plasmid and in the P<sub>AOX1</sub> (reverse). One or multiple tandem  
18  
19 integration of the expression cassette delivers a defined band with both primer  
20  
21 combinations, whereas the single copy integration only displays a band with the  
22  
23 control primer setting. The strain was confirmed by previously measured intracellular  
24  
25 fluorescence level as a single copy strain due to its low value in the  
26  
27 fluorescent/transformant landscape and selected as calibrator for the establishment  
28  
29 of the *GFP* and *ARG4* standard curves. In addition the integration of further  
30  
31 expression cassettes in other orientations or in other loci was excluded by southern  
32  
33 blot analysis (data not shown). The standard curves cover a copy quantity range from  
34  
35 1.2 x 10<sup>4</sup> to 3.1 x 10<sup>6</sup> and 1.1 x 10<sup>5</sup> to 3.1 x 10<sup>6</sup> copies per reaction for SYBR Green  
36  
37 and TaqMan detection, respectively.  
38  
39  
40  
41  
42  
43

44 For calculations with the Sequence Detection Software SDS v1.2 (Applied  
45  
46 Biosystems), the copy quantity equation as described in [24] was used, considering a  
47  
48 *P. pastoris* genome size of 9.7 Mbp [25] since at the time when this work was  
49  
50 performed the genome sequence was not public available. Resulting in 94,000  
51  
52 copies of the genome present in 1 ng of haploid *P. pastoris* genomic DNA.  
53  
54 Meanwhile the published sequence (GenBank: FN392319.1, FN392320.1,  
55  
56 FN392321.1, FN392322.1, FN392323.1, FN392324.1 and FN392325.1) which still  
57  
58 contains a few gaps shows a similar size of about 9.4 Mbp [26].  
59  
60

The mean Ct values were plotted against the  $\log^{10}$  of their initial template copy quantity and standard curves were generated by a linear regression of the plotted points. The PCR efficiency was calculated from the slope of each standard curve by the following equation [24]:

$$E = 10^{-1/\text{slope}} - 1$$

Samples displaying Ct values lower than 26 and standard deviations lower than 0.3 were used for copy number determinations according to absolute and relative calculations.

### Absolute and relative quantification

For correct determination of the starting copy quantity regardless of precise amounts and qualities of input genomic DNAs, the reference gene *ARG4* was also quantified in parallel. The normalized copy number calculated by absolute quantification is given by the following equation [11]:

$$\text{Copy number}_{\text{TARGET GENE}} = \frac{\text{Copy quantity}_{\text{TARGET GENE}}}{\text{Copy quantity}_{\text{ARG4GENE}}}$$

Relative quantification of the copy number was performed according to the  $2^{-\Delta\Delta Ct}$  method [27], where  $\Delta\Delta Ct = \Delta Ct$  of target –  $\Delta Ct$  of calibrator;  $\Delta Ct = Ct$  of target or calibrator –  $Ct$  of reference (*ARG4*);  $E =$  PCR efficiency.

$$\text{Copy number}_{\text{TARGET GENE}} = (1 + E)^{-\Delta\Delta Ct}$$

A strain with known copy number (in our case *P. pastoris* X-33 pAOX D2), ideally with one copy of the target gene served as a calibrator strain.

## RESULTS AND DISCUSSION

The quality of the DNA sample is one key step for successful quantification via RT-PCR [13]. Therefore the search of a gDNA preparation method which delivers suitable gDNA with a fast isolation protocol was an important step. *P. pastoris* standard fast isolation protocols using commercially available chemicals delivered high yields of gDNA, but also caused DNA degradation (Figure 1). Excellent results were obtained using glass beads and phenol/chloroform extraction as described by Hoffmann and Winston [23].

In total, six *P. pastoris* strains carrying the GFP-Zeocin expression construct were chosen as models for the establishment of quantitative RT-PCR in *P. pastoris*. gDNA of these *P. pastoris* strains, two under the control of the wild type  $P_{AOX1}$  (named X-33 pAOX D2 and E2) and four under the control of the deletion variant  $P_{AOX1d6^*}$  (named X-33 pAOXd6\* A5, D10, F9 and F2) were analyzed (Table 1). Assuming that GFP expression and therefore the measured fluorescence intensity increases with the number of integrated copies we expected that *P. pastoris* X-33 pAOX D2 was a single copy strain. This strain was used for the generation of the standard curve required for the real time PCR experiments.

gDNA standard curves were generated in triplicate determinations and showed to be linear over a range from  $1.2 \times 10^4$  to  $3.1 \times 10^6$  copies (based on an estimated 9.7 Mbp full genome size, see material and methods) in the case of SYBR Green detection. A tighter linear range was detected with the TaqMan method in our hands. The standard curves showed high efficiency ( $>0.92$ ) and were approximately equal for the target and the reference gene, which is a prerequisite for copy number calculations using the relative  $2^{-\Delta\Delta C_t}$  method [27]. As at the time when we started these studies no reports for real time PCR performed in *P. pastoris* were available,

1 we decided to use the *ARG4* gene as an endogenous control [28] as this is a single  
2 copy gene of the haploid host. In this comparative technique, the amount of the  
3 target gene *GFP* is normalized by using the reference gene *ARG4* and set relative to  
4 the calibrator, which was in our case the single copy X-33 pAOX D2 strain (Figure 2).  
5 Hence, the *GFP/ARG4* ratio of a sample was normalized by the *GFP/ARG4* ratio of  
6 the calibrator.  
7

8  
9 Two detection strategies were followed: 1) sequence-specific TaqMan probes and 2)  
10 SYBR Green. It was tested if the use of SYBR Green leads to similar results as the  
11 more expensive TaqMan probe based detection system, which is described to be the  
12 more sensitive detection method [29] (Table 1). In the range of one to five copies of  
13 the expression cassette the results were similar, no matter if SYBR Green or TaqMan  
14 probe was used. Important to mention are the higher Ct values obtained with the  
15 TaqMan detection method which is in line with the already mentioned narrow linear  
16 range compared with the SYBR Green method. In any case absolute or relative  
17 quantification delivered consistent results, although a modest deviation for the multi  
18 copy strain was detected with the TaqMan method.  
19

20 The quantitation of integrated expression cassettes showed that pAOX D2 and  
21 pAOXd6\* A5, although displaying different GFP intensities, are single copy  
22 transformants (or have identically integrated expression cassettes). These results  
23 confirm that the deletion d6\* leads to a decrease of promoter activity to about 60% [3]  
24 under the described cultivation conditions.  
25

26 Sustaining the aim of a universal tool for copy number determination, which fits for  
27 numbers of different expression constructs, good results have been obtained by  
28 using common sequences (e.g.  $P_{AOX1}$ ), which are present in the integrated  
29 expression cassette and in the *P. pastoris* haploid genome. This setting resolves the  
30 identification of one single copy strain (calibrator), since the wild type strain already  
31

1  
2 contains one copy of the sequence, and designed primers and probes can be used  
3  
4 as universal tools.  
5

6  
7 To show the importance of this method in a laboratory routine, a real case scenario  
8  
9 has been followed. The already mentioned benefits of helper protein co-expression  
10  
11 have been focused in a so called non-rational approach. Therefore, a reporter strain  
12  
13 was transformed with a *P. pastoris* genomic DNA library cloned under the control of  
14  
15 the  $P_{GAP}$ . The *P. pastoris* reporter strain harbored 7-8 copies of a *CalB* expression  
16  
17 cassette under the control of the  $P_{AOX1}$  (named *P. pastoris* pPICK-*CalB* H5), and  
18  
19 displayed rather low levels of active enzyme in the culture supernatant compared to  
20  
21 single copy strains. Previous experiments showed that certain bottlenecks during the  
22  
23 secretion pathway led to a limited growth of the *P. pastoris CalB* strains and low  
24  
25 expression levels of the target gene, which could be at least in part be removed by  
26  
27 co-expression of *PpPDI* (S. Abad unpublished data). Following this outcome, we  
28  
29 aimed at identifying additional putative helper proteins by a second transformation  
30  
31 and individual co-overexpression of *P. pastoris* genomic DNA library clones.  
32  
33 Therefore genomic *P. pastoris* DNA fragments had been ligated behind the GAP  
34  
35 promoter of the pGAPZB expression plasmid. Putative new helper proteins from this  
36  
37 library were expected to lead to an increased *CalB* expression which can be  
38  
39 determined by measuring the lipase activity in the culture supernatant of the reporter  
40  
41 strain.  
42  
43  
44  
45  
46  
47  
48

49 In total 3250 transformants were screened, which seemed to be an acceptable  
50  
51 number for our initial purpose since four candidates (*CalB* H5/2-L18, H5/9-H21,  
52  
53 H5/10-H22, H5/10-F22) displayed higher *CalB* expression levels as the reporter  
54  
55 strain. These improved transformants were chosen for a re-screen and the  
56  
57 expression improvement was confirmed for all candidates. Interestingly, none of the  
58  
59 chosen co-expression strains delivered higher expression values than single copy  
60



1  
2 CalB strains, and no reproducibility of the results was possible after transformation of  
3  
4 the recovered construct into the reporter strain. The copy number determination  
5  
6 performed with the SYBR Green method using primers designed for P<sub>AOX1</sub> sequence  
7  
8 and the *P. pastoris* KM71H strain as calibrator explained the described results. The  
9  
10 chosen “improved” transformants showed lower numbers of integrated CalB  
11  
12 expression cassettes. The transformants H5/2-L18 and H5/10-F22 harbored four and  
13  
14 two copies, respectively while H5/9-H21 and H5/10-H22 displayed only one CalB  
15  
16 expression cassette (Table1).  
17  
18  
19  
20  
21  
22

### 23 **CONCLUDING REMARKS**

24  
25 Quantitative PCR allows the determination of the number of integrated expression  
26  
27 cassettes of *P. pastoris*. Strains harboring up to five copies were analyzed and  
28  
29 delivered consistent results with two different detection methods. However, the  
30  
31 narrow linear range of the TaqMan method should be taken into account for further  
32  
33 experiments. This PCR based procedure provides a reliable but faster and simple  
34  
35 alternative to standard southern blot analysis. It also allows correlations between  
36  
37 integration events into the genome and the respective expression levels which is  
38  
39 highly relevant for expression studies employing different promoters, redesigned  
40  
41 genes or alternative signal sequences. Due to its simplicity it can be applied  
42  
43 routinely..  
44  
45  
46  
47  
48

49 Absolute and relative quantification showed to have no influence in the final outcome  
50  
51 of the experiment, whereas the latter demands more precise parameters. The house-  
52  
53 keeping gene *ARG4* showed to be a reliable endogenous control. However, caution  
54  
55 should be taken when an Arg<sup>-</sup> auxotroph strain is used as platform for expression  
56  
57 experiments.  
58  
59  
60

1  
2 As demonstrated by application of the described method multiple transformation  
3 steps can lead to the potential loss of previously integrated expression cassettes,  
4 especially if the same locus is targeted for the genomic integration. Gene dosage  
5 control of previously integrated expression cassettes after every further  
6 transformation step is essential.  
7  
8  
9  
10  
11  
12

## 13 14 15 16 **ACKNOWLEDGMENTS**

17  
18 The authors acknowledge the FFG, the Province of Styria and SFG for financial  
19 support.  
20  
21  
22  
23  
24  
25  
26  
27

## 28 **CONFLICT OF INTEREST STATEMENT**

29  
30 The authors have declared no conflict of interest.  
31  
32

## 33 **REFERENCES**

- 34  
35  
36 [1] Macauley-Patrick, S., Fazenda, M. L., McNeil, B., Harvey, L. M., Heterologous protein  
37 production using the *Pichia pastoris* expression system. *Yeast* 2005, 22, 249-270.  
38  
39 [2] Lin-Cereghino, J., Cregg, J. M., Heterologous protein expression in the methylotrophic  
40 yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* 2000, 24, 45-66.  
41  
42 [3] Hartner, F. S., Ruth, C., Langenegger, D., Johnson, S. N., *et al.*, Promoter library designed  
43 for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res.* 2008, 36, 76.  
44  
45 [4] Zheng, H., Wang, X., Chen, J., Zhu, K., *et al.*, Expression, purification, and  
46 immobilization of His-tagged D-amino acid oxidase of *Trigonopsis variabilis* in *Pichia*  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2 [5] Hohenblum, H., Gasser, B., Maurer, M., Borth, N., Mattanovich, D., Effects of gene  
3 dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*.  
4 *Biotechnol. Bioeng.* 2004, 85, 367-375.  
5  
6  
7  
8  
9 [6] Gasser, B., Maurer, M., Gach, J., Kunert, R., Mattanovich, D., Engineering of *Pichia*  
10 *pastoris* for improved production of antibody fragments. *Biotechnol. Bioeng.* 2006, 94, 353-  
11 361.  
12  
13  
14  
15  
16 [7] Inan, M., Aryasomayajula, D., Sinha, J., Meagher, M. M., Enhancement of protein  
17 secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. *Biotechnol.*  
18 *Bioeng.* 2006, 93, 771-778.  
19  
20  
21  
22  
23 [8] Tsai, C. W., Duggan, P. F., Shimp, R. L., Miller, L. H., Narum, D. L., Overproduction of  
24 *Pichia pastoris* or *Plasmodium falciparum* protein disulfide isomerase affects expression,  
25 folding and O-linked glycosylation of a malaria vaccine candidate expressed in *P. pastoris*. *J.*  
26 *Biotechnol.* 2006, 121, 458-470.  
27  
28  
29  
30  
31  
32  
33 [9] Vad, R., Nafstad, E., Dahl, L. A., Gabrielsen, O. S., Engineering of a *Pichia pastoris*  
34 expression system for secretion of high amounts of intact human parathyroid hormone. *J.*  
35 *Biotechnol.* 2005, 116, 251-260.  
36  
37  
38  
39  
40 [10] Ballester, M., Castello, A., Ibanez, E., Sanchez, A., Folch, J. M., Real-time quantitative  
41 PCR-based system for determining transgene copy number in transgenic animals.  
42 *Biotechniques* 2004, 37, 610-613.  
43  
44  
45  
46  
47 [11] Laurendeau, I., Bahuau, M., Vodovar, N., Larramendy, C., *et al.*, TaqMan PCR-based  
48 gene dosage assay for predictive testing in individuals from a cancer family with INK4 locus  
49 haploinsufficiency. *Clin. Chem.* 1999, 45, 982-986.  
50  
51  
52  
53  
54 [12] Ginzinger, D. G., Gene quantification using real-time quantitative PCR: an emerging  
55 technology hits the mainstream. *Exp. Hematol.* 2002, 30, 503-512.  
56  
57  
58  
59  
60

- 1  
2 [13] Cankar, K., Stebih, D., Dreo, T., Zel, J., Gruden, K., Critical points of DNA  
3  
4 quantification by real-time PCR-effects of DNA extraction method and sample matrix on  
5  
6 quantification of genetically modified organisms. *BMC Biotech.* 2006, 6, 37.  
7  
8  
9 [14] Bennett, R. P., Cox, C. A., Hoeffler, J. P., Fusion of green fluorescent protein with the  
10  
11 Zeocin<sup>TM</sup>-resistance marker allows visual screening and drug selection of transfected  
12  
13 eukaryotic cells. *Biotechniques* 1998, 24, 478-482.  
14  
15  
16 [15] Schroer, K., Luef, K. P., Hartner, F. S., Glieder, A., Pscheidt, B., Engineering the *Pichia*  
17  
18 *pastoris* methanol oxidation pathway for improved NADH regeneration during whole-cell  
19  
20 biotransformation. *Metab. Eng.* 2009, 8-17.  
21  
22  
23 [16] Marx, H., Mecklenbrauker, A., Gasser, B., Sauer, M., Mattanovich, D., Directed gene  
24  
25 copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA  
26  
27 locus. *FEMS Yeast Res* 2009, 9, 1260-1270.  
28  
29  
30 [17] Cramer, A., Whitehorn, E. A., Tate, E., Stemmer, W. P., Improved green fluorescent  
31  
32 protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* 1996, 14, 315-319.  
33  
34  
35 [18] Liu, Z., Pscheidt, B., Avi, M., Gaisberger, R., *et al.*, Laboratory evolved biocatalysts for  
36  
37 stereoselective syntheses of substituted benzaldehyde cyanohydrins. *ChemBioChem* 2008, 9,  
38  
39 58-61.  
40  
41  
42 [19] Cregg J.M., R. K. A., Transformation, *Methods in Molecular Biology* 1998; 103, pp. 27-  
43  
44 39.  
45  
46  
47 [20] Weis, R., Luiten, R., Skranc, W., Schwab, H., *et al.*, Reliable high-throughput screening  
48  
49 with *Pichia pastoris* by limiting yeast cell death phenomena. *FEMS Yeast Res.* 2004, 5, 179-  
50  
51 189.  
52  
53  
54 [21] Wach, A., Brachat, A., Alberti-Segui, C., Rebischung, C., Philippsen, P., Heterologous  
55  
56 HIS3 marker and GFP reporter modules for PCR-targeting in *Saccharomyces cerevisiae*.  
57  
58 *Yeast* 1997, 13, 1065-1075.  
59  
60

- 1  
2 [22] Zhang, N., Suen, W. C., Windsor, W., Xiao, L., *et al.*, Improving tolerance of *Candida*  
3  
4 *antarctica* lipase B towards irreversible thermal inactivation through directed evolution.  
5  
6 *Protein Eng.* 2003, *16*, 599-605.  
7  
8  
9 [23] Hoffman, C. S., Winston, F., A 10 minute DNA preparation from yeast efficiently  
10  
11 releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 1987, *57*, 267-  
12  
13 272.  
14  
15  
16 [24] Lee, C., Kim, J., Shin, S. G., Hwang, S., Absolute and relative qPCR quantification of  
17  
18 plasmid copy number in *Escherichia coli*. *J. Biotechnol.* 2006, *123*, 273-280.  
19  
20  
21 [25] Ohi, H., Okazaki, N., Uno, S., Miura, M., Hiramatsu, R., Chromosomal DNA patterns  
22  
23 and gene stability of *Pichia pastoris*. *Yeast* 1998, *14*, 895-903.  
24  
25  
26 [26] De Schutter, K., Lin, Y. C., Tiels, P., Van Hecke, A., *et al.*, Genome sequence of the  
27  
28 recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* 2009, *27*, 561-566.  
29  
30  
31 [27] Livak, K. J., Schmittgen, T. D., Analysis of relative gene expression data using real-time  
32  
33 quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 2001, *25*, 402-408.  
34  
35  
36 [28] Lin Cereghino, G. P., Lin Cereghino, J., Sunga, A. J., Johnson, M. A., *et al.*, New  
37  
38 selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of  
39  
40 *Pichia pastoris*. *Gene* 2001, *263*, 159-169.  
41  
42  
43 [29] Bubner, B., Baldwin, I. T., Use of real-time PCR for determining copy number and  
44  
45 zygosity in transgenic plants. *Plant Cell Rep.* 2004, *23*, 263-271.  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**FIGURE LEGENDS**

**Figure 1:** Comparison of *P. pastoris* gDNA quality of diverse isolation methods.

G: GeneRuler 1kb (Fermentas),  $\lambda$ :  $\lambda$ -HindIII Standard (Fermentas), 1: 5  $\mu$ L of gDNA isolated using Y-PER (Novagen) and chloroform/phenol extraction followed by ethanol precipitation, 2: 5  $\mu$ L of gDNA isolated according to Easy DNA<sup>TM</sup>-Kit (Invitrogen), 3: 5  $\mu$ L of gDNA isolated using a E.Z.N.A.<sup>®</sup> Yeast DNA Kit, 4: 5  $\mu$ L of gDNA isolated using glass beads and phenol/chloroform extraction as described by Hoffmann and Winston [23].

**Figure 2:** RT-PCR determination using SYBR Green and TaqMan methods. Target and reference gene amplification of *P. pastoris* GFP-Zeocin strains and standard curve. A: SYBR Green detection, B: TaqMan detection.

**Table 1:** Copy number calculation according to absolute (Abs. Q) and relative (Rel. Q) quantification.

<i>Pichia pastoris</i> strain	SYBR Green <sup>a)</sup>		TaqMan <sup>b)</sup>		Resulting copy number	Measured Expression GFP: [RFU], CalB: [mU/L]	Expected copy number <sup>e)</sup>
	Abs. Q	Rel. Q	Abs Q.	Rel Q.			
X-33 pAOX D2	1.0	1.0	0.8	1.0	1 <sup>c)</sup>	987 ± 81	1
X-33 pAOX E2	1.7	1.8	1.6	2.0	2 <sup>c)</sup>	2068 ± 219	2
X-33 pAOXΔ6* A5	1.1	1.1	1.0	1.2	1 <sup>c)</sup>	422 ± 22	1
X-33 pAOXΔ6* D10	1.9	2.1	2.0	2.4	2 <sup>c)</sup>	1063 ± 189	2
X-33 pAOXΔ6* F9	2.0	2.1	1.8	2.2	2 <sup>c)</sup>	925 ± 58	2
X-33 pAOXΔ6* F2	4.6	4.8	4.0	5.0	4-5 <sup>c)</sup>	1678 ± 437	4-5
KM71H pPICK-CalB H5	8.9	8.6	-	-	7-8 <sup>d)</sup>	0.033 ± 0.005	-
KM71H CalB H5/2-L18	5.7	5.8	-	-	4-5 <sup>d)</sup>	0.390 ± 0.051	-
KM71H CalB H5/9-H21	2.2	2.3	-	-	1 <sup>d)</sup>	0.853 ± 0.086	-
KM71H CalB H5/10-H22	2.6	2.7	-	-	1-2 <sup>d)</sup>	0.876 ± 0.054	-
KM71H CalB H5/10-F22	3.1	3.1	-	-	2 <sup>d)</sup>	0.955 ± 0.075	-

<sup>a)</sup> ARG(sense/antisense): 5'-TCCTCCGGTGGCAGTTCTT-3'/5'-TCCATTGACTCCCGTTTTGAG-3' and

GFP(sense/antisense): 5'-AAATTTAAGGGTAAGCTTTCCGTATG-3'/5'-ATGGTGATGTTAATGGGCACAA-3' or

P<sub>AOX1</sub>(sense/antisense): 5'-GAAGCTGCCCTGTCTTAAACCTT-3'/5'-CAAAGCTTGTC AATTGGAACCA-3'

<sup>b)</sup> GFP-probe: 5'-ATCACCTTCACCCTCTCCACTGACAGAAAA-3', ARG-probe: 5'-TTCAACACTGGCTTCAAAGTCAAACGTTGAA-3'

<sup>c)</sup> based on absolute and relative quantifications results with SYBR Green and TaqMan methods.

<sup>d)</sup> based on results of absolute and relative quantifications with SYBR Green method. The endogenous P<sub>AOX1</sub> sequence has been subtracted.

<sup>e)</sup> based on fluorescence measurement assuming a linear correlation between GFP intensity and the number of integrated copies.

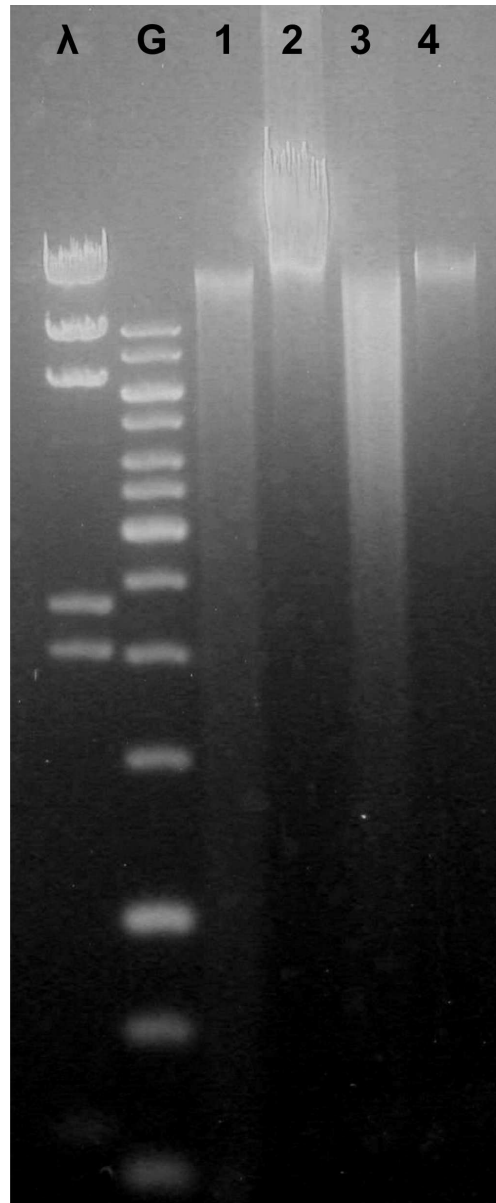


Figure 1: Comparison of *P. pastoris* gDNA quality of diverse isolation methods.  
G: GeneRuler 1kb (Fermentas),  $\lambda$ :  $\lambda$ -HindIII Standard (Fermentas), 1: 5  $\mu$ L of gDNA isolated using Y-PER (Novagen) and chloroform/phenol extraction followed by ethanol precipitation, 2: 5  $\mu$ L of gDNA isolated according to Easy DNATM-Kit (Invitrogen), 3: 5  $\mu$ L of gDNA isolated using a E.Z.N.A.® Yeast DNA Kit, 4: 5  $\mu$ L of gDNA isolated using glass beads and phenol/chloroform extraction as described by Hoffmann and Winston [23].  
86x208mm (300 x 300 DPI)



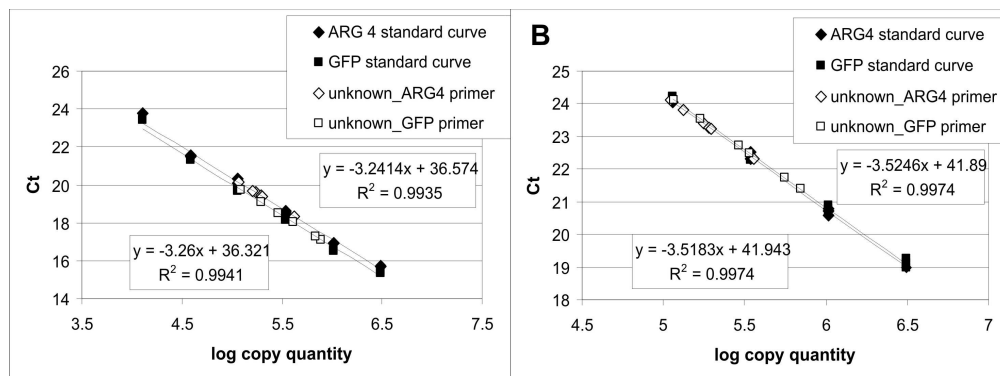


Figure 2: RT-PCR determination using SYBR Green and TaqMan methods. Target and reference gene amplification of *P. pastoris* GFP-Zeocin strains and standard curve. A: SYBR Green detection, B: TaqMan detection.  
249x92mm (300 x 300 DPI)