**TECHNISCHE
UNIVERSITÄT
WIEN**
Vienna University of Technology

DISSERTATION

Methods of improving *Trichoderma reesei* as a whole cell biocatalyst

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften

unter der Leitung von
Ao.Univ.Prof. Mag. Dr.rer.nat. Robert Mach

am Institut für
Verfahrenstechnik, Umwelttechnik und techn. Biowissenschaften
E166

eingereicht an der
Technischen Universität Wien
Fakultät für Technische Chemie

von

Dipl.-Ing. Birgit Jovanović
Matrikelnummer 0225219
Traiskirchnerstraße 19
2512 Tribuswinkel

Wien, am 14. November 2013

Kurzfassung

Trichoderma reesei ist ein filamentöser Ascomycet, der aufgrund seiner Fähigkeit, große Mengen an Cellulasen und Hemicellulasen zu produzieren, von weitreichendem biotechnologischem Interesse ist. Desweiteren ist der Pilz dadurch in der Lage, auch komplexe Substrate aus Lignocellulose zu verwerten, wodurch sich eine Verwendung als Ganzzellbiokatalyst anbietet, die als Edukt billige und nachwachsende Abfallprodukte aus der Landwirtschaft einsetzt. Ein Schwerpunkt dieser Arbeit war, die Möglichkeit zu untersuchen, den künstlichen Süßstoff Erythritol mit *T. reesei* herzustellen, wobei als Substrat Weizenstroh eingesetzt wurde, das mit einem alkalischen Organosolvprozess vorbehandelt worden war. Dazu wurde zuerst das produzierende Enyzm Eryhrosereduktase identifiziert und charakterisiert, danach wurden Überexpressionsstämme des Wildtyps sowie des Mutantenstamms Rut-C30 hergestellt und deren Erythritolproduktion untersucht. Ein zweiter Aspekt, zur Verbesserung der Anwendung von *T. reesei* als Ganzzellbiokatalyst, war die *in vivo*-Footprinting-Technik soweit zu verbessern, dass auch Änderungen in der Methylierbarkeit von *cis*-Elementen sichtbar werden, die von den Kulturbedingungen abhängen. Dazu wurde einerseits die Durchführung der Analyse verbessert, außerdem wurde ein Programm mit dem Namen ivFAST geschrieben, das eine entsprechende Datenauswertung erlaubt.

Abstract

Trichoderma reesei is a filamentous ascomycot with a very high potential to produce cellulases and hemicellulases, which gives it great importance in various biotechnological applications. It also makes the fungus capable of degrading complex lignocellulosic substrates, which makes it attractive for usage as whole cell biocatalyst, utilizing cheap and sustainable biowaste material as starting material. One focus of this work was to investigate the possibility of producing the artificial sweetener erythritol in *T. reesei*, using wheat straw, pretreated by an alkaline organosolve process, as substrate. Therefore, first the producing enzyme erythrose reductase was identified and characterized, and then an overexpression strains from the wild-type as well as from the mutant strain Rut-C30 was constructed in order to enhance production. A second aspect of improving the usability of *T. reesei* as a whole cell catalyst was to generally gain new insights in gene regulation by improving the *in vivo* footprinting technique in a way that even condition-dependent differences in the methylateability of *cis* sites become visible. Therefore, the conduction of the analysis was improved, and additionally, a program that provides the necessary data evaluation, called ivFAST, was written.

Contents

Chapter 1

Introduction

The filamentous ascomycot *Trichoderma reesei* was discovered during World War II on the Solomon Islands, where it disintegrated the cotton canvas of military tents. First identified as *T. viride* QM6a, it took 20 years to recognize the isolate as own species, named *T. reesei* [1]. Remarkably, the originally isolated strain QM6a with its high potential to produce cellulases is the ancestor strain of all biotechnologically used strains. For a long time only the anmorph of *Trichoderma reesei* was known, until in 1996 Kuhls *et. al.* [2] identified the telemorph *Hypocrea jecorina*. *T. reesei* is a very strong producer of cellulases and hemicellulases, which makes it interesting for industrial applications. A genome analysis using the JGI Genome Portal revealed 10 celluloytic and 16 xylanolytic enzyme-encoding genes [3]. The range of applications for these native lignocellulose-degrading enzymes reaches from pulp and paper industry [4, 5, 6] over food and feed [7, 8, 9] and textile industry [10, 11, 12] as well as biofuel production [13, 14, 15]. A *T. reesei* mutant of special interest for industrial applications is Rut-C30, a carbon catabolite derepressed, cellulase-hyperproducing mutant obtained from QM6a by one UV irradiation step and two rounds of treatment with nitrosoguanidine (NTG). Rut-C30 is the parental strain for many production strains.

In this work we focused on two aspects on how to improve *T. reesei* as whole cell catalyst. One aim was to investigate the possibility of producing erythritol with *T. reesei*, the other aim was to provide an improved *in vivo* footprinting technique that allows detailed studies on condition-dependent differences in the acessability of *cis* elements of promoters, leading to further insights on gene regulation.

1.1 Production of erythritol with whole cell catalysis in *T. reesei*

1.1.1 Properties and applications of erythritol

Erythritol is a four-carbon sugar alcohol with a sweetness of 60-80 % of sucrose (in a 10 % (w/v) solution), and is therefore mainly used as low-calorie sweetener. Its extremely low energy yield of only up to 0.2 cal/g (compared to about 2 cal/g for other polyols) results from the fact that erythritol is barely metabolized by the human body, but excreted unchanged with the urine. Blood glucose and insulin levels are not affected by the uptake of erythritol [16], which makes it also suitable to be used with diabetes. A great advantage over other polyols used as sweeteners is that erythritol causes much less laxative distress. As a very small molecule, it is easily absorbed in the upper intestine and therefore, barely fermented by colon-inhabiting microorganisms [17]. Also, erythritol is not assimilated by *Streptococcus mutans* and therefore is non-cariogenic.

As a naturally occurring substance (e.g. in beer, sake, wine, soy sauce, water melon, pear, and grape with levels up to 0.13 % [18]), erythritol has already been consumed by humans for a long time. Intensive studies have been conducted on the safety of erythritol. Tests on cell cultures (bacterial and mammalian) and animals provided no evidence for carcinogenic, mutagenic or teratogenic potential. With an LD_{50} of 5 g/kg and only general symptoms caused by hypertonic solutions, erythritol can be classified as non-toxic [18].

Aside from its usage as sweetener, erythritol also has applications as flavor enhancer, formulation aid, humectant, stabilizer, and thickener. It has some favorable physical and chemical properties for using it as food additive, which are high thermal stability (no decomposition or colorization when kept for 1 h at 200 ℃, makes it suitable for bakery), acid stability (makes it suitable for soft drinks), better crystallizeability and less hygroscopy than sucrose, and a very weak aftertaste. Its negative enthalpy of solution can be used in candies and chewing gums to create a chilling sensation [19].

1.1.2 Current production methods of erythritol

Erythritol is solely produced in biotechnological processes. The chemical synthesis methods are not only all rather complicated and expensive, but also yield equimolar amounts of byproducts like ethylene glycol or threitol, which are hard to separate. Instead, industrial scale production is done in fermentative processes with osmotolerant yeasts in media with high osmotic pressure.

The first erythritol-producing yeasts were isolated by Hajny in 1964 from fresh pollen and assumed to belong to the genus *Torula* [20]. Later, other microorganisms, like *Aureobasidium* sp., *Candida magnoliae*, *Moniliella* sp., *Pichia* sp., *Pseudozyma tsukubaensis* [21], *Trichosporon* sp., or *Yarrowia lipolytica* [22], have been found to produce reasonable amounts of erythritol [16]. Not all of them are applicable for industrial scale production due to the heavy formation of byproducts such as glycerol or ribitol. Industrially used organisms are usually mutagenized by UV irradiation, treatment with NTG or ethylmethane sulfonate (EMS) to improve features like osmotolerance, leading to higher erythritol formation, foaming under aerobic culture condition, and production of byproducts. Examples are the mutant *Aureobasidium* sp. SN124A, received by UV irradiation and NTG treatment, with a yield of 47.6 %, reported by Ishizuka *et. al.* in 1989 [23], or the mutant M2 from *C. magnoliae* KFCC 11023, obtained by EMS treatment, with a yield of 43 % based on glucose conversion to erythritol, reported by Yang *et. al.* in 1999 [24]. In a recent work from 2011, Savergave *et. al.* reported an erythritol production without formation of other polyols in the *C. magnoliae* NCIM 3470 mutant R23 with a yield of 31.1 % on glucose [25].

All of the production methods of erythritol mentioned above have the disadvantage of using glucose or sucrose as carbon source in high concentrations to obtain high osmotic pressures, which is not favorable from a (socio-)economical point of view. It would be an interesting perspective to produce the erythritol instead from a renewable biowaste source that is not in competition with usage as food. There are already some studies on erythritol production from crude glycerol obtained as waste product from biodiesel production with *Y. lipolytica* [22, 26, 27], where in 2012 by Tomaszewska *et. al.* a yield of up to 49 % could be reached [26]. The field of useable alternative substrate could be broadened by investigating other groups than yeast-like fungi for their ability to produce erythritol. A lot of filamentous fungi are capable of degrading plant waste material, as already mentioned above for *T. reesei*.

1.1.3 Wheat straw as renewable substrate

Wheat straw is a byproduct of food production, and not in competition with it. With its easy availability and its cheap price, it is an attractive substrate for bioindustry. Wheat straw consists of about one third cellulose, one fourth hemicellulose (mainly glucoronoarabinoxylan, see Fig. 1.1) and one fifth lignin (see Fig. 1.2). Since most microorganisms are not capable of degrading lignin, this structural component can hinder the degradation of the cellulose and the hemicellulose due to spatially blocking enzyme access.
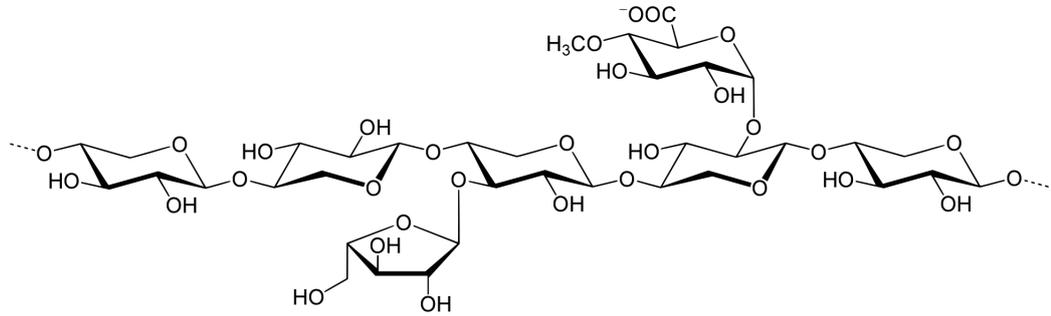
Figure 1.1: Structure of glucuronoarabinoxylane

In order to make the valuable carbon sources better available, the lignin can be removed in a pretreatment of the wheat straw. Such processes are already well established in pulp and paper industry, but normally they include aggressive chemicals and toxic catalysts, preventing later microbial growth on such treated straw. An alternative process was patented by the company Annikki in 2012 [28], which uses only sodium hydroxide and a short chained alcohol at moderate temperatures for the extraction of the lignin. By using ethanol as organic solvent, thoroughly washing is sufficient to make the pretreated wheat straw suitable as substrate, even for species sensitive to alcohol like *T. reesei*.

Figure 1.2: Composition of wheat straw

1.1.4 Characterization of erythrose reductases

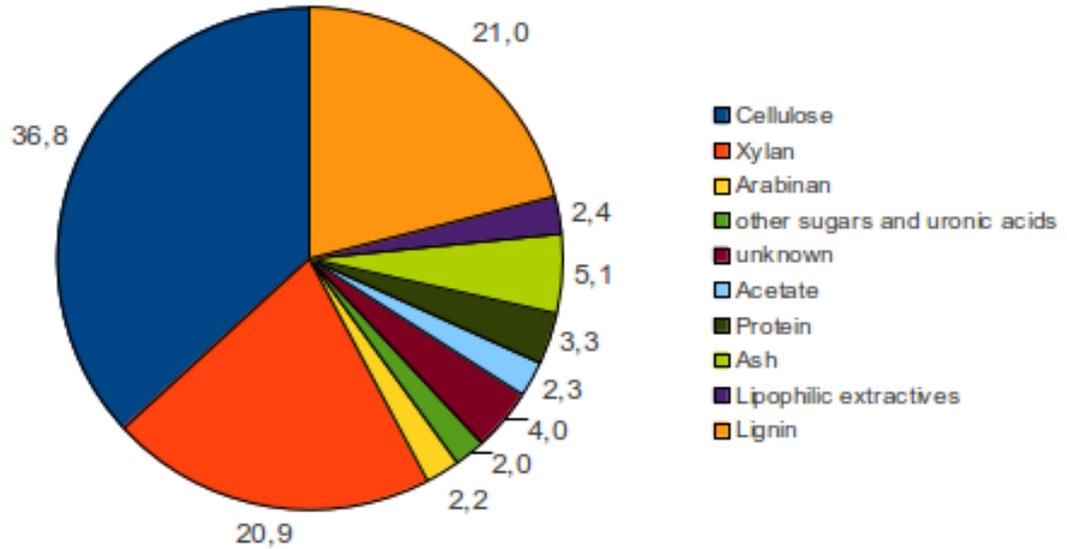Erythritol is produced as a side-product of the pentose phosphate pathway (PPP). The PPP intermediate erythrose-4-phosphate is dephosphorelated and subsequently reduced by an erythrose reductase (see Fig. 1.3). Several such erythrose reductases have been characterized for yeasts, but no such studies have been carried out for filamentous fungi before. In 'Characterization of erythrose reductases from filamentous fungi' ([29], see Appendix A.1 for full text) we identified the erythrose reductases (termed Err1) from *T. reesei*, *Aspergillus niger*, and *Fusarium graminearum* by *in silico* analysis, and characterized the heterogously in *Escherichia coli* expressed proteins by means of enzyme assays.

1.1.5 Production of erythritol in *T. reesei*

After the identification of the erythrose reductase encoding gene (*err1*), the gene has been overexpressed in the *T. reesei* wild type and the mutant strain Rut-C30 ([30] submitted to Microbial Cell Factories, see Appendix A.2 for full text). Therefore, two different constructs were introduced. One with the *err1* under control of the native constitutive pyruvat kinase promoter (p*pki*), named QPEC1 (derived from the wild-type) and RPEC1 (derived from Rut-C30), the other under control of the native β-xylosidase promoter (p*bxl1*), named QBEC2 (derived from the wild-type) and RBEC2 (derived from Rut-C30). The best recombinant strains were identified by transcript analysis of *err1*. The erythritol formation was first studied in shake flasks with D-xylose as easily metabolizeable substrate. Afterwards, the strains were cultivated in bio-reactors with pretreated wheat straw as sole carbon source.

1.2 Improvement of *in vivo* footprinting for condition-dependent identification of *cis* elements

In vivo footprinting allows to identify DNA positions where a protein is bound under the analyzed cultivation conditions. This is of great importance for studies of regulatory mechanisms and transcription factors, which effect gene transcription by binding to specific sites (*cis* elements) in the upstream regulatory region (URR) of a gene.
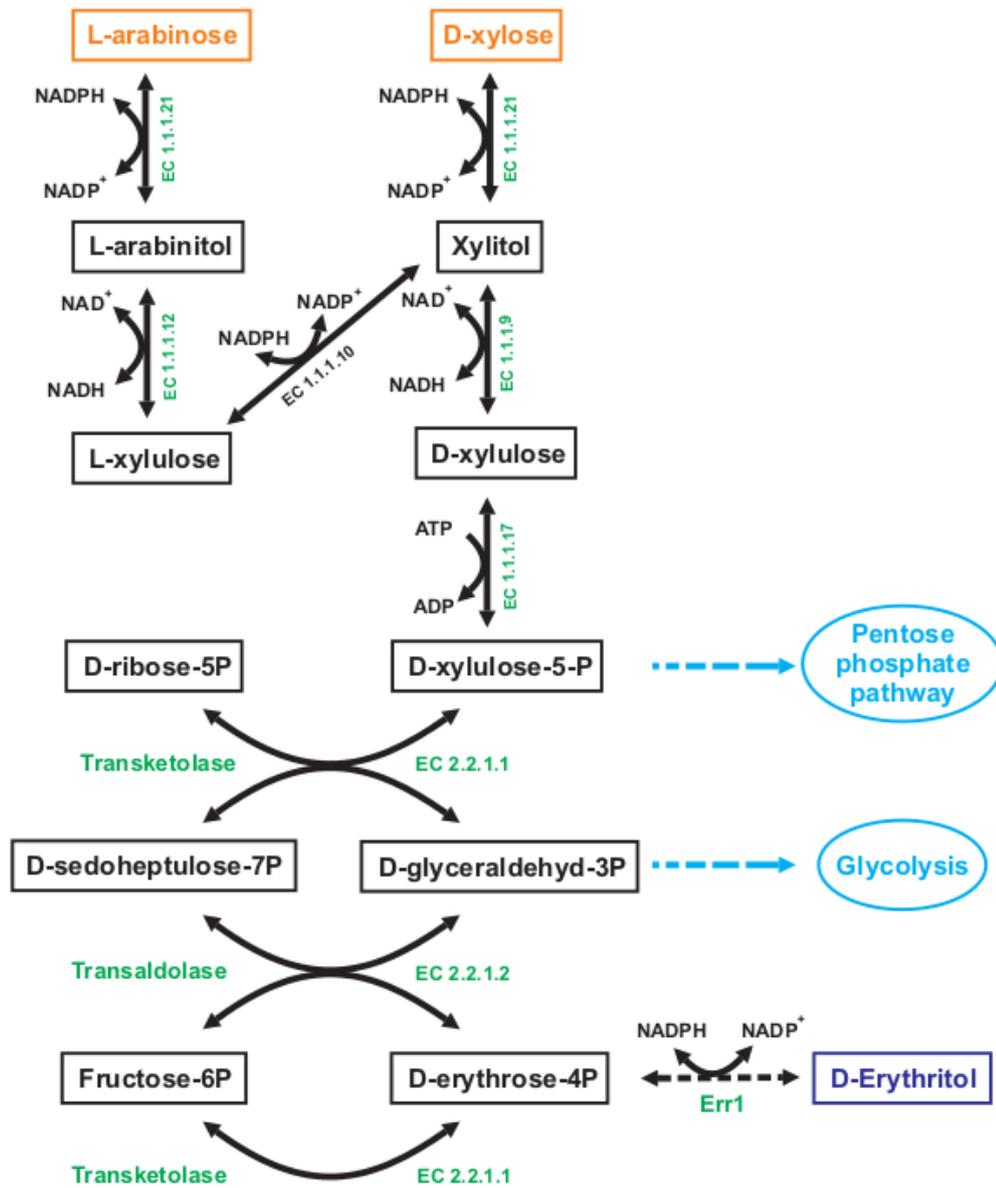
Figure 1.3: Pathway of erythritol production starting from D-xylose and L-arabinose as substrate.

1.2.1 History of *in vivo* footprinting

The first footprinting method to detect protein-DNA interactions was published in 1985 by Jackson and Felsenfeld [31]. It was an *in vitro* method, using the accessibility of DNA, derived from nuclear digests, by DNase I to identify sites blocked by bound proteins. An alternative procedure is to treat the DNA with an alkylating reagent like dimethylsulfate (DMS) and subsequently cut the DNA at the methylated sites, as published originally in 1985 by Ginger *et. al.* [32] and Ephrussi *et. al.* [33]. Compared to DNase I, DMS has the advantage that it easily penetrates intact cells (even cells with cellwalls) and therefore, can be used for *in vivo* methylation in all kinds of organisms. The disadvantage of DMS is that it only methylates purines, so the received information is less complete. A significant improvement of the footprinting technique could be achieved by the introduction of ligation-mediated polymerase chain reaction (LMPCR), originally published in 1989 by Mueller and Wold [34] and Pfeifer *et. al.* [35]. Briefly, LMPCR consists of five steps, which are (1) specific cleavage of the DNA, generating 5'-phosphorelated molecules, (2) primer extension with a specific primer 1, generating a blunt end on one side, (3) blunt end ligation of an asymmetric linker, (4) exponential amplification by PCR with a specific primer 2 and a linker primer, which hybridizes to the ligated sequence, (5) few cycles with a labeled primer 3, which can be used for visualization afterwards. At the early times, the primer labeling was done with radioactive isotopes, and the amplificated fragments were separated on slab gels. Problems with weak or missing bands on the one hand and extra bands on the other hand could be encountered in 1992 by Garrity and Wold [36] by the use of *Thermococcus litoralis* DNA polymerase (Vent polymerase) instead of the originally used Sequenase and *Taq*. Vent polymerase does not possess a terminal deoxynucleotidyltransferase activity, whereas Sequenase (used for the first strand synthesis reaction) adds an extra base to about 50 % of the products and *Taq* (used for the PCR and the labeling reaction) adds an extra base even to about 95 % of the products, leading to serious errors and bad reproducibility. Further method improvements were achieved by the introduction of fluorescence labeling as reported by Huang and Sun in 1999 [37] instead of the radioactive labeling, and the use of automated capillary sequencers instead of slab gels as reported by Trotha *et. al.* in 2002 [38]. Concerning the application of *in vivo* footprinting for filamentous fungi some additional terms have to be considered. Due to the apically growth of the mycelium, it contains cells of different ages, which might be in different physiological condition. Wolschek *et. al.* reported in 1998 [39] the use of germinating spores with germ tubes containing less than 10 nuclei in liquid culture for the use

of *in vivo* footprinting with DMS in *A. niger*, *A. nidulans*, and *Penicillium chrysogenum*.

1.2.2 Regulation of cellulolytic and xylanolytic enzymes in *T. reesei*

As already mentioned above, *T. reesei* is a strong producer of cellulases and hemicellulases. The 10 cellulases produced are two cellobiohydrolases (CBHI and CBHII), six endo-glucanases (EGI - EGVI) and two β-glucosidases (BGLI and BGLII). The most important of the 16 hemicellulases are three endo-xylanases (XYNI - XYNIII) and one β-xylosidase (BXL1). Whereas the cellulase genes are all induced in a common way by cellulose and its derivatives (*e. g.* cellobiose, sophorose), lactose and L-sorbose, suggesting a common regulatory mechanism, the xylanolytic enzymes have different main inducers. *xyn1* is induced by D-xylose and repressed by D-glucose [40], *xyn2* is partially constitutively expressed and further induced by xylobiose, xylan, cellulose and sophorose [41], whereas *xyn3* is induced by sophorose and L-sorbose, but not by D-xylose or its oligomers [42].

Several transcription factors have been identified to participate in the expression of the cellulolytic and xylanolytic genes. The main transcription activator of hydrolase genes is Xyr1 [43], a zinc binuclear cluster protein, which is known to bind to 5'-GGC(A/T)$_3$-3'-motifs, which can also occur in inverted repeats [44]. Another major transcription factor is CreI, a Cys_2His_2 zinc finger protein responsible for carbon catabolite repression with the binding sequence 5'-GGRG-3' [45]. Further to mention are ACEI, a protein with three Cys_2His_2-type zinc finger motifs, binding to the *cis* element 5'-AGGCA-3', which represses the expression of all major cellulolytic and xylanolytic genes in the presence of sophorose and cellulose [46], and ACEII, a zinc binuclear cluster protein, which binds to the *cis* element 5'-GGCTAATAA-3' and enhances expression of cellulase-encoding genes and *xyn2* in the presence of cellulose [47].

1.2.3 An improved *in vivo* footprinting technique

In our publication 'A highly sensitive *in vivo* footprinting technique for condition-dependent identification of *cis* elements' ([48], see Appendix A.3 for full text), we conducted an *in vivo* methylation of DNA from *T. reesei* with DMS. The methylated DNA was extracted and cut with HCl. For the LMPCR, an improved protocol with optimized cycling conditions was established. The subsequent labeling of the DNA was done with 6-FAM, and the

DNA fragments were analyzed with a capillary sequencer using an internal size standard, provided by a sequencing company, which ensured constant quality of the sequencing data and the possibility for high throughput. For the evaluation of the sequencing data, we developed a program called ivFAST (*in vivo* Footprinting Analysis Software Tool), which does a pairwise comparison of footprinting results from different conditions (under consideration of replicates) to find protected or hypersensitive site. For easy visual interpretation, the results are not only given in numbers, but are also displayed as a heat map with three color shades for each protected or hypersensitive sites. A detailed manual of ivFAST is given in Appendix A.4.

Chapter 2

Results and Discussion

2.1 Characterization of erythrose reductases from filamentous fungi

In [29], we characterized the Err1 from *T. reesei*, *A. niger*, and *F. graminearum*, expressed as GST-fusion proteins in *E. coli*. The DNA sequences, used for expression plasmid construction, were identified by *in silico* analysis, starting from the sequence of ER3 from *Trichosporonoides megachiliensis* [49]. After isolation of the fusion proteins first the optimal assay parameters with erythrose as substrate and NADPH as co-substrate were determined. We found a temperature of 40 ℃ and a pH of 6.5 to be the favorable conditions. Investigation of the substrate specificity with various aldehydes and ketons ranging from C2 to C6 as substrate and NADPH as co-substrate showed that the examined erythrose reductases clearly preferred aldehydes over ketones. Concerning the size of the substrate, C3 and C4 molecules were the preferred ones, followed by decreasing activities with C2 and C5, whereas no activity could be observed for C6. We also tested NADH as co-substrate (with erythrose as substrate), but found no activity with it, so all three enzymes strictly require NADPH as co-substrate. This behavior has also been observed for erythrose reductases from *C. magnoliae* [50]. Whereas the Err1 from *T. reesei* and *A. niger* showed overall activities in the same order of magnitude, with slight differences in substrate preference, the Err1 from *F. graminearum* generally achieved only about one tenth of the activity compared to the other two enzymes.

2.2   Production of erythritol with *T. reesei*

After the characterization of the three Err1, we focused on the production of erythritol in *T. reesei*. In [30], we examined the native erythritol production in the wild-type strain and in the mutant strain Rut-C30, as well as the production in the *err1* overexpression strains QPEC1, RPEC1, QBEC2, and RBEC2, created during the study. Cultivation in shake flasks on the easily metabolizeable substrate D-xylose showed increased erythritol formation in the constitutive overexpression strains QPEC1 and RPEC1 compared to their respective parental strains. On this medium, the wild-type strain and QPEC1 performed slightly better than the mutant strain Rut-C30 and RPEC1. A different situation was observed for cultivation on pretreated wheat straw in bioreactors. Here, the wild-type and its two overexpression strains QPEC1 and QBEC2 showed no noteworthy difference in erythritol production. RPEC1 and RBEC2, on the other hand, showed a clearly enhanced production of erythritol. Also, Rut-C30 and its derived strains generally produced erythritol in an about 10-fold amount compared to the wild-type. It can be assumed that the carbon catabolite-derepressed, cellulase-hyperproducing mutant Rut-C30 is better capable of utilizing the complex substrate wheat straw than the wild-type. Altogether, we made a proof of concept that *T. reesei* natively produces erythritol and that overexpression of *err1* led to an increased production of erythritol. At the moment, the produced amount of erythritol is far from industrial applications, but with further strain improvements there is still potential to further increase the production. A promising approach would be to introduce a transporter for erythritol in *T. reesei*, because in contrast to yeasts, *T. reesei* does not secrete the produced erythritol, but stores it only intracellularly. This would not only improve the chemical balance for the enzymatic reaction, but also would change the osmotic conditions in favor to erythritol production. Another possibility would be to overexpress the L-arabinitol dehydrogenase and the D-xylulose reductase to enforce the flux in the PPP and thus, provide more substrate for erythritol production.

2.3   Identification of regulatory regions with *in vivo* footprinting

As described in [48], with the new technique we were able to confirm already known binding sites for Xyr1 and CreI in the URR of the *xyn1* gene in *T. reesei*. A comparison of *in vivo* footprinting results from mycelia under repressing conditions (D-glucose) and inducing conditions (D-xylose) was

done in the wild-type strain and in a *xyr1* deletion strain. As reference mycelia replaced to medium without carbon source was taken. As expected, we found signals at the Xyr1 binding site in the wild-type strain, but not in the deletion strain, whereas for the CreI binding site both strains yielded signals. Another test of the technique was done using the URR of the *cbh2* gene of *T. reesei*. Comparing again repressing (D-glucose) and inducing (D-xylose) conditions and no carbon source as reference, we could identify the *cbh2*-activating element (CAE) [51], which consists of a putative Xyr1 binding site and an overlapping CCAAT-box. The CCAAT-box hereby yielded strongly glucose dependent signals, whereas the Xyr1 binding site showed no condition dependent differences, leading to the assumption that the carbon source specific response is due to the CCAAT-box, while Xyr1 binds permanently. Additionally, two further Xyr1 binding sites, so far only known from *in silico* analysis, were shown to be active in a condition dependent way, as well as a not yet verified CreI binding site depending on glucose. Finally, we investigated the URR of the *xyn2* gene, which has a similar architecture as the URR of *cbh2*. Similar to the CAE, here we have the xylanase activating element (XAE) [52], also comprising a Xyr1 binding site and a CCAAT-box. The XAE is located close to the second Xyr1 binding site. Upstream of the XAE, an AGAA-box, mediating repression, is located. All these elements we could identify by *in vivo* footprinting under repressing (D-glucose), inducing (D-xylose), and reference (no carbon source) conditions. We also identified a second AGAA-box 4 bp upstream of the first one, which fits the assumption that the transcription factor binding to it is supposed to bind as a dimer. Furthermore, a not yet recognized Cre1 binding site, depending on glucose, was identified, as well as a palindromic Xyr1 binding site, exhibiting condition-dependent differences. Aside from this, two so far completely unknown regions could be identified, which reacted in a condition-dependent way.

The application of the new technique, incorporating capillary gel electrophoresis with an internal standard and data evaluation with ivFAST, offers the possibility of establishing an open end database for each URR of interest, leading to completely new insights on known and new *cis* elements.

Appendix A

Publications (full text)

A.1 Characterization of erythrose reductases from filamentous fungi

AMB Express
a SpringerOpen Journal

ORIGINAL ARTICLE   **Open Access**

Characterization of erythrose reductases from filamentous fungi

Birgit Jovanović, Robert L Mach and Astrid R Mach-Aigner[*]

Abstract

Proteins with putative erythrose reductase activity have been identified in the filamentous fungi *Trichoderma reesei*, *Aspergillus niger,* and *Fusarium graminearum* by in silico analysis. The proteins found in *T. reesei* and *A. niger* had earlier been characterized as glycerol dehydrogenase and aldehyde reductase, respectively. Corresponding genes from all three fungi were cloned, heterologously expressed in *Escherichia coli,* and purified. Subsequently, they were used to establish optimal enzyme assay conditions. All three enzymes strictly require NADPH as cofactor, whereas with NADH no activity could be observed. The enzymatic characterization of the three enzymes using ten substrates revealed high substrate specificity and activity with D-erythrose and D-threose. The enzymes from *T. reesei* and *A. niger* herein showed comparable activities, whereas the one from *F. graminearum* reached only about a tenth of it for all tested substrates. In order to proof in vivo the proposed enzyme function, we overexpressed the erythrose reductase-encoding gene in *T. reesei*. An increased production of erythritol by the recombinant strain compared to the parental strain could be detected.

Keywords: *Trichoderma reesei*; *Aspergillus niger*; *Fusarium graminearum*; Erythrose reductase; Erythritol

Introduction

Erythritol is a four-carbon sugar alcohol, which is applied as flavour enhancer, formulation aid, humectants, stabilizer, thickener, and as low-calorie sweetener, of which the latter is the main utilization. It has a natural occurrence in several foods including beer, sake, wine, soy sauce, water melon, pear and grape (O'Donnell and Kearsley 2012; Sreenath and Venkatesh 2008) and is well tolerated by the human body (Munro et al. 1998). Erythritol can be chemically synthesized from dialdehyde starch with a nickel catalyst at high temperatures, but this process is not stereospecific and low in yield, and therefore, not industrialized (Moon et al. 2010). Instead erythritol is produced in biotechnological processes using osmophilic yeasts obtained by random mutagenesis as *Aureobasidium* sp. (Ishizuka et al. 1989; Sasaki et al. 1990), *Trichosporonoides* sp. (Suh et al. 1999), (*Torula* sp. Oh et al. 2001), and *Candida magnoliae* (Koh et al. 2003; Ryu et al. 2000). As substrate a highly concentrated glucose solution (typically 40% (w/v)) is applied, which is

gained from chemically and enzymatically hydrolyzed wheat- and cornstarch. It serves as carbon source and causes high osmotic pressure, which pushes the yeast to produce the osmolyte erythritol (reviewed by (Moon et al. 2010)).

Even though the production of erythritol and the according enzyme, erythrose reductase, have been well studied in yeasts, no such enzymes have yet been identified in filamentous fungi. For this study the filamentous ascomycota *Trichoderma reesei* (telemorph *Hypocrea jecorina*, (Kuhls et al. 1996)), *Aspergillus niger,* and *Fusarium graminearum* (telemorph *Gibberella zeae*) were chosen because of their great importance in biotechnology. The (hemi)cellulases of *T. reesei* are widely used in pulp and paper production (Buchert et al. 1998; Noé P. 1986; Welt 1995), food and feed industry (Galante 1993; Lanzarini 1989; Walsh et al. 1993), textile industry (Koo 1994; Kumar 1994; Pedersen 1992), and more recently, for 2nd generation biofuel (cellulose ethanol) production (Hahn-Hägerdal et al. 2006; Himmel et al. 2007; Ragauskas et al. 2006). *A. niger* is used for the production of organic acids, as citric acid and gluconic acid (Ruijter et al. 2002), for heterologous protein expression Archer and Turner (2006), as well as production of pectinases Bussink et al. (1992; Delgado et al. 1992;

* Correspondence: astrid.mach-aigner@tuwien.ac.at
Department for Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Str. 1a, Wien A-1060, Austria

Parenicová et al. 2000) and hemicellulases, such as xylanases and arabinases Gielkens et al. (1997; van Peij et al. 1997). *F. graminearum* is a well studied filamentous fungus because of its relevance as plant pathogen that can infect numerous plants like cereals, but also dicotyledons (Pirgozliev et al. 2003; Urban et al. 2002). Additionally, it is also used in biotechnological applications such as heterologous protein expression (Royer et al. 1995).

In contrast to yeasts, the use of filamentous fungi offers the interesting perspective of using non-food plant biomass (e.g. lignocellulose) as substrate. By secretion of xylanolytic enzymes, these fungi are capable of degrading xylans into their major monomers D-xylose and L-arabinose. They can be directly metabolized to D-xylose-5-phosphate to supplement the pentose phosphate pathway (PPP), from which erythritol is a side product. D-xylulose-5-phosphate and D-ribose-5-phosphate are transferred by a transketolase to D-sedoheptulose-7-phosphate and D-glyceraldehyde-3-phosphate, which are further processed by a transaldolase to fructose-6-phosphate and D-erythrose-4-phosphate. A schematic drawing of the according pathway is given in Additional file 1. Erythritol is formed by dephosphorylation of D-erythrose-4-phosphate and the following reduction:

$$D-Erythrose + NADPH \xrightarrow{Reductase} Erythritol + NADP^+$$

The characterization of the enzyme performing this reduction, namely erythrose reductase, has been done for some yeasts e. g. by (Lee et al. (2010), (Lee et al. 2003), (Ookura et al. 2005)), but until now no such enzyme has been identified for the above-mentioned filamentous fungi.

In this study, we identified by in silico analysis proteins in *T. reesei*, *A. niger*, and *F. graminearum* exhibiting a high sequence similarity to the erythrose reductase (ER1) from *Trichosporonoides megachiliensis*. Accordingly, in this manuscript the corresponding proteins from the three organisms are referred to the term Err1 (Erythrose reductase 1) for easier reading. The respective genes were cloned and their protein products were heterologously expressed and purified. All three putative Err1 proteins were characterized in enzymatic assays with respect to their substrate specificity to D-erythrose and nine other potential substrates. In order to do this, the optimal assay conditions (temperature and pH) for all three enzymes were determined before, and then their usages of the different substrates were tested. Finally, we aimed to prove the function of the putative erythrose reductase in vivo. Therefore, the corresponding *T. reesei* enzyme was overexpressed in this fungus

and the production of erythritol in the recombinant strain was compared to the parental strain.

Materials and methods
Strains and cultivation conditions
The *T. reesei* strain QM6aΔtmus53 (Steiger et al. 2011), the *A. niger* strain N400 (CBS 120.49), and the *F. graminearum* strain PH1 (NRRL31084) were maintained on malt extract (MEX) agar, complete medium agar (Pontecorvo et al. 1953), and small nutrient agar (Brunner et al. 2007), respectively. The recombinant *T. reesei* strain PEC1, produced during this study, was maintained on MEX agar containing hygromycin B.

Cultivation in shakeflasks was performed in 1-l-Erlenmeyer flasks containing 250 ml (Mandels-Andreotti (MA) medium Mandels 1985) supplemented with 1% (w/v) D-xylose. For inoculation 10^9 conida per litre were used. Growth conditions were pH 5, 30℃, and 160 rpm shaking rate. For harvesting mycelia, samples of 60 ml were drawn after 24 h and 30 h. For short-term storage, mycelia were shock-frozen and kept in liquid nitrogen.

Plasmid construction
The in silico identified *err1* genes from *T. reesei*, *A. niger*, and *F. graminearum* were amplified from cDNA. The cDNA was generated as described below in the according section. Primers were used to introduce restriction sites adjacent to the gene. Primer sequences are given in Table 1. The PCR products were subcloned into pJET-1.2 (Thermo Scientific, Waltham, MA, USA), using chemically competent *Escherichia coli* TOP 10 (Invitrogen, Life Technologies Ltd, Paisley, UK) for plasmid replication.

For the construction of pGEX-err1T, pGEX-err1A, and pGEX-err1F the *err1* gene was excised from pJET-1.2 by *Eco*RI/*Bam*HI digestion and inserted into pGEX-4T-2 (GE Healthcare Life sciences, Little Chalfont, Buckinghamshire, UK).

For the construction of pBJ-PEC1 the vector pRLM$_{ex30}$ Mach et al. (1994), which contains the *hph* gene flanked by the *pki* promoter and the *cbh2* terminator, was used. The *hph* gene was removed by *Nsi*I/*Xba*I digestion and subsequently, *err1*, which was excised from JET-1.2 also by *Nsi*I/*Xba*I digestion, was inserted.

Protoplast transformation
The protoplast transformation of *T. reesei* was performed as described by (Gruber et al. 1990). 5 µg of the plasmid pBJ-PEC1 and 1 µg pAN7, which confers hygromycin B resistance (Punt et al. 1987), were co-transformed into the fungal genome.

DNA analysis
Fungal genomic DNA was isolated by phenol-chloroform extraction, using a FastPrep®-24 (MP Biomedicals, Santa

Table 1 Oligonucleotides used during the study

Name	Sequence (5' - 3')[a]	Usage
err1_A.nig_BamHI_f	ATATA**GGATCC**ATGTCTCTCGGAAAGAAGGTTACTCTC	pGEX-err1A
err1_A.nig_NotI_r	TATAT**GCGGCCGC**TTAAACAATCACCTTATGACCAGCAGGC	pGEX-err1A
err1_T.ree_BamHI_f	ATATA**GGATCC**ATGTCTTCCGGAAGGACC	pGEX-err1T
err1_T.ree_NotI_r	TATAT**GCGGCCGC**TTACAGCTTGATGACAGCAGTG	pGEX-err1T
err1_F.gra_BamHI_f	ATATA**GGATCC**ATGTCTTTCGGTCGAACTGTCACTC	pGEX-err1F
err1_F.gra_NotI_r	TATAT**GCGGCCGC**TTACAGCTTGAGAACAACCTGGTGG	pGEX-err1F
err1_XbaI_f	ATATA**TCTAGA**ATGTCTTCCGGAAGGACC	Vector construction for fungal transformation
err1_Nsi_r	TATAT**ATGCAT**TTACAGCTTGATGACAGCAGTG	
qerr1_f	CTTTACCATTGAGCACCTCGACG	RT-qPCR *err1*
qerr1_r	GGTCTTGCCCTGCTTCTTGG	RT-qPCR *err1*
qact1_f	TGAGAGCGGTGGTATCCACG	RT-qPCR *act1*
qact1_r	GGTACCACCAGACATGACAATGTTG	RT-qPCR *act1*
qsar1_f	TGGATCGTCAACTGGTTCTACGA	RT-qPCR *sar1*
qsar1_r	GCATGTGTAGCAACGTGGTCTTT	RT-qPCR *sar1*

[a] restriction enzyme sites are given in bold letters.

Ana, CA, USA) for cell disruption. Therefore about 100 mg of mycelia was transfered to 400 µl DNA extraction buffer (0.1 M Tris–HCl pH 8.0, 1.2 M NaCl, 5 mM EDTA) and grounded with glass beads (0.37 g Ø 0.01 – 0.1 mm, 0.25 g Ø 1 mm, 1 piece Ø 3 mm) using the FastPrep. Afterwards, the mixture was immediately put on 65°C, supplemented with 9 µM RNase A, and incubated for 30 min. Then 200 µl of phenol (pH 7.9) and 200 µl of a chloroform-isoamyl alcohol-mixture (25:1) were added, and vigorous mixing followed each addition. Phases were separated by centrifugation (12000 g, 10 min, 4°C) and the aqueous phase was transferred into a new vial. DNA was precipitated by addition of the 0.7-fold volume of isopropanol. After 20 min incubation at room temperature the DNA was separated by centrifugation (20000 g, 20 min, 4°C) and washed with 500 µl ethanol (70%). The air-dried DNA pellet was solubilised in 50 µl Tris–HCl (10 mM, pH 7.5) at 60°C.

RNA isolation and cDNA synthesis
RNA extraction from fungal mycelia was performed with peqGOLD TriFast™ (peqlab, Erlangen, Germany) according to the manufacturer's procedure, using a FastPrep®-24 (MP Biomedicals, Santa Ana, CA, USA) for cell disruption. RNA quantity and quality were determined with a NanoDrop 1000 (Thermo Scientific, Waltham, MA, USA). A 260 nm/280 nm ratio of at least 1.8 was stipulated for further sample processing.

cDNA synthesis was performed with RevertAid™ H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA) according to the manufacturer's procedure, using 0.5 µg of RNA.

Transcript analysis
Quantitative PCR (qPCR) analysis was performed in a Rotor-Gene Q cycler (Qiagen, Hilden, Germany). The qPCR amplification mixture had a total volume of 15 µl, containing 7.5 µl 2× IQ SYBR Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA), 100 nM forward and reverse primer, and 2 µl cDNA (diluted 1:100). Primer sequences are given in Table 1. As reference genes *act1* and *sar1* were used (Steiger et al. 2010). All reactions were performed in triplicates. For each gene a no-template control and a no-amplification control (0.01% SDS added to the reaction mixture) was included in each run. The cycling conditions for *act1* and *err1* comprised 3 min initial denaturation and polymerase activation at 95°C followed by 40 cycles of 15 s at 95°C, 15 s at 59°C, and 15 s at 72 s. For *sar1* different cycling conditions were applied: 3 min initial denaturation and polymerase activation at 95°C followed by 40 cycles of 15 s at 95°C and 120 s at 64 s. PCR efficiency was calculated from the Rotor-Gene Q software. Relative expression levels were calculated using the equation

$$relative\ transcript\ ratio = E_r^{C(r)} \cdot E_t^{-C(t)} \cdot E_{ro}^{-C(ro)} \cdot E_{to}^{C(to)}$$

where E is cycling efficiency, C is the threshold cycling number, r is the reference gene, t the target gene and o marks the sample which is taken for normalization Pfaffl (2001).

Glutathione S-transferase (GST): Err1 fusion proteins
GST fusion proteins of the erythrose reductases from *T. reesei*, *A. niger*, and *F. graminearum* were expressed using plasmids pGEX-err1T, pGEX-err1A, and pGEX-

err1F, respectively, in *E. coli* BL21(DE3)pLysS (Promega, Madison, WI, USA). The protein expression was done in shakeflasks on lysogeny broth supplemented with 100 μg/ml ampicillin at 37°C and 200 rpm. For induction 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added when the culture reached an OD_{600} between 0.7 and 0.8. Cells were harvested 3 h after induction by centrifugation, resuspended in phosphate buffered saline supplemented with 1% Triton X-100, and sonicated using a Sonifier® 250 Cell Disruptor (Branson, Danbury, CT, USA) (power 70%, duty cycle 40%, power for 10 s, pause for 50 s, 10 cycles, on ice). Insoluble compounds were separated by centrifugation (2600 g, 10 min, 4°C). Purification of the proteins was performed using GSTrap™ FF (GE Healthcare Life sciences, Little Chalfont, Buckinghamshire, UK) according to standard procedures. The purified protein solutions were stored at 4°C. There was no considerable loss of activity observed within one month under these storage conditions. The addition of glycerol must be avoided because it has an influence on the enzymatic assay described later.

SDS-PAGE analysis

For the SDS-PAGE analysis a 10% polyacrylamide gel with a tris-glycine buffer (25 mM Trizma® base (Sigma Aldrich, St. Louis, MO, USA), 1.9 mM glycine, 0.5% SDS) was used. Gel casting and running the gel was done with the Mini-PROTEAN® Tetra Cell system (Bio-Rad Laboratories, Hercules, CA, USA). From all three protein expressions 2 μl of the crude extract, 2 μl of the flow-through, and 12 μl of the wash solution, respectively, were applied on the gel. Of the eluted protein from *A. niger* 2 μl, from *F. graminearium* 12 μl, and from *T. reesei* 1 μl were applied. All the samples were supplemented with 4 μl 4x Laemmli sample buffer (Bio-Rad Laboratories, Hercules, CA, USA), filled up with distilled water to a final volume of 16 μl, and incubated for 10 min at 95°C for denaturation. After denaturation, samples were kept on ice until application on the gel. For protein size estimation 2.5 μl of PageRuler™ Prestained ProteinLadder (Thermo Scientific, Waltham, MA, USA) were used. The electrophoresis was carried out at a constant voltage of 160 V. Staining of the gels was done with PageBlue Protein Staining Solution (Thermo Scientific, Waltham, MA, USA) according to the manufacturers protocol.

Enzymatic assay

Enzymatic analysis was performed according to a slightly modified, previously by Lee et al. (2003) described protocol. The reducing reaction was performed in a total volume of 1 ml containing 50 mM Sorenson's phosphate buffer (pH 6.5), 160 μM NADPH or NADH, 100 μl purified GST::Err1 fusion protein, and 10 mM substrate. As substrates L-arabinose, dihydroxyacetone (DHA), D-erythrose, D-glucose, L-glyceraldehyde, glyoxal, methylglyoxal, D-threose, D-xylose, and D-xylulose were used. In a spectrophotometer the consumption of NADPH or NADH over time was followed at 340 nm at the indicated temperature. After 1 min incubation without substrate the reaction was started by adding 100 μl 100 mM substrate.

The oxidizing reaction was performed in a total volume of 1 ml containing 50 mM Tris/HCl (pH 9.0), 400 μM NADP⁺, 200 μl purified GST::Err1 fusion protein, and 10 mM erythritol. In a spectrophotometer the formation of NADPH over time was followed at 340 nm at a temperature of 40°C. After 1 min incubation without substrate the reaction was started by adding 100 μl 100 mM erythritol.

Enzymatic assays were performed in triplicates. Activity is defined in katal (kat), and 1 katal is the conversion of 1 mol substrate per second. The specific activity k_{cat} is defined as 1 katal per mol enzyme and the catalytic efficacy is defined as k_{cat}/K_m.

Gas chromatography (GC) analysis

Mycelia were ground under liquid nitrogen. The powder was suspended in 3 ml distilled water and sonicated using a Sonifier® 250 Cell Disruptor (Branson, Danbury, CT, USA) (power 70%, duty cycle 40%, power for 3 min, on ice). Insoluble compounds were separated by centrifugation (20000 g, 10 min, 4°C). Sample preparation for GC was done in triplicates as follows: 300 μl of the supernatant, supplemented with 10 ng sorbitol as internal standard, was gently mixed with 1.2 ml ethanol (96%) and incubated for 30 min at room temperature for protein precipitation. The precipitant was separated by centrifugation (20000 g, 10 min, 4°C). Samples were dried under vacuum and thereafter silylated (50 μl pyridine, 250 μl hexamethyldisilazane, 120 μl trimethylsilyl chloride). For quantitative erythritol determination a GC equipment (Agilent Technologies, Santa Clara, CA, USA) with a HP-5-column (30 m, inner diameter 0.32 mm, film 0.26 μm) (Agilent Technologies, Santa Clara, CA, USA) was used. The mobile phase consisted of helium with a flow of 1.4 l/min, the column temperature was as follows: 150°C for 1 min, ramping 150 − 220°C (ΔT 4°C/min), ramping 220–320°C (ΔT 20°C/min), 320°C for 6.5 min. Detection was performed with FID at 300°C. The retention times were determined using pure standard substances.

Results

Identification of putative erythrose reductase proteins by in silico analysis

Ookura et al. (2005) biochemically characterized three isoenzymes of the erythrose reductase (ER1, ER2, and

ER3) from the industrial erythritol production strain *Trichosporonoides megachiliensis* SNG-42. The protein sequences of ER1 (NCBI accession number BAD90687.1), ER2 (NCBI accession number BAD90688.1), and ER3 (NCBI accession number BAD90689.1) were compared with the NCBI database using BLASTP to find proteins with similar sequence in the filamentous fungi *T. reesei*, *A. niger*, and *F. graminearum*. The following proteins were found in these organisms: for *T. reesei* the NADP-dependent glycerol dehydrogenase (GLD1) (NCBI accession number ABD83952.1, query coverage 98%, max. ident. 50%, E-value 5e-87); for *A. niger* the aldehyde reductase 1 (Alr1) CBS 513.88 (NCBI accession number XP_001394119.2, query coverage 98%, max. ident. 49%, E-value 8e-88); and for *F. graminearum* a hypothetical protein FG04223.1 (NCBI accession number XP_384399.1, query coverage 98%, max. ident. 48%, E-value 1e-98). Query results are given relative to ER3, which showed a

slightly better match with the protein found for *T. reesei* than ER1 and ER2. Figure 1 shows the phylogram of the above-mentioned protein sequences. The protein found for *T. reesei* has originally been described as glycerol dehydrogenase Liepins et al. (2006), but was not tested with D-erythrose or erythritol as substrate. So the high sequence similarity to ER3 led us to the assumption that this protein might have an erythrose reductase activity. The corresponding enzyme from *A. niger*, Alr1, was only generally recognized as a NADPH-dependent member of the aldo-keto reductase superfamily, but no physiological function was identified up to now. For the *F. graminearum* protein no function was proposed so far.

Purification of heterologously expressed Err1 proteins

The corresponding structural genes of the before identified proteins from *T. reesei*, *A. niger*, and *F. graminearum* (termed from now on Err1) were heterologously expressed
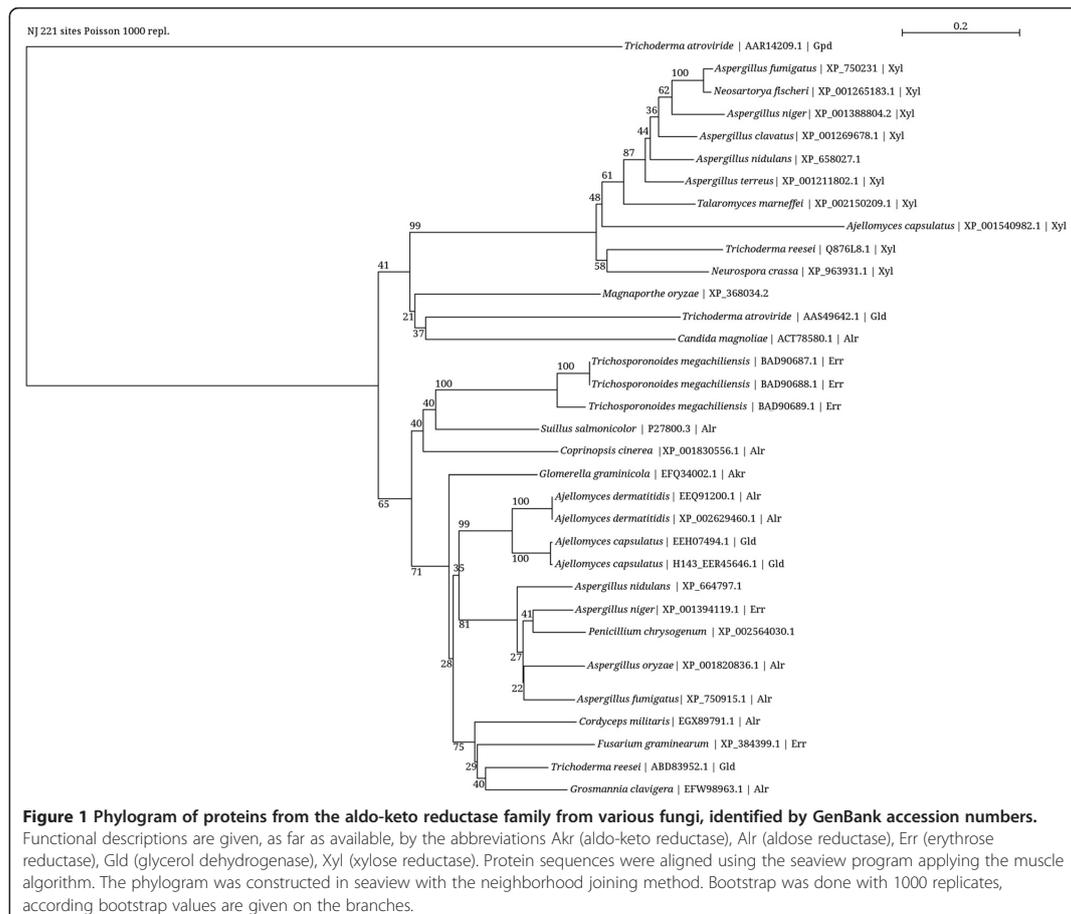

Figure 1 Phylogram of proteins from the aldo-keto reductase family from various fungi, identified by GenBank accession numbers.
Functional descriptions are given, as far as available, by the abbreviations Akr (aldo-keto reductase), Alr (aldose reductase), Err (erythrose reductase), Gld (glycerol dehydrogenase), Xyl (xylose reductase). Protein sequences were aligned using the seaview program applying the muscle algorithm. The phylogram was constructed in seaview with the neighborhood joining method. Bootstrap was done with 1000 replicates, according bootstrap values are given on the branches.
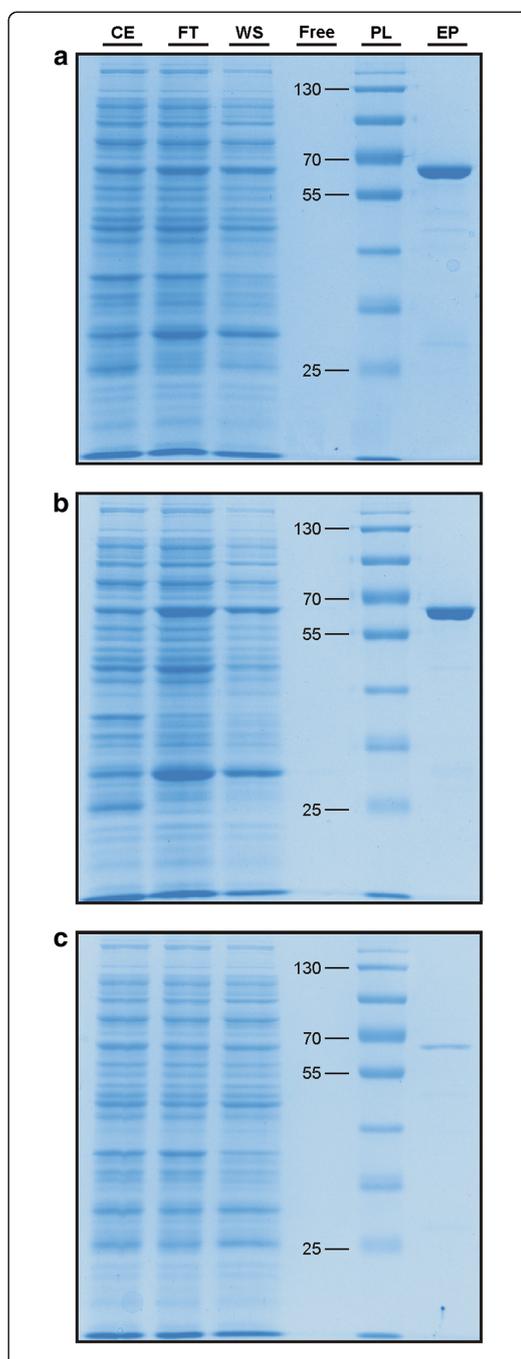
Figure 2 SDS-PAGE analyses of the purification of the three GST:Err1 fusion proteins. Applied were the crude extract (CE), the flow-through from application of the crude extract (FT), the wash solution (WS) and the eluated protein (EP) resulting from expression of GST proteins fused to the Err1 from *T. reesei* **(a)**, *A. niger* **(b)**, and *F. graminearum* **(c)**. A prestained protein ladder (PL) was used for estimation of protein size; indicated sizes are given in kDa.

in *E. coli*. Therefore, the fungi were grown on rich medium. Subsequently, RNA was isolated and reversely transcribed into cDNA, which was used as a template for amplification of the respective *err1* genes. Cloning into a pGEX vector allowed the expression as GST fusion proteins. After induction using IPTG, the *E. coli* cells were decomposed and the three GST::Err1 fusion proteins were isolated via a corresponding purification system. Soluble enzyme expression and purification of all three proteins at their correct, calculated size (64, 64, and 65 kDa, for the protein from *T. reesei, A. niger,* and *F. graminearum,* respectively) was confirmed using SDS-PAGE (Figure 2).

Optimal parameters for the erythrose reductase enzyme assay

Since neither of the proteins has yet been characterized using D-erythrose as a substrate, the optimal parameters for the enzymatic assay had to be determined. Enzyme assays were performed using the proteins heterologously expressed in *E. coli*.

For the reducing reaction, which converts D-erythrose to erythritol, a previous study reported a pH of 7.0 for the ER from *C. magnoliae* Lee et al. (2003). For the GLD1 from *T. reesei* Liepins et al. (2006) also reported a pH optimum of 7.0, but this was determined with different substrates. Therefore, Sorenson's phosphate buffers from pH 6.0 to 8.0 were tested in steps of 0.5 pH units. Online resource 2 depicts the measured progression of the absorption caused by NADPH consumption. We found that a pH of 6.5 is clearly favorable for the Err1 from *T. reesei* (Additional file 2a). The enzymes from *A. niger* and *F. graminearum* showed strongest decrease in absorbance at pH 7.0, but the differences between varying pH conditions were neglible for both (Additional file 2b and Additional file 2c). Therefore, the temperature optimization was carried out at pH 6.5 for all three enzymes from 10°C to 50°C (in steps of 10°C). For Err1 from *T. reesei* we found an increase in activity between 10°C and 40°C, whereas 40°C and 50°C already yielded almost identical activities (Additional file 2d). The enzyme from *A. niger* showed only slightly better performance at 50°C compared to 40°C (Additional file 2e). For the Err1 from *F. graminearum* enzyme denaturation occurred most probably at 50°C, which can be deduced from the early loss of activity at a still high NADPH concentration (Additional file 2f). Since the improvement in Err1

activity using 50°C instead of 40°C was negligibly anyway and with respect to better enzyme stability, 40°C was chosen for further measurements.

Testing the three enzymes under optimized conditions with NADH instead of NADPH as co-factor for neither of them yielded a detectable activity. This is in accordance with former reports on the *T. reseei* enzyme, which showed activity only under consumption of NADPH, but not with NADH Liepins et al. (2006).

For the oxidizing reaction, which converts erythritol to D-erythrose under consumption of NADP$^+$, former studies proposed a pH of about 9 for similar reactions Colowick (1963). Consequently, Tris/HCl buffers of pH 8.0, 8.5, and 9.0 (equals the upper range of this buffer system) were tested at an assay temperature of 40°C. Only at pH 9 the oxidation of erythritol was the favored direction of the reaction, however, it proceeded much slower than the inverse reaction described before. At pH 8.5 an oscillating reaction was observed, whereas at pH 8.0 the equilibrium was completely on the reducing side of the reaction (data not shown).

Altogether, we suggest the usage of a buffers system at pH 6.5 and a temperature of 40°C for the erythrose reductase assay.

Substrate specificity and activity of Err1
Substrates were chosen in order to cover molecules from 2 to 6 carbon atoms (C2 – C6) on the one hand, and aldehydes and ketones on the other hand: the dialdehyde glyoxal (C2), the keto-aldehyde methylglyoxal (C3), the trioses DHA and L-glyceraldehyde, the aldotetroses D-erythrose and D-threose, the aldopentoses L-arabinose and D-xylose, the ketopentose D-xylulose, and the aldohexose D-glucose.

The three enzymes showed some differences in both, substrate specificity as well as in total activity. But for all of them the activity using DHA, D-glucose, D-xylose, and D-xylulose was too low to evaluate the kinetics parameters. Consequently, these substances will be neglected in the further discussion.

The Err1 from *T. reesei* seemed to slightly favor D-threose over the other substrates, but showed only slight differences in K_m considering the standard deviations (Table 2). The turnover number (k_{cat}) on the other hand was for methylglyoxal and L-glyceraldehyde higher than for D-erythrose, followed by D-threose in the fourth place and here the differences were considerably. Looking at the catalytic efficacy (k_{cat}/K_m), D-threose performed a little bit better than L-glyceraldehyde and D-erythrose, only to be seconded by methylglyoxal (Table 2). Altogether, the enzyme had a similar good performance for D-erythrose and D-threose and therefore, obviously here lacks stereospecificity. L-glyceraldehyde had the lowest specificity considering K_m, but the second

Table 2 Substrate specificity of Err1 from *T. reesei*

Substrate[a]	K_m [µM]	k_{cat} [kat/mol]	k_{cat}/K_m [1/(mM·s)]
L-arabinose	124.56 ± 9.78[b]	3.21 ± 0.22	25.80 ± 0.23
Dihydroxyaceton	n.d.[c]	n.d.	n.d.
D-erythrose	134.52 ± 9.34	36.51 ± 2.13	271.41 ± 3.00
D-glucose	n.d.	n.d.	n.d.
L-glyceraldehyde	158.04 ± 5.00	47.89 ± 1.86	303.02 ± 2.18
Glyoxal	102.74 ± 9.76	18.84 ± 1.34	183.41 ± 4.40
Methylglyoxal	131.86 ± 1.84	72.58 ± 0.28	550.41 ± 5.55
D-threose	94.07 ± 2.46	29.03 ± 0.89	308.59 ± 1.36
D-xylose	n.d.	n.d.	n.d.
D-xylulose	n.d.	n.d.	n.d.

[a] listed in alphabetical order.
[b] mean of three replicates and standard deviation is given.
[c] means not detectable.

best k_{cat}. The catalytic efficacy was about the same as for D-threose. Glyoxal had a K_m between D-threose and D-erythrose, but k_{cat} and catalytic efficacy were lower than for both, D-erythrose and D-threose. The same is true for L-arabinose, only that k_{cat}, and therefore also k_{cat}/K_m, was much lower (about 10-fold) than for the other substrates.

Referring to K_m, the *A. niger* enzyme clearly preferred D-erythrose (Table 3). On the other hand, k_{cat} and the catalytic efficacy were comparably low for D-erythrose, but very high for D-threose. With methylglyoxal the best performance was achieved, but with relatively low specificity. A similar result was found for L-glyceraldehyde, which performed second best considering k_{cat} and showed a similar K_m. For the *A. niger* enzyme glyoxal reached a k_{cat} higher than that of D-erythrose, but with a worse K_m, so k_{cat}/K_m was still higher for D-erythrose. The utilization of L-arabinose led to similar kinetic parameters as obtained with the *T. reesei* enzyme.

The Err1 from *F. graminearum* slightly favored methylglyoxal over D-erythrose looking at the K_m, but

Table 3 Substrate specificity of Err1 from *A. niger*

Substrate[a]	K_m [µM]	k_{cat} [kat/mol]	k_{cat}/K_m [1/(mM·s)]
L-arabinose	286.66 ± 27.06[b]	7.32 ± 0.51	25.55 ± 0.64
Dihydroxyaceton	n.d.[c]	n.d.	n.d.
D-erythrose	139.39 ± 6.45	24.95 ± 1.05	179.00 ± 0.76
D-glucose	n.d.	n.d.	n.d.
L-glyceraldehyde	319.28 ± 4.12	143.23 ±1.12	448.61 ± 2.29
Glyoxal	330.95 ± 3.06	49.09 ± 0.68	148.34 ± 0.68
Methylglyoxal	352.81 ± 24.42	196.04 ± 13.43	555.66 ± 0.39
D-threose	279.50 ± 7.89	108.44 ± 1.98	387.97 ± 3.87
D-xylose	n.d.	n.d.	n.d.
D-xylulose	n.d.	n.d.	n.d.

[a] listed in alphabetical order.
[b] mean of three replicates and standard deviation is given.
[c] means not detectable.

again the difference was not significant (Table 4). In k_{cat} and catalytic efficacy D-erythrose was also only excelled by methylglyoxal and L-glyceraldehyde. D-threose and glyoxal had a similar turnover rate, but considering K_m the specificity was much higher for D-threose. For L-arabinose no measurable activity was found. Generally, it is notable that for all substrates k_{cat} and catalytic efficacy were much lower (more then 10-fold) if the *F. graminearum* enzyme was used compared to those from the other two species.

Overexpression of *err1* in *T. reesei* proves its function in vivo

To investigate if Err1 in vivo really has the proposed functionality, an according overexpression strain of *T. reesei* was constructed and its production of erythritol was compared to its parental strain. For constant expression we have put the *T. reesei err1* gene under control of the constitutive *pki* promoter and transformed the construct into the fungal genome. The strains received from protoplast transformation were analyzed by PCR with regard to the presence of the vector construct. Positive ones were screened for *err1* expression based on transcript analysis, and the one with the highest increase in *err1* expression compared to its parental strain (named PEC1) was chosen for further characterization. Both, the parental and recombinant strain were grown on D-xylose as carbon source for 30 h in shakeflasks. Samples were drawn after 24 h and 30 h, and subsequently used for RT-qPCR and GC analysis. RT-qPCR confirmed a considerably elevated transcript level of *err1* in the recombinant strain compared to the parental strain, which was already observed during the above-mentioned screening process (Figure 3a).

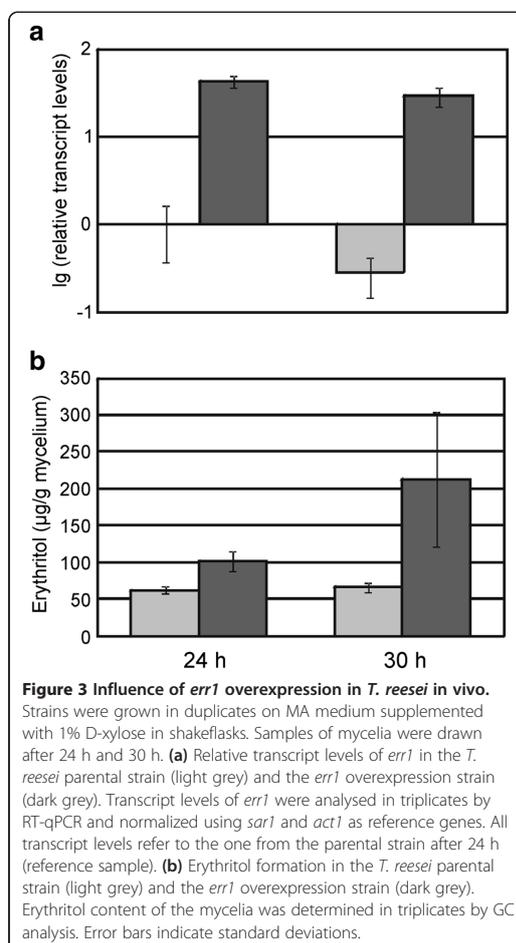GC analysis of the intracellular erythritol concentration of both strains demonstrated that the *err1*

Figure 3 Influence of *err1* overexpression in *T. reesei* in vivo. Strains were grown in duplicates on MA medium supplemented with 1% D-xylose in shakeflasks. Samples of mycelia were drawn after 24 h and 30 h. **(a)** Relative transcript levels of *err1* in the *T. reesei* parental strain (light grey) and the *err1* overexpression strain (dark grey). Transcript levels of *err1* were analysed in triplicates by RT-qPCR and normalized using *sar1* and *act1* as reference genes. All transcript levels refer to the one from the parental strain after 24 h (reference sample). **(b)** Erythritol formation in the *T. reesei* parental strain (light grey) and the *err1* overexpression strain (dark grey). Erythritol content of the mycelia was determined in triplicates by GC analysis. Error bars indicate standard deviations.

Table 4 Substrate specificity of Err1 from *F. graminearum*

Substrate[a]	K_m [μM]	k_{cat} [kat/mol]	k_{cat}/K_m [1/(mM·s)]
L-arabinose	n.d.[b]	n.d.	n.d.
Dihydroxyaceton	n.d.	n.d.	n.d.
D-erythrose	227.61 ± 8.81[c]	3.72 ± 0.09	16.36 ± 0.23
D-glucose	n.d.	n.d.	n.d.
L-glyceraldehyde	298.72 ± 88.4	8.54 ± 2.28	28.57 ± 0.83
Glyoxal	535.16 ± 6.42	2.90 ± 0.07	5.42 ± 0.06
Methylglyoxal	214.32 ± 7.64	6.76 ± 0.24	31.55 ± 0.02
D-threose	380.48 ± 18.65	2.91 ± 0.00[d]	7.64 ± 0.37
D-xylose	n.d.	n.d.	n.d.
D-xylulose	n.d.	n.d.	n.d.

[a] listed in alphabetical order.
[b] means not detectable.
[c] mean of three replicates and standard deviation is given.
[d] means < 0.01.

overexpression strain indeed was able to produce more erythritol than its parental strain. After 24 h, the erythritol concentration in the recombinant strain was 1.6-fold higher than in the parental strain, and after 30 h it was even 3.2-fold, respectively (Figure 3b).

Discussion

Based on the protein sequences of the known erythrose reductases from *Trichosporonoides megachiliensis* SNG-42 (Ookura et al. 2005), we identified by in silico analysis candidate proteins for Err1 in *T. reesei*, *A. niger*, and *F. graminearum*. In vitro analysis of these proteins by an enzyme assay confirmed for all of them a high substrate specificity and turnover rate for D-erythrose. Out of ten tested aldehydes and ketones, ranging from C2 to C6, only methylglyoxal and L-glyceraldehyde partly showed better performance or substrate specificity than D-erythrose and

its diastereomer D-threose. For the cell toxin methylglyoxal it is known that aldehyde reductases show considerable activity for it, and convert it to hydroxyacetone (95%) and D-lactaldehyde (5%) (Thornalley 1996). But the main detoxification of methylglyoxal is done by the glyoxalase system, consisting of glyoxalase I and II and catalytic amounts of reduced glutathione. These enzymes belong to superfamily cl14632, whereas Err1 belongs to superfamily cl00470 and utilizes NADPH as cofactor. Therefore, it is very unlikely that Err1 belongs to the glyoxalase system. Interestingly, the good performance of erythrose reductase with glyceraldehydes, which was observed in this study, was also reported by (Lee et al. 2003) for *C. magnoliae*.

Neither of the tested Err1 proteins from the three fungi has a clear specificity for D-erythrose over D-threose or vice versa. In case of the Err1 from *T. reesei* D-erythrose showed a higher turnover number than D-threose, but the differences in K_m were not substantially. The Err1 from *A. niger* on the one hand clearly preferred D-erythrose considering K_m, but on the other hand, the turnover number was considerably higher for D-threose. Only the enzyme from *F. graminearum* has a slight preference for D-erythrose, which is reflected by both characteristic numbers, K_m and k_{cat}. Since Err1 takes various short-chained aldehydes as substrate it is not surprising that it utilizes the diastereomers D-erythrose and D-threose in a similar manner.

Aside from D-erythrose (C4), D-threose (C4), L-glyceraldehyde (C3) and methylglyoxal (C3) also glyoxal (C2) caused distinct activity. The enzymes from *T. reesei* and *A. niger* also showed measurable activity with the C5-sugar L-arabinose, but it was much lower than the activity of the substrates mentioned before. With D-xylose, the other C5-aldehyde tested, only a poor activity of these two enzymes was detected, which turned out to be too low to calculate kinetic parameters. The C6-sugar D-glucose showed no activity at all. It can therefore be proposed that Err1 is limited to substrates with a chain length ≤ 5 C-atoms, with best performance for 3 and 4 C-atoms. The two ketones analyzed, DHA and D-xylulose, showed no measurable activity. This leads to the assumption that only aldehydes are suitable substrates, which is in accordance with the previous general assignment of the *A. niger* enzyme as aldehyde reductase.

The Err1 from *T. reesei* and *A. niger* performed quite similar (activity is in the same order of magnitude), whereas the enzyme from *F. graminearum* showed much lower activity (about one tenth) for all substrates. Also, the latter was found to be less temperature-stable than the other ones, as the loss of activity was visible within minutes if kept at 50°C.

Comparing the kinetic parameters using D-erythrose as substrate and NADPH as co-factor, a ten times higher K_m was observed for the Err1 proteins from *T. reesei*

and *A. niger* characterized in this study than for ER1 and ER2 from *C. magnoliae* (Lee et al. 2010). The k_{cat} of Err1 from *T. reesei* and *A. niger* is in the same order of magnitude as ER2, resulting in a 10-fold higher catalytic efficacy of ER2. The strict requirement of NADPH as cofactor is in accordance with results for *C. magnoliae* (Lee et al. 2010). However, the presence of erythrose reductase activity in these filamentous fungi is an important prerequisite for the possibility of developing production strategies using non-food plant biomass. Notably, the enhanced *err1* expression in a recombinant *T. reesei* strain led to an increased formation of erythritol. Even if the yield is not at the level of the yeast production strains, it should be considered that these strains have already undergone extensive mutagenesis and were screened for erythritol production. Any kind of engineering steps are still open in order to increase erythritol production in filamentous fungi. As this is an attractive alternative that would use cheap and sustainable starting materials an according patent was issued (Mach-Aigner et al. 2012).

Finally, the recombinant *T. reesei* strain, which over-expressed *err1*, and its parental strain demonstrated functionality of the erythrose reductase in vivo. This emphasizes that the earlier characterizations of the enzyme from *T. reesei* as Gld1 (Liepins et al. 2006) and the one from *A. niger* as Alr1 missed an important biological function of the enzyme. In summary, all three levels of investigation (in silico, in vitro, and in vivo) have provided evidence that the proteins identified are catalyzing the side reaction of the PPP, in which D-erythrose is converted to erythritol and vice versa. Altogether, this supports their capability to function as erythrose reductases.

Additional files

Additional file 1: Schematic drawing of the metabolic pathway concerning erythritol as a side product of the phosphate pathway.

Additional file 2: Determination of the optimal pH and temperature for assaying Err1 activity from filamentous fungi. Collecting the absorbance data was restarted 60 s after the enzyme reaction was started by addition of D-erythrose and was continued over the time indicated in s. Different pH conditions (6.0, dark blue; 6.5, orange; 7.0, yellow; 7.5, light blue; 8.0, dark red) at 40°C (a, b, c) and different temperatures (10°C, dark blue; 20°C, orange; 30°C, yellow; 40°C, light blue; 50°C, dark red) at pH 6.5 (d, e, f) were tested using GST-fusion proteins of Err1 from *T. reesei* (a, d), *A. niger* (b, e), and *F. graminearum* (c, f).

References
Archer DB, Turner G (2006) Genomics of protein secretion and hyphal growth in *Aspergillus*. In: Brown AJP (ed) The Mycota, vol XIII. Springer, Berlin Heidelberg

Brunner K, Lichtenauer AM, Kratochwill K, Delic M, Mach RL (2007) Xyr1 regulates xylanase but not cellulase formation in the head blight fungus *Fusarium graminearum*. Curr Genet 52(5–6):213–220

Buchert J, Oksanen T, Pere J, Siika-aho M, Suurnäkki A, Viikari L (1998) Applications of *Trichoderma reesei* enzymes in the pulp and paper industry. In: Harman GE, Kubicek CP (eds) *Trichoderma & Gliocladium*, vol 2. Taylor & Francis Ltd., London, UK, pp 343–357

Bussink H, van den-Hombergh J, van den-Ijssel P, Visser J (1992) Characterization of polygalacturonase-overproducing *Aspergillus niger* transformants. Appl Microbiol Biotechnol 37:324–329

Colowick SP (ed) (1963) Preparation and assay of enzymes (continued). Acad. Press, New York, NY, USA

Delgado L, Trejo B, Huitron C, Aguilar G (1992) Pectin lyase from *Aspergillus* sp. CH-Y-1043. Appl Microbiol Biotechnol 39:515–519

Galante YM, Monteverdi R, Inama S, Caldini C, De-Conti A, Lavelli V, Bonomi F (1993) New applications of enzymes in wine making and olive oil production. Italian Biochem Soc Trans 4:34

Gielkens MM, Visser J, de-Graaff LH (1997) Arabinoxylan degradation by fungi: characterization of the arabinoxylan-arabinofuranohydrolase encoding genes from *Aspergillus niger* and *Aspergillus tubingensis*. Curr Genet 31(1):22–29

Gruber F, Visser J, Kubicek CP, de Graaff LH (1990) The development of a heterologous transformation system for the cellulolytic fungus *Trichoderma reesei* based on a *pyrG*-negative mutant strain. Curr Genet 18(1):71–76

Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund MF, Liden G, Zacchi G (2006) Bio-ethanol - the fuel of tomorrow from the residues of today. Trends Biotechnol 24(12):549–556

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315(5813):804–807

Ishizuka H, Wako K, Kasumi T, Sasaki T (1989) Breeding of a mutant of *Aureobasidium* sp. with high erythritol production. J Ferm Bioeng 68(5):310–314

Koh E-S, Lee T-H, Lee D-Y, Kim H-J, Ryu Y-W, Seo J-H (2003) Scale-up of erythritol production by an osmophilic mutant of *Candida magnoliae*. Biotechnol Lett 25:2103–2105

Koo H, Ueda M, Wakida T, Yoshimura Y, Igarashi T (1994) Cellulase treatment of cotton fabrics. Textile Res J 64(2):70–74

Kuhls K, Lieckfeldt E, Samuels GJ, Kovacs W, Meyer W, Petrini O, Gams W, Börner T, Kubicek CP (1996) Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. Proc Natl Acad Sci U S A 93(15):7755–7760

Kumar A, Lepola M, Purtell C (1994) Enyme finishing of man-made cellulosic fabrics. Text Chem Color 26(10):25–28

Lanzarini G, Pifferi PG (1989) Enzymes in the fruit juice industry. In: Cantarelli C, Lanzarini G (eds) Biotechnology Applications in Beverage Production. Elsevier Science, London, UK, pp 189–222

Lee D-H, Lee Y-J, Ryu Y-W, Seo J-H (2010) Molecular cloning and biochemical characterization of a novel erythrose reductase from *Candida magnoliae* JH110. Microb Cell Fact 9(43):43

Lee J-K, Kim S-Y, Ryu Y-W, Seo J-H, Kim J-H (2003) Purification and characterization of a novel erythrose reductase from *Candida magnoliae*. Appl Environ Microbiol 69(7):3710–3718. doi:10.1128/AEM.69.7.3710-3718.2003

Liepins J, Kuorelahti S, Penttilä M, Richard P (2006) Enzymes for the NADPH-dependent reduction of dihydroxyacetone and D-glyceraldehyde and L-glyceraldehyde in the mould *Hypocrea jecorina*. FEBS J 273:4229–4235

Mach RL, Schindler M, Kubicek CP (1994) Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expression signals. Curr Genet 25(6):567–570

Mach-Aigner A, Mach RL, Pucher M (2012) Method for the production of erythritol. European Patent EP 20100183799:5.4

Mandels M (1985) Applications of cellulases. Biochem Soc Trans 13(2):414–416

Moon H-J, Jeya M, Kim I-W, Lee J-K (2010) Biotechnological production of erythritol and its applications. Appl Microbiol Biotechnol 86:1017–1025. doi:10.1007/s00253-010-2496-4

Munro IC, Bernt WO, Borzelleca JF, Flamm G, Lynch BS, Kennepohl E, Bär EA, Modderman J (1998) Erythritol: An interpretive summary of biochemical, metabolic, toxicological and clinical data. Food Chem Toxicol 36:1139–1174

Noé P, Chevalier J, Mora F, Comtat J (1986) Action of enzymes in chemical pulp and fibers. Part II: Enzymatic beating. J Wood Chem Technol 6:167–184

O'Donnell K, Kearsley M (eds) (2012) Sweetners and sugar alternatives in food technology. John Wiley & Sons, Ltd., Hoboken, NJ, USA

Oh D-K, Cho C-H, Lee J-K, Kim S-Y (2001) Increased erythritol production in fed-batch cultures of *Torula* sp. by controlling glucose concentration. JIM&B 26:248–252

Ookura T, Azuma K, Isshiki K, Taniguchi H, Kasumi T, Kawamura Y (2005) Primary structure analysis and functional expression of erythrose reductases from erythritol-producing fungi (*Trichosporonoides megachiliensis* SNG-42). Biosci Biotechnol Biochem 69(5):944–951

Parenicová L, Benen JA, Kester HC, Visser J (2000) *pgaA* and *pgaB* encode two constitutively expressed endopolygalacturonases of *Aspergillus niger*. Biochem J 345(3):637–644

Pedersen GP, Screws GA, Cereoni DA (1992) Biopolishing of cellulosic fabrics. Can Text J:31–35

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29(9):e45

Pirgozliev S, Edwards S, Hare M, Jenkinson P (2003) Strategies for the control of *Fusarium* head blight in cereals. Eur J Plant Pathol 109(7):731–742. doi:10.1023/a:1026034509247

Pontecorvo G, Roper JA, Hemmons LM, Macdonald KD, Bufton AW (1953) The genetics of *Aspergillus nidulans*. Adv Genet 5:141–238

Punt PJ, Oliver RP, Dingemanse MA, Pouwels PH, van den-Hondel CAMJJ (1987) Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. Gene 56(1):117–124

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311(5760):484–489

Royer JC, Moyer DL, Reiwitch SG, Madden MS, Jensen EB, Brown SH, Yonker CC, Johnstone JA, Golightly EJ, Yoder WT, Shuster JR (1995) *Fusarium graminearum* A 3/5 as a novel host for heterologous protein production. Nat Biotech 13(12):1479–1483

Ruijter GJG, Kubicek CP, Visser J (2002) Production of organic acids by fungi. In: Osiewacz HD (ed) The Mycota, vol X, Industrial applications. Springer, Berlin Heidelberg, Germany, pp 213–230

Ryu Y-W, Park CY, Park JB, Kim S-Y, Seo J-H (2000) Optimization of erythritol production by *Candida magnoliae* in fed-batch culture. JIM&B 25:100–103

Sasaki T, Kasumi T, Kubo N, Kainuma K, Wako K, Ishizuka H, Kawaguchi G, Oda T (1990) Novel *Aureobasidium* sp: microorganisms, method for obtaining the same and method for preparing erythritol with the same., USA Patent US4939091 A, 3.7

Sreenath K, Venkatesh YP (2008) Analysis of erythritol in foods by polyclonal antibody-based indirect competitive ELISA. Anal Bioanal Chem 391:609–615

Steiger MG, Mach RL, Mach-Aigner AR (2010) An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*). J Biotechnol 145(1):30–37. doi:10.1016/j.jbiotec.2009.10.012

Steiger MG, Vitikainen M, Uskonen P, Brunner K, Adam G, Pakula T, Penttilä M, Saloheimo M, Mach RL, Mach-Aigner AR (2011) Transformation system for *Hypocrea jecorina* (*Trichoderma reesei*) that favors homologous integration and employs reusable bidirectionally selectable markers. Appl Environ Microbiol 77(1):114–121. doi:10.1128/AEM.02100-10

Suh SH, Kim DC, Cho YJ, Jeon YJ, Lee JH (1999) Process for producing erythritol using mutant *Trichosporonoides*. Patent US5962287 A, 5.10.1999, USA

Thornalley PJ (1996) Pharmacology of methylglyoxal: Formation, modification of proteins and nucleic acids, and enzymatic detoxification - A role in pathogenesis and antiproliferative chemotherapy. Gen Pharmac 27(4):565–573

Urban M, Daniels S, Mott E, Hammond-Kosack K (2002) *Arabidopsis* is susceptible to the cereal ear blight fungal pathogens *Fusarium graminearum* and *Fusarium culmorum*. Plant J 32(6):961–973. doi:10.1046/j.1365-313X.2002.01480.x

van-Peij NN, Brinkmann J, Vrsanska M, Visser J, de-Graaff LH (1997) β-Xylosidase
   activity, encoded by *xlnD*, is essential for complete hydrolysis of xylan by
   Aspergillus niger but not for induction of the xylanolytic enzyme spectrum.
   Eur J Biochem 245(1):164–173
Walsh GA, Power RF, Headon DR (1993) Enzymes in the animal-feed industry.
   Trends Biotechnol 11(10):424–430
Welt T, Dinus RJ (1995) Enzymatic deinking - a review. Prog Paper Recycling
   4(2):36–47

A.2 Erythritol production on wheat straw using *Trichoderma reesei*

Erythritol production on wheat straw using Trichoderma reesei

Birgit Jovanović[1] , Robert L Mach[1] and Astrid R Mach-Aigner[*1]

[1]Department for Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Str. 1a, A-1060 Wien, Austria

Email: Birgit Jovanović - birgit.jovanoivc@tuwien.ac.at; Robert L Mach - robert.mach@tuwien.ac.at; Astrid R Mach-Aigner[*]- astrid.mach-aigner@tuwien.ac.at;

[*]Corresponding author

Abstract

Background Erythritol is a four-carbon sugar alcohol, which is mainly used as artificial sweetener. Current production methods employ osmophilic yeasts, which are cultivated in highly concentrated sugar solutions. With regard to (socio)economical issues, the production with biomass-degrading fungi like *Trichoderma reesei*, using non-food lignocellulosic biomass such as wheat straw as substrate, would be an attractive alternative. In a previous study we identified the erythrose reductase-encoding gene (*err1*) in three biomass-degrading filamentous fungi and characterized the activity of these enzymes towards erythrose-to-erythritol conversion.

Results In this study we overexpressed the *err1* gene in the *T. reesei* wild-type and in the cellulase hyperproducing, carbon catabolite derepressed strain Rut-C30 in order to investigate the possibility of producing erythritol with *T. reesei*. Two different promoters were used for *err1* overexpression in both strains, a constitutive (the native pyruvat kinase (*pki*) promoter) and an inducible one (the native β-xylosidase (*bxl1*) promoter). The derived recombinant strains were precharacterized by analysis of *err1* transcript formation on D-xylose and xylan. Based on this, one strain of each type was chosen for further investigation for erythritol production in shake flasks and in bioreactor experiments. For the latter, we used wheat straw pretreated by an alkaline organosolve process as lignocellulosic substrate. Shake flask experiments on D-xylose showed increased erythritol formation for both, the wild-type and the Rut-C30 overexpression strain compared to their respective parental strain. Bioreactor cultivations on wheat straw did not increase erythritol formation in the wild-type overexpression strain. However, *err1* overexpression in Rut-C30 led to a clearly higher erythritol formation on wheat straw.

Conclusions In this study we demonstrated the possibility of producing erythritol with the biomass-degrading fungus *T. reesei* and the usability of pretreated wheat straw as sole carbon source. The native formation of erythritol could be increased by overexpression of the *err1* gene. The strain Rut-C30 turned out to be the favourable host strain for production of erythritol if wheat straw is the chosen substrate.

Keywords

Erythritol, Erythrose reductase, *Trichoderma reesei*, wheat straw, lignocellulose

Background

Erythritol is a four-carbon sugar alcohol, which is applied as flavor enhancer, formulation aid, humectant, stabilizer, thickener, and as low-calorie sweetener, of which the latter is the main utilization.

Compared with other polyols yielding about 2 cal/g, erythritol yields only up to 0.2 cal/g, which is due to the fact that erythritol does not undergo systemic metabolism in the human body but is excreted unchanged in the urine [1]. Additionally, as a small molecule, it is easily absorbed already in the upper intestine and therefore, causes less digestive distress than other sweeteners [2]. Since erythritol is not assimilated by *Streptococcus mutans* it is noncariogenic. Furthermore it has some favorable physical and chemical properties: it is thermally stable (no decomposition or colorization at 200 ℃ for 1 h), better crystallizeable than sucrose, and less hygroscopic [3]. The negative enthalpy of solution leads to a cooling effect when dissolved. The sweetness of erythritol is plain with very weak after-taste. In a 10 % (w/v) solution it has 60-80 % the sweetness of sucrose. It has a natural occurrence in several foods including beer, sake, wine, soy sauce, water melon, pear, and grape. The tolerance of erythritol by animals and humans was intensively studied [4]. No adverse toxicological effects were observed. Also no carcinogenic, mutagenic or teratogenic potential or effects on fertility could be detected. Therefore, erythritol is a sugar substitute with a growing market and optimization of its production remains an issue.

Current biotechnological production of erythritol use osmophilic yeasts like *Aureobasidium* sp., *Trichosporonoides* sp., *Torula* sp., and *Candida magnolia*. As substrate a highly concentrated glucose (typically 40 % (w/v)) solution is applied, which is gained from chemically and enzymatically hydrolyzed wheat- and cornstarch. The hydrolyzed starch serves as carbon source and causes a high osmotic pressure that pushes the yeast to produce the osmolyte erythritol [1]. Although these processes reach 40 % (w/w) yields of D-glucose to erythritol conversion, they depend on D-glucose as starting material. With regard to (socio)economical issues, starch-derived D-glucose is not a preferable substrate. Therefore, it would be an interesting alternative to use organisms that can utilize non-food, lignocellulosic biomass for the production of erythritol.

In a previous work [5] we characterized the erythrose reductases (Err1) from the filamentous ascomycota *Trichoderma reesei* (telemorph *Hypocrea jecorina* [6]), *Aspergillus niger*, and *Fusarium graminearium* (telemorph *Gibberella zeae*), which are all very potent degraders of biomass. It turned out that the Err1 of *T. reesei* and *A. niger*

showed comparable activities, whereas the Err1 from *F. graminearium* had a considerably lower activity [5]. In the present study we focused on the potential of producing erythritol in *T. reesei* from lignocellulosic biomass. The native lignocellulose-degrading enzymes of the fungus have already broad application in industry, *i.e.* in pulp and paper [7–9], food and feed [10–12], and textile industries [13–15] as well as in biofuel production [16–18]. As such a strong producer of cellulases and hemicellulases (a genome-wide search using the JGI Genome Portal revealed for *T. reesei* 10 celluloytic and 16 xylanolytic enzyme-encoding genes [19]) it is likely that *T. reesei* is able to grow on cheap biowaste material like wheat straw as the sole carbon source. This is supported by former reports on *T. reesei* capable of growing on lignocellulosic material [20, 21].

In this study we used wheat straw that was pretreated by an alkaline organosolve process [22] to remove the lignin up to a residual concentration of about 1 % (w/w), which makes the cellulose and hemicellulose more easily accessible for the fungus. We investigated a *T. reesei* wild-type strain and the strain Rut-C30. Rut-30 is a cellulase hyperproducing, carbon catabolite derepressed mutant, which is the parental strain of most industrially used *T. reesei* strains. In both strains the *err1* gene was overexpressed using either the native, constitutive promoter from the pyruvat kinase encoding gene (*pki*) or the native, inducible promoter from the β-xylosidase 1 encoding gene (*bxl1*). The overexpression strains were screened for enhanced *err1* transcript formation and the best ones where then cultivated on D-xylose and wheat straw for investigating their erythritol production capacity.

Results

Characterization of err1 overexpression strains

Protoplast transformation of the wild-type strain with the plasmid pBJ-PEC1, introducing *err1* under the constitutive *pki* promoter of *T. reesei*, yielded 8 recombinant strains (named QPEC1-#). With the plasmid pBJ-BEC2, introducing *err1* under the inducible *bxl1* promoter of *T. reesei*, 3 recombinant strains (named QBEC2-#) were received. Biolistic transformation of Rut-C30 with the plasmid pBJ-PEC1 yielded 12 recombinant strains (named RPEC1-#), and with the plasmid pBJ-BEC2 20 recombinant strains (named RBEC2-#) were ob-

tained. Stable insertion of the plasmid into the fungal genome was confirmed by isolation of genomic DNA and a following PCR amplifying a fragment including the introduced promoter and *err1*. After two rounds of spore streak outs, 3-7 recombinant strains of all four types were chosen for further characterization according to their growth. The selected recombinant strains were cultivated in shake flasks on D-xylose as well as birch-wood xylan followed by transcript analysis of *err1*. From each type, the strain with the highest transcript rate was chosen for further characterization (Fig. 1). From now on strains were termed QPEC1, QBEC2, RPEC1, and RBEC2, respectively. A determination of the copy number of the newly introduced *err1* in the four finally selected recombinant strains was performed by Southern blot analysis (Fig. 2). Ectopic *in tandem* integration, which is the most common in *T. reesei*, was observed in all four strains. For QPEC1 and RPEC1 more than 5 additional copies were estimated, for QBEC2 and RBEC2 1-2 additional copies were estimated.

Increased production of erythritol on D-xylose

In order to get first insight in the native erythritol formation in the parental strains and the effect of the *err1* overexpression, the strains QPEC1 and RPEC1, as well as their respective parental strains, were cultivated in shake flasks. For this first experiment D-xylose was used as carbon source as all strains grow well on this carbon source, and on the other hand as the monomer of the xylan-backbone it is a main component of lignocellulose, which is aimed to be used finally. Samples were taken after biomass formation was observed and analyzed by gas chromatography (GC) for erythritol production. Separate analysis of the supernatant and the mycelia revealed that no erythritol could be found in the supernatant. The erythritol concentrations detected in the mycelia are presented in Fig. 3. For the wild-type and QPEC1 we could demonstrate, that the overexpression strain contained clearly more erythritol than the parental strain, with an increase of 1.6-fold (24 h) and 3.2-fold (30 h) (Fig. 3a). For Rut-C30 and RPEC1 the increase of intracellular erythritol concentration in the *err1* overexpression strain are not that explicit compared to the wild-type and to QPEC1 (Fig. 3b). After 30 h and 36 h the increase in the recombinant strain is 1.2-fold and 1.4-fold, respectively, compared to the parental strain. Compared with the wild-type, both Rut-C30 and RPEC1 contained slightly less erythritol. However, this observation was considered as a preliminary result because the advantages of using Rut-C30 are not necessarily that pronounced on D-xylose than on a lignocellulosic substrate, which finally should be used according to the aim of this study.

Erythritol formation by the wild-type and its err1 overexpression strains on pretreated wheat straw

Experiments to investigate the growth ability on pretreated wheat straw and corresponding erythritol production on this substrate were performed by cultivation in a bioreactor starting with the wild-type strain and its respective *err1* overexpression strains, QPEC1 and QBEC2. All three strains were able to grow on wheat straw as sole carbon source, even if inoculated directly with conidia and not with pregrown fungal mycelium. Microscopic analysis of samples taken 8 h after inoculation already showed a high germination rate. Further microscopic samples taken during the fermentation process showed good mycelial growth, strongly branched hyphae, and disappearance of the straw, which is due to enzymatic degradation by the fungus. Samples for investigation of erythritol production were taken 48 h and 72 h after inoculation. Since the cultivation broth contained aside from the mycelia also wheat straw as insoluble compound, it was not possible to separate the mycelia for analysis. Therefore, the whole samples were analyzed for erythritol content. Sodium hydroxide soluble protein (SSP) was determined and was used as an indicator for the biomass concentration. From the SSP one can conclude that the strains have a similar growth behavior (Fig. 4a). The xylanase activity was similar in the wild-type and in QBEC2, but clearly increased in QPEC1 after 72 h (Fig. 4b). In contrast to the results from the shake flask experiments on D-xylose, no increase in production of erythritol in the recombinant strains could be found by GC analysis (Fig. 4c), even though transcript analysis of *err1* showed a slight increase in for the recombinant strains after 48 h, and an even more pronounced one after 72 h (Fig. 4d). Summarizing, these strains can grow on wheat straw and metabolize derived monosaccharides to erythritol. However, overexpression of *err1* did not enhance erythritol formation on wheat straw.

Erythritol formation by Rut-C30 and its err1 over-expression strains on pretreated wheat straw

As a cellulase hyperproducing strain, Rut-C30 can be expected to better utilize lignocellulosic substrates compared to the wild-type strain. Indeed, an analog experiment to the one described above, using Rut-C30, RPEC1, and RBEC2 showed more promising results as increased erythritol production in the overexpression strains was observed (data not shown). Consequently, a more extensive study drawing samples every 12 h, starting 18 h after inoculation, was conducted with these strains again cultivated then in a bioreactor on pretreated wheat straw. The SSP indicated an similar growth behavior for all strains, whereupon RPEC1 after 42 h slightly dropped a little bit behind the others (Fig. 5a). The same pattern could be observed even more clearly for the xylanase activities (Fig. 5b). The course of erythritol concentration is depicted in Fig. 5c. One can observe that the parental strain Rut-C30 started a slightly faster with erythritol formation. All strains reached their maximum erythritol production after 42 h, whereupon the *err1* overexpression strains showed increased formation compared to their parental strain. Even though RPEC1 and RBEC2 shared nearly the same maximum erythritol concentration, they differed in their time course of production. Erythritol formation by RBEC2 rose faster in the beginning, but also dropped faster after having reached the maximum. After 66 h the erythritol concentration dropped for all three strains to a nearly equal level, so it seems that the overexpression of *err1* does not only boost the formation of erythritol but might also trigger the erythritol consumption of this storage compound when conditions (*e.g.* carbon source availability) become less favorable. It should also be noticed that the amount of erythritol produced by the recombinant strains was about 10-fold higher compared to the wild-type at the peak of production. The transcript analysis showed constant expression of *err1* for RPEC1 and an increasing expression for RBEC2, which is in good accordance with the type of promoters used. The expression of *err1* in the parental strain first decreased until it reaches a minimum at 42 h after inoculation. Afterwards, it slightly reincreased, but always remained lower than in the overexpression strains (Fig. 5d).

Discussion

The by an alkaline organosolve process pretreated wheat straw [22] used in our experiments, turned out to be a very well utilizeable substrate for *T. reesei* cultivation. In contrary to other pretreatment processes, this method does not require any chemicals or catalysts that subsequently inhibit fungal growth. The alcohol, which is used in the process as organic solvent, can be sufficiently removed by washing. The achieved removal of lignin (up to a residual share of 1 %) makes the utilizeable cellulose and hemicellulose enough accessible for the fungus so that even direct inoculation with conidia was possible with this substrate as sole carbon source.

The comparison of the recombinant strains with their respective parentals showed that the overexpression of *err1* was successful an led to an increase in erythritol formation. In case of the wild-type and its recombinant strains this effect was more pronounced in shake flask cultivations on D-xylose, whereas Rut-C30 and its recombinant strains yielded better results in the bioreactor cultivation on pretreated wheat straw. Not only the relative increase of erythritol concentration in the recombinant strains compared to the parental strain was higher, but also the total amount of erythritol produced was about 10-fold increased compared to the wild-type and its recombinant strains. This observation can be explained by the fact that Rut-C30 is a cellulase hyperproducing, carbon catabolite derepressed strain, which makes it very likely that it better utilizes a complex substrate like wheat straw. This assumption is supported by the observed increased biomass formation and enhanced xylanase activity produced. Concerning the promoters used, the constitutive *pki* promoter seems to be favorable, since the erythritol production peak was slightly higher and this high level remained for a longer period (54 h). It should be mentioned that an even higher maximum might occur between the samples taken. However, cultivation time turned out to be an important factor for the erythritol formation, since after the peak of production, the erythritol concentration drops about as fast as it rises in the beginning. Accordingly, the elimination of the back reaction can be considered as one of the main targets of further strain improvement. Since we found that in *T. reesei* erythritol is not exported to the media, but accumulated in the cell, presumably, the most efficient way to prevent the back reaction would be to force the fungus to secrete the erythritol. This strategy would also be fa-

4

vorable in consideration of the osmotic balance of the cell. Taking into account that in erythritol production methods using yeasts, erythritol can be found in the supernatant (see *e.g.* [23–25]), in yeasts must exist a transport system for erythritol that probably can be introduced into *T. reesei*. Another strategy to improve erythritol formation could be to reduce the accumulation of other polyols. This would on the one hand provide additional starting material for the erythritol production, and at the same time it would prevent an additional rise of the intracellular osmotic pressure by these substances.

GC analysis of the cultivation broth of Rut-C30 and its recombinant strains grown on wheat straw revealed especially a high accumulation of arabinitol, but also considerable amounts of xylitol (data not shown). Both substances are metabolites in the interconversion of the pentoses derived from lignocellulose degradation (*i.e.* L-arabinose, D-xylose) (Fig. 6). Overexpression of the L-arabinitol dehydrogenase and the D-xylulose reductase in *T. reesei* might help here to enforce the flux of these two major substrates into the pentose phosphate pathway (PPP) and thus enhance erythritol formation, which is a side product of the PPP. Even if the amounts of erythritol produced by now in *T. reesei* do not reach the current production standards with yeasts, it must be taken into consideration, that these yeast strains are highly mutagenized, and subsequently selected for high erythritol production. Additional metabolic engineering as described above and strain screening might lead to competitive production levels in biomass-degrading fungi like *T. reesei*, with the advantage of using cheap and sustainable substrates.

Conclusions

In the present study we demonstrated that the production of erythritol on the renewable, non-food substrate wheat straw, using *T. reesei* is possible. The alkaline organosolve pretreatment process used for the wheat straw is compatible for subsequent fungal growth and provides an easily utilizeable substrate. Moreover, strain modification by overexpression of *err1* led to increased erythritol formation on this substrate.

Methods

Strains and cultivation conditions

The *T. reesei* strains QM6aΔ*tmus53* [26] and Rut-C30 (ATCC 56765), which was derived from the wild-type strain QM6a by one UV-light and two N-methyl-N'-nitro-N-nitrosoguanidine mutation steps [27], were maintained on malt extract (MEX) agar. The recombinant *T. reesei* strains QPEC1, QBEC2, RPEC1, and RBEC2 generated during this study, were maintained on MEX agar containing 250 μl/l hygromycin B (Merck, Darmstadt, Germany).

Purification of transformed strains by streak out of spores was done on MEX agar containing 250 μl/l hygromycin B and 500 μl/l IGEPAL® CA-630 (Sigma-Aldrich, St. Louis, MO, USA).

Cultivation in shake flasks was performed in 250-ml-Erlenmeyer flasks containing 50 ml Mandels-Andreotti (MA) medium [28] supplemented with 1 % (w/v) D-xylose or 1 % (w/v) birch-wood xylan. For inoculation 10^9 conidia per liter were used. Growth conditions were pH 5, 30 ℃, and 160 rpm shaking rate. Mycelia and supernatant were seperated by filtration. For short-term storage, harvested mycelia were shock-frozen and kept in liquid nitrogen, supernatants were kept at -20 ℃.

Plasmid construction

The *err1* gene and the promoter region of *bxl1* (1.5 kbp upstream *bxl1*, p*bxl1*) from *T. reesei* were amplified from cDNA, which was generated as described below in the according section. Primers were used to introduce restriction sites adjacent to the gene. Primer sequences are given in Table 1. The PCR product was subcloned into pJET-1.2 (Thermo Scientific, Waltham, MA, USA), using chemically competent *E. coli* TOP 10 (Invitrogen, Life Technologies Ltd, Paisley, UK) for plasmid replication.

For the construction of pBJ-PEC1 the vector pRLM$_{ex30}$ [29], which contains the *hph* gene flanked by the *pki* promoter (p*pki*) and the *cbh2* terminator, was used. The *hph* gene was removed by *Nsi*I/*Xba*I digestion and subsequently, *err1* that was excised from pJET-1.2 also by *Nsi*I/*Xba*I digestion, was inserted.

For the construction of pBJ-BEC2 p*pki* was excised from pBJ-PEC1 with *Xho*I/*Xba*I digestion and replaced by p*bxl1*, excised from pJET-1.2 with *Sal*I/*Xba*I digestion.

Protoplast transformation

For QM6aΔ*tmus53* protoplast transformation was performed as described in [30]. 5 µg of either pBJ-PEC1 or pBJ-BEC2 and 1 µg pAN7, which confers hygromycin B resistance [31], were co-transformed into the fungal genome.

Biolistic transformation

Rut-C30 was transformed with the Biolistic® PDS-1000/He Particle Delivery System (Bio-Rad Laboratories, Hercules, CA, USA) according to a modified protocol originally described in [32]. 5 µg of either pBJ-PEC1 or pBJ-BEC2 and 1 µg pAN7, which confers hygromycin B resistance [31], were co-transformed into the fungal genome.

DNA isolation

Fungal genomic DNA was isolated by phenol-chloroform extraction, using a FastPrep®-24 (MP Biomedicals, Santa Ana, CA, USA) for cell disruption. About 100 mg of mycelia was transferred to 400 µl DNA extraction buffer (0.1 M Tris-HCl pH 8.0, 1.2 M NaCl, 5 mM EDTA) and grounded with glass beads (0.37 g Ø 0.01 – 0.1 mm, 0.25 g Ø 1 mm, 1 piece Ø 3 mm) using the FastPrep. Afterwards, the mixture was immediately put on 65 ℃, supplemented with 9 µM RNase A, and incubated for 30 min. Then 200 µl of phenol (pH 7.9) and 200 µl of a chloroform-isoamyl alcohol-mixture (25:1) were added, with vigorous mixing following each addition. Phases were separated by centrifugation (12000 g, 10 min, 4 ℃) and the aqueous phase was transferred into a new vial. DNA was precipitated by addition of the 0.7-fold volume of isopropanol to the aqueous phase. After 20 min incubation at room temperature (RT) the DNA was separated by centrifugation (20000 g, 20 min, 4 ℃) and washed with 500 µl ethanol (70 %). The air-dried DNA pellet was solubilised in 50 µl Tris-HCl (10 mM, pH 7.5) at 60 ℃.

RNA isolation and cDNA synthesis

RNA extraction from fungal mycelia was performed with peqGOLD TriFast™(peqlab, Erlangen, Germany) according to the manufacturer's procedure, using a FastPrep®-24 (MP Biomedicals) for cell disruption. RNA quantity and quality were determined with a NanoDrop 1000 (Thermo Scientific).

A 260 nm/280 nm ratio of at least 1.8 was stipulated for further sample processing. cDNA synthesis was performed with RevertAid™H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer's procedure using 0.5 µg of RNA.

Transcript analysis

RT-qPCR analysis was performed in a Rotor-Gene Q cycler (Qiagen, Hilden, Germany). The qPCR amplification mixture had a total volume of 15 µl, containing 7.5 µl 2x IQ SYBR Green Supermix (Bio-Rad Laboratories), 100 nM forward and reverse primer, and 2 µl cDNA (diluted 1:100). Primer sequences are given in Table 1. As reference genes *act1* and *sar1* were used [33]. All reactions were performed in triplicates. For each gene a no-template control and a no-amplification control (0.01 % SDS added to the reaction mixture) was included in each run. The cycling conditions for *act1* and *err1* comprised 3 min initial denaturation and polymerase activation at 95 ℃, followed by 40 cycles of 15 s at 95 ℃, 15 s at 59 ℃ and 15 s at 72 s. For *sar1* different cycling conditions were applied: 3 min initial denaturation and polymerase activation at 95 ℃, followed by 40 cycles of 15 s at 95 ℃, and 120 s at 64 s. PCR efficiency was calculated from the Rotor-Gene Q software. Relative expression levels were calculated using the equation

$$relative\ transcript\ ratio = E_r^{C_r} \cdot E_t^{-C_t} \cdot E_{r_0}^{-C_{r_0}} \cdot E_{t_0}^{C_{t_0}}, \tag{1}$$

where E is cycling efficiency, C is the threshold cycling number, r is the reference gene, t the target gene and a 0 marks the sample which is used as the reference [34].

Probe preparation for Southern blot analysis

For the probe preparation 500 ng of *err1* cDNA, 5 µl 10x Klenow buffer (Thermo Scientific), and 6.5 µl 100 µM random hexamer primer (Thermo Scientific) were filled up with double distilled water (ddH₂O) to a final volume of 39 µl and incubated at 95 ℃ for 5-10 min. The reaction mixture was put on ice and 5 µl Biotin PCR Labeling Mix (Jena Bioscience, Jena, Germany) and 1 µl Klenow fragment exo- (Thermo Scientific) were added. The mixture was filled up with ddH₂O to a final volume of 50 µl and incubated at 37 ℃ for 24 h. For DNA precipitation 10 µl LiCl

(4 M) and 200 μl ethanol (96 %) were added. After incubation at RT for 15 min, at ice for 15 min, and at -20 ℃ for 1 h, the DNA was separated by centrifugation with 20000 g at 4 ℃ for 30 min. The pellet was washed with 500 μl ethanol (70 %), followed by centrifugation with 20000 g at 4 ℃ for 10 min. After drying the pellet at 50 ℃ for about 10 min, it was dissolved in 100 μl ddH$_2$O. The quality of the probe was tested by agarose gel electrophoresis, and the concentration was determined with the NanoDrop 1000 (Thermo Scientific).

Southern blot analysis

For the Southern blot 15 μg chromosomal DNA of each strain used in this study was digested in a triple digestion with 5 μl of each *Nde*I, *Sal*I, and *Bgl*II (each 10 U/μl, Thermo Scientific), using 10x Buffer O (Thermo Scientific). The reaction mixtures were filled up with ddH$_2$O to a final volume of 100 μl and then split into 20 μl aliquots for digestion at 37 ℃ over night (o/n). After digestion, samples were incubated at 70 ℃ for 20 min and the completion of digestion controlled by agarose gel electrophoresis. Digestion aliquots of each sample were pooled and concentrated to a final volume of 10 - 20 μl and applied to a 1 % agarose gel. As length standard 5 μl Gene Ruler 1 kb DNA Ladder (Thermo Scientific) was used. Using the Mini-Sub® Cell system (Bio-Rad Laboratories) the gel was run at 80 V for 1 h in TAE buffer, and afterwards incubated in 0.4 M NaOH and 0.6 M NaCl, and then in 0.5 M Tris (pH 7.5) and 1.5 M NaCl for 30 min each. The DNA was transferred to a Biodyne B membrane (Pall Corporation, Port Washington, NY, USA) by a capillary blot with 10x saline-sodium citrate (SSC) buffer (1.5 M NaCl, 0.15 M sodium citrate, pH 7.2) o/n. After blotting, the membrane was incubated in 0.4 M NaOH, and then in 0.2 M Tris (pH 7.5) for 1 min each. Cross-linking was performed with a GS Gene Linker UV chamber (Bio-Rad Laboratories) using program C3 and 150 mJoule on the wet membrane. For pre-hybridization, the membrane was incubated at 65 ℃ for 3 h in 20 ml Southern blot hybridization buffer (25 % (v/v) 20x SSC, 10 % (v/v) 50x Denhardt's solution, 0.2 % (v/v) EDTA (0.5 M, pH 8.0), 0.05 M NaH$_2$PO$_4$, 0.1 % (w/v) SDS, 0.5 % (w/v) BSA), supplemented with 100 μg/ml single stranded salmon sperm DNA and freshly denaturated (10 min at 95 ℃) probe. For hybridization, the membrane was incubated at 65 ℃ o/n in

10 ml Southern blot hybridization buffer. The membrane was washed twice at RT for 5 min in 50 ml 2x SSC supplemented with 0.1 % (w/v) SDS, followed by washing twice at 65 ℃ for 15 min with 50 ml 0.1x SSC, supplemented with 0.1 % (w/v) SDS. After incubation at RT for 10 min in Southern blot blocking solution (125 mM NaCl, 17 mM Na$_2$HPO$_4$, 8 mM NaH$_2$PO$_4$, 0.5 % (w/v) SDS, pH 7.2), the membrane was incubated light-protected at RT for 30 min in Southern blot blocking solution supplemented with 1 μg/ml Dylight 650-labeled Streptavidin (Thermo Scientific). The membrane was washed 4 times light-protected at RT for 10 min in 50 ml 1:10 diluted Southern blocking solution and then scanned with Typhoon FLA 9500 (GE Healthcare Life Sciences, Buckinghamshire, England) set for Alexa Fluor 647 at 1000 V.

Cultivation in bioreactors

Cultivation was performed in 2-l-bench top bioreactors (Bioengineering AG, Wald, Swiss), containing 1.3 l fermentation medium ((NH$_4$)$_2$SO$_4$ 3.50 g/l, KH$_2$PO$_4$ 5.00 g/l, MgSO$_4$·7H$_2$O 1.25 g/l, NaCl 0.625 g/l, peptone from Casein 1.25 g/l, Tween® 80 0.625 g/l), supplemented with 1.5 ml/l trace element solution (FeSO$_4$·7H$_2$O 0.90 mM, MnSO$_4$·H$_2$O 0.50 mM, ZnSO$_4$·7H$_2$O 0.24 mM, CaCl$_2$·7H$_2$O 0.68 mM), 1.7 % (w/v) wheat straw (pretreated by an alkaline organosolv process for lignin removal [22] (Annikki, Graz, Austria)), and Antifoam Y-30 Emulsion (1 ml/bioreactor). For inoculation 10^9 conidia per liter were used. Agitation rate was 500 rpm, temperature was 28 ℃, and aeration rate was 0.5 vvm.

GC analysis

Mycelia from shake flask cultures were ground under liquid nitrogen. The powder was suspended in 3 ml distilled water and sonicated using a Sonifier® 250 Cell Disruptor (Branson, Danbury, CT, USA) (power 70 %, duty cycle 40 %, power for 3 min, on ice). Insoluble compounds were separated by centrifugation (20000 g, 10 min, 4 ℃), the clear supernatant was used for further processing.

Supernatants from shake flask cultures were used directly for further processing.

For samples from cultivation in bioreactors 30 ml of the whole cultivation broth were first mechanically disrupted with a potter for 1 min, then sonicated,

and afterwards centrifuged as described above for mycelia from shake flask cultures.

Sample preparation for GC was done in triplicates as follows: 300 μl of the clear supernatant (prepared as described above), supplemented with 10 ng myo-inositol as internal standard, was gently mixed with 1.2 ml ethanol (96 %) and incubated for 30 min at RT for protein precipitation. The precipitate was separated by centrifugation (20000 g, 10 min, 4 ℃). Samples were dried under vacuum and thereafter silylated (50 μl pyridine, 250 μl hexamethyldisilazane, 120 μl trimethylsilyl chloride). For quantitative erythritol determination a GC equipment (Agilent Technologies, Santa Clara, CA, USA) with a HP-5-column (30 m, inner diameter 0.32 mm, film 0.26 μm) (Agilent) was used. The mobile phase consisted of helium with a flow of 1.4 l/min, the column temperature was as follows: 150 ℃ for 1 min, ramping 150 – 220 ℃ (ΔT 4 ℃/min), ramping 220 - 320 ℃ (ΔT 20 ℃/min), 320 ℃ for 6.5 min. Detection was performed with FID at 300 ℃. The retention times were determined using pure standard substances.

Sodium hydroxide soluble protein (SSP)

2 ml cultivation broth were centrifuged at 20000 g for 10 min at 4 ℃. The supernatant was discarded and the pellet resuspended in 3 ml 0.1 M NaOH before sonication with a Sonifier® 250 Cell Disruptor (Branson) (power 70 %, duty cycle 40 %, power 20 s, pause 40 s, 10 cycles, on ice). The sonicated samples were incubated for 3 h at RT. After centrifugation (20000 g, 10 min, 4 ℃) the supernatant was used to determine protein concentration with a Bradford assay. Therefore, 20 μl diluted sample (1:10 - 1:100) were added to 1 ml 1:5 diluted Bradford Reagens (Bio-Rad Laboratories) and incubated for exactly 10 min at RT before messuring the absorption on a V-630 UV-Vis spectrophotometer (Jasco, Tokio, Japan) at 595 nm. As standard bovine serum albumin in concentrations from 10 - 100 μg/ml was used.

Competing interests

A European patent entitled 'Method for the production of erythritol' (no. EP20100183799, 5.4.2012) [35] was issued.

Author's contributions

BJ generated and characterized the recombinant strains, conducted the cultivations and analyses and drafted the manuscript. ARMA contributed to the manuscript, the design of the study, and supervision of experiments. RLM contributed to the design of the study. All authors critically read the manuscript.

References

1. Moon HJ, Jeya M, Kim IW, Lee JK: **Biotechnological production of erythritol and its applications**. *Appl Microbiol Biotechnol* 2010, **86**(4):1017–1025, [http://dx.doi.org/10.1007/s00253-010-2496-4].

2. Livesey G: **Tolerance of low-digestible carbohydrates: a general view**. *British Journal of Nutrition* 2001, **85**:7–16.

3. Kasumi T: **Fermentative production of polyols and utilization for food and other products in Japan**. *JARQ* 1995, **29**:49–55.

4. Munro I, Bernt W, Borzelleca J, Flamm G, Lynch B, Kennepohl E, Bär E, Modderman J: **Erythritol: an interpretive summary of biochemical, metabolic, toxicological and clinical data**. *Food and Chemical Toxicology* 1998, **36**(12):1139 – 1174, [http://www.sciencedirect.com/science/article/pii/S027869159800091X].

5. Jovanovic B, Mach R, Mach-Aigner A: **Characterization of erythrose reductases from filamentous fungi**. *AMB Express* 2013, **3**:43, [http://www.amb-express.com/content/3/1/43].

6. Kuhls K, Lieckfeldt E, Samuels GJ, Kovacs W, Meyer W, Petrini O, Gams W, Börner T, Kubicek CP: **Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina***. *Proceedings of the National Academy of Sciences* 1996, **93**(15):7755–7760, [http://www.pnas.org/content/93/15/7755.abstract].

7. Buchert J, Oksanen T, Pere J, Siika-aho M, Suurnäkki A, Viikari L: **Applications of *Trichoderma reesei* enzymes in the pulp and paper industry**. In *Trichoderma & Gliocladium, Volume 2*. Edited by Harman G, Kubicek C, London, UK: Taylor & Francis Ltd. 1998:343–357.

8. Noé P, Chevalier J, Mors F, Comtat J: **Action of xylanases on chemical pulp fibers Part II : Enzymatic beating**. *Journal of Wood Chemistry and Technology* 1986, **6**(2):167–184, [http://www.tandfonline.com/doi/abs/10.1080/02773818608085222].

9. Welt T, Dinus R: **Enzymatic deinking - a review**. *Progress in Paper Recycling* 1995, **4**(2):36–47.

10. Galante Y, R M, S I, Caldini C, de Conti A, Lavelli V, Bonomi F: **New applications of enzymes in wine making and olive oil production**. *Italian Biochem Soc trans* 1993, **4**:34.

11. Lanzarini G, Pifferi P: **Enzymes in the fruit juice industry**. In *Biotechnology Applications in Beverage Production*, Elsevier Applied Food Science Series. Edited by Cantarelli C, Lanzarini G, Springer Netherlands 1989:189–222, [http://dx.doi.org/10.1007/978-94-009-1113-0_13].

12. Walsh G, Power R, Headon D: **Enzymes in the animal-feed industry**. *Trends Biotechnol* 1993, **11**(10):424–430.

13. Koo H, Ueda M, Wakida T, Yoshimura Y, Igarashi T: **Cellulase treatment of cotton fabrics**. *Textile Research Journal* 1994, **64**(2):70–74, [http://trj.sagepub.com/content/64/2/70.abstract].

14. Kumar A, Lepola M, Purtell C: **Enyme finishing of man-made cellulosic fabrics**. *Textile Chemist and Colorist* 1994, **26**(10):25–28.

15. Pedersen GP, Screws GA, Cereoni DA: **Biopolishing of cellulosic fabrics**. *Can Text J* 1992.

16. Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund M, Liden G, Zacchi G: **Bio-ethanol - the fuel of tomorrow from the residues of today**. *Trends Biotechnol* 2006, **24**(12):549–556.

17. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD: **Biomass recalcitrance: engineering plants and enzymes for biofuels production**. *Science* 2007, **315**(5813):804–807.

18. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T: **The path forward for biofuels and biomaterials**. *Science* 2006, **311**(5760):484–489, [http://www.sciencemag.org/content/311/5760/484.abstract].

19. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, Leon ALd, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)**. *Nature Biotechnology* 2008, **26**(5):553–560.

20. Acebal C, Castillon M, Estrada P, Mata I, Costa E, Aguado J, Romero D, Jimenez F: **Enhanced cellulase production from *Trichoderma reesei* QM 9414 on physically treated wheat straw**. *Applied Microbiology and Biotechnology* 1986, **24**(3):218–223, [http://dx.doi.org/10.1007/BF00261540].

21. Dashtban M, Kepka G, Seiboth B, Qin W: **Xylitol production by genetically engineered *Trichoderma reesei* strains using barley straw as feedstock**. *Applied Biochemistry and Biotechnology* 2013, **169**(2):554–569, [http://dx.doi.org/10.1007/s12010-012-0008-y].

22. Fackler K, Ters T, Ertl O, Messner K: **Method for lignin recovery**. *Patent WO/2012/027767* 2012, [http://www.sumobrain.com/patents/WO2012027767.html].

23. Ryu YW, Park CY, Park JB, Kim SY, Seo JH: **Optimization of erythritol production by *Candida magnoliae* in fed-batch culture**. *Journal of Industrial Microbiology and Biotechnology* 2000, **25**(2):100–103, [http://dx.doi.org/10.1038/sj.jim.7000039].

24. Rymowicz W, Rywińska A, Marcinkiewicz M: **High-yield production of erythritol from raw glycerol in fed-batch cultures of *Yarrowia lipolytica***. *Biotechnology Letters* 2009, **31**(3):377–380, [http://dx.doi.org/10.1007/s10529-008-9884-1].

25. Sawada K, Taki A, Yamakawa T, Seki M: **Key role for transketolase activity in erythritol production by *Trichosporonoides megachiliensis* SN-G42**. *Journal of Bioscience and Bioengineering* 2009, **108**(5):385 – 390, [http://www.sciencedirect.com/science/article/pii/S1389172309002497].

26. Steiger M, Vitikainen M, Uskonen P, Brunner K, Adam G, Pakula T, Penttilä M, Saloheimo M, Mach R, Mach-Aigner AR: **Transformation system for *Hypocrea jecorina* (*Trichoderma reesei*) that favors homologous integration and employs reusable bidirectionally selectable markers**. *Appl Environ Microbiol* 2011, **77**:114–121.

27. Montenecourt BS, Eveleigh DE: **Selective screening methods for the isolation of high yielding cellulase mutants of *Trichoderma reesei***. In *Hydrolysis of cellulose: Mechanisms of enzymatic and acid catalysis*. Edited by Brown RD, Jurasek L, Washington, D. C.: American Chemical Society 1979:289–301, [http://pubs.acs.org/doi/abs/10.1021/ba-1979-0181.ch014].

28. Mandels M: **Applications of cellulases**. *Biochem Soc Trans* 1985, **13**(2):414–416.

29. Mach R, Schindler M, Kubicek CP: **Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expression signals**. *Curr Genet* 1994, **25**(6):567–570.

30. Gruber F, Visser J, Kubicek CP, de Graaff LH: **The development of a heterologous transformation system for the cellulolytic fungus *Trichoderma reesei* based on a *pyrG*-negative mutant strain**. *Curr Genet* 1990, **18**:71–76.

31. Punt PJ, Oliver RP, Dingemanse MA, Pouwels PH, van den Hondel CA: **Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli***. *Gene* 1987, **56**:117–124, [http://www.sciencedirect.com/science/article/pii/0378111987901648].

32. Te'o V, Bergquist P, Nevalainen K: **Biolistic transformation of *Trichoderma reesei* using the Bio-Rad seven barrels Hepta Adaptor system**. *Journal of Microbiological Methods* 2002, **51**(3):393 – 399, [http://www.sciencedirect.com/science/article/pii/S0167701202001264].

33. Steiger MG, Mach RL, Mach-Aigner AR: **An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*)**. *Journal of Biotechnology* 2010, **145**:30–37, [http://www.sciencedirect.com/science/article/pii/S0168165609004775].

34. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR**. *Nucleic Acids Res* 2001, **29**(9):45.

35. Mach R, Mach-Aigner A: **Method for the production of erythritol**. *Patent EP20100183799* 2012, [http://www.google.de/patents/EP2436772A1?cl=en].

Figures

Figure 1 - Transcript analysis of err1 in parental and recombinant T. reesei strains

The *T. reesei* wild-type strain (WT) and preselected recombinant strains derived from transformation of the wild-type (a, b) or of Rut-C30 (c, d), which are expressing *err1* either under the constitutive *pki* promoter (a, c) or under the inducible *bxl1* promoter (b, d), respectively, were cultivated in shake flasks on D-xylose (blue bars) for 30 h (wild-type) or 72 h (Rut-C30) and on birch-wood xylan (red bars) for 48 h (wild-type) or 72 h (Rut-C30). Strains chosen for further experiments are framed in yellow. The transcript analysis was performed by qPCR using *sar1* and *act1* as genes for data normalization and levels always refer to the wild-type strain on the respective carbon source (indicated by a blue and red asterisk). Results are given as relative transcript ratios in logarithmic scale (lg). The values are means from three measurements. Error bars indicate standard deviations. Biological experiments (cultivations) were performed in duplicates.

Figure 2 - Southern Blot analysis of parental and err1 overexpression T. reesei strains

On an agarose gel *Nde*I/*Sal*I/*Bgl*II-digested DNA from the wild-type (WT) strain bearing the native *err1* (2340 bp), the thereof derived *err1* overexpression strains QPEC1 (containing the native *err1* (2340 bp) and n+1 inserted fragments (5410 bp)) and QBEC2 (containing the native *err1* (2340 bp) and n+1 inserted fragments (6626 bp)), Rut-C30 bearing the native *err1* (2340 bp), and the thereof derived *err1* overexpression strains RPEC1 (containing the native *err1* (2340 bp) and n+1 inserted fragments (5410 bp)) and RBEC2 (containing the native *err1* (2340 bp) and n+1 inserted fragments (6626 bp)) was separated. n means the band intensity in relation to the native *err1*-containing band. A 1 kb DNA ladder (L) was used for estimation of DNA fragment size; indicated sizes are given in bp. As probe a biotin-labeled fragment containing the structural *err1* gene was used. For visualization Dylight 650-labeled streptavidin was applied and the membrane was scanned with a Typhoon FLA 9500.
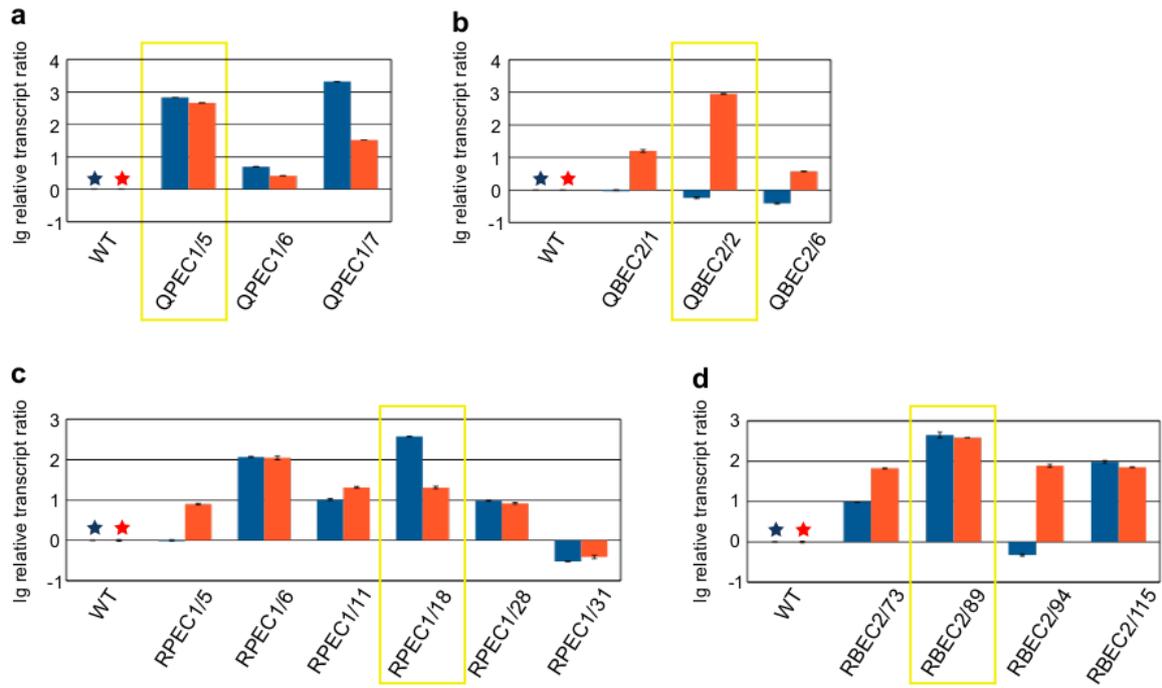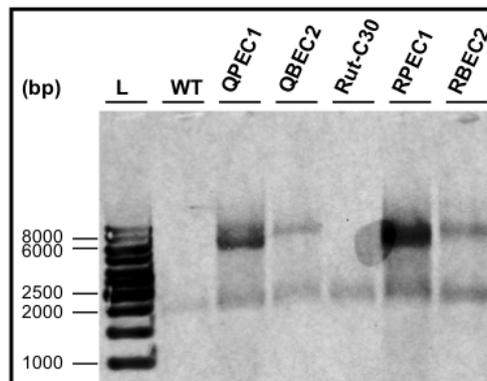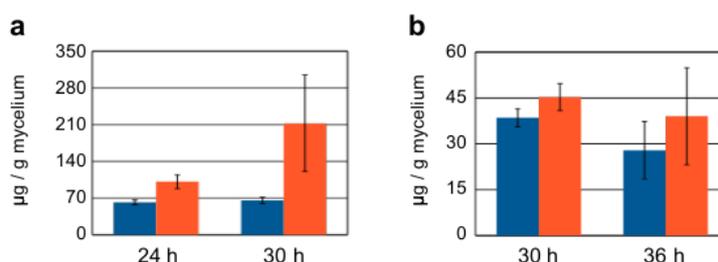
Fig. 1

Fig. 2

11

Fig. 3

Figure 3 - Erythritol production on D-xylose
The *T. reesei* (a) wild-type strain (blue bars) and the thereof derived *err1* overexpression strain QPEC1 (red bars) as well as (b) Rut-C30 (blue bars) and the thereof derived *err1* overexpression strain RPEC1 (red bars) were cultivated in shake flasks on D-xylose. Samples were taken after the indicated time and erythritol concentration was determined by GC-analysis from cell free extracts. Biological experiments (cultivations) were performed in duplicates. Standard deviations were obtained from two biological duplicates and measurements in triplicates each.

Figure 4 - Cultivation of the wild-type and err1 overexpression strains on wheat straw
The *T. reesei* wild-type strain (blue bars), and the *err1* overexpression strains QPEC1 (red bars) and QBEC2 (yellow bars) were cultivated in bench-top bioreactors on pretreated wheat straw. Samples were taken after 48 and 72 h. (a) Sodium soluble protein concentration (given in g/l cultivation broth) was measured in triplicates in cultivation broth samples after cell disruption to indicate biomass formation. Standard deviations were below 5 %. (b) Xylanase activity (given in U/l cultivation broth) was measured in triplicates in the cultivation supernatants. Standard deviations were below 5 %. (c) Erythritol concentration (given in mg/l cultivation broth) was measured in triplicates by GC in cultivation broth samples after cell disruption. Error bars indicate standard deviations. (d) Transcript analysis of *err1* (given in as relative transcript ration in logarithmic scale (lg)) was performed by qPCR in triplicates using *sar1* and *act1* as genes for data normalization and levels always refer to the wild-type strain cultivated for 48 h (as indicated by an asterisk). Error bars indicate standard deviations. Biological experiments (cultivations) were performed in duplicates.

Figure 5 - Cultivation of Rut-C30 and err1 overexpression strains on wheat straw
Rut-C30 (blue bars), and the *err1* overexpression strains RPEC1 (red bars) and RBEC2 (yellow bars) were cultivated in bench-top bioreactors on pretreated wheat straw. Samples were taken after 18, 30, 42, 54, 66, and 72 h. (a) Sodium soluble protein concentration (given in g/l cultivation broth) was measured in triplicates in cultivation broth samples after cell disruption to indicate biomass formation. Standard deviations were below 5 %. (b) Xylanase activity (given in U/l cultivation broth) was measured in triplicates in the cultivation supernatants. Standard deviations were below 5 %. (c) Erythritol concentration (given in mg/l cultivation broth) was measured in triplicates by GC in cultivation broth samples after cell disruption. Error bars indicate standard deviations. (d) Transcript analysis of *err1* (given in as relative transcript ration in logarithmic scale (lg)) was performed by qPCR in triplicates using *sar1* and *act1* as genes for data normalization and levels always refer to Rut-C30 cultivated for 18 h (as indicated by an asterisk). Error bars indicate standard deviations. Biological experiments (cultivations) were performed in duplicates.

Fig. 4
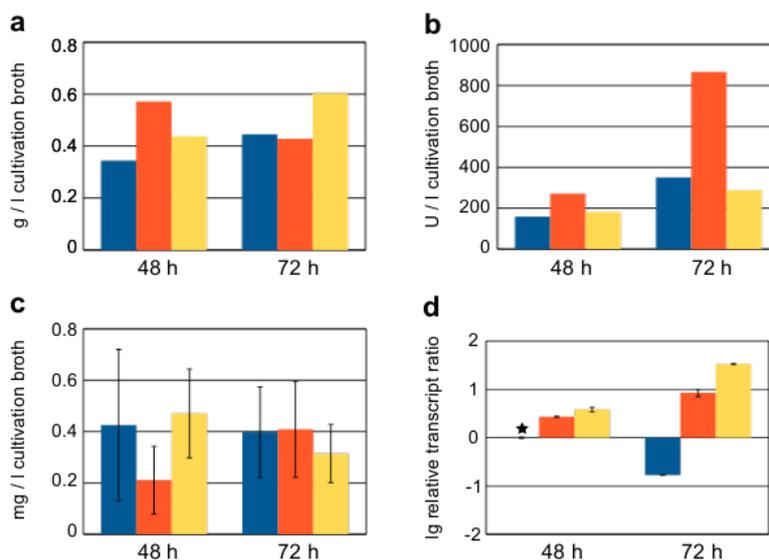
Figure 6 - **Schematic drawing of metabolic pathways of pentoses and erythritol in T. reesei**
Metabolites are given in boxes. Monomeric sugars derived from hydrolytic lignocellulose degradation by
T. reesei are given in orange. The target substance, erythritol, is given in purple. Enzyme names and EC
numbers are given in green. Adjacent pathways are indicated in blue. Dashed arrows indicate (possible)
involvement of more than one enzyme.

Tables

Table 1 - **Oligonucleotides used during the study**

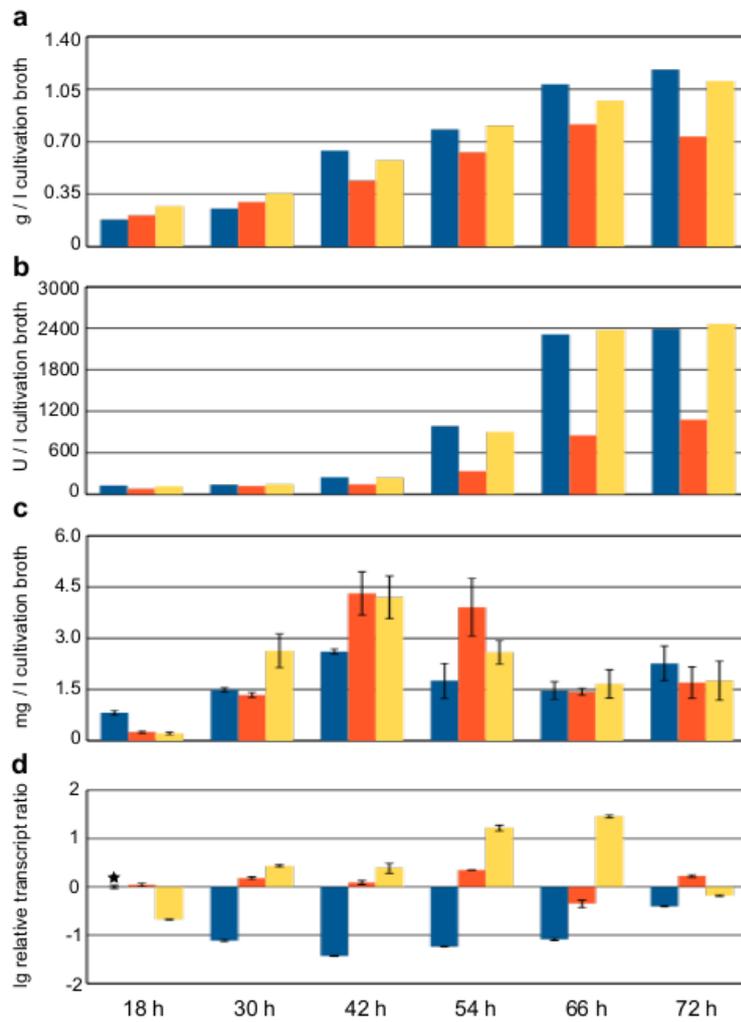| Table 1 - Oligonucleotides used during the study		
Name	**Sequence (5' - 3')**	**Usage**
pbxl1_SalI_EcoRI_f	ATATA GTCGAC GAATTC AGCTTGTCTGCCTTGATTACCATCC	Vector construction
pbxl1_XbaI_r	ATATA TCTAGA TGCGTCCGGCTGTCCTTC	Vector construction
err1_XbaI_f	ATATA TCTAGA ATGTCTTCCGGAAGGACC	Vector construction
err1_Nsi_r	TATAT ATGCAT TTACAGCTTGATGACAGCAGTG	Vector construction
ppki_f	GCACGCATCGCCTTATCGTC	PCR test
qerr1_f	CTTTACCATTGAGCACCTCGACG	RT-qPCR
qerr1_r	GGTCTTGCCCTGCTTCTTGG	RT-qPCR
qact1_f	TGAGAGCGGTGGTATCCACG	RT-qPCR
qact1_r	GGTACCACCAGACATGACAATGTTG	RT-qPCR
qsar1_f	TGGATCGTCAACTGGTTCTACGA	RT-qPCR
qsar1_r	GCATGTGTAGCAACGTGGTCTTT	RT-qPCR
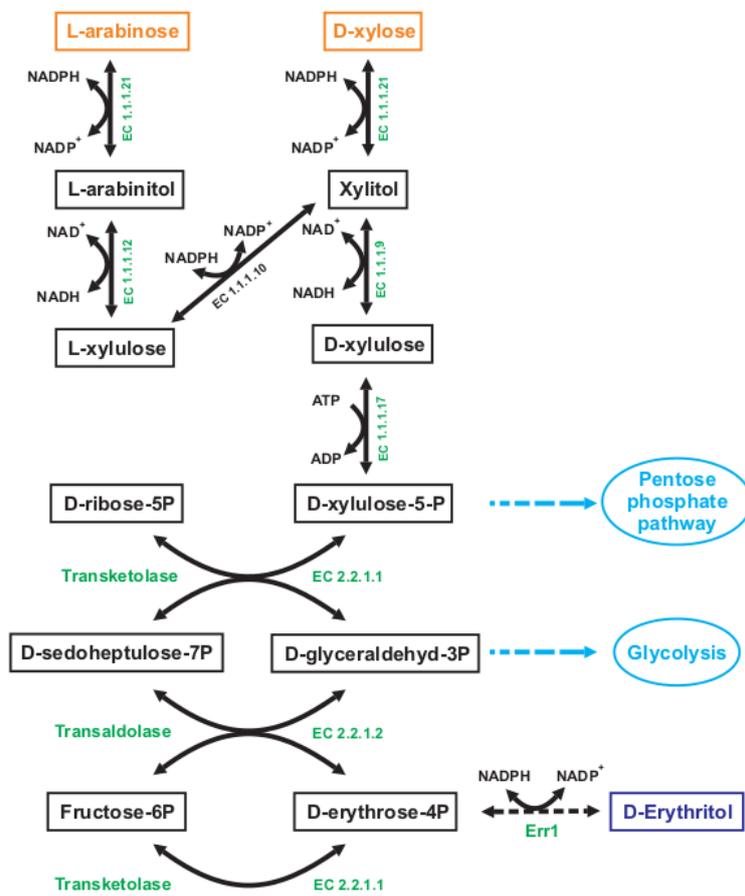
13

Fig. 5

Fig. 6

A.3 A highly sensitive *in vivo* footprinting technique for condition-dependent identification of *cis* elements

A highly sensitive *in vivo* footprinting technique for condition-dependent identification of *cis* elements

Rita Gorsche[1], Birgit Jovanovic[1], Loreta Gudynaite-Savitch[2], Robert L. Mach[1] and Astrid R. Mach-Aigner[1,*]

[1]Research Division Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Str. 1 a, A-1060 Vienna, Austria and [2]Department of Biology, University of Ottawa, Gendron Hall, 30 Marie Curie, Ottawa, ON, K1N6N5, Canada

ABSTRACT

Knowing which regions of a gene are targeted by transcription factors during induction or repression is essential for understanding the mechanisms responsible for regulation. Therefore, we re-designed the traditional *in vivo* footprinting method to obtain a highly sensitive technique, which allows identification of the *cis* elements involved in condition-dependent gene regulation. Data obtained through DMS methylation, HCl DNA cleavage and optimized ligation-mediated PCR using fluorescent labelling followed by capillary gel electrophoresis are analysed by ivFAST. In this work we have developed this command line-based program, which is designed to ensure automated and fast data processing and visualization. The new method facilitates a quantitative, high-throughput approach because it enables the comparison of any number of *in vivo* footprinting results from different conditions (e.g. inducing, repressing, de-repressing) to one another by employing an internal standard. For validation of the method the well-studied upstream regulatory region of the *Trichoderma reesei xyn1* (endoxylanase 1) gene was used. Applying the new method we could identify the motives involved in condition-dependent regulation of the *cbh2* (cellobiohydrolase 2) and *xyn2* (endoxylanase 2) genes.

INTRODUCTION

The sequence-specific binding of transcription factors to the DNA is a key element of transcriptional regulation (1–3). Therefore, the knowledge of which areas of an upstream regulatory region (URR) are specifically targeted by proteins is essential for the further understanding of regulatory mechanisms. For this purpose *in vivo* and *in vitro* footprinting methods employing nucleases such as DNaseI (4–7) or alkylating agents such as dimethylsulfate (DMS) (8,9) are routinely used to detect protein–DNA interactions. DMS treatment of DNA leads to methylation of guanine and adenine residues, with each guanine or adenine residue of purified DNA having the same probability of being methylated. When used for *in vivo* footprinting DMS readily penetrates living cells. There, protein–DNA interactions cause either a decreased accessibility of certain G or A residues to DMS (protection) or an increased reactivity (hypersensitivity) (10).

The URRs of eukaryotic DNA are complex and include a number of different recognition sites that can be targeted by multiple transcription factors at a time (2). Furthermore, the important regulatory elements are often hundreds of bases away from the transcription start (1), necessitating the coverage of large regions in the footprinting reactions. Additionally, various genes and transcription factors are grouped together in regulons. Elucidating the binding characteristics of transcription factors as well as the transcriptional regulation and interdependencies in regulons requires the analysis of footprinting patterns of the URRs of a number of different genes under various different conditions. Therefore, a standardized, high-throughput approach to traditional *in vivo* footprinting allowing parallel investigation of a number of conditions and/or isolates is necessary.

The original protocol for DMS *in vivo* footprinting was already established in 1985 (8,9) and has been improved upon since then by adding ligation-mediated PCR (LM-PCR) (11). LM-PCR quantitatively maps single-strand DNA breaks having phosphorylated 5′-ends within single-copy DNA sequences. Briefly, it involves blunt-end ligation of an asymmetric double-stranded linker onto the 5′-end of each, before cleaved, blunt-ended DNA molecule. This linker adds a common and known

*To whom correspondence should be addressed. Tel: +43 664 60588 7253; Fax: +43 1 5880117299; Email: astrid.mach-aigner@tuwien.ac.at

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

sequence to all 5′-ends allowing exponential PCR amplification of an adjacent, unknown genomic sequence (12). Furthermore, optimizing the polymerase and cycling conditions (13), and adapting the method to different kinds of cells, from cell lines (8,11,14,15) and yeast (9) to filamentous fungi (16), was achieved. Nevertheless, due to the use of polyacrylamide gels and radioactive labelling of the DNA fragments the resulting protocol was laborious, used hazardous substances, yielded results of strongly varying quality, and consequently, was not yet suitable for high-throughput projects.

The use of fluorescent labels and separation of DNA fragments by capillary sequencer has meanwhile been introduced to a number of similar techniques, such as RFLP (17), AFLP (18), *in vitro* DNaseI footprinting (19) or chromatin analysis (20,21). In 2000, an approach applying automated LM-PCR with infrared fluorochrome-labelled primers and a LI-COR DNA sequencer for detection was used to compare *in vivo* to *in vitro* UV-treated DNA (22). In this study we employed [6-FAM]-labelling of the DNA fragments in DMS *in vivo* footprinting and analysis via capillary sequencer employing an internal size standard. Moreover, we made use of analysis by a certified sequencing service, which guarantees stable and controlled analysis conditions. This resulted in a fast and sensitive way to analyse fragment size as well as peak intensities in a large number of samples, providing an excellent tool for comparison of URRs in a number of different isolates and different conditions. The final step to an automated high-throughput *in vivo* footprinting technique is the manner in which the acquired data is processed. Traditional *in vivo* footprinting employs visual comparison to align sequences with band patterns and densitometric measurements to determine band intensities [e.g. (11,23–25)]. For standardized comparison of multiple samples from different experiments, a computational processing of the analysis data is paramount. Therefore, we developed a data analysis tool (termed ivFAST) that plots normalized peak area ratios against sequence data and automatically determines which bases are protected from or hypersensitive to methylation by DMS.

To test the new method we examined part of the Xyr1/Cre1 regulon of *Trichoderma reesei* (teleomorph *Hypocrea jecorina*). *Trichoderma reesei* is a filamentous ascomycete of great industrial importance because of its high potency in secretion of hydrolases. Xyr1 is recognized as the essential activator for most hydrolytic-enzyme encoding genes in *T. reesei*, e.g. *cbh1*, *cbh2* (Cellobiohydrolases I and II-encoding) and *egl1* (Enoglucanase I-encoding), as well as *xyn1* and *xyn2* (Xylanases I and II-encoding) (26,27). Previous footprinting experiments identified a 5′-GGC $(T/A)_3$-3′-motif as the Xyr1-binding site in the URRs of *cbh2*, *xyn1*, *xyn2* and *xyn3* (28–31). Cre1, on the other hand, is characterized as a repressor responsible for mediating carbon catabolite repression of hydrolytic-enzyme encoding genes (32), such as *cbh1* and *xyn1* (33,34). 5′-SYGGRG-3′ was found to be the consensus sequence for Cre1-binding (35).

In this study, the URR of the above-mentioned *xyn1* gene was used to validate the method. By using

traditional, gel-based *in vivo* footprinting next to the new, software-based method we found that the new method allows not only a comparison of *in vivo* methylated samples to naked DNA (i.e. *in vitro* methylated, genomic DNA used as a reference), but is sensitive enough for a comparison of *in vivo* methylated samples with each other. This we demonstrate by applying the new method to the URRs of the *cbh2* and *xyn2* gene. These URRs are of similar architecture bearing the so-called cellulase-activating element [CAE; 5′-ATTGGGT AATA-3′; (31)] or xylanase-activating element [XAE; 5′-GGGTAAATTGG-3′; (30)], respectively, of which both were previously identified as essential for gene regulation. By employing the new method we have detected the following motifs: (i) the CAE and the XAE, (ii) other generally known, but in these URRs so far unrecognized motifs (such as Xyr1- or Cre1-binding sites) and (iii) so far unknown motifs.

MATERIALS AND METHODS

Strains and growth conditions

The ascomycete *H. jecorina* (*T. reesei*) QM9414 [ATCC 26921; a cellulase hyper-producing mutant derived from wild-type strain QM6a (36)] and an according *xyr1* deletion strain (23) were used in this study and were maintained on malt agar. For replacement experiments mycelia were pre-cultured in 1-l-Erlenmeyer flasks on a rotary shaker (180 rpm) at 30°C for 18 h in 250 ml of Mandels-Andreotti (MA) medium (37) supplemented with 1% (w/v) glycerol as sole carbon source. An amount of 10^9 conidia per litre (final concentration) were used as inoculum. Pre-grown mycelia were washed and equal amounts were re-suspended in 20 ml of MA media containing 1% (w/v) glucose, 0.5 mM D-xylose, 1.5 mM sophorose as sole carbon source or no carbon source, respectively, and incubated for 3 h (growth conditions) or 5 h (resting cell conditions). For *in vitro* DNA methylation mycelium grown on rich medium (3% malt extract, 1% glucose, 1% peptone) was used.

In vivo methylation of genomic DNA

Methylation of DNA *in vivo* was performed according to Wolschek *et al.* (16). An amount of 40 μl of DMS in 2 ml MES (200 mM, pH 5.5) were added to 20 ml of fungal culture and incubated at 30°C and 180 rpm for 2 min. Methylation was stopped with 100 ml of ice-cold TLEβ buffer [10 mM Tris pH 8, 1 mM EDTA, 300 mM LiCl, 2% (v/v) β-mercaptoethanol]. Mycelia were harvested, washed with TLEβ buffer and distilled water, and frozen in liquid nitrogen. DNA extraction was performed according to standard protocol (38). The DNA was cleaved at methylated purines by incubating 100 μl of DNA solution (∼100 μg) with 6.3 μl HCl (0.5 M) on ice for 1.5 h (39). The DNA was precipitated with 25 μl sodium acetate (3 M, pH 5) and 500 μl ethanol, dissolved in 250 μl bi-distilled water and incubated at 90°C for 30 min with 10 μl NaOH (1 M). After addition of 25 μl Tris (1 M, pH 7.5) and adjustment of the pH to 7.5, the DNA fragments were again precipitated with sodium acetate and ethanol, dissolved in

100 µl Tris (10 mM, pH 7.5) and purified using the QIAquick Nucleotide Removal Kit (Qiagen, Hilden, Germany).

In vitro methylation of genomic DNA

For *in vitro* methylation genomic DNA extracted from mycelium grown on full medium was methylated according to Mueller *et al.* (14). An amount of 100 µl of DNA solution (~100 µg) was incubated with 400 µl of DMS reaction buffer (0.05 M sodium cacodylate, 0.001 M EDTA, pH 8) and 2 µl of DMS (1:20 dilution in bi-distilled water) at room temperature for 5 min. The reaction was stopped by adding 50 µl of stop solution (1.5 M sodium acetate, 1 M β-mercaptoethanol). The DNA was precipitated twice with sodium acetate and ethanol and dissolved in 100 µl Tris (10 mM, pH 7.5). Cleavage of the DNA was performed as described above. This DNA was used as one reference and we refer to it using the term 'naked DNA' throughout the manuscript.

Traditional, gel-based analysis of DNA fragments via LM-PCR

LM-PCR was performed using Vent Polymerase [New England Biolabs (NEB), Ipswich, MA] as described by Garrity and Wold (13). End-labelling of RG72-2 using γ-^{32}P-ATP was done according to Mueller and Wold (11) and resulting DNA fragments were extracted with phenol/chloroform/isoamylalcohol (25:24:1, vol/vol) and precipitated with ethanol. The DNA pellet was re-suspended in 10 µl of loading dye (0.05% bromophenol blue, 0.05% xylene cyanol, 20 mM EDTA), heated at 95°C for 5 min and loaded on a 6% polyacrylamide sequencing gel.

Generation of DNA fragments via modified LM-PCR

LM-PCR was modified from the original protocol of Mueller and Wold (11) and the adaptation of Wolschek *et al.* (16). First-strand synthesis was performed in a 30 µl reaction mixture containing 1× buffer (NEB), 0.01 µM oligo 1, 0.2 mM dNTPs, 1 U Vent polymerase (NEB) and 300–400 ng DNA template. The following PCR program was performed: denaturation at 95°C for 5 min, annealing at 55.5°C for 30 min and elongation at 75°C for 10 min. For the annealing of the linker oligonucleotides 21 µmol each of oligo-long and oligo-short in 400 µl of Tris (0.25 M, pH 7.7) were heated at 95°C for 5 min and slowly cooled to 30°C (0.01°C/s). For ligation of the linker the sample was put on ice and 4 µl of T4 ligase buffer [10×, Promega Corporation (PC), Madison, WI, USA], 4 µl of linker and 1.5 U of T4 DNA ligase (Promega) were added. After incubation at 17°C overnight the DNA fragments were precipitated with sodium acetate, ethanol and 10 µg of tRNA, and dissolved in 10 µl of Tris (10 mM pH 7.5).

Amplification of the DNA fragments was performed in a 25 µl reaction mixture containing 10 µl sample DNA, 1× buffer (NEB), 0.2 mM dNTPs, 0.2 µM oligo 2, 0.2 µM oligo-long, and 1 U Vent polymerase (NEB). The PCR program was the following: initial denaturation at 95°C for 2.5 min followed by 17 cycles of 1 min at 95°C, 2 min at 60.5°C and 3 min at 75°C.

For the labelling reaction 1 U of Vent polymerase (NEB) and oligo 3 (5′-[6-FAM]-labelled, 0.2 µM final concentration) were added and the following PCR program was performed: initial denaturation at 95°C for 2.5 min, followed by five cycles of 1 min at 95°C, 2 min at 63.5°C and 3 min at 75°C.

All LM-PCR reactions were performed in triplicates.

Separation of 6-FAM-labelled DNA fragments

Separation of the fluorescently labelled DNA-fragments via capillary gel electrophoresis (CGE) was performed by Microsynth AG (Balgach, Switzerland) on an ABI 3730 XL Genetic Analyser (Life Technologies Corporation, Carlsbad, CA, USA) using GeneScan™ 600-LIZ as internal size standard (Life Technologies). Data from DNA fragment analysis, i.e. peak area values and DNA fragment length, was determined using Peak Scanner™ Software v1.0 (Life Technologies).

Analysis of peak data

To improve sample throughput the analysis of CGE data were automated using ivFAST (*in vivo* footprinting analysis software tool). This software tool was developed and used for the first time during this work. It is a command line-based program, written in Java 6. For the heatmap creation the JHeatChart library (http://www.javaheatmap.com/) was used. This is a Java library for generating heatmap charts for output as image files, which is open source under an LGPL license (http://www.gnu.org/licenses/lgpl-3.0.en.html). ivFAST reads in plain text files containing the CGE results from a specified folder, as well as a DNA sequence file in FASTA format. Given a start point in the DNA sequence and a direction, the program maps the measured peaks to the given sequence and removes background peaks not matching an A or G in the sequence (according to the default setting). The peak area of valid peaks is normalized against total peak area and the share of standard peaks in total peak area to account for variance in the CGE analysis. In addition, normalization against the ratio of unincorporated primers to total peak area is used to account for differences in PCR efficiency. From sample replicates (at least duplicates) the mean peak area and the sample variance (based on a Student's distribution) is calculated for each peak. To determine whether peaks differ significantly from sample to sample their 95% confidence intervals (two-sided) for the mean of the sample replicates are checked to be non-overlapping (pairwise comparison of samples). If this criterion is fulfilled, the quotient of the mean peak areas of sample to reference sample is calculated. From the result of this calculation a text file as well as a heatmap is created, where protected bases with quotients <1 are printed in three shades of red and hypersensitive bases with quotients >1 are printed in three shades of blue. The ivFAST manual, which explains how the software works and how to use it, is included in the software package. From there, the step-by-step

conversion of the data, the according algorithms and the normalization of data can be inferred in all details.

A minimum of two replicates needs to be available to run the software. The authors recommend using (at least) three replicates, which was done throughout this study.

RNA-extraction and reverse transcription

Harvested mycelia were homogenized in 1 ml of peqGOLD TriFast DNA/RNA/protein purification system reagent (PEQLAB Biotechnologie, Erlangen, Germany) using a FastPrep FP120 BIO101 ThermoSavant cell disrupter (Qbiogene, Carlsbad, USA). RNA was isolated according to the manufacturer's instructions, and the concentration was measured using the NanoDrop 1000 (Thermo Scientific, Waltham, USA).

After treatment with DNase I (Fermentas, part of Thermo Fisher Scientific, St. Leon-Rot, Germany), synthesis of cDNA from 0.45 µg mRNA was carried out using the RevertAid[TM] H Minus First Strand cDNA Synthesis Kit (Fermentas); all reactions were performed according to the manufacturer's instructions.

Quantitative PCR analysis

All quantitative PCRs (qPCRs) were performed in a Rotor-Gene Q cycler (QIAGEN). All reactions were performed in triplicate. The amplification mixture (final volume 15 µl) contained 7.5 µl 2× ABsolute[TM] QPCR SYBR® Green Mix (ABgene, part of Thermo Fisher Scientific, Cambridge, UK), 100 nM forward and reverse primer and 2.0 µl cDNA (diluted 1:100). Primer sequences are provided in Table 1. Each run included a template-free control and an amplification-inhibited control (0.015% SDS added to the reaction mixture). The cycling conditions were comprised of a 15 min initial polymerase activation at 95°C, followed by 40 cycles of 15 s at 95°C, 15 s at 60°C (*xyn2, xyr1* and *act*) and 15 s at 72°C; for *sar1*, following the initial activation/denaturation, we ran 40 cycles of 15 s at 95°C and 120 s at 64°C. All PCR efficiencies were >90%. Data analysis, using *sar1* and *act* as reference genes, and calculations using REST 2009 were performed as published previously (40).

RESULTS AND DISCUSSION

Development of an improved, software-based *in vivo* footprinting technique

Motivation for method design
Improving the original *in vivo* footprinting protocol was necessary for a number of reasons. Besides the fact that switching from radioactive to fluorescent labelling is preferable for safety reasons, detection of labelled DNA fragments by CGE instead of densitometric analysis of a sequence gel is significantly faster, more accurate and more sensitive, especially since the use of a commercial sequencing service ensures stability and reproducibility of the fragment length analysis. A further goal of the method improvement was to permit the analysis of a large sample set simultaneously, as well as to enable comparisons of samples based on varying reference samples.

Finally, an increase in sensitivity compared with the original protocol was anticipated.

Method description and optimization
The main steps of the procedure are depicted in Figure 1. First, fungal mycelia were incubated under different cultivation conditions of interest (inducing, repressing, de-repressing). The *in vivo* methylation of fungal mycelia was performed as described before using DMS (16). DNaseI cannot enter the fungal cell and therefore, was not used for *in vivo* footprinting in this study. DNA extraction of genomic DNA was followed by DNA cleavage using HCl, which led to DNA breaks at methylated guanine and adenine residues. Next, LM-PCR was applied because it is a sensitive and specific technique for visualization of *in vivo* footprints. To determine the optimal number of cycles for the amplification and labelling reaction in the LM-PCR, reactions with 17 and 20 cycles for the amplification step, and 5, 10, 15 and 20 cycles for the labelling reaction were conducted. Samples obtained by *in vivo* methylation and subsequent extraction and cleavage of genomic DNA from fungal mycelia (*in vivo* methylated samples) as well as *in vitro* methylated, fungal genomic DNA (naked DNA) as a reference were used as templates. For the amplification step 20 cycles turned out to be too many, because even though differences in peak area values between naked DNA and *in vivo* methylated samples could be detected, *in vivo* methylated samples from different cultivation conditions did not show any significant differences (data not shown). This suggested that the reaction had already reached the end of the exponential phase and the concentrations of DNA fragments had started to level. When stopping the reaction after 17 cycles clear differences between samples from different cultivation conditions can be detected (data not shown), consequently it was chosen. As for the labelling reaction, samples with five and 10 cycles showed an increase in peak area values, while the peak area values did not increase for 15 and 20 cycles (data not shown), indicating that fewer cycles are sufficient to produce clear fluorescence signals. A comparison of reactions with five and 10 cycles again showed that an increase in cycles resulted in a decrease in distinction of different cultivation conditions (data not shown). Consequently, five cycles were chosen as optimal for the labelling reaction.

Development of ivFAST
Performing footprinting reactions of large sample sets simultaneously requires a software-based data analysis. Therefore, in this work we developed a software tool to facilitate data analysis. First, the peak area values and DNA fragment lengths are extracted from the *.fsa-files received from the custom service after CGE (e.g. Supplementary Figure S1) to plain text files. The essential steps of the data analysis are incorporated into a command line-based program: i.e. plotting against the DNA sequence, normalization of peak area values and filtering statistically significantly different bases (protected or hypersensitive) according to a chosen reference sample (compare flowchart in Figure 1). This software tool is easy to use and permits analysis of a dataset and

Table 1. Oligonucleotides used in this study

Name	Sequence 5′– 3′	Usage
RG53	GAATTCAGATC	*iv*-FP, oligo-short
RG54	GCGGTGACCCGGGAGATCTGAATTC	*iv*-FP, oligo-long
RG67	AAGTCATTGCACTCCAAGGC	*iv*-FP, *xyn1* oligo 1 fw
RG68	CCTCTTCACATCATGATTTGAGC	*iv*-FP, *xyn1* oligo 1 rev
RG69	ATTCTGCAGCAAATGGCCTCAAGCAAC	*iv*-FP, *xyn1* oligo 2 fw
RG70	CAAGTGAGGTTGAAAGCGGCTCGTA	*iv*-FP, *xyn1* oligo 2 rev
RG71	[6-FAM]CTGCAGCAAATGGCCTCAAGCAACTACG	*iv*-FP, *xyn1* oligo 3 fw
RG72	[6-FAM]GAGGTTGAAAGCGGCTCGTACAGTATCC	*iv*-FP, *xyn1* oligo 3 rev
RG72-2	GAGGTTGAAAGCGGCTCGTACAGTATCC	*iv*-FP, *xyn1* oligo 3 rev
RG97	AAGCGCTAATGTGGACAGGATT	*iv*-FP, *cbh2* oligo 1 fw
RG98	CAATACACAGAGGGTGATCTTAC	*iv*-FP, *cbh2* oligo 1 rev
RG99	CATTAGCCTCAAGTAGAGCCTATTTCCTC	*iv*-FP, *cbh2* oligo 2 fw
RG100	GCCTCTTCAGGTGAGCTGCTG	*iv*-FP, *cbh2* oligo 2 rev
RG101	[6-FAM]GCCTCAAGTAGAGCCTATTTCCTCGCC	*iv*-FP, *cbh2* oligo 3 fw
RG102	[6-FAM]CTTCAGGTGAGCTGCTGTGAGACCATG	*iv*-FP, *cbh2* oligo 3 rev
RG127	GTTCCGATATATGAGATTGCCAAG	*iv*-FP, *xyn2* oligo 1 fw
RG128	GTTGATGTCTTCTTGCTTCAGC	*iv*-FP, *xyn2* oligo 1 rev
RG129	AGCCGTTATTCAGACAATGTATGTGCCG	*iv*-FP, *xyn2* oligo 2 fw
RG130	GGAGTTGTTGTGTCTTTTGGGCTTGG	*iv*-FP, *xyn2* oligo 2 rev
RG131	[6-FAM]CCGTTATTCAGACAATGTATGTGCCGGGC	*iv*-FP, *xyn2* oligo 3 fw
RG132	[6-FAM]GTTGTTGTGTCTTTTGGGCTTGGAGGGG	*iv*-FP, *xyn2* oligo 3 rev
act fw	TGAGAGCGGTGGTATCCACG	*act* qPCR
act rev	GGTACCACCAGACATGACAATGTTG	*act* qPCR
sar1 fw	TGGATCGTCAACTGGTTCTACGA	*sar1* qPCR
sar1 rev	GCATGTGTAGCAACGTGGTCTTT	*sar1* qPCR
cbh2 fw	CTATGCCGGACAGTTTGTGGTG	*cbh2* qPCR
cbh2 rev	GTCAGGCTCAATAACCAGGAGG	*cbh2* qPCR
xyn1 fw	CAGCTATTCGCCTTCCAACAC	*xyn1* qPCR
xyn1 rev	CAAAGTTGATGGGAGCAGAAG	*xyn1* qPCR
xyn2 fw	GGTCCAACTCGGGCAACTTT	*xyn2* qPCR
xyn2 rev	CCGAGAAGTTGATGACCTTGTTC	*xyn2* qPCR

visualization of the results in a very short time, i.e. data analysis starting from obtained CGE results can be done in 10 min per sample (given that three replicates are used). ivFAST is freely available at http://www.vt.tuwien.ac.at/ biotechnology_and_microbiology/gene_technology/mach_ aigner_lab/EN/. From there, both the software and a detailed manual can be downloaded. The manual explains how to use the software and how it works including the step-by-step conversion of the data, the according algorithms and the normalization of data. On the one hand, ivFAST actually determines the precise intensity of protection or hypersensitivity and yields as output the exact number given in a text file. On the other hand, ivFAST also displays results in a gradual mode of visualization (three shades for each, protection and hypersensitivity, of which the range is manually adjustable) and yields as output a heatmap as *.png-file for graphic display of results.

Validation of the newly developed *in vivo* footprinting technique

Comparison of the new technique to traditional in vivo *footprinting*
As a first attempt, the newly developed, software-based technique was compared with the traditional, gel-based *in vivo* footprinting approach. Because the URR of *xyn1* is well-studied and the *cis* elements involved and the contacting *trans* factors are widely known, it was chosen for a comparative investigation of both techniques side-by-side.

As Xyr1 is the main transactivator of the *xyn1* gene expression, an URR part covering two Xyr1-binding sites [previously proven functional by deletion analysis (34,41)] was analysed. Using traditional *in vivo* footprinting, the protection of some bases could only be detected when compared with naked DNA, whereas no condition-specific differences (regardless if repressing or inducing) were found (Figure 2). In contrast, the new technique generally yielded more protection/hypersensitivity signals compared with the gel if samples from *in vivo* footprinting were compared with naked DNA (Figure 2, G/ND, XO/ND). Most strikingly, the new technique also displays signals if *in vivo* footprinting results from inducing conditions (D-xylose) were compared with those from repressing conditions (glucose) (Figure 2, XO/G). Summarizing, the traditional, gel-based method and the comparison to naked DNA applying the new method revealed a similar *in vivo* footprinting pattern under repressing and inducing conditions. However, only the new method detects clear induction-specific differences, which are in good accordance with *xyn1* transcript data (Supplementary Figure S2a).

Reproducibility of the new technique
In order to test the reproducibility of the method, *in vivo* footprinting of samples from two different conditions (repressing and inducing) and from two biological replicates of each was performed. The original trace data of these samples and—as a reference—of naked DNA (performed also in duplicates) is pictured in Figure 3a. Comparing the electropherograms of the replicates, it becomes clear that
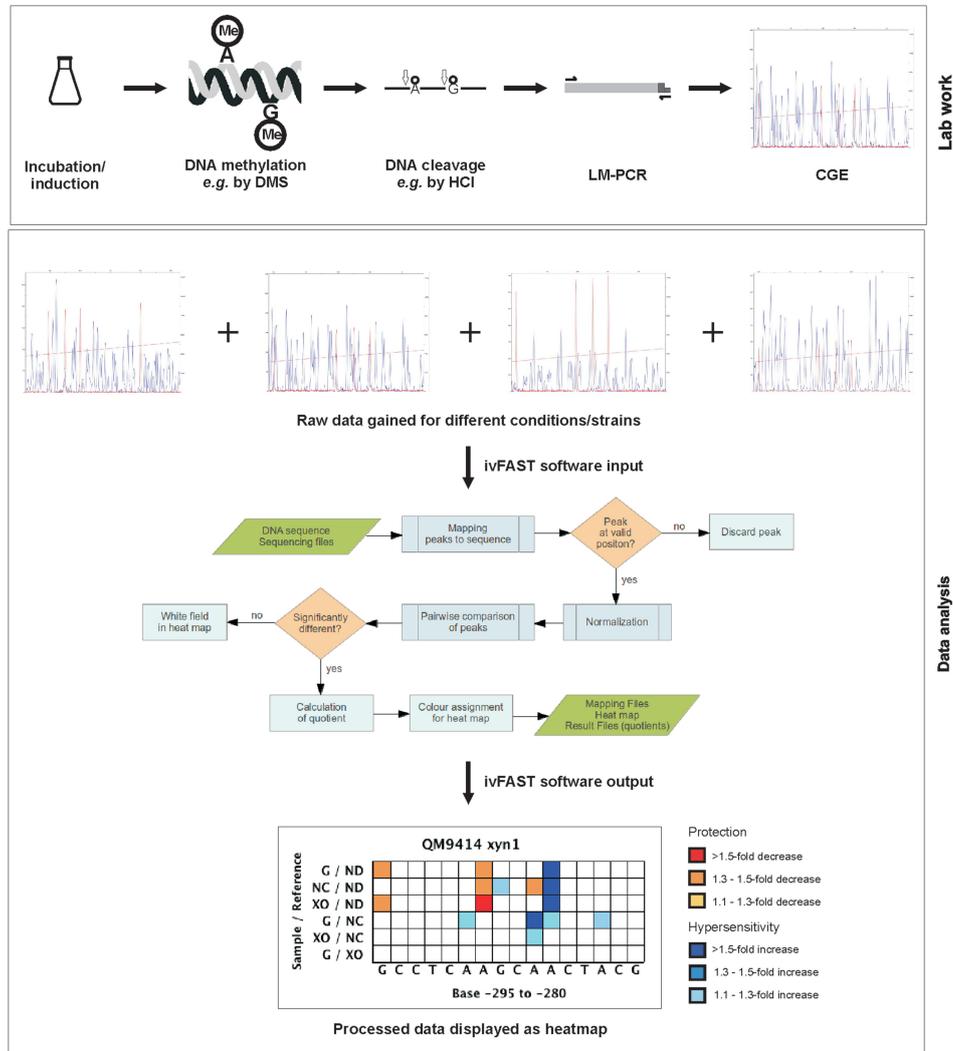
Figure 1. Schematic presentation of the workflow and generation of final data. The main steps of the software-based, high-throughput *in vivo* footprinting method comprise growing/incubating the microorganism under conditions to be investigated (e.g. inducing conditions), *in vivo* DNA methylation using e.g. DMS, DNA extraction, DNA cleavage by e.g. HCl followed by LM-PCR and CGE. A subset of CGE analyses results to be compared (raw data) are submitted to electronic data analysis using the ivFAST software for generation of the results displayed as final heatmap (processed data output). The steps of processing the data by the ivFAST software can be inferred from the flowchart (for more details see the ivFAST manual). Heatmap: *x*-axis gives the analysed DNA sequence; *y*-axis shows which samples are referred to each other (e.g. G/ND means 'glucose repressing conditions referred to naked DNA'); only signals that are statistically different are considered; protected bases are highlighted in red shades and hypersensitive bases are highlighted in blue shades; 1.1- to 1.3-fold difference between compared conditions is shown in light shaded colour, 1.3- to 1.5-fold difference between compared conditions is shown in middle shaded colour and >1.5-fold difference between compared conditions is shown in dark shaded colour.

their peak pattern is the same. The peak pattern of the naked DNA strongly differs from both types of *in vivo* footprinting samples (repressing/inducing condition). If the *in vivo* footprinting sample from repressing conditions (glucose) is compared with the one from inducing conditions (D-xylose), slight differences in certain peak ratios can be observed. These findings support the above-mentioned conclusion that strong differences can be detected comparing *in vivo* footprinting samples with naked DNA, but also detection of condition-dependent
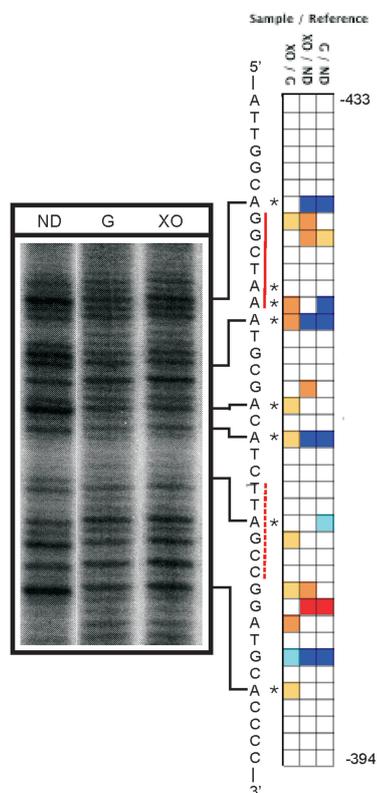
Figure 2. Comparison of the traditional *in vivo* footprinting to the newly developed method. *In vivo* footprinting analysis of the coding strand of a *xyn1* URR (−433- to −394-bp upstream from ATG) covering two Xyr1-binding sites, which are indicated by red lines (solid, site is located on the coding strand; dashed, site is located on the non-coding strand), was performed. *Trichoderma reesei* cultivated on glucose (G) or D-xylose (XO) followed by DMS-induced *in vivo* methylation and naked DNA as a reference (ND) were analysed. Left side shows a gel obtained by the traditional method. Asterisks indicate protected bases. Right side shows a heatmap yielded by the newly developed method.

differences (comparing *in vivo* footprinting results from a certain condition to another) are now possible. Using these raw data for analysis applying ivFAST, a heatmap for each replicate is obtained (Figure 3b). They yield a similar signal pattern, regardless if the *in vivo* footprinting samples are referred to naked DNA (Figure 3b; compare G1/ND1 and XO1/ND1 with G2/ND2 and XO2/ND2) or to each other (Figure 3b; compare G1/XO1 with G2/XO2). Most importantly, the heatmap that results from referring the same type of replicate (glucose, D-xylose, naked DNA) to each other (Figure 3c) is given. As expected hardly any signal is yielded in this case supporting a sufficient reproducibility of the method.

Verification of signals yielded by the new technique
In order to test the reliability of the signal yielded by the new technique, we used the wild-type and an isogenic *xyr1*

deletion strain for *in vivo* footprinting analyses. Xyr1 is the main transactivator of *xyn1* gene expression (26). Consequently, a region of the *xyn1* URR covering a functional binding site for Xyr1 was chosen for investigation. The consensus sequence for Xyr1 DNA binding [5′-GGC(A/T)₃-3′] was previously investigated by EMSA and *in vitro* footprinting (28). As a control, the investigated region also includes a functional binding site for another transcription factor involved in *xyn1* gene regulation, namely Cre1 (34), which is still intact (Figure 4a). The consensus sequence for Cre1 DNA binding (5′-SYGGRG-3′) was previously investigated by EMSA and *in vitro* footprinting (35). As before, the strains were cultivated on glucose (repressing condition) or D-xylose (inducing condition). As mentioned above, again, reference to naked DNA generally highlights a high number of bases as protected or hypersensitive, but does not provide a condition-specific pattern (Figure 4b). A direct comparison of *in vivo* footprinting results (repressing conditions referred to inducing conditions) of the wild-type (Figure 4b, G/XO) with those of the *xyr1* deletion strain (Figure 4b, Δxyr1-G/Δxyr1-XO) revealed that while the hypersensitivity at the Xyr1-binding site disappears, the protection at the Cre1-binding site is increased in the deletion strain. This observation is not unexpected as the activator Xyr1 is not contacting this regulatory region in the deletion strain, and Cre1, which was shown to be involved in chromatin packaging (42), can now deploy its repressor function unrestrainedly.

Applying the new *in vivo* footprinting technique to previously identified URRs

In vivo *footprinting of the URR of the* cbh2 *gene*
In 1998 the CAE in *T. reesei* was reported to be crucial for regulation of *cbh2* gene expression, encoding a major cellulase (31). Meanwhile, Xyr1 was identified as the major transactivator of most hydrolase-encoding genes including *cbh2* (26,28,43). Allowing one mismatch in the Xyr1-binding motif reveals that the CAE consists of a putative Xyr1-binding site and an overlapping CCAAT-box, which is a common *cis* element in URRs of eukaryotes. Therefore, we analysed an URR including the CAE as well as two additional, *in silico* identified Xyr1-binding sites (allowing one mismatch) and a putative Cre1-binding site (Figure 5a). We performed *in vivo* footprinting of mycelia from repressing conditions (glucose), inducing conditions (sophorose), and used the sample gained from incubation without carbon source as the reference condition.

The CCAAT-box within the CAE and an adjacent A-stretch reacts strongly glucose-dependent (Figure 5c, G/NC, SO/G), while the Xyr1-binding site within the CAE does not yield condition-specific differences (Figure 5b). These two observations might suggest that a carbon source-specific response is mediated via the CCAA T-box, while Xyr1 binds permanently. The latter assumption is in good accordance with the finding that no *de novo* synthesis of Xyr1 is necessary for an initial induction of target genes suggesting that Xyr1 is always available in the cell at a low level (44). However, the new method
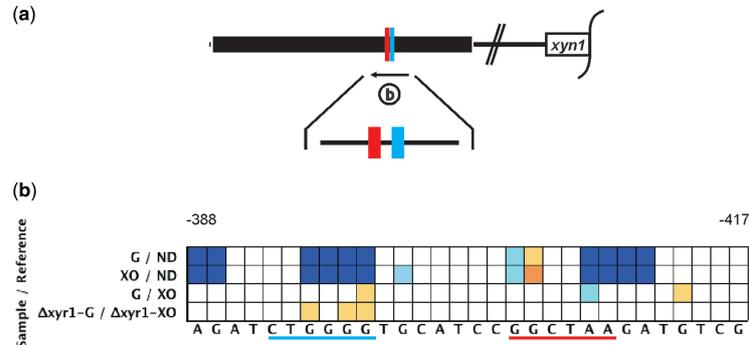
Figure 3. Comparison of two biological replicates analysed by the newly developed *in vivo* footprinting method. *In vivo* footprinting analysis of the non-coding strand of a *xyn1* URR (−388- to −417-bp upstream from ATG) covering a Cre1-binding site (underlined in blue) and a Xyr1-binding site (underlined in red) was performed. *Trichoderma reesei* cultivated on glucose (G) or D-xylose (XO) followed by DMS-induced *in vivo* methylation and naked DNA as a reference (ND) were analysed. (**a**) Original data of two biological replicates (Replicates 1 and 2) obtained after CGE displayed as electropherograms next to each other. Peaks in the electropherograms of the glucose replicates are marked by the corresponding DNA bases for easier orientation. (**b**) Analysed data of two biological replicates (indicated by the numbers 1 and 2) using ivFAST displayed as heatmaps under each other. (**c**) Analysed data of two biological replicates (indicated by the numbers 1 and 2) using ivFAST, if one replicate refers to the other, displayed as a heatmap.

demonstrates that the two additional, *in silico* identified Xyr1-binding sites are active, but seem to be contacted in a condition-dependent way (Figure 5b and c). This coincides with findings that *cbh2* induction by sophorose goes along with increased *xyr1* transcript formation (45).

Finally, the condition-dependent comparison reveals a not yet verified single Cre1-binding site as active regulatory element giving glucose-dependent signals (Figure 5b). Transcript analysis of *cbh2* is complementary to *in vivo* footprinting data, e.g. the induction-dependent

Figure 4. Comparison of *in vivo* footprinting results of a deletion and a parental strain. (**a**) *In vivo* footprinting analysis of a *xyn1* URR (−388- to −417-bp upstream from ATG) covering a Cre1-binding site (underlined in blue) and a Xyr1-binding site (underlined in red) was performed. (**b**) The non-coding strand was analysed after incubation of the *T. reesei* parental and a *xyr1* deletion strain (Δxyr1) on glucose (G) or D-xylose (XO) followed by DMS-induced *in vivo* methylation. Naked DNA was used as a reference (ND).
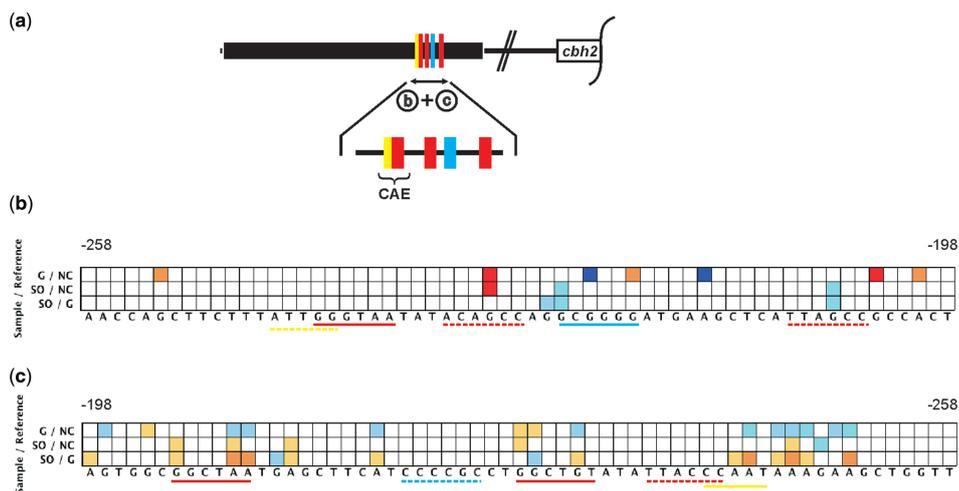

Figure 5. *In vivo* footprinting analysis of the *T. reesei cbh2* URR. (**a**) A *cbh2* URR covering the cellulase activating element (CAE) comprising a CCAAT-box (yellow) and a Xyr1-binding site (red), two additional Xyr1-binding sites and a Cre1-binding site (blue) was investigated (−258- to −198-bp upstream from ATG). The coding strand (**b**) and the non-coding strand (**c**) were analysed after incubation of *T. reesei* on sophorose (SO), glucose (G) or without carbon source (NC) followed by DMS-induced *in vivo* methylation.

(sophorose) protection of the activator's (Xyr1)-binding sites or the repression-dependent (glucose) protection of the repressor's (Cre1)-binding site (Supplementary Figure S2b).

In vivo *footprinting of the URR of the* xyn2 *gene*
The URR of the *xyn2* gene, whose product is the main endo-xylanase of *T. reesei*, has a similar architecture as the one of *cbh2*. In 2003 the XAE comprising a CCAAT-box adjacent to a Xyr1-binding site was reported to be essential for *xyn2* expression (30). The XAE is located close to a second Xyr1-binding site (bearing two mismatches) (Figure 6a, IV, and V). Both Xyr1-binding sites need to

be intact for binding Xyr1 *in vitro* and *in vivo* (46). Upstream of the XAE an AGAA-box has before been described as a *cis* element mediating repression (46,47) (Figure 6a, III). We performed *in vivo* footprinting of mycelia from repressing conditions (glucose), inducing conditions (D-xylose), and the reference condition (without carbon source). On the one hand we confirmed the above-mentioned, previously identified *cis* elements, and additionally, revealed condition-dependent contacting by their *trans* factors (Figure 6b and c).

Interestingly, the new *in vivo* footprinting method identified a second AGAA-box, which is located 4-bp upstream of the first one and arranged as inverted repeat
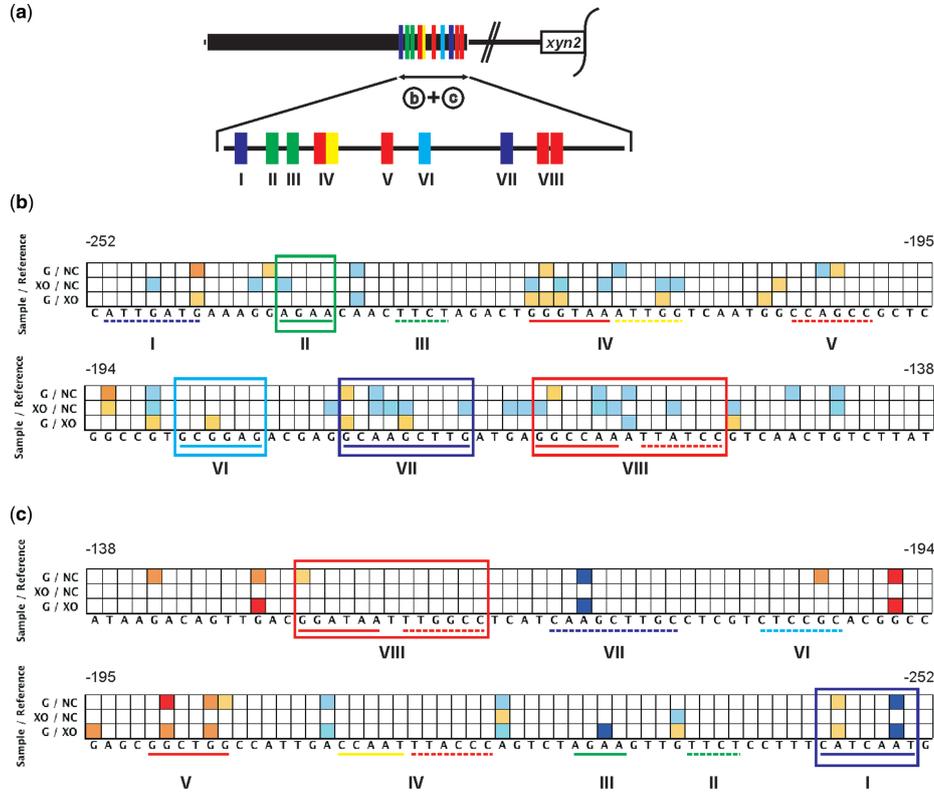
Figure 6. *In vivo* footprinting analysis of the *T. reesei xyn2* URR. (**a**) A *xyn2* URR showing a high number of cis elements [AGAA-box (green), CCAAT-box (yellow), Xyr1-binding site (red) and Cre1-binding site (blue)] was investigated (−252- to −138-bp upstream from ATG). The coding strand (**b**) and the non-coding strand (**c**) were analysed after incubation of *T. reesei* on glucose (G), D-xylose (XO) or without carbon source (NC) followed by DMS-induced *in vivo* methylation. Newly identified motifs are given in frame, motifs with DNA sequence not reported before are given in purple.

(Figure 6b, II). The occurrence of the AGAA-motif as a palindrome is in accordance to an earlier report that this *cis* element is contacted by a basic helix–loop–helix transcription factor, which canonically binds as dimer (47). Also a yet not recognized, single Cre1-binding site could be identified (Figure 6b, VI) exhibiting a glucose-dependent protection (Figure 6b, G/XO; 6c, G/NC). Additionally, a palindromic Xyr1-binding site spaced by 1 bp was revealed, of which both sites yield condition-specific differences (Figure 6b, c, VIII).

However, *in vivo* footprinting of this region highlighted two more regions, which are contacted in a condition-dependent way. The first one, 5′-ATTGATG-3′ (−251 to −245 bp), yields signals on both investigated strands (Figure 6b, c, I) and bears an unusual TCAAT-box (Figure 6c, I). The second one, 5′-GCAAGCTTG-3′ (−177 to −169 bp), also yields signals on both investigated strands and contains an octameric palindrom (CAAGCT TG) overlapping with an Ace1-binding site [5′-AGGCA-3′, (48)] (Figure 6b, c, VII). Ace1 is a narrow domain

transcription factor functioning as repressor of cellulase and xylanase expression (48). A sound interpretation of transcript analysis (Supplementary Figure S2c) compared with *in vivo* footprinting data in this case is difficult to provide because too many new motifs, of which the regulatory function is unknown, were identified. However, induction- or repression-dependent protection/hypersensitivity was observed indicating regulatory functionality.

Comparison of regulatory and non-regulatory regions
In order to validate the false positive signal rate of the method we performed footprinting analyses of longer upstream sequences of the above-described genes, i.e. *xyn1*, *xyn2* and *cbh2*. The analysed fragments cover regions previously reported to be regulatory and non-regulatory each (29–31,34,41,49). The heatmaps obtained by referring *in vivo* footprinting results from repressing and inducing conditions to each other are provided in Supplementary Figures S3–S5, respectively. To get additional indication on protein–DNA interaction, the reference of *in vivo* footprinting samples to naked DNA is also

included. It is noteworthy that previously identified motifs (Figures 4–6) also show protection/hypersensitivity when *in vivo* footprinting samples were compared with naked DNA (Supplementary Figures S3–S5). Most of the additionally detected signals can be assigned to known motifs (details are described in the respective legends to Supplementary Figures S3–S5), whereas long sequence stretches without any known motif hardly yielded signals. In case of *xyn2,* the two newly identified, before unknown motifs (compare Figure 6, I, VII) are also displayed by the comparison to naked DNA (Supplementary Figure S5).

Potential of the *in vivo* footprinting method

As already outlined, the software-based *in vivo* footprinting method presented in this study provides the possibility to identify *cis* elements in a fast and sensitive way. We are convinced that this method is a very useful tool for a broad range of investigations concerning regulatory elements, not only in filamentous fungi, but in all organisms. It is important to note that in this study footprinting was performed with DMS followed by HCl DNA cleavage because this a good practice in case of filamentous fungi. However, the proposed approach, in particular DNA fragment analysis and data analysis by ivFAST, is not limited to a certain footprinting or DNA cleavage reagent. The ivFAST manual explains how to adjust parameters in order to analyse data obtained from other footprinting techniques. Compared with the traditional *in vivo* footprinting approach our method employs an internal standard. This allows a comparison of the URR of a gene from any number of conditions or strains, cell lines or tissues without necessity for generating all data at the same time. Because of this and the fact that the method is highly robust (biological and technical replicates do not show relevant differences) it is possible to establish an open-end database for each URR of interest. Furthermore, generated datasets provide a quantitative insight into *trans* factor/*cis* element interaction depicted by a gradual display of results (heatmap). The new footprinting method allows the identification of new variants of already known *cis* elements and of completely new motifs. This is achieved by shuffling the respective, pairwise comparisons of conditions or cells of interest. Including data from *trans* factor deletion strains/cells in such a database makes the assignment of the according *cis* element possible. It is noteworthy that some improvements of the described technique are also useful for *in vitro* footprinting of purified proteins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [50,51].

ACKNOWLEDGEMENTS

The authors thank John Tomashek for a critical discussion on the present study.

REFERENCES

1. Narlikar,L. and Ovcharenko,I. (2009) Identifying regulatory elements in eukaryotic genomes. *Brief. Funct. Genomic Proteomic*, **8**, 215–230.
2. Dynan,W.S. and Tjian,R. (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature*, **316**, 774–778.
3. Maniatis,T., Goodbourn,S. and Fischer,J.A. (1987) Regulation of inducible and tissue-specific gene expression. *Science*, **236**, 1237–1245.
4. Jackson,P.D. and Felsenfeld,G. (1985) A method for mapping intranuclear protein-DNA interactions and its application to a nuclease hypersensitive site. *Proc. Natl Acad. Sci. USA*, **82**, 2296–2300.
5. Jackson,P.D. and Felsenfeld,G. (1987) *In vivo* footprinting of specific protein-DNA interactions. *Methods Enzymol.*, **152**, 735–755.
6. Kemper,B., Jackson,P.D. and Felsenfeld,G. (1987) Protein-binding sites within the 5′ DNase I-hypersensitive region of the chicken alpha D-globin gene. *Mol. Cell. Biol.*, **7**, 2059–2069.
7. Zinn,K. and Maniatis,T. (1986) Detection of factors that interact with the human beta-interferon regulatory region *in vivo* by DNAase I footprinting. *Cell*, **45**, 611–618.
8. Ephrussi,A., Church,G.M., Tonegawa,S. and Gilbert,W. (1985) B lineage–specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science*, **227**, 134–140.
9. Giniger,E., Varnum,S.M. and Ptashne,M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, **40**, 767–774.
10. Shaw,P.E. and Stewart,A.F. (2001) Identification of protein-DNA contacts with dimethyl sulfate. Methylation protection and methylation interference. *Methods Mol. Biol.*, **148**, 221–227.
11. Mueller,P.R. and Wold,B. (1989) *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science*, **246**, 780–786.
12. Drouin,R., Therrien,J.P., Angers,M. and Ouellet,S. (2001) *In vivo* DNA analysis. *Methods Mol. Biol.*, **148**, 175–219.
13. Garrity,P.A. and Wold,B.J. (1992) Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and *in vivo* footprinting. *Proc. Natl Acad. Sci. USA*, **89**, 1021–1025.
14. Mueller,P.R., Salser,S.J. and Wold,B. (1988) Constitutive and metal-inducible protein:DNA interactions at the mouse metallothionein I promoter examined by *in vivo* and *in vitro* footprinting. *Genes Dev.*, **2**, 412–427.
15. Gilmour,D.S. and Fan,R. (2009) Detecting transcriptionally engaged RNA polymerase in eukaryotic cells with permanganate genomic footprinting. *Methods*, **48**, 368–374.
16. Wolschek,M.F., Narendja,F., Karlseder,J., Kubicek,C.P., Scazzocchio,C. and Strauss,J. (1998) *In situ* detection of protein-DNA interactions in filamentous fungi by *in vivo* footprinting. *Nucleic Acids Res.*, **26**, 3862–3864.
17. Trotha,R., Reichl,U., Thies,F.L., Sperling,D., König,W. and König,B. (2002) Adaption of a fragment analysis technique to an automated high-throughput multicapillary electrophoresis device for the precise qualitative and quantitative characterization of microbial communities. *Electrophoresis*, **23**, 1070–1079.
18. Hartl,L. and Seefelder,S. (1998) Diversity of selected hop cultivars detected by fluorescent AFLPs. *Theor. Appl. Genet.*, **96**, 112–116.
19. Zianni,M., Tessanne,K., Merighi,M., Laguna,R. and Tabita,F.R. (2006) Identification of the DNA bases of a DNase I footprint by

the use of dye primer sequencing on an automated capillary DNA analysis instrument. *J. Biomol. Tech.*, **17**, 103–113.

20. Ingram,R., Gao,C., Lebon,J., Liu,Q., Mayoral,R.J., Sommer,S.S., Hoogenkamp,M., Riggs,A.D. and Bonifer,C. (2008) PAP-LMPCR for improved, allele-specific footprinting and automated chromatin fine structure analysis. *Nucleic Acids Res.*, **36**, e19.

21. Ingram,R., Tagoh,H., Riggs,A.D. and Bonifer,C. (2005) Rapid, solid-phase based automated analysis of chromatin structure and transcription factor occupancy in living eukaryotic cells. *Nucleic Acids Res.*, **33**, e1.

22. Dai,S.M., Chen,H.H., Chang,C., Riggs,A.D. and Flanagan,S.D. (2000) Ligation-mediated PCR for quantitative *in vivo* footprinting. *Nat. Biotechnol.*, **18**, 1108–1111.

23. Brewer,A.C., Marsh,P.J. and Patient,R.K. (1990) A simplified method for *in vivo* footprinting using DMS. *Nucleic Acids Res.*, **18**, 5574.

24. Dimitrova,D., Giacca,M. and Falaschi,A. (1994) A modified protocol for *in vivo* footprinting by ligation-mediated polymerase chain reaction. *Nucleic Acids Res.*, **22**, 532–533.

25. Granger,S.W. and Fan,H. (1998) *In vivo* footprinting of the enhancer sequences in the upstream long terminal repeat of Moloney murine leukemia virus: differential binding of nuclear factors in different cell types. *J. Virol.*, **72**, 8961–8970.

26. Stricker,A.R., Grosstessner-Hain,K., Würleitner,E. and Mach,R.L. (2006) Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryot. Cell*, **5**, 2128–2137.

27. Stricker,A.R., Mach,R.L. and de Graaff,L.H. (2008) Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). *Appl. Microbiol. Biotechnol.*, **78**, 211–220.

28. Furukawa,T., Shida,Y., Kitagami,N., Mori,K., Kato,M., Kobayashi,T., Okada,H., Ogasawara,W. and Morikawa,Y. (2009) Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. *Fungal Genet. Biol.*, **46**, 564–574.

29. Rauscher,R., Würleitner,E., Wacenovsky,C., Aro,N., Stricker,A.R., Zeilinger,S., Kubicek,C.P., Penttilä,M. and Mach,R.L. (2006) Transcriptional regulation of *xyn1*, encoding xylanase I, in *Hypocrea jecorina*. *Eukaryot. Cell*, **5**, 447–456.

30. Würleitner,E., Pera,L., Wacenovsky,C., Cziferszky,A., Zeilinger,S., Kubicek,C.P. and Mach,R.L. (2003) Transcriptional regulation of *xyn2* in *Hypocrea jecorina*. *Eukaryot. Cell*, **2**, 150–158.

31. Zeilinger,S., Mach,R.L. and Kubicek,C.P. (1998) Two adjacent protein binding motifs in the *cbh2* (cellobiohydrolase II-encoding) promoter of the fungus *Hypocrea jecorina* (*Trichoderma reesei*) cooperate in the induction by cellulose. *J. Biol. Chem.*, **273**, 34463–34471.

32. Ilmén,M., Thrane,C. and Penttilä,M. (1996) The glucose repressor gene *cre1* of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form. *Mol. Gen. Genet.*, **251**, 451–460.

33. Ilmén,M., Onnela,M.L., Klemsdal,S., Keranen,S. and Penttilä,M. (1996) Functional analysis of the cellobiohydrolase I promoter of the filamentous fungus *Trichoderma reesei*. *Mol. Gen. Genet.*, **253**, 303–314.

34. Mach,R.L., Strauss,J., Zeilinger,S., Schindler,M. and Kubicek,C.P. (1996) Carbon catabolite repression of xylanase I (*xyn1*) gene expression in *Trichoderma reesei*. *Mol. Microbiol.*, **21**, 1273–1281.

35. Strauss,J., Mach,R.L., Zeilinger,S., Hartler,G., Stoffler,G., Wolschek,M. and Kubicek,C.P. (1995) Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*. *FEBS Lett.*, **376**, 103–107.

36. Mäntylä,A.L., Rossi,K.H., Vanhanen,S.A., Penttilä,M.E., Suominen,P.L. and Nevalainen,K.M. (1992) Electrophoretic karyotyping of wild-type and mutant *Trichoderma longibrachiatum* (*reesei*) strains. *Curr. Genet.*, **21**, 471–477.

37. Mandels,M. (1985) Applications of cellulases. *Biochem. Soc. Trans.*, **13**, 414–416.

38. Gruber,F., Visser,J., Kubicek,C.P. and de Graaff,L.H. (1990) The development of a heterologous transformation system for the cellulolytic fungus *Trichoderma reesei* based on a *pyrG*-negative mutant strain. *Curr. Genet.*, **18**, 71–76.

39. Maxam,A.M. and Gilbert,W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.*, **65**, 499–560.

40. Steiger,M.G., Mach,R.L. and Mach-Aigner,A.R. (2010) An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*). *J. Biotechnol.*, **145**, 30–37.

41. Zeilinger,S., Mach,R.L., Schindler,M., Herzog,P. and Kubicek,C.P. (1996) Different inducibility of expression of the two xylanase genes *xyn1* and *xyn2* in *Trichoderma reesei*. *J. Biol. Chem.*, **271**, 25624–25629.

42. Zeilinger,S., Schmoll,M., Pail,M., Mach,R.L. and Kubicek,C.P. (2003) Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. *Mol. Genet. Genomics*, **270**, 46–55.

43. Stricker,A.R., Steiger,M.G. and Mach,R.L. (2007) Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. *FEBS Lett.*, **581**, 3915–3920.

44. Mach-Aigner,A.R., Pucher,M.E., Steiger,M.G., Bauer,G.E., Preis,S.J. and Mach,R.L. (2008) Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. *Appl. Environ. Microbiol.*, **74**, 6554–6562.

45. Derntl,C., Gudynaite-Savitch,L., Calixte,S., White,T., Mach,R.L. and Mach-Aigner,A.R. (2013) Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used *Trichoderma* strains. *Biotechnol. Biofuels*, **6**, 62.

46. Stricker,A.R., Trefflinger,P., Aro,N., Penttilä,M. and Mach,R.L. (2008) Role of Ace2 (Activator of Cellulases 2) within the *xyn2* transcriptosome of *Hypocrea jecorina*. *Fungal Genet. Biol.*, **45**, 436–445.

47. Mach-Aigner,A.R., Grosstessner-Hain,K., Pocas-Fonseca,M.J., Mechtler,K. and Mach,R.L. (2010) From an electrophoretic mobility shift assay to isolated transcription factors: a fast genomic-proteomic approach. *BMC Genomics*, **11**, 644.

48. Aro,N., Ilmén,M., Saloheimo,A. and Penttilä,M. (2003) ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Appl. Environ. Microbiol.*, **69**, 56–65.

49. Stangl,H., Gruber,F. and Kubicek,C.P. (1993) Characterization of the *Trichoderma reesei cbh2* promoter. *Curr. Genet.*, **23**, 115–122.

50. Andrianopoulos,A. and Timberlake,W.E. (1994) The *Aspergillus nidulans abaA* gene encodes a transcriptional activator that acts as a genetic switch to control development. *Mol. Cell Biol.*, **14**, 2503–2515.

51. Metz,B., Seidl-Seiboth,V., Haarmann,T., Kopchinskiy,A., Lorenz,P., Seiboth,B. and Kubicek,C.P. (2011) Expression of biomass-degrading enzymes is a major event during conidium development in *Trichoderma reesei*. *Eukaryot. Cell*, **10**, 1527–1535.

A.4 ivFAST manual

**TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology**

ivFAST

Version 1.0

Contents

1   What is ivFAST?

ivFAST (*in vivo* Footprinting Analysis Software Tool) is a command line-based program to map peaks obtained from capillary gel electrophoresis to a DNA sequence and to make a pairwise comparison of the peak areas from different samples for each sequence position. The mapping and the results of the calculations are provided as text files. For easier visual analysis of the data generated a heatmap is created, comprising three color levels for protected and hypersensitive bases.

2   How does ivFAST work?

ivFAST conducts the following steps:

1. Data import
   ivFAST imports two types of files: a FASTA-file, containing the according DNA sequence, and the sequencing files, containing the measured peaks with retention time and area, given in plain text format. For each sequenced sample three sequencing files have to be provided: one containing all sample peaks sequenced (also the primer artefacts), one containing only the internal size standard peaks, and one containing only the sample peaks to be analyzed.

2. Sequence processing
   The input sequence is always given as the coding strand from 5' to 3'. Therefore it has to be transferred to the analyzed strand and direction, which is 3' to 5' on the non-coding strand for forward primed sequencing, and 3' to 5' on the coding strand for reverse primed sequencing.

3. Mapping
   The program starts the mapping by assigning the first peak to a sequence position specified by the user. Further positions are calculated from the pairwise distances of consecutive peaks, which are rounded to integers. If the calculated position matches a base in the sequence, listed in the configuration file under 'validBases', it is taken for further calculation. Valid bases are bases, at which DNA can be cut dependent on the methylation and cleavage method used. If the position matches a base not listed in 'validBases' it is treated as background noise and removed from further processing. The result of the mapping is outputted as plain text file. Peaks identified as background are indicated with an asterisk. Background peaks occur frequently at the beginning, but seldom at the end of the analyzed region. If it is likely, that a peak is falsely identified as background peak, an additional notification is displayed on the command line, so the user can make manual corrections in the according sequencing file containing the peaks to evaluate (the file containing all peaks does not need to be changed).

4. Calculation
   First the single peak areas of the peaks to be analyzed are normalized. The normalization factor is calculated for each sample in three parts. The first is the share of the sample peaks in all peaks (including standard

peaks), which accounts for slightly varying ratios of sample amount to standard amount. The second is the share of true sample peaks (without primer artifacts) to all sample peaks, which accounts for different reaction efficiencies in the previous PCR. The third is the sum of all areas (standard and sample), which accounts for differences in overall fluorescent signal due to varying CGE analysis. With x_a an area from the file containing all sample peaks, x_s an area from the file containing the standard peaks, and x_p an area from the file containing the sample peaks to be analyzed, the normalized peak area of x_p^i is defined as

$$\hat{x}_p = x_p \ / \left(\frac{\sum_{i=0}^{m} x_a^i}{\sum_{i=0}^{m} x_a^i + \sum_{i=0}^{m} x_s^i} \cdot \frac{\sum_{i=l}^{m} x_p^i}{\sum_{i=0}^{m} x_p^i} \cdot \left(\sum_{i=0}^{m} x_a^i + \sum_{i=0}^{m} x_s^i \right) \right), \quad (1)$$

with m the total number of peaks in this sample and l indicating the first peak with a size greater than the primer length (as defined in the configuration file).

After normalization the n replicates of a sample (belonging to the same condition) are grouped together and for each normalized peak $\hat{x}_p$ the sample mean

$$\overline{x}_p = \frac{1}{n} \sum_{i=1}^{n} \hat{x}_p^i, \quad (2)$$

the sample variance

$$s_p^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\hat{x}_p^i - \overline{x}_p)^2, \quad (3)$$

and the confidence interval for the true mean μ_p

$$\overline{x}_p - \frac{t \cdot s_p}{\sqrt{n}} \leq \mu_p \leq \overline{x}_p + \frac{t \cdot s_p}{\sqrt{n}} \quad (4)$$

are calculated, based on the Student's t-distribution, with t obtained from the tabularized values of the confidence interval $F_{n-1}(t) = 0.95$ (recommended value, but adjustable in the configuration file).

Now the program does a pairwise comparison between the sample means $\overline{x}_p$ of the user-defined pairs of sample (S) and reference (R) conditions where for each peak it is checked, whether the two sample means $\overline{x}_p(S)$ and $\overline{x}_p(R)$ can be said to be different. As criterion for this non-overlapping confidence intervals of the sample means

$$\left(\overline{x}_p(S) \pm \frac{t(S) \cdot s_p(S)}{\sqrt{n(S)}} \right) \cap \left(\overline{x}_p(R) \pm \frac{t(R) \cdot s_p(R)}{\sqrt{n(R)}} \right) = 0 \quad (5)$$

was defined. If this is fulfilled, then the quotient sample/reference $\frac{\overline{x}_p(S)}{\overline{x}_p(R)}$ is built and assigned a color in the heatmap.

5. Output of the result
   The values of the calculated quotients $\frac{\overline{x}_p(S)}{\overline{x}_p(R)}$ are written in a plain text file (one for each user-defined reference condition). Based on these a heatmap

is created (one for all user-defined pairs of conditions S/R) where protected bases (with S/R < 1) are depicted in shades of red, hypersensitive bases (with S/R > 1) in shades of blue (Fig. 1).

Legend to ivFAST heatmap

Protection		Hypersensitivity	
■	>1.5-fold decrease	■	>1.5-fold increase
■	1.3 - 1.5-fold decrease	■	1.3 - 1.5-fold increase
■	1.1 - 1.3-fold decrease	■	1.1 - 1.3-fold increase

Figure 1: Legend heatmap

3   How to use ivFAST

3.1   Running the program

- Open the command line.

- Change to the directory where ivFAST.jar is located.

- Call the program with

  java -jar ivFAST.jar -d DIRECTORY [-s START] [-sd STARDIAGRAM] [-e END] [-p PAIRS]

  (arguments in brackets are optional)

  -d Directory that contains the input files.
     The program first performs a check to see if the file names comply with the specifications. If they do, it will try to read them. Make sure the directory contains only valid input files and only one sequence file, otherwise the program will produce a failure message and cancel the run.

  -s Sequence position where the program should start the mapping.
     The first peak of each data file will be assigned to this position. For files with direction argument 'f' (forward priming) a positive integer is expected, counting forward from the first character of the provided sequence. For files with direction argument 'r' (reverse priming) a negative integer is expected, counting backwards from the last character of the provided sequence.

  -sd Sequence position where the heatmap will start.
     The absolute value of STARDIAGRAM must be $\geq$ the absolute value of START. For files with direction argument 'f' (forward priming) a positive integer is expected, counting forward from the first character of the provided sequence. For files with direction argument 'r'

(reverse priming) a negative integer is expected, counting backwards from the last character of the provided sequence.

-e Sequence position where the heatmap will end.
For files with direction argument 'f' (forward priming) a positive integer is expected, counting forward from the first character of the provided sequence. For files with direction argument 'r' (reverse priming) a negative integer is expected, counting backwards from the last character of the provided sequence.

-p Pairs of sample and references that should be calculated.
Syntax: Sample#/Reference#, Sample#/Reference#, ...
Samples are sorted alphabetically according to the content of the file name field ⟨CONDITION⟩ (see section 3.2.1), and numbered with integers, starting with 0.

- If all arguments have already been provided in the command line, the program will start immediately. Otherwise it will ask the user to provide the missing information. For the found conditions a numbered list will be provided in this case.

- The output of the calculation will be provided in a newly created subdirectory named with date and time, localized in the same directory as ivFAST.jar.

3.2 File specifications

3.2.1 Input files

All files must be plain text files and placed in one directory, which is given as parameter to the program. No other files are allowed in this directory.

Sequence file   The user has to provide exactly one DNA sequence file, containing the DNA sequence to be analyzed, named

sequence_⟨NAME⟩.txt

There are no restrictions for the field ⟨NAME⟩. The content of this file must have a fasta-like format. This means that the first line is reserved for descriptions of any kind, whereas all following lines are expected to contain the DNA sequence. Line breaks within the sequence are ignored. Allowed characters are only 'A', 'C', 'G' and 'T' (not case-sensitive). General characters like 'N', 'Y', etc. are not accepted.

Data files   The data files contain the measured fragment size and area from the capillary gel electrophoresis. Each peak has to be a separate line, where first the size and then, separated by a space or a tab, the area is given. As decimal separator '.' and ',' are both accepted. The file must not contain any headers.
The data files must be named

⟨ORGANISM⟩_⟨GENE⟩_⟨f|r⟩_⟨CONDITION⟩_⟨REPLICATE⟩_⟨a|p|s⟩.txt

The fields ⟨ORGANISM⟩ and ⟨GENE⟩ can contain any characters except '_', which is reserved as separator, but they must be identical for all data files in the directory (case-sensitive). The content of these two fields will show up as title of the heatmap.

The field ⟨f|r⟩ has to contain either 'f', if the forward primer was used for the analysis, or 'f', if the reverse primer was used. In case it is 'f', the program will start counting from the beginning of the provided sequence and will take the complement of it, to obtain the complementary strand to the primer. In case it is 'r', the program will start counting from the end of the provided sequence and will take the reverse of it, to obtain the complementary strand to the primer.

The field ⟨CONDITION⟩ can contain any characters except '_' and indicates the different conditions that should be compared. The content of this field will be used to label the rows of the heatmap.

The field ⟨REPLICATE⟩ is used to distinguish replicates obtained from the same conditions. Since the program has to calculate the standard deviation for each condition, at least two files must be provided for each ⟨CONDITION⟩, whose names only differ in the field ⟨REPLICATE⟩. Any characters except '_' can be used in this field.

The field ⟨a|p|s⟩ is used to distinguish the files containing all sample peaks ('a'), only the sample peaks to be analysed ('p'), or only the standard peaks ('s'). For each replicate of a condition all these three fields have to be provided.

3.2.2   Output files

All output files are generated in a new directory named with date and time of the run.

For each given data file a mapping file is produced, which documents how the measured fragment sizes have been mapped to the provided sequence. Additionally it also contains the normalized areas for each peak. Asterisk indicate background peaks, i.e. peaks that occur at bases not listed in the field 'validBases' in the configuration file, and therefore are neglected in the further calculation.

The result of the calculation is provided both as heatmap and as text files. There will be one heatmap for all compared pairs of sample and reference, but one text file for each reference. The given deviations represent the fraction of normalized sample area divided by normalized reference area. Additionally a file is generated that contains the sum over all areas for each sample, as well as the mean and standard deviation of this sum over all samples. A standard deviation of up to 15 % is acceptable to ensure proper normalization.

3.2.3   Configuration file

The configuration file config.properties is located in the folder config, which has to be placed in the same directory as ivFAST.jar. It contains calculation parameters, that can be adjusted by the user, which are: the type of bases at which the DNA can be cut, the probability for the t-distribution and the color ranges used in the heatmap. The default file content is:

6

```
validBases=AG
primerLength=40
probability=0.95
lowerRange=1.1
middleRange=1.3
upperRange=1.5
```

The field 'validBases' specifies at which bases the DNA can be cut and is used by the program to find the valid positions for the mapping in the provided DNA sequence. Valid entries are $\in \{A,C,G,T\}$, listed without deliminator. In the default specification it is assumed that the DNA can be cut at the bases 'A' and 'G', so each 'A' and 'G' in the DNA sequence is treated as a valid position for the mapping, whereas peaks mapped to a 'C' or 'T' are treated as background peaks.

The field 'primerLength' specifies the length of the primers used for the PCR. It is used to define a cut-off size for small fragment artefacts in the sequencing, that do not correspond to meaningful peaks. Varying this parameter affects the normalization (see chapter 2, Calculation for details).

The field 'probability' (α) specifies the width of the confidence interval (see eqn. 4) and is given as two-sided probability

$$\alpha = F_{n-1}(t) - F_{n-1}(-t) = 0.95 \tag{6}$$

(see also Fig. 2). It must hold a value $\in (0, 1)$. Varying this parameter affects the probability of two peaks being considered as different (see chapter 2, Calculation for details)
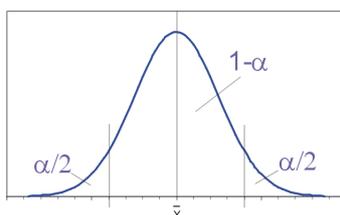

Figure 2: Confidence interval of $\overline{x}$ (sample mean) with probability α. 1-α and $\alpha/2$ denote the portion of the area included in the according region. (Source: Philipendula at the German language Wikipedia, GNU-FDL)

The fields 'lowerRange', 'middleRange' and 'upperRange' are used to define the color ranges of the heatmap (see Tab. 1 and Fig. 1) and can take values > 1, with the restraint lowerRange < middleRange < upperRange.

3.3   Manual adjustments

Background peaks occur frequently for small fragment sizes, but are unlikely for longer fragments. Normally, when a peak in the higher fragment size region is mapped to an invalid position (i.e. considered as background), this is a mapping failure. It occurs due to small inaccuracies in the size determination that can accumulate, when some bases are skipped between two peaks. The program will

Color	RGB code	Range
dark red	(255,0,0)	$S/R < (upperRange)^{-1}$
middle red	(230,140,80)	$(upperRange)^{-1} \leq S/R < (middleRange)^{-1}$
light red	(245,208,115)	$(middleRange)^{-1} \leq S/R < (lowerRange)^{-1}$
light blue	(120,194,240)	$lowerRange < S/R \leq middleRange$
middle blue	(14,121,242)	$middleRange < S/R \leq upperRange$
dark blue	(0,0,255)	$upperRange < S/R$

Table 1: Color definitions of the heatmap.

automatically detetect such possible mapping failures and produce a warning message on the comand-line. But the program is not able to correct these issues automatically, since it cannot judge wether it is a real failure or not. Therefore the user has to correct the according size values in the seqencing file manually (only in the file with the suffix 'p') and repeat the run.

Here is an example, how this is done:
Assuming we have an excerpt of a sequence, with the masses obtained from the sequencing file, as shown in Table 2. Given is the way how it should be mapped. But when calculating the difference between the two masses it is 3.29, which would be rounded to 3, so the calculated position for the second peak would not be the 'G' but the 'T', since the program counts the rounded mass difference between two peaks forward to determine the next position.

A	C	C	T	G
136,83				140,12

Table 2: Example for manual corrections.

In order to get the correct result, the user has to change the values of the size in the sequencing file manually, e.g. to 136.72 and 140.22. Now the difference is 3.50, which will be rounded to 4 and therefore will yield the correct mapping. But be careful not to alter the distances to the next surrounding peaks when changing the values!

After all changes have been done and saved, the program has to be run a second time in order to get the correct mapping results.

4   Credits

This program was written at the Gene Technology Group at the Institute of Chemical Engineering, Vienna University of Technology. For citation please refer to the original paper 'A highly sensitive *in vivo* footprinting technique for

condition-dependent identification of *cis* elements', written by Rita Gorsche, Birgit Jovanović, Loreta Gudynaite-Savitch, Robert L. Mach, and Astrid R. Mach-Aigner, submitted to Nucleic Acids Research.

9

Appendix B

Curiculum vitae

09/1993 – 06/1997	Bundesrealgymnasium, Baden Frauengasse
09/1997 – 06/2002	Höhere Bundes-Lehr- und Versuchsanstalt
	für chemische Industrie, Wien Rosensteingasse
	Ausbildungsschwerpunkt Biochemie, Bio- und Gentechnologie
	Diplomarbeit: „Einsatz der 2D-Elektrophorese
	zum Monitoring von rekombinanten Proteinen
	während eines Fermentationsprozesses"
10/2002 – 06/2008	Diplomstudium Technische Physik an der TU Wien
	Diplomarbeit: „Masses of anti-de Sitter spacetimes"
	Verleihung des Grades Dipl.Ing. am 26.06.2008
08/2008 - 09/2009	Berufstätigkeit
10/2009 – 01/2010	Masterstudium Biotechnologie und Bioanalytik
	an der TU Wien
03/2010 – 11/2013	Dr.-Studium der technischen Wissenschaften,
	Technische Chemie an der TU Wien
	Dissertation: „Methods of improving *Trichoderma*
	reesei as a whole cell biocatalyst"

Appendix C

List of Publications

D-xylose concentration-dependent hydrolase expression profiles and the function of CreA and XlnR in *Aspergillus niger*
Astrid R. Mach-Aigner, Jimmy Omony, Birgit Jovanović, Anton J. B. van Boxtel, and Leo H. de Graaff
Applied Environmental Microbiology 2012, 78:9, pages 3145-3155
doi:10.1128/AEM.07772-11

Characterization of erythrose reductases from filamentous fungi
Birgit Jovanović, Astrid R. Mach-Aigner, and Robert L. Mach
AMB Express 2013, 3:1
doi:10.1186/2191-0855-3-43

A highly sensitive *in vivo* footprinting technique for condition-dependent identification of *cis* elements
Rita Gorsche, Birgit Jovanović, Loreta Gudynaite-Savitch, Robert L. Mach, and Astrid R. Mach-Aigner
Nucleic Acids Research 2013
doi:10.1093/nar/gkt883

Erythritol production on wheat straw using *Trichoderma reesei*
Birgit Jovanović, Astrid R. Mach-Aigner, and Robert L. Mach
Submitted to Microbial Cell Factories 2013

Appendix D

Bibliography

[1] *Trichoderma & Gliocladium*. Taylor & Francis Ltd. 1998.

[2] Kuhls K, Lieckfeldt E, Samuels GJ, Kovacs W, Meyer W, Petrini O, Gams W, Börner T, Kubicek CP: **Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina***. *Proceedings of the National Academy of Sciences* 1996, **93**(15):7755–7760, [http://www.pnas.org/content/93/15/7755.abstract].

[3] Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, Leon ALd, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)**. *Nature Biotechnology* 2008, **26**(5):553–560.

[4] Buchert J, Oksanen T, Pere J, Siika-aho M, Suurnäkki A, Viikari L: **Applications of *Trichoderma reesei* enzymes in the pulp and paper industry**. In *Trichoderma & Gliocladium, Volume 2*. Edited by Harman G, Kubicek C, London, UK: Taylor & Francis Ltd. 1998:343–357.

[5] Noé P, Chevalier J, Mors F, Comtat J: **Action of xylanases on chemical pulp fibers Part II : Enzymatic beating**. *Journal of*

Wood Chemistry and Technology 1986, **6**(2):167–184, [http://www.tandfonline.com/doi/abs/10.1080/02773818608085222].

[6] Welt T, Dinus R: **Enzymatic deinking - a review**. *Progress in Paper Recycling* 1995, **4**(2):36–47.

[7] Galante Y, R M, S I, Caldini C, de Conti A, Lavelli V, Bonomi F: **New applications of enzymes in wine making and olive oil production**. *Italian Biochem Soc trans* 1993, **4**:34.

[8] Lanzarini G, Pifferi P: **Enzymes in the fruit juice industry**. In *Biotechnology Applications in Beverage Production*, Elsevier Applied Food Science Series. Edited by Cantarelli C, Lanzarini G, Springer Netherlands 1989:189–222, [http://dx.doi.org/10.1007/978-94-009-1113-0_13].

[9] Walsh G, Power R, Headon D: **Enzymes in the animal-feed industry**. *Trends Biotechnol* 1993, **11**(10):424–430.

[10] Koo H, Ueda M, Wakida T, Yoshimura Y, Igarashi T: **Cellulase treatment of cotton fabrics**. *Textile Research Journal* 1994, **64**(2):70–74, [http://trj.sagepub.com/content/64/2/70.abstract].

[11] Kumar A, Lepola M, Purtell C: **Enyme finishing of man-made cellulosic fabrics**. *Textile Chemist and Colorist* 1994, **26**(10):25–28.

[12] Pedersen GP, Screws GA, Cereoni DA: **Biopolishing of cellulosic fabrics**. *Can Text J* 1992.

[13] Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund M, Liden G, Zacchi G: **Bio-ethanol - the fuel of tomorrow from the residues of today**. *Trends Biotechnol* 2006, **24**(12):549–556.

[14] Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD: **Biomass recalcitrance: engineering plants and enzymes for biofuels production**. *Science* 2007, **315**(5813):804–807.

[15] Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T: **The path forward for biofuels and biomaterials**. *Science* 2006, **311**(5760):484–489, [http://www.sciencemag.org/content/311/5760/484.abstract].

[16] Moon HJ, Jeya M, Kim IW, Lee JK: **Biotechnological production of erythritol and its applications**. *Appl Microbiol Biotechnol* 2010, **86**(4):1017–1025, [http://dx.doi.org/10.1007/s00253-010-2496-4].

[17] Livesey G: **Tolerance of low-digestible carbohydrates: a general view**. *British Journal of Nutrition* 2001, **85**:7–16.

[18] Munro I, Bernt W, Borzelleca J, Flamm G, Lynch B, Kennepohl E, Bär E, Modderman J: **Erythritol: an interpretive summary of biochemical, metabolic, toxicological and clinical data**. *Food and Chemical Toxicology* 1998, **36**(12):1139 – 1174, [http://www.sciencedirect.com/science/article/pii/S027869159800091X].

[19] Kasumi T: **Fermentative production of polyols and utilization for food and other products in Japan**. *JARQ* 1995, **29**:49–55.

[20] Hajny GJ, Smith JH, Garver JC: **Erythritol production by a yeast-like fungus**. *Applied Microbiology* 1964, **12**(3):240–246, [http://aem.asm.org/content/12/3/240.abstract].

[21] Jeya M, Lee KM, Tiwari M, Kim JS, Gunasekaran P, Kim SY, Kim IW, Lee JK: **Isolation of a novel high erythritol-producing *Pseudozyma tsukubaensis* and scale-up of erythritol fermentation to industrial level**. *Applied Microbiology and Biotechnology* 2009, **83**(2):225–231, [http://dx.doi.org/10.1007/s00253-009-1871-5].

[22] Rymowicz W, Rywińska A, Marcinkiewicz M: **High-yield production of erythritol from raw glycerol in fed-batch cultures of *Yarrowia lipolytica***. *Biotechnology Letters* 2009, **31**(3):377–380, [http://dx.doi.org/10.1007/s10529-008-9884-1].

[23] Ishizuka H, Wako K, Kasumi T, Sasaki T: **Breeding of a mutant of *Aureobasidium* sp. with high erythritol production**. *Journal of Fermentation and Bioengineering* 1989, **68**(5):310 – 314, [http://www.sciencedirect.com/science/article/pii/0922338X89900032].

[24] Yang SW, Park JB, Soo Han N, Ryu YW, Seo JH: **Production of erythritol from glucose by an osmophilic mutant of *Candida magnoliae***. *Biotechnology Letters* 1999, **21**(10):887–890, [http://dx.doi.org/10.1023/A%3A1005566420982].

[25] Savergave LS, Gadre RV, Vaidya BK, Narayanan K: **Strain improvement and statistical media optimization for enhanced**

erythritol production with minimal by-products from **Candida magnoliae mutant R23**. *Biochemical Engineering Journal* 2011, **55**(2):92–100, [http://www.sciencedirect.com/science/article/pii/S1369703X11000799].

[26] Tomaszewska L, Rywińska A, Gładkowski W: **Production of erythritol and mannitol by *Yarrowia lipolytica* yeast in media containing glycerol**. *Journal of Industrial Microbiology & Biotechnology* 2012, **39**(9):1333–1343, [http://dx.doi.org/10.1007/s10295-012-1145-6].

[27] Rywińska A, Tomaszewska L, Rymowicz W: **Erythritol biosynthesis by *Yarrowia lipolytica* yeast under various culture conditions**. *African Journal of Microbiology Research* 2013, **7**(27):3511–3516.

[28] Fackler K, Ters T, Ertl O, Messner K: **Method for lignin recovery**. *Patent WO/2012/027767* 2012, [http://www.sumobrain.com/patents/WO2012027767.html].

[29] Jovanovic B, Mach R, Mach-Aigner A: **Characterization of erythrose reductases from filamentous fungi**. *AMB Express* 2013, **3**:43, [http://www.amb-express.com/content/3/1/43].

[30] Jovanovic B, Mach R, Mach-Aigner A: **Erythritol production on wheat straw using *Trichoderma reesei*** 2013.

[31] Jackson PD, Felsenfeld G: **A method for mapping intranuclear protein-DNA interactions and its application to a nuclease hypersensitive site**. *Proceedings of the National Academy of Sciences* 1985, **82**(8):2296–2300, [http://www.pnas.org/content/82/8/2296.abstract].

[32] Giniger E, Varnum SM, Ptashne M: **Specific {DNA} binding of GAL4, a positive regulatory protein of yeast**. *Cell* 1985, **40**(4):767 – 774, [http://www.sciencedirect.com/science/article/pii/0092867485903368].

[33] Ephrussi A, Church G, Tonegawa S, Gilbert W: **B lineage–specific interactions of an immunoglobulin enhancer with cellular factors in vivo**. *Science* 1985, **227**(4683):134–140, [http://www.sciencemag.org/content/227/4683/134.abstract].

[34] Mueller P, Wold B: **In vivo footprinting of a muscle specific enhancer by ligation mediated PCR**. *Science* 1989,

246(4931):780–786, [http://www.sciencemag.org/content/246/4931/780.abstract].

[35] Pfeifer GP, Drahovsky D: **DNA methyltransferase polypeptides in mouse and human cells**. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1986, **868**(4):238 – 242, [http://www.sciencedirect.com/science/article/pii/016747818690059X].

[36] Garrity PA, Wold BJ: **Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and in vivo footprinting.** *Proceedings of the National Academy of Sciences* 1992, **89**(3):1021–1025, [http://www.pnas.org/content/89/3/1021.abstract].

[37] Huang J, Sun M: **A modified AFLP with fluorescence-labelled primers and automated DNA sequencer detection for efficient fingerprinting analysis in plants**. *Biotechnology Techniques* 1999, **13**:277–278.

[38] Trotha R, Reichl U, Thies FL, Sperling D, König W, König B: **Adaption of a fragment analysis technique to an automated high-throughput multicapillary electrophoresis device for the precise qualitative and quantitative characterization of microbial communities**. *Electrophoresis* 2002, **23**(7-8):1070–1079, [http://dx.doi.org/10.1002/1522-2683(200204)23:7/8<1070::AID-ELPS1070>3.0.CO;2-H].

[39] Wolschek MF, Narendja F, Kubicek CP, Strauss J, Karlseder J, Scazzocchio C: **In situ detection of protein—DNA interactions in filamentous fungi by in vivo footprinting**. *Nucleic Acids Research* 1998, **26**(16):3862–3864, [http://nar.oxfordjournals.org/content/26/16/3862.abstract].

[40] Mach RL, Strauss J, Zeilinger S, Schindler M, Kubicek CP: **Carbon catabolite repression of xylanase I (*xyn1*) gene expression in *Trichoderma reesei***. *Molecular Microbiology* 1996, **21**(6):1273–1281.

[41] Zeilinger S, Mach RL, Schindler M, Herzog P, Kubicek CP: **Different inducibility of expression of the two xylanase genes *xyn1* and *xyn2* in *Trichoderma reesei***. *Journal of Biological Chemistry* 1996, **271**(41):25624–25629, [http://www.jbc.org/content/271/41/25624.abstract].

[42] Xu J, Nogawa M, Okada H, Morikawa Y: **Regulation of *xyn3* gene expression in *Trichoderma reesei* PC-3–7**. *Applied Microbiology and Biotechnology* 2000, **54**(3):370–375, [http://dx.doi.org/10.1007/s002530000410].

[43] Astrid R Mach-Aigner MGSGEBSJP Marion E Pucher, Mach RL: **Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina***. *Appl Environ Microbiol* 2008.

[44] Furukawa T, Shida Y, Kitagami N, Mori K, Kato M, Kobayashi T, Okada H, Ogasawara W, Morikawa Y: **Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei***. *Fungal Genetics and Biology* 2009, **46**(8):564 – 574, [http://www.sciencedirect.com/science/article/pii/S1087184509000772].

[45] Strauss J, Mach RL, Zeilinger S, Hartler G, Stöffler G, Wolschek M, Kubicek C: **Crel, the carbon catabolite repressor protein from *Trichoderma reesei***. {*FEBS*} *Letters* 1995, **376**(1–2):103 – 107, [http://www.sciencedirect.com/science/article/pii/0014579395012555].

[46] Saloheimo A, Aro N, Ilmén M, Penttilä M: **Isolation of the *ace1* gene encoding a Cys$_2$-His$_2$ transcription factor involved in regulation of activity of the cellulase promoter *cbh1* of *Trichoderma reesei***. *Journal of Biological Chemistry* 2000, **275**(8):5817–5825, [http://www.jbc.org/content/275/8/5817.abstract].

[47] Aro N, Saloheimo A, Ilmén M, Penttilä M: **ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei***. *Journal of Biological Chemistry*.

[48] Gorsche R, Jovanovic B, Gudynaite-Savitch L, Mach RL, Mach-Aigner AR: **A highly sensitive in vivo footprinting technique for condition-dependent identification of cis elements**. *Nucleic Acids Research* 2013, [http://nar.oxfordjournals.org/content/early/2013/10/03/nar.gkt883.abstract].

[49] Ookura T, Azuma K, Isshiki K, Taniguchi H, Kasumi T, Kawamura Y: **Primary structure analysis and functional expression of erythrose reductases from erythritol-producing fungi (*Trichosporonoides megachiliensis* SNG-42)**. *Bioscience, Biotechnology, and Biochemistry* 2005, **69**(5):944–951.

[50] Lee DH, Lee YJ, Ryu YW, Seo JH: **Molecular cloning and bio-chemical characterization of a novel erythrose reductase from *Candida magnoliae* JH110**. *Microbial Cell Factories* 2010, **9**:43, [http://www.microbialcellfactories.com/content/9/1/43].

[51] Zeilinger S, Mach RL, Kubicek CP: **Two adjacent protein binding motifs in the *cbh2* (cellobiohydrolase II-encoding) promoter of the fungus *Hypocrea jecorina* (*Trichoderma reesei*) cooper-ate in the induction by cellulose**. *Journal of Biological Chemistry* 1998, **273**(51):34463–34471, [http://www.jbc.org/content/273/51/34463.abstract].

[52] Würleitner E, Pera L, Wacenovsky C, Cziferszky A, Zeilinger S, Kubicek CP, Mach RL: **Transcriptional regulation of *xyn2* in *Hypocrea jecorina***. *Eukaryotic Cell* 2003, **2**:150–158, [http://ec.asm.org/content/2/1/150.abstract].