



TECHNISCHE
UNIVERSITÄT
WIEN

Vienna University of Technology

Diese Dissertation wurde begutachtet von

Ao. Univ. Prof. Mag. Dr. Robert L. Mach

Ao. Univ. Prof. Mag. Dr. Hubertus Haas

Privatdoz. Mag. Dr. Irina Druzhinina

DISSERTATION

Trichoderma reesei:

New insights into gene regulation and potential use as cell factory

ausgeführt zum Zwecke der Erlangung des akademischen Grades einer Doktorin der technischen
Wissenschaften unter der Leitung von

Ao. Univ. Prof. Mag. Dr. rer. nat. Robert L. Mach

und

DI Dr. techn. Astrid R. Mach-Aigner

E166

Institut für Verfahrenstechnik, Umwelttechnik und Technische Biowissenschaften

eingereicht an der Technischen Universität Wien

Fakultät für Technische Chemie

von

DI Rita Gorsche

Matrikelnummer: 0325052

Leysnerstraße 1/14

1140 Wien

Wien, am

TABLE OF CONTENTS

ZUSAMMENFASSUNG	2
ABSTRACT	3
INTRODUCTION	4
<i>Trichoderma reesei</i>	4
Regulation of cellulase and hemicellulase-encoding genes	4
Xylanase regulator 1 (Xyr1).....	4
Other important transcription factors	5
Protein-DNA interactions within the promoter regions of hydrolytic enzyme-encoding genes.....	6
<i>In vivo</i> footprinting	8
<i>Trichoderma</i> as whole-cell factory	9
AIM OF THE PHD THESIS	10
CONCLUSION	11
REFERENCES	14
PUBLICATIONS.....	20
PART I: New Insights Into Gene Regulation	20
PART II: <i>Trichoderma</i> as Cell Factory	69
CURRICULUM VITAE.....	91
LIST OF PUBLICATIONS.....	92

ZUSAMMENFASSUNG

Der filamentöse Ascomyzet *Trichoderma reesei* ist vor allem für seine großartigen Eigenschaften in der industriellen Produktion von Proteinen bekannt. *Trichoderma reesei* wird vor allem für die industrielle Produktion von Zellulasen, Hemizellulasen und anderen biopolymer-abbauenden Enzymen eingesetzt. Daher wurden in der vorliegenden Arbeit die folgenden Themen behandelt, um weitere Einblicke in die Regulation der Genexpression und den möglichen Einsatz von *Trichoderma reesei* als Zellfabrik zu erhalten.

Die Regulation der Zellulase- und Hemizellulase-codierenden Gene wird von verschiedenen unterschiedlichen Transkriptionsfaktoren beeinflusst. Um die Induktionsmuster der Proteinexpression zu verstehen ist die Untersuchung von Protein-DNA-Interaktionen *in vivo* von größter Bedeutung. Daher verbesserten wir die traditionelle Methode des „*in vivo* footprinting“, um eine reproduzierbare, stabile und zuverlässige Methode für die Identifikation von Protein-DNA-Interaktionen zu erhalten. In Kombination mit der neuen computerunterstützten, automatisierten Datenanalyse gewährleistet dieser Ansatz eine objektive Auswertung von Hochdurchsatzprojekten.

Eine einzige Punktmutation im Transaktivator Xyr1 führt zu einer Deregulation der Xylanaseexpression, während die Induktion der Zellulaseexpression weiterhin durch Sophorose induzierbar bleibt. CD-Analysen von Xyr1 und Xyr1_{A824V} zeigen eine Änderung der Sekundärstruktur und einen Verlust der Strukturänderung als Reaktion auf die Gegenwart bestimmter Zuckermoleküle nahe. Diese Ergebnisse legen den Schluss nahe, dass Xyr1 eine sogenannte „nuclear receptor“-ähnliche Domäne besitzt.

Trichoderma reesei ist ein wichtiger industrieller Produzent von hydrolytischen Enzymen. Daher wäre auch ein Einsatz als Zellfabrik denkbar. Daher wurde zum ersten Mal die heterologe Expression einer bakteriellen Zwei-Enzymkaskade in *Trichoderma reesei* durchgeführt. Das Ziel war die Produktion von N-Acetylneuraminsäure, einem wertvollen Rohstoff für die pharmazeutische Industrie, durch Fermentation auf kolloidalem Chitin, einem billigen und ergiebigen Rohstoff.

ABSTRACT

The filamentous ascomycete *Trichoderma reesei* is known for its high protein production and capabilities of secretion. *Trichoderma reesei* has been extensively used in industry for the production of cellulases, hemicellulases, and other biopolymer-degrading enzymes. Therefore, the following topics are being addressed in order to obtain further insights into gene regulation and the potential use of *Trichoderma reesei* as a cell factory.

The regulation of cellulase- and hemicellulase-encoding genes is subject to a number of different transcription factors. In order to understand the induction patterns responsible, the study of protein-DNA interactions *in vivo* is highly relevant. Therefore we have improved traditional *in vivo* footprinting to obtain a reproducible, stable and reliable method to identify protein-DNA interactions. The combination with automated data analysis ensures objective evaluation of high-throughput projects.

A single point mutation in Xyr1 was shown to cause strong deregulation of xylanase expression, while cellulase expression remained inducible by sophorose. CD analysis of the wild-type Xyr1 and Xyr1_{A824V} suggests a change in secondary structure as well as loss of structural response to the present carbohydrate. This points towards the presence of a nuclear receptor-like domain in Xyr1.

Since *Trichoderma reesei* is an important industrial producer of hydrolytic enzymes, this suggests the potential use of *Trichoderma reesei* as cell factory. For the first time the heterologous expression of a bacterial two-enzyme cascade was attempted, which has resulted in the production of N-acetyl neuraminic acid, a valuable chemical for the pharmaceutical industry, through fermentation of colloidal chitin, a cheap and abundant carbon source.

INTRODUCTION

Trichoderma reesei

The filamentous ascomycete *Trichoderma reesei* (teleomorph *Hypocrea jecorina*) [35] is known for its high secretion of a wide variety of hydrolytic enzymes for the degradation of biopolymers such as cellulose and hemicelluloses. Today hydrolytic enzymes produced with the help of *Trichoderma reesei* (*T. reesei*) are used for the improvement of fiber properties and enzymatic bleaching in the pulp and paper industry [6], for enzymatic saccharification and improvement of filter properties in the food and feed industry [17], for enzymatic stone-washing of denim fabrics and bio-finishing of non-denim cellulosic fabric in the textile industry [18], and for the saccharification of biomass for the production of biofuels [17].

Regulation of cellulase and hemicellulase-encoding genes

The genome of *T. reesei* includes genes encoding for 10 cellulases and 16 hemicellulases [44]. The most prominent among the cellulose-degrading enzymes are the two cellobiohydrolases CBHI and CBHII (EC 3.2.1.91) [48, 81], the five endo- β -1,4-glucanases EGI to EGV (EC 3.2.1.4) [56, 63, 65, 67] and the two β -glucosidases BGLI and BGLII (EC 3.2.1.21) [5, 7]. The most abundant xylan degrading enzymes secreted are two endo- β -1,4-xylanases XYNI and XYNII (EC 3.2.1.8) [82], and one β -xylosidase BXLI (EC 3.2.1.37) [42].

The expression of these enzymes is induced by degradation products of cellulose and xylan, with the transcription of cellulase-encoding genes being induced by cellulose, cellobiose or the transglycosylation product sophorose [25, 74, 84]. In case of the transcription of xylanase-encoding genes, *xyn1* can be induced by xylan, cellulose and D-xylose at low concentrations, while *xyn2* is always present at a low basal level and can be further induced by xylan, xylobiose, sophorose, cellobiose and low concentrations of D-xylose [41, 88]. Transcription of *bxll* is induced by xylobiose [43]. In addition the cellulase- and hemicellulase-encoding genes are subject to carbon catabolite repression (CCR) mediated by the carbon catabolite repressor 1 (Cre1) [24, 75].

Xylanase regulator 1 (Xyr1)

Xyr1 is a Gal4-like Zn₂Cys₆ binuclear cluster protein that can bind GGC(T/A)₃ motifs as homo- or heterodimer and functions as a wide-domain activator of all important hydrolytic enzyme-encoding genes, with the exception of *bgl2* [16, 76]. Transcriptional analysis of a *xyr1* disruption strain shows no transcript formation for either the major cellulase-encoding genes *cbh1*, *cbh2* and *egl1*, or the major xylanase-encoding genes *xyn1* and *xyn2* under repressing or inducing conditions. In addition the

expression of the D-xylose reductase 1 (Xyl1), which catalyzes the first step in the D-xylose catabolism, is strictly dependent on the presence of Xyr1 [39].

The expression of Xyr1 itself is subject to CCR mediated by Cre1 in the presence of D-glucose. While the inducers of xylanase expression, namely D-xylose or xylobiose, do not show any inducing effect on the *xyr1* transcript formation [39], sophorose, a strong inducer of cellulase expression, also increases the formation of *xyr1* transcript [12]. This rise in *xyr1* transcript is directly connected to an increase in the expression of the two main cellulase genes *cbh1* and *cbh2*, which suggests a strong dependence of cellulase expression on the amount of Xyr1 present. In contrast, the expression of the two major xylanases *xyn1* and *xyn2* does not directly depend on the amount of Xyr1 but is rather contingent on additional mechanisms [12].

The single point mutation A824V was shown to cause an increase and a strong deregulation of xylanase expression as well as elevated levels of cellulase expression, which remained further inducible by sophorose. According to *in silico* analysis this mutation is located in an α -helix located in the so-called fungal transcription factor regulatory middle homology region of Xyr1 and might cause this effect based on a change in secondary structure [12].

Other important transcription factors

There are a number of important transcription factors involved in the regulation of hydrolytic enzyme-encoding gene expression in *T. reesei*. The most prominent and well-characterized ones are the following:

- Carbon Catabolite Repressor 1 (Cre1):
In *T. reesei* CCR in the presence of high concentrations of easily utilizable monosaccharides such as D-glucose or D-xylose is mediated via the repressor protein Cre1. The consensus sequence for Cre1-binding is 5'-SYGGRG-3' and Cre1-binding is dependent on dimerization and phosphorylation [9, 24, 75]. The promoter regions of *cbh1* and *xyn1* include Cre1-binding sites in tandem and inverted repeats respectively, which are responsible for transcriptional repression. Transcript analysis of RutC-30 (a Cre1-negative strain) and CK11 (*cre1*-retransformation strain of RutC-30) showed derepression/repression of *xyr1* transcription in the presence of glucose, which can in turn explain the presence of CCR for genes like *cbh2* or *xyn2*, whose promoter regions do not have double Cre1-binding sites [23, 37, 39]. In addition Cre1 was found to be important for nucleosome positioning in the *cbh1* promoter region [60].
- Activator of Cellulases 1 (Ace1):
This transcription factor is a repressor of hydrolytic enzyme expression. Transcript analysis of an

ace1 deletion strain caused earlier induction and higher expression of the cellulase-encoding genes *cbh1*, *cbh2*, *egl1* and *egl2*, as well as the xylanase-encoding genes *xyn1* and *xyn2* [3, 64]. Ace1 was shown to directly compete with Xyr1 in binding to one of the Xyr1-binding sites in the *xyn1* promoter. This suggests that the repressor function is mediated by a heterodimer of Ace1 and Xyr1, probably together with an additional protein, while *xyn1* transcriptional activation is caused by the binding of a Xyr1 homodimer [59].

- Activator of Cellulases 2 (Ace2):

This transcription factor is an activator of cellulase gene expression. Transcript analysis of an *ace2* deletion strain showed delayed induction of *cbh1*, *cbh2*, *egl1* and *egl2* under inducing conditions as well as overall lower enzyme production. Furthermore, under inducing conditions (xylan, xylobiose) the expression of *xyn2* was initially higher, but yielded lower final xylanase activities than the parental strain, hinting at an essential role of Ace2 in the low constitutive transcript level of *xyn2* as well as the induction by xylobiose. In contrast, Ace2 has no influence on *xyn1* transcript formation or induction by sophorose [3, 79]. Ace2 was shown to bind both Xyr1-binding sites within the *xyn2* promoter as dimer and only when it is phosphorylated [79].

- Hap2/3/5-Complex

The Hap2/3/5-complex in *T. reesei* is a homologue of the Hap- and AnCF-protein complexes in *Saccharomyces cerevisiae* and *Aspergillus nidulans* respectively [45, 53, 55, 73]. It binds to the CCAAT motif in the promoter regions of hydrolytic enzyme encoding genes and is required for transcription. Studies of nucleosome positioning in the *cbh2* promoter suggest that the Hap2/3/5-complex is involved in the opening of the chromatin structure necessary for transcriptional activation [89-91].

- Xylanase promoter-binding protein 1 (Xpp1)

In vivo footprinting and EMSA analyses of the *xyn2* promoter area suggested the binding of a putative repressor protein to an AGAA-box [87]. Affinity chromatography and subsequent MS analysis resulted in the identification of three DNA-binding proteins, which are potentially involved in *xyn2* gene regulation [40, 77].

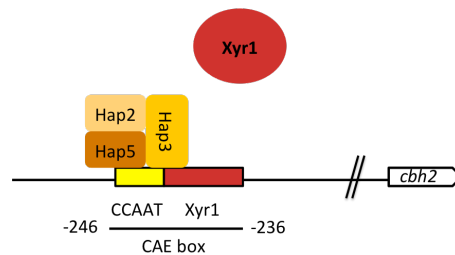
Protein-DNA interactions within the promoter regions of hydrolytic enzyme-encoding genes

In order to understand specific induction patterns within the Xyr1 regulon, the study of protein-DNA interactions within the promoter regions of the respective genes can provide valuable insights. Past efforts in that direction using deletion analysis, site mutagenesis, traditional *in vivo* footprinting and EMSA experiments have already yielded the following results:

- *cbh2*:

The area in the *cbh2* promoter essential for transcriptional activation was identified as the so-called *cbh2*-activating element (CAE) from position -246 to -236, which is bound by the Hap2/3/5-complex via the CCAAT-box as well as Xyr1 via the Xyr1-binding site (see Figure 1). A mutation in either of these binding sites does not completely abolish transcription, which points to an interaction between these proteins. In addition, binding of the Hap2/3/5-complex to the CAE element has been shown to play an important role in nucleosome positioning and transcriptional activation of *cbh2* [72, 89-91].

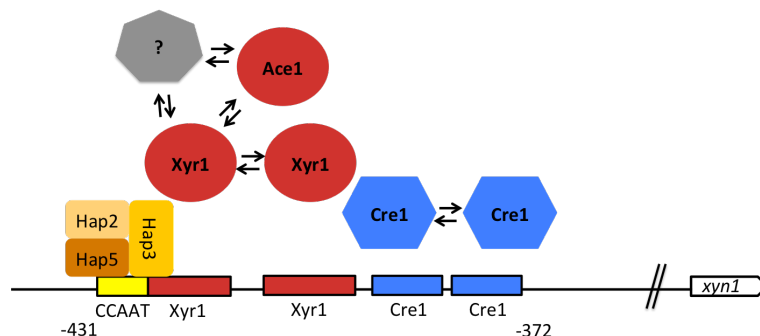
Figure 1: Schematic representation of protein-DNA interactions within the *cbh2* promoter



- *xyn1*:

Regarding the *xyn1* promoter region the motifs identified were positioned from -372 to -431 and binding of the Hap2/3/5-complex to a CCAAT-box was observed (see Figure 2). Additionally, two Xyr1-binding sites positioned as inverted repeat 10 bp apart were shown to play a role in the regulation of *xyn1* gene expression. While Xyr1 binds to both motifs, Ace1 competes Xyr1-binding exclusively on the downstream site. This suggests that the binding of a Xyr1 homodimer results in activation of transcription, while the binding of an Ace1/Xyr1 heterodimer, together with a possible third interaction partner, leads to repression. In addition an inverted repeat of two Cre1-binding sites was shown to be contacted by Cre1 under glucose repressing conditions [37, 59, 78].

Figure 2: Schematic representation of protein-DNA interactions within the *xyn1* promoter

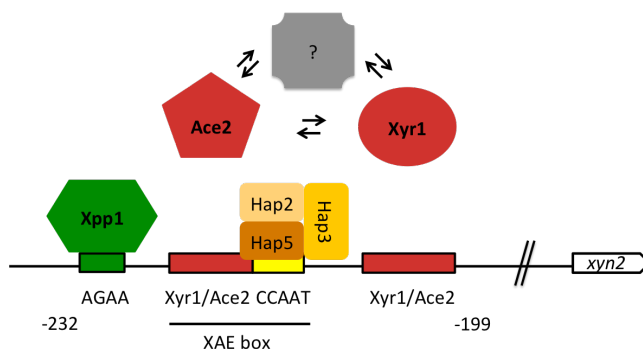


- *xyn2*:

The area from -199 to -232 includes the previously identified motifs (see Figure 3). Among those

was once more the Hap2/3/5-complex binding to the CCAAT-box. In addition, a double Xyr1-site, arranged as inverted repeat, was shown to be contacted by Xyr1 as well as Ace2. In the case of Ace2 this contact is strictly dependent on dimerization and phosphorylation. This suggests an intriguing interplay between Xyr1 and Ace2 combined with the possible recruiting of additional still unidentified proteins in the *xyn2* transcriptosome. An AGAA-box was additionally identified as being contacted by a putative repressor protein (Xylanase promoter-binding protein 1 (Xpp1)) under glucose repressing conditions [3, 40, 78, 79, 87].

Figure 3: Schematic representation of protein-DNA interactions within the *xyn2* promoter



***In vivo* footprinting**

One way to shed light on specific factors involved in gene regulation is the identification of distinct areas of the upstream regulatory region (URR) of a gene, which are specifically bound by regulatory proteins. This can be accomplished by various *in vivo* and *in vitro* footprinting methods, which rely on nucleases, such as DNaseI [26, 27, 30, 92] or alkylating agents, such as dimethylsulfate (DMS) [15, 21], to identify changes in DNA accessibility due to protein-DNA interaction. In the case of DMS *in vivo* footprinting the original method was already established in 1985 [15, 21] and has since been improved upon by including ligation-mediated PCR (LM-PCR) [50], optimizing the polymerase and cycling conditions [19], and applying the protocol to different kinds of cells, from cell lines [15, 20, 49, 50] and yeast [21], to filamentous fungi [86]. Nevertheless, the resulting method still suffered from the fact that the use of polyacrylamide gels and radioactive labeling of the DNA fragments constituted a rather laborious protocol. This used hazardous materials and yielded results of strongly varying quality which were not suitable for high-throughput projects.

The fact that the URRs of eukaryotic genes are rather complex and are made up of various sites that are often targeted by various transcription factors at a time further complicates the identification of regulatory patterns [14]. Since a number of genes and transcription factors are grouped together in regulons, shedding light on the specific factors involved in transcriptional regulation of such requires the simultaneous analysis of footprinting patterns of the URRs of a number of different genes under

various conditions. This calls for a standardized, high-throughput approach to allow the analysis of a great number of various footprinting patterns at a time.

***Trichoderma* as whole-cell factory**

Trichoderma reesei has been extensively used for the industrial production of hydrolytic enzymes. Today, cellulases are the third largest group of industrially produced enzymes worldwide, with *T. reesei* being their principal producer [85]. The strains used in industry are the result of a long history of strain improvement and mutations and reach protein production and secretion on a scale of 100 g/l [69]. In contrast to other industrial microorganisms, *T. reesei* is cultivated on cheap and simple media such as mixtures of cellulose and xylan in plant materials from agricultural waste [38], or the industrial byproduct lactose [70]. In addition to their native hydrolytic enzymes, a number of heterologous enzymes from closely related donor organisms, such as α -amylase or phospholipase from *Aspergillus* *sp.*, are industrially produced using *T. reesei* as host organism [1].

On lab scale a number of different heterologous proteins have successfully been expressed in *T. reesei* [11, 22, 28, 29, 31-34, 42, 47, 51, 52, 54, 57, 58, 61, 62, 66, 68, 83], though efforts only recently went into metabolic engineering and the use of recombinant *T. reesei* strains for the production of valuable chemicals [8, 10].

AIM OF THE PHD THESIS

Trichoderma reesei is an important industrial producer of hydrolytic enzymes owing to its great secretory capacity and simple cultivation. Keeping in mind that for the hydrolysis of cellulose and hemicelluloses in industrial applications a definite composition of the hydrolytic enzyme mixture is essential and it is therefore important to produce these enzymes as specifically as possible, further insights into the regulation of the expression of these enzymes are necessary. In addition, the fact that *T. reesei* is such a successful producer of both homologous and heterologous proteins raises the question if it would not also be suitable as a whole cell catalyst.

Therefore in the course of this thesis the following three topics were addressed:

- I. In order to gain more insight into the Xyr1 regulon, which encompasses all major cellulases and hemicellulases, a high-throughput *in vivo* footprinting method needed to be established. The goal was a highly reproducible, reliable and objective method based on fluorescent labeling of DNA fragments, standardized capillary gel electrophoresis and automated computer-assisted data analysis.
- II. An A824V mutation in Xyr1, the main transcriptional activator of hydrolytic enzyme-encoding genes, causes a significantly different phenotype concerning hydrolytic enzyme-encoding gene expression. In light of this, the question was raised whether there is a mechanistic explanation for this effect as a result of change in protein structure and reaction to carbon sugars.
- III. *T. reesei* is widely used for the industrial production of hydrolytic enzymes, largely due to its great secretory capacity. Keeping in mind that it can utilize a number of very cheap complex carbon sources, such as cellulose or chitin, the possibility of using *T. reesei* as cell factory was addressed. So far only single heterologous proteins have been expressed in *T. reesei*, so the expression of a bacterial 2-enzyme cascade was attempted in order to produce a highly valuable chemical building block from a renewable, cheap feedstock.

CONCLUSION

As can be seen from the publications included in this thesis the questions mentioned above were successfully addressed, which led to new insights into gene regulation and the potential use of *T. reesei* as cell factory.

Ad I.)

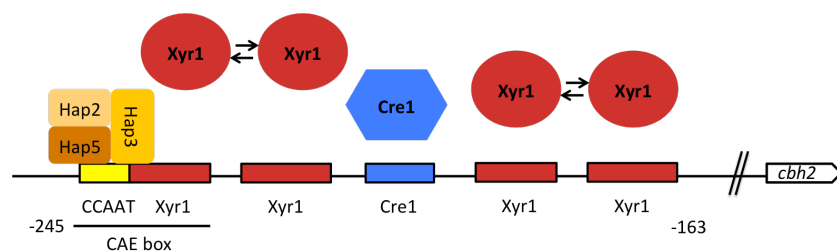
As previously mentioned an improvement of the original *in vivo* footprinting protocol is vital for various reasons. Firstly, fluorescent labeling is preferable to radioactive labeling simply because of safety considerations. In addition analysis of fluorescently labeled DNA fragments by a commercial sequencing service is important to ensure stability and reproducibility of the results. Furthermore, computer-assisted and automated data analysis provides for objective results, which are directly displayed as heatmaps.

The newly optimized *in vivo* footprinting method has proved to be more sensitive than the original protocol, which facilitates the comparison of samples from different conditions, as opposed to simple comparison to naked DNA. After ensuring reproducibility, stability and reliability of the footprinting patterns based on the URR of *xyn1* the newly improved protocol was applied to the URRs of *cbh2* and *xyn2*. For all three regions investigated previous findings were confirmed and new sites were identified as follows:

- *cbh2*:

In vivo footprinting results for repressing and inducing conditions, D-glucose and sophorose respectively, yielded a number of new condition-dependent signals. Glucose-dependent signals in the CCAAT-box but not in the adjacent Xyr1-site suggest condition-dependent binding of the Hap2/3/5-complex, while Xyr1 is permanently bound. A second Xyr1-double site was also identified, as well as a single Cre1-site which yielded glucose-dependent signals.

Figure 4: Updated schematic representation of protein-DNA interactions within the *cbh2* promoter

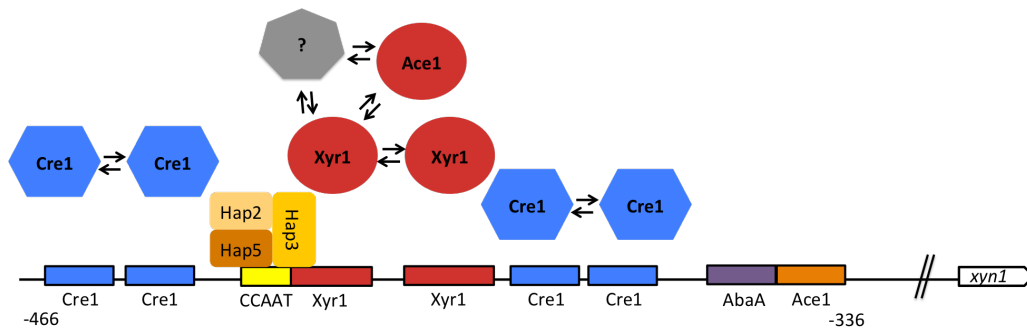


- *xyn1*:

The promoter region of *xyn1* was investigated using D-glucose as repressing and a low D-

xylose concentration as inducing condition. In addition to the regulatory sites identified previously, a second putative Cre1-double site as well as an Ace1-site yielded signals. Furthermore, a GTAATG-site was found that suggests the possible involvement of an AbaA-homolog in the transcriptional regulation of *xyn1*. In *Aspergillus nidulans* AbaA was shown to be involved in conidiophore development [2]. This might be connected to the expression of hydrolytic enzymes during conidia development in *T. reesei* [46].

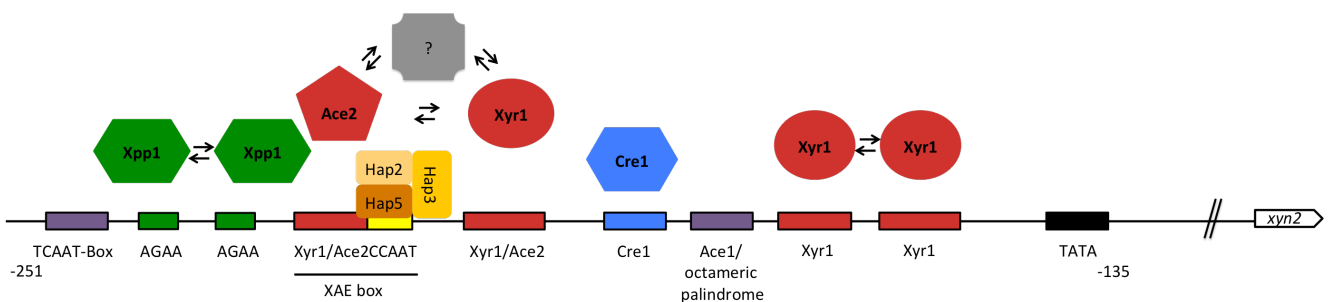
Figure 5: Updated schematic representation of protein-DNA interactions within the *xyn1* promoter



- *xyn2*:

For the analysis of the *xyn2* promoter region D-glucose as repressing condition and a low D-xylose concentration as an inducing condition were investigated. In addition to the regulatory elements mentioned earlier, a number of new putative protein binding sites was identified. A second AGAA-box arranged as an inverted repeat to the one previously described yielded condition-dependent signals, suggesting the binding of the putative repressor protein Xpp1 as a dimer. A second Xyr1-double site was also found, as well as a single putative Cre1-site, which showed condition-dependent differences. Furthermore, two additional possible protein binding motifs were found, namely both an area containing an unusual TCAAT-box and an octameric palindrome overlapping with an Ace1-binding site. Finally, the TATA-box exhibited especially strong signals.

Figure 6: Updated schematic representation of protein-DNA interactions within the *xyn2* promoter



Ad II.)

The A824V mutation in Xyr1 described earlier causes a change in the protein's secondary structure. A comparison of the circular dichroism (CD) spectra of wild-type Xyr1 and Xyr1_{A824V} indicated a loss in helical content and reduced protein-DNA complex formation in the mutant. Keeping in mind that this mutation leads to increased xylanase expression *in vivo*, it must be concluded that the change in DNA-binding affinity does not directly cause this change in phenotype. Due to the fact that this point mutation led to a general deregulation of xylanase-encoding genes while the inducibility of cellulase gene expression by sophorose was retained [12], a possible conformational change of Xyr1 in the presence of certain metabolites was considered. Analysis of CD spectra showed a variation in the response to D-xylose and D-glucose for the wild-type Xyr1, while this allosteric response was lost for Xyr1_{A824V}.

These results suggest the presence of a nuclear receptor-like domain in Xyr1, which might react to the presence of certain carbohydrates with a change in structure and presentation of the transactivating domain. This carbohydrate-specific response is lost in the presence of the A824V point mutation.

Ad III.)

Chitin is the second most abundant form of biomass available and is composed of β -(1,4)-linked N-acetylglucosamine (GlcNAc) [4]. *T. reesei* has been found to have a number of genes encoding chitinolytic enzymes and is able to successfully degrade chitin down to the monomer GlcNAc [13, 44, 71]. The heterologous expression of two bacterial enzymes, namely GlcNAc-2-epimerase from *Anabaena sp.* CH1 (EC 5.1.3.8) [36] and NeuNAc synthase from *Campylobacter jejuni* (EC 2.5.1.56) [80], under the control of the constitutive pyruvate kinase (*pki*) promoter resulted in a recombinant strain whose native chitinolytic potential was complemented to yield the valuable N-acetyl-neuraminic acid (NeuNAc). In a fermentation process this recombinant strain was able to utilize colloidal chitin as sole carbon source and the biosynthesis of NeuNAc on a scale of several μg per g mycelium (dry weight) was achieved.

Undoubtedly, these results do not constitute a competitive production process, but this work can be seen as successful proof-of-concept. It is indeed possible to engineer a two-step bacterial enzyme cascade into a saprophytic fungus to obtain a whole-cell catalyst for the production of a valuable chemical building block for the pharmaceutical industry from an abundant waste material using a simple and cheap cultivation method.

REFERENCES

1. AMFEP. *List of commercial enzymes*. 2009 [20/01/2013]; Available from: <http://www.amfep.org/content/list-enzymes>.
2. Andrianopoulos, A. and W.E. Timberlake, The *Aspergillus nidulans* abaA gene encodes a transcriptional activator that acts as a genetic switch to control development. *Mol Cell Biol*, 1994. 14(4): p. 2503-15.
3. Aro, N., et al., ACEI of *Trichoderma reesei* is a repressor of cellulase and xylanase expression. *Appl Environ Microbiol*, 2003. 69(1): p. 56-65.
4. Ballenweg, S., *Roempp Online*. 2005, Thieme Chemistry, Stuttgart.
5. Barnett, C.C., R.M. Berka, and T. Fowler, Cloning and amplification of the gene encoding an extracellular beta-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. *Biotechnology (N Y)*, 1991. 9(6): p. 562-7.
6. Buchert, J., Oksanen, T. et al., Applications of *Trichoderma reesei* enzymes in the pulp and paper industry, in *Trichoderma & Gliocladium*, C.P. Kubicek, Harman, G.E., Editor. 1998, Taylor & Francis: London. p. 343-363.
7. Chen, H., M. Hayn, and H. Esterbauer, Purification and characterization of two extracellular beta-glucosidases from *Trichoderma reesei*. *Biochim Biophys Acta*, 1992. 1121(1-2): p. 54-60.
8. Chen, X., et al., Overexpression of bacterial ethylene-forming enzyme gene in *Trichoderma reesei* enhanced the production of ethylene. *International journal of biological sciences*, 2010. 6(1): p. 96-106.
9. Cziferszky, A., R.L. Mach, and C.P. Kubicek, Phosphorylation positively regulates DNA binding of the carbon catabolite repressor Cre1 of *Hypocrea jecorina* (*Trichoderma reesei*). *J Biol Chem*, 2002. 277(17): p. 14688-94.
10. Dashtban, M., et al., Xylitol production by genetically engineered *Trichoderma reesei* strains using barley straw as feedstock. *Applied biochemistry and biotechnology*, 2013. 169(2): p. 554-69.
11. De Faria, F.P., et al., Expression and processing of a major xylanase (XYN2) from the thermophilic fungus *Hemicella grisea* var. *thermoidea* in *Trichoderma reesei*. *Letters in applied microbiology*, 2002. 34(2): p. 119-123.
12. Derntl, C., et al., Mutation of the Xylanase regulator 1 causes a glucose blind hydrolase expressing phenotype in industrially used *Trichoderma* strains. *Biotechnol Biofuels*, 2013. 6(1): p. 62.
13. Donzelli, B.G., G. Ostroff, and G.E. Harman, Enhanced enzymatic hydrolysis of langostino shell chitin with mixtures of enzymes from bacterial and fungal sources. *Carbohydr Res*, 2003. 338(18): p. 1823-33.
14. Dynan, W.S. and R. Tjian, Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature*, 1985. 316(6031): p. 774-8.
15. Ephrussi, A., et al., B lineage--specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science*, 1985. 227(4683): p. 134-40.
16. Furukawa, T., et al., Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in *Trichoderma reesei*. *Fungal Genet Biol*, 2009. 46(8): p. 564-74.
17. Galante, Y.M., de Conti, A., Monteverdi, R., Application of *Trichoderma* enzymes in the food and feed industries, in *Trichoderma & Gliocladium*, C.P. Kubicek, Harman, G.E., Editor. 1998, Taylor & Francis: London. p. 327-342.

18. Galante, Y.M., de Conti, A., Monteverdi, R., Application of *Trichoderma* enzymes in the textile industry, in *Trichoderma & Gliocladium*, C.P. Kubicek, Harman, G.E., Editor. 1998, Taylor & Francis: London. p. 311-325.
19. Garrity, P.A. and B.J. Wold, Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and *in vivo* footprinting. *Proceedings of the National Academy of Sciences of the United States of America*, 1992. 89(3): p. 1021-5.
20. Gilmour, D.S. and R. Fan, Detecting transcriptionally engaged RNA polymerase in eukaryotic cells with permanganate genomic footprinting. *Methods*, 2009. 48(4): p. 368-74.
21. Giniger, E., S.M. Varnum, and M. Ptashne, Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, 1985. 40(4): p. 767-74.
22. Harkki, A., et al., A novel fungal expression system: secretion of active calf chymosin from the filamentous fungus *Trichoderma reesei*. *Bio/technology*, 1989. 7: p. 596-603.
23. Ilmén, M., et al., Functional analysis of the cellobiohydrolase I promoter of the filamentous fungus *Trichoderma reesei* *Mol Gen Genet*, 1996. 253(3): p. 303-14.
24. Ilmén, M., C. Thrane, and M. Penttilä, The glucose repressor gene *cre1* of *Trichoderma*: isolation and expression of a full-length and a truncated mutant form. *Mol Gen Genet*, 1996. 251(4): p. 451-60.
25. Ilmén, M., et al., Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Appl Environ Microbiol*, 1997. 63(4): p. 1298-306.
26. Jackson, P.D. and G. Felsenfeld, A method for mapping intranuclear protein-DNA interactions and its application to a nuclease hypersensitive site. *Proceedings of the National Academy of Sciences of the United States of America*, 1985. 82(8): p. 2296-300.
27. Jackson, P.D. and G. Felsenfeld, *In vivo* footprinting of specific protein-DNA interactions. *Methods in enzymology*, 1987. 152: p. 735-55.
28. Joutsjoki, V.V., et al., Secretion of the *Hormoconis resinae* glucoamylase P enzyme from *Trichoderma reesei* directed by the natural and the *cbh1* gene secretion signal. *FEMS microbiology letters*, 1993. 112(3): p. 281-6.
29. Joutsjoki, V.V., T.K. Torkkeli, and K.M. Nevalainen, Transformation of *Trichoderma reesei* with the *Hormoconis resinae* glucoamylase P (*gamP*) gene: production of a heterologous glucoamylase by *Trichoderma reesei*. *Current genetics*, 1993. 24(3): p. 223-8.
30. Kemper, B., P.D. Jackson, and G. Felsenfeld, Protein-binding sites within the 5' DNase I-hypersensitive region of the chicken alpha D-globin gene. *Molecular and cellular biology*, 1987. 7(6): p. 2059-69.
31. Kiiskinen, L.L., et al., Expression of *Melanocarpus albomyces* laccase in *Trichoderma reesei* and characterization of the purified enzyme. *Microbiology*, 2004. 150(Pt 9): p. 3065-74.
32. Kontkanen, H., T. Reinikainen, and M. Saloheimo, Cloning and expression of a *Melanocarpus albomyces* steryl esterase gene in *Pichia pastoris* and *Trichoderma reesei*. *Biotechnology and bioengineering*, 2006. 94(3): p. 407-15.
33. Kontkanen, H., et al., Characterization of *Melanocarpus albomyces* steryl esterase produced in *Trichoderma reesei* and modification of fibre products with the enzyme. *Applied microbiology and biotechnology*, 2006. 72(4): p. 696-704.
34. Kontkanen, H., et al., Novel *Coprinosia cinerea* polyesterase that hydrolyzes cutin and suberin. *Applied and environmental microbiology*, 2009. 75(7): p. 2148-57.
35. Kuhls, K., et al., 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*, 1996. 93(15): p. 7755-60.

36. Lee, Y.C., H.C. Chien, and W.H. Hsu, Production of N-acetyl-D-neuraminic acid by recombinant whole cells expressing *Anabaena sp.* CH1 N-acetyl-D-glucosamine 2-epimerase and *Escherichia coli* N-acetyl-D-neuraminic acid lyase. *J Biotechnol*, 2007. 129(3): p. 453-60.
37. Mach, R.L., et al., Carbon catabolite repression of xylanase I (*xynI*) gene expression in *Trichoderma reesei*. *Mol Microbiol*, 1996. 21(6): p. 1273-81.
38. Mach, R.L. and S. Zeilinger, Regulation of gene expression in industrial fungi: *Trichoderma*. *Appl Microbiol Biotechnol*, 2003. 60(5): p. 515-22.
39. Mach-Aigner, A.R., et al., Transcriptional regulation of *xyrI*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. *Appl Environ Microbiol*, 2008. 74(21): p. 6554-62.
40. Mach-Aigner, A.R., et al., From an electrophoretic mobility shift assay to isolated transcription factors: a fast genomic-proteomic approach. *BMC Genomics*, 2010. 11: p. 644.
41. Mach-Aigner, A.R., M.E. Pucher, and R.L. Mach, D-Xylose as a repressor or inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*). *Appl Environ Microbiol*, 2010. 76(6): p. 1770-6.
42. Margolles-Clark, E., et al., Cloning of genes encoding alpha-L-arabinofuranosidase and beta-xylosidase from *Trichoderma reesei* by expression in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, 1996. 62(10): p. 3840-6.
43. Margolles-Clark, E., M. Ilmén, and M. Penttilä, Expression patterns of ten hemicellulase genes of the filamentous fungus *Trichoderma reesei* on various carbon sources. *J Biotechnol*, 1997. 57: p. 167-179.
44. Martinez, D., et al., Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol*, 2008. 26(5): p. 553-60.
45. McNabb, D.S. and I. Pinto, Assembly of the Hap2p/Hap3p/Hap4p/Hap5p-DNA complex in *Saccharomyces cerevisiae*. *Eukaryot Cell*, 2005. 4(11): p. 1829-39.
46. Metz, B., et al., Expression of biomass-degrading enzymes is a major event during conidium development in *Trichoderma reesei*. *Eukaryot Cell*, 2011. 10(11): p. 1527-35.
47. Miettinen-Oinonen, A., et al., Overexpression of the *Aspergillus niger* pH 2.5 acid phosphatase gene in a heterologous host *Trichoderma reesei*. *Journal of biotechnology*, 1997. 58(1): p. 13-20.
48. Mong Chen, C., M. Gritzali, and D.W. Stafford, Nucleotide Sequence and Deduced Primary Structure of Cellobiohydrolase II from *Trichoderma Reesei*. *Nat Biotech*, 1987. 5(3): p. 274-278.
49. Mueller, P.R., S.J. Salser, and B. Wold, Constitutive and metal-inducible protein:DNA interactions at the mouse metallothionein I promoter examined by *in vivo* and *in vitro* footprinting. *Genes & development*, 1988. 2(4): p. 412-27.
50. Mueller, P.R. and B. Wold, *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science*, 1989. 246(4931): p. 780-6.
51. Murray, P., et al., Expression in *Trichoderma reesei* and characterisation of a thermostable family 3 beta-glucosidase from the moderately thermophilic fungus *Talaromyces emersonii*. *Protein expression and purification*, 2004. 38(2): p. 248-57.
52. Nyysönen, E., et al., Efficient production of antibody fragments by the filamentous fungus *Trichoderma reesei*. *Bio/technology*, 1993. 11(5): p. 591-5.
53. Olesen, J.T. and L. Guarente, The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex. *Genes Dev*, 1990. 4(10): p. 1714-29.

54. Paloheimo, M., et al., High-yield production of a bacterial xylanase in the filamentous fungus *Trichoderma reesei* requires a carrier polypeptide with an intact domain structure. *Applied and environmental microbiology*, 2003. 69(12): p. 7073-82.
55. Papagiannopoulos, P., et al., The hapC gene of *Aspergillus nidulans* is involved in the expression of CCAAT-containing promoters. *Mol Gen Genet*, 1996. 251(4): p. 412-21.
56. Penttilä, M., et al., Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the *endoglucanase I* gene. *Gene*, 1986. 45(3): p. 253-63.
57. Penttilä, M., Heterologous protein production in *Trichoderma*, in *Trichoderma & Gliocladium*, G.E.K. Harman, C.P., Editor. 1998, Taylor & Francis Ltd: London. p. 365-382.
58. Poidevin, L., et al., Heterologous production of the *Piromyces equi* cinnamoyl esterase in *Trichoderma reesei* for biotechnological applications. *Letters in applied microbiology*, 2009. 49(6): p. 673-8.
59. Rauscher, R., et al., Transcriptional regulation of *xyn1*, encoding xylanase I, in *Hypocrea jecorina*. *Eukaryot Cell*, 2006. 5(3): p. 447-56.
60. Ries, L., et al., The role of CRE1 in nucleosome positioning within the *cbh1* promoter and coding regions of *Trichoderma reesei*. *Appl Microbiol Biotechnol*, 2013.
61. Saarelainen, R., et al., Expression of Barley Endopeptidase B in *Trichoderma reesei*. *Applied and environmental microbiology*, 1997. 63(12): p. 4938-40.
62. Salles, B.C., et al., Identification of two novel xylanase-encoding genes (*xyn5* and *xyn6*) from *Acrophialophora nainiana* and heterologous expression of *xyn6* in *Trichoderma reesei*. *Biotechnology letters*, 2007. 29(8): p. 1195-201.
63. Saloheimo, A., et al., A novel, small endoglucanase gene, *egl5*, from *Trichoderma reesei* isolated by expression in yeast. *Mol Microbiol*, 1994. 13(2): p. 219-28.
64. Saloheimo, A., et al., 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*. *J Biol Chem*, 2000. 275(8): p. 5817-25.
65. Saloheimo, M., et al., EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. *Gene*, 1988. 63(1): p. 11-22.
66. Saloheimo, M. and M.-L. Niku-Paavola, Heterologous production of a ligninolytic enzyme: expression of the *Phlebia radiata* laccase gene in *Trichoderma reesei*. *Bio/technology*, 1991. 9(10): p. 987-990.
67. Saloheimo, M., et al., cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *Eur J Biochem*, 1997. 249(2): p. 584-91.
68. Schmoll, M., et al., Recombinant production of an *Aspergillus nidulans* class I hydrophobin (DewA) in *Hypocrea jecorina* (*Trichoderma reesei*) is promoter-dependent. *Applied microbiology and biotechnology*, 2010. 88(1): p. 95-103.
69. Schuster, A. and M. Schmoll, Biology and biotechnology of *Trichoderma*. *Appl Microbiol Biotechnol*, 2010. 87(3): p. 787-99.
70. Seiboth, B., et al., The D-xylose reductase of *Hypocrea jecorina* is the major aldose reductase in pentose and D-galactose catabolism and necessary for beta-galactosidase and cellulase induction by lactose. *Mol Microbiol*, 2007. 66(4): p. 890-900.
71. Seidl, V., et al., A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS J*, 2005. 272(22): p. 5923-39.
72. Stangl, H., F. Gruber, and C.P. Kubicek, Characterization of the *Trichoderma reesei cbh2* promoter. *Curr Genet*, 1993. 23(2): p. 115-22.

73. Steidl, S., et al., AnCF, the CCAAT binding complex of *Aspergillus nidulans*, contains products of the hapB, hapC, and hapE genes and is required for activation by the pathway-specific regulatory gene amdR. *Mol Cell Biol*, 1999. 19(1): p. 99-106.
74. Sternberg, D. and G.R. Mandels, Induction of cellulolytic enzymes in *Trichoderma reesei* by sophorose. *J Bacteriol*, 1979. 139(3): p. 761-9.
75. Strauss, J., et al., Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*. *FEBS Lett*, 1995. 376(1-2): p. 103-7.
76. Stricker, A.R., et al., Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. *Eukaryot Cell*, 2006. 5(12): p. 2128-37.
77. Stricker, A.R. and R.L. Mach, Gene Regulation of Hydrolase Expression: Lessons from the Industrially Important Fungus *Trichoderma reesei* (*Hypocrea jecorina*), in *Current Advances in Molecular Mycology*, Y. Gherbawy, R.L. Mach, and M. Rai, Editors. 2008, Nova Science Publishers, Inc.: NY.
78. Stricker, A.R., R.L. Mach, and L.H. de Graaff, Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). *Appl Microbiol Biotechnol*, 2008. 78(2): p. 211-20.
79. Stricker, A.R., et al., Role of Ace2 (Activator of Cellulases 2) within the *xyn2* transcriptosome of *Hypocrea jecorina*. *Fungal Genet Biol*, 2008. 45(4): p. 436-45.
80. Sundaram, A.K., et al., Characterization of N-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. *Biochem J*, 2004. 383(Pt 1): p. 83-9.
81. Teeri, T., Salovouri, I., Knowles, J., The molecular cloning of the major cellulase gene from *Trichoderma reesei*. *Biotechnology*, 1983. 1: p. 696-699.
82. Törrönen, A., et al., The two major xylanases from *Trichoderma reesei*: characterization of both enzymes and genes. *Biotechnology (N Y)*, 1992. 10(11): p. 1461-5.
83. Uusitalo, J.M., et al., Enzyme production by recombinant *Trichoderma reesei* strains. *Journal of biotechnology*, 1991. 17(1): p. 35-49.
84. Vaheri, M.P., Leisola, M., Kaupinnen, V., Transglycosylation products of the cellulase system of *Trichoderma reesei*. *Biotechnol Lett*, 1979. 1: p. 41-46.
85. Wilson, D.B., Cellulases and biofuels. *Curr Opin Biotechnol*, 2009. 20(3): p. 295-9.
86. Wolschek, M.F., et al., *In situ* detection of protein-DNA interactions in filamentous fungi by *in vivo* footprinting. *Nucleic acids research*, 1998. 26(16): p. 3862-4.
87. Würleitner, E., et al., Transcriptional regulation of *xyn2* in *Hypocrea jecorina*. *Eukaryot Cell*, 2003. 2(1): p. 150-8.
88. Zeilinger, S., et al., Different inducibility of expression of the two xylanase genes *xyn1* and *xyn2* in *Trichoderma reesei*. *J Biol Chem*, 1996. 271(41): p. 25624-9.
89. Zeilinger, S., R.L. Mach, and C.P. Kubicek, 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*, 1998. 273(51): p. 34463-71.
90. Zeilinger, S., et al., The *Hypocrea jecorina* HAP 2/3/5 protein complex binds to the inverted CCAAT-box (ATTGG) within the *cbh2* (cellobiohydrolase II-gene) activating element. *Mol Genet Genomics*, 2001. 266(1): p. 56-63.
91. Zeilinger, S., et al., Nucleosome transactions on the *Hypocrea jecorina* (*Trichoderma reesei*) cellulase promoter *cbh2* associated with cellulase induction. *Mol Genet Genomics*, 2003. 270(1): p. 46-55.

92. Zinn, K. and T. Maniatis, Detection of factors that interact with the human beta-interferon regulatory region *in vivo* by DNAase I footprinting. *Cell*, 1986. 45(4): p. 611-8.

PUBLICATIONS

PART I: New Insights Into Gene Regulation

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

(published in *Nucleic Acids Research*, 2013, 1-12)

incl. Supplementary Data and *ivFAST* User Manual

Publication II: Characterization of a nuclear receptor-like domain of the Xylanase regulator 1, a Gal4-like zinc binuclear cluster regulatory protein

(submitted at the Journal of Biological Chemistry)

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

Rita Gorsche¹, Birgit Jovanovic¹, Loreta Gudynaite-Savitch², Robert L. Mach¹ and Astrid R. Mach-Aigner^{1,*}

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

Received May 16, 2012; Revised August 29, 2013; Accepted September 9, 2013

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 10⁹ 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-³²P-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 RevertAidTM 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× ABsoluteTM 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	<i>iv</i> -FP, oligo-short
RG54	GCGGTGACCCGGGAGATCTGAATTC	<i>iv</i> -FP, oligo-long
RG67	AAGTCATTGCACTCCAAGGC	<i>iv</i> -FP, <i>xyn1</i> oligo 1 fw
RG68	CCTCTTCACATCATGATTTGAGC	<i>iv</i> -FP, <i>xyn1</i> oligo 1 rev
RG69	ATTCTGCAGCAAATGGCCTCAAGCAAC	<i>iv</i> -FP, <i>xyn1</i> oligo 2 fw
RG70	CAAGTGAGGTTGAAAGCGGCTCGTA	<i>iv</i> -FP, <i>xyn1</i> oligo 2 rev
RG71	[6-FAM]CTGCAGCAAATGGCCTCAAGCAACTACG	<i>iv</i> -FP, <i>xyn1</i> oligo 3 fw
RG72	[6-FAM]GAGGTTGAAAGCGGCTCGTACAGTATCC	<i>iv</i> -FP, <i>xyn1</i> oligo 3 rev
RG72-2	GAGGTTGAAAGCGGCTCGTACAGTATCC	<i>iv</i> -FP, <i>xyn1</i> oligo 3 rev
RG97	AAGCGCTAATGTGGACAGGATT	<i>iv</i> -FP, <i>cbh2</i> oligo 1 fw
RG98	CAATACACAGAGGTTGATCTTAC	<i>iv</i> -FP, <i>cbh2</i> oligo 1 rev
RG99	CATTAGCCTCAAGTAGAGCCTATTTCCCTC	<i>iv</i> -FP, <i>cbh2</i> oligo 2 fw
RG100	GCCTCTTCAGGTGAGCTGCTG	<i>iv</i> -FP, <i>cbh2</i> oligo 2 rev
RG101	[6-FAM]GCCTCAAGTAGAGCCTATTTCCCTCGCC	<i>iv</i> -FP, <i>cbh2</i> oligo 3 fw
RG102	[6-FAM]CTTCAGGTGAGCTGTGAGACCATG	<i>iv</i> -FP, <i>cbh2</i> oligo 3 rev
RG127	GTTCCGATATATGAGATTGCCAAG	<i>iv</i> -FP, <i>xyn2</i> oligo 1 fw
RG128	GTTGATGTCTTCTTGCTTCAGC	<i>iv</i> -FP, <i>xyn2</i> oligo 1 rev
RG129	AGCCGTTATTCAGACAATGTATGTGCCG	<i>iv</i> -FP, <i>xyn2</i> oligo 2 fw
RG130	GGAGTTGTTGTGTCTTTGGGCTTGG	<i>iv</i> -FP, <i>xyn2</i> oligo 2 rev
RG131	[6-FAM]CCGTTATTCAGACAATGTATGTGCCGGGG	<i>iv</i> -FP, <i>xyn2</i> oligo 3 fw
RG132	[6-FAM]GTTGTTGTGTCTTTGGGCTTGGAGGGG	<i>iv</i> -FP, <i>xyn2</i> oligo 3 rev
act fw	TGAGAGCGGTGGTATCCACG	<i>act</i> qPCR
act rev	GGTACCACCAGACATGACAATGTTG	<i>act</i> qPCR
sar1 fw	TGGATCGTCAACTGGTTCTACGA	<i>sar1</i> qPCR
sar1 rev	GCATGTGTAGCAACGTGGTCTTT	<i>sar1</i> qPCR
cbh2 fw	CTATGCCGGACAGTTTGTGGTG	<i>cbh2</i> qPCR
cbh2 rev	GTCAGGCTCAATAACCAGGAGG	<i>cbh2</i> qPCR
xyn1 fw	CAGCTATTCGCCTTCCAACAC	<i>xyn1</i> qPCR
xyn1 rev	CAAAGTTGATGGGAGCAGAAG	<i>xyn1</i> qPCR
xyn2 fw	GGTCCAACCTCGGGCAACTTT	<i>xyn2</i> qPCR
xyn2 rev	CCGAGAAGTTGATGACCTTGTTT	<i>xyn2</i> 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

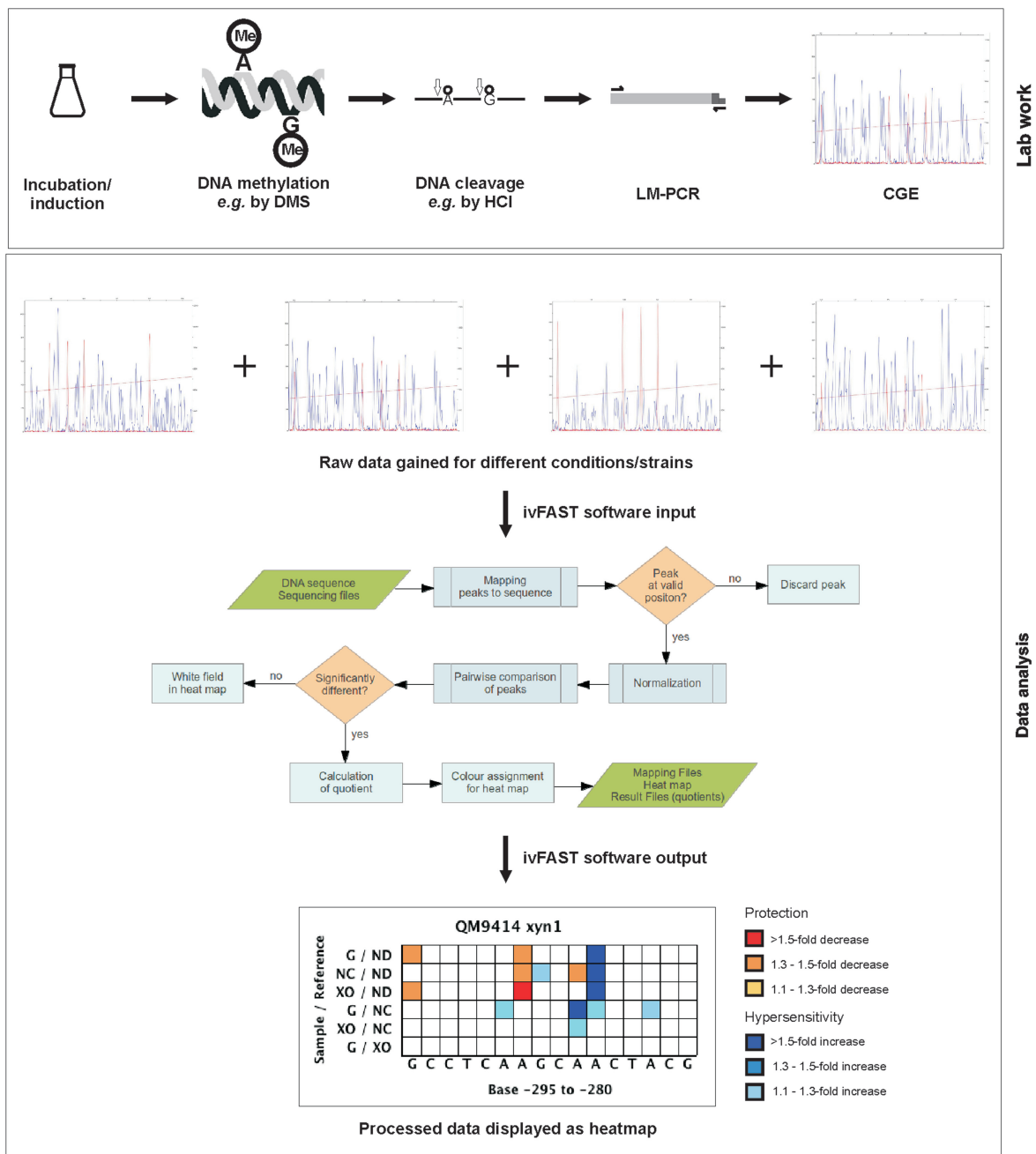


Figure 1. Schematic presentation of the workflow and generation of final data. The main steps of the software-based, high-throughput *in vivo* footprinting method comprise growing/incubating the microorganism under conditions to be investigated (e.g. inducing conditions), *in vivo* DNA methylation using 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

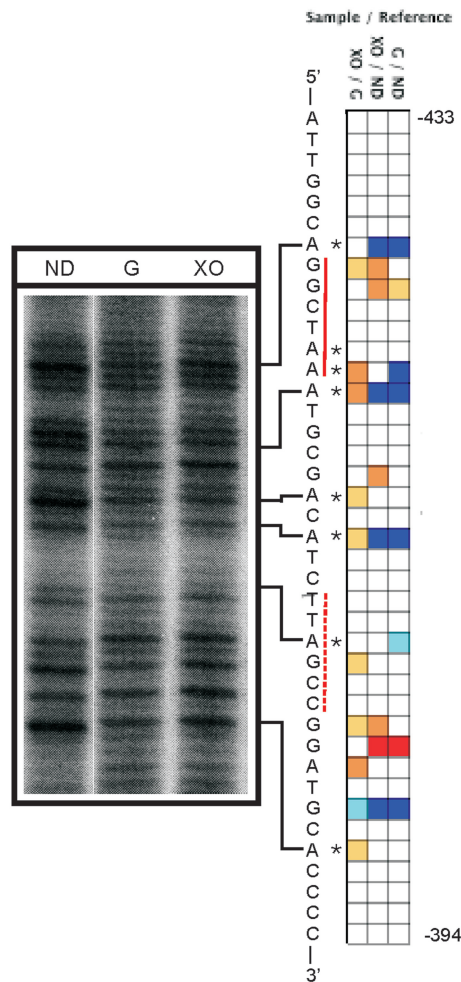


Figure 2. Comparison of the traditional *in vivo* footprinting to the newly developed method. *In vivo* footprinting analysis of the coding strand of a *xyn1* URR (–433- to –394-bp upstream from ATG) covering two Xyr1-binding sites, which are indicated by red lines (solid, site is located on the coding 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 is given (Figure 3c). 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, $\Delta xyr1$ -G/ $\Delta xyr1$ -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 CCAAT-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

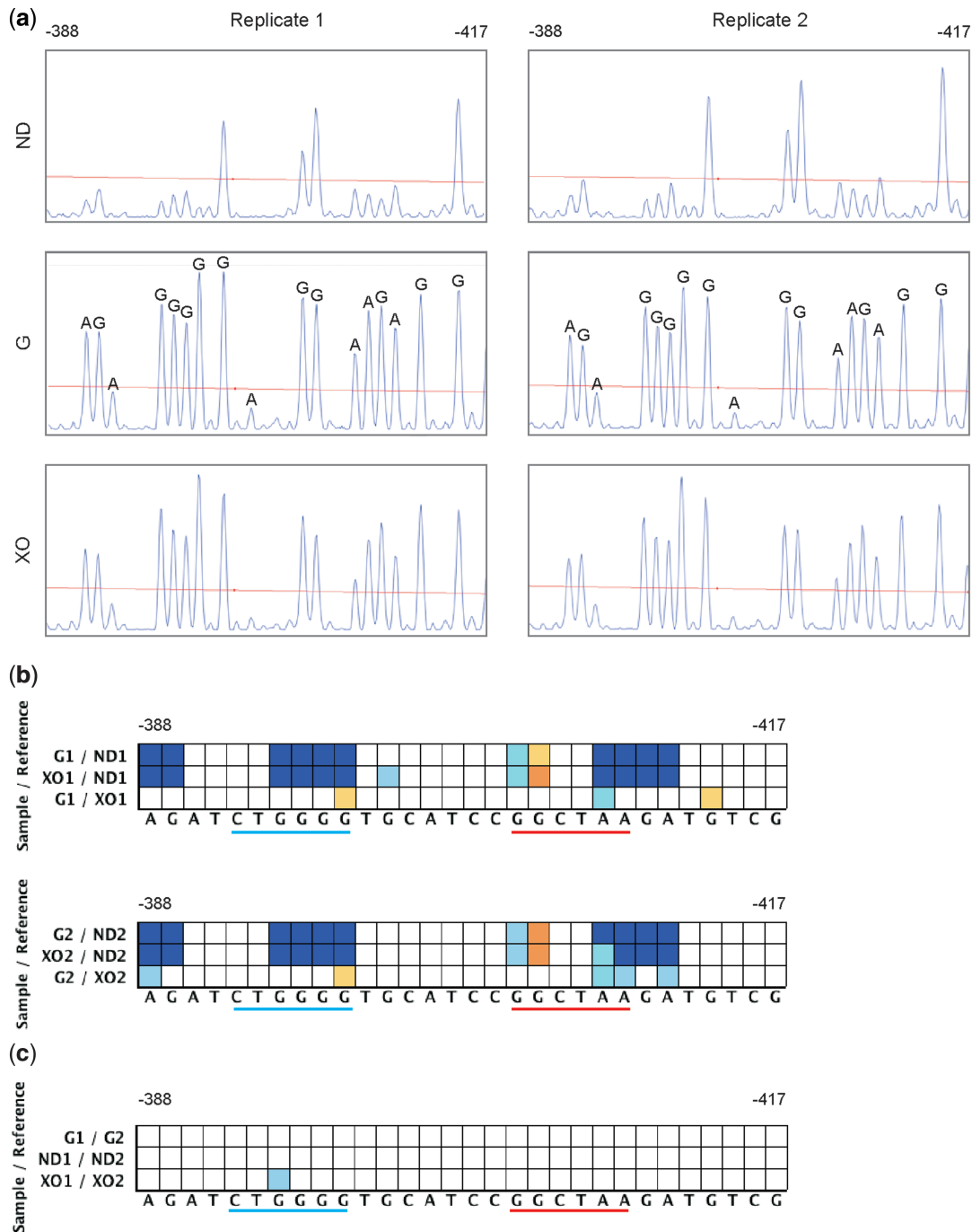


Figure 3. Comparison of two biological replicates analysed by the newly developed *in vivo* footprinting method. *In vivo* footprinting analysis of the non-coding strand of a *xynI* URR (–388- to –417-bp upstream from ATG) covering a Cre1-binding site (underlined in blue) and a Xyr1-binding site (underlined 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 *xyn1* 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

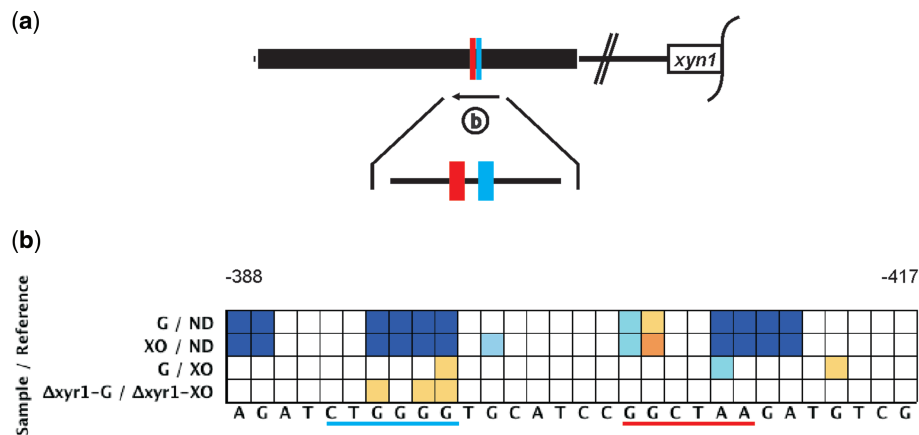


Figure 4. Comparison of *in vivo* footprinting results of a deletion and a parental strain. (a) *In vivo* footprinting analysis of a *xyn1* URR (–388- to –417-bp upstream from ATG) covering a Cre1-binding site (underlined in blue) and a Xyr1-binding site (underlined in red) was performed. (b) The non-coding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

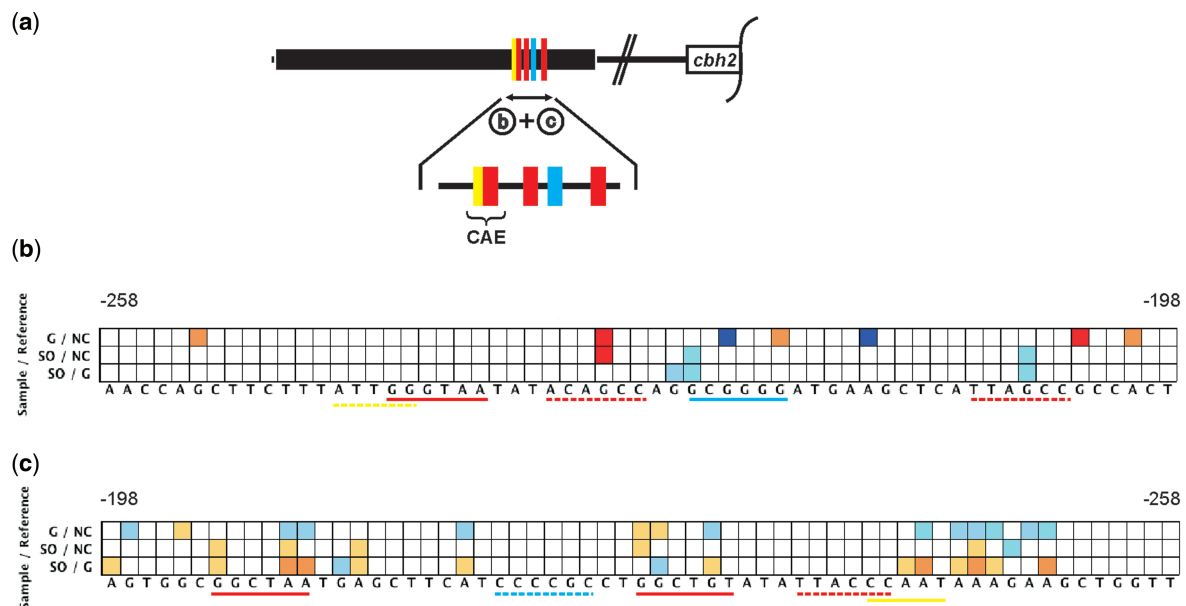


Figure 5. *In vivo* footprinting analysis of the *T. reesei* *cbh2* URR. (a) A *cbh2* URR covering the cellulase activating element (CAE) comprising a CCAAT-box (yellow) and a Xyr1-binding site (red), two additional Xyr1-binding sites and a Cre1-binding site (blue) was investigated (–258- to –198-bp 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

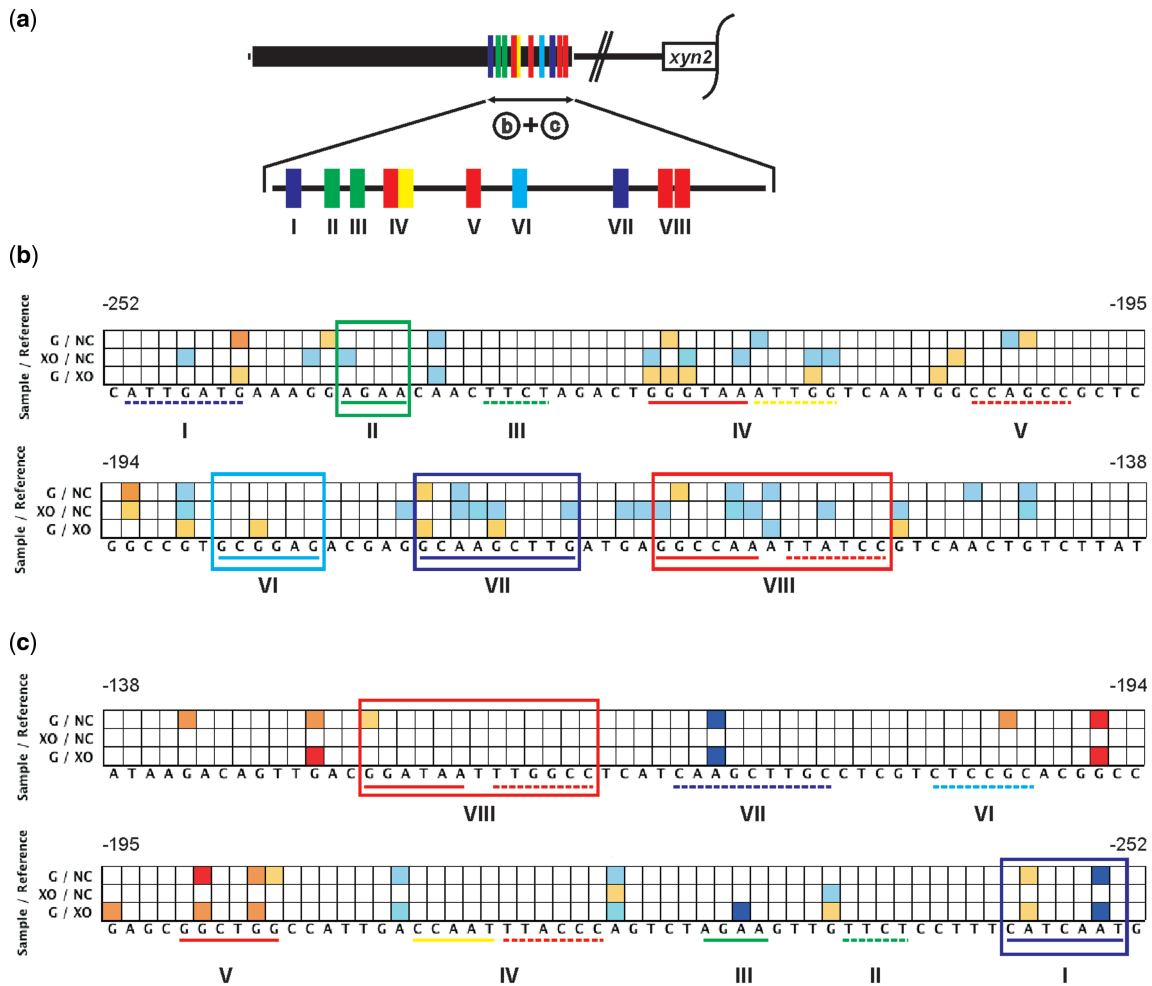


Figure 6. *In vivo* footprinting analysis of the *T. reesei* *xyn2* URR. (a) A *xyn2* URR showing a high number of cis elements [AGAA-box (green), CCAAT-box (yellow), Xyr1-binding site (red) and Cre1-binding site (blue)] was investigated (−252- to −138-bp upstream from ATG). The coding strand (b) and 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 (CAAGCTTG) overlapping with an Acel1-binding site [5'-AGGCA-3', (48)] (Figure 6b, c, VII). Acel 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.

FUNDING

Austrian Science Fund FWF [P20192 to R.L.M., V232-B20 to A.R.M.-A.]; Vienna University of Technology [AB-Tec doctoral program]; Iogen Corp. (to R.L.M.). Funding for open access charge: Austrian Science Fund FWF.

Conflict of interest statement. None declared.

REFERENCES

- Narlikar, L. and Ovcharenko, I. (2009) Identifying regulatory elements in eukaryotic genomes. *Brief. Funct. Genomic Proteomic*, **8**, 215–230.
- Dynan, W.S. and Tjian, R. (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature*, **316**, 774–778.
- Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) Regulation of inducible and tissue-specific gene expression. *Science*, **236**, 1237–1245.
- 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.
- Jackson, P.D. and Felsenfeld, G. (1987) *In vivo* footprinting of specific protein-DNA interactions. *Methods Enzymol.*, **152**, 735–755.
- 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.
- Zinn, K. and Maniatis, T. (1986) Detection of factors that interact with the human beta-interferon regulatory region *in vivo* by DNase I footprinting. *Cell*, **45**, 611–618.
- 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.
- Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, **40**, 767–774.
- 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.
- Mueller, P.R. and Wold, B. (1989) *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science*, **246**, 780–786.
- Drouin, R., Therrien, J.P., Angers, M. and Ouellet, S. (2001) *In vivo* DNA analysis. *Methods Mol. Biol.*, **148**, 175–219.
- 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.
- 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.
- Gilmour, D.S. and Fan, R. (2009) Detecting transcriptionally engaged RNA polymerase in eukaryotic cells with permanganate genomic footprinting. *Methods*, **48**, 368–374.
- 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.
- 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.
- Hartl, L. and Seefelder, S. (1998) Diversity of selected hop cultivars detected by fluorescent AFLPs. *Theor. Appl. Genet.*, **96**, 112–116.
- 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., Wacenovskiy, 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., Wacenovskiy, 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.

Supplementary figure legends

Supplementary figure S1. An example of a CGE result provided by custom service as *.fsa-file. (a) CGE results of the *cbh2* coding strand following *in vivo* footprinting after replacement on glucose for 3 h (blue), GeneScan 500-ROX internal size standard (red). (b) A zoom into the region of interest including the CAE element is given. The peaks are labelled with the corresponding sequence.

Supplementary figure S2. Analysis of transcript level ratios. *T. reesei* was pre-cultured on glycerol and was thereafter transferred to MA medium without a carbon source (NC), or to MA medium containing glucose (G) as the sole carbon source or to MA media containing sophorose (SO) or D-xylose (XO) as an inducer. Incubation was performed for 3 or 5 h. Transcript levels of the genes *xyn1* (a), *cbh2* (b), and *xyn2* (c) were normalized against *act* and *sar1* transcript levels, and calculated using REST 2009. Transcript level ratios always refer to the sample without carbon source, which is indicated by an asterisk. Results are given in logarithmic scale (lg). The values are means of two independent biologic experiments measured in triplicates. Error bars indicate standard deviations.

Supplementary figure S3. *In vivo* footprinting analysis of the *T. reesei xyn1* 5' upstream region. (a) A *xyn1* fragment covering a non-regulatory region (white bar) and a functional regulatory region (grey bar) bearing five already characterised, functional regulatory elements, *i.e.* a CCAAT-box (yellow), two Xyr1-binding sites (red), and two Cre1-binding sites (blue) (shown in lane SA) was investigated. 5'-UR, 5' upstream region; DA, regulatory region identified by promoter deletion analyses (34). SA, functional binding sites identified by site mutagenesis (29); AF, fragment analyzed during the present study (indicated by a black arrow). (b) The non-coding strand (-270 to -482 bp upstream from ATG) of the *T. reesei* parental strain cultivated on glucose (G) or D-xylose (XO) followed by DMS-induced *in vivo* methylation and naked DNA as a reference (ND) was analyzed. Grey arrow indicates the border between the previously identified non-regulatory and the regulatory region. For comparison to naked DNA the gradual visualization was set as follows: protected bases are highlighted in red shades and hypersensitive bases are highlighted in blue shades; 2.5-fold to 5.0-fold difference between compared conditions is shown in light shaded colour, 5.0-fold to 7.5-fold difference between compared conditions is shown in middle shaded colour, and more than 7.5-fold difference between compared conditions is shown in dark shaded colour. For comparison of *in vivo* footprinting samples the gradual visualisation kept at default settings (see manuscript and legend to main figure 1).

Additionally detected signals could mostly be assigned to putative regulatory elements *i.e.* two Cre1-binding sites (blue), an Ace1-site (orange) (48), and an AbaA-binding element (lavender) previously identified to be responsible for the determination of conidiophore development in *A. nidulans* (50). The latter finding is in good accordance with recent reports on expression of cellulases and hemicellulases as a major event during conidia development in *T. reesei* (51).

Supplementary figure S4. *In vivo* footprinting analysis of the *T. reesei cbh2* 5' upstream region. (a) A *cbh2* fragment covering a non-regulatory region (white bar) and a functional regulatory region (grey bar) bearing two already characterised, functional regulatory elements, *i.e.* a CCAAT-box (yellow) and

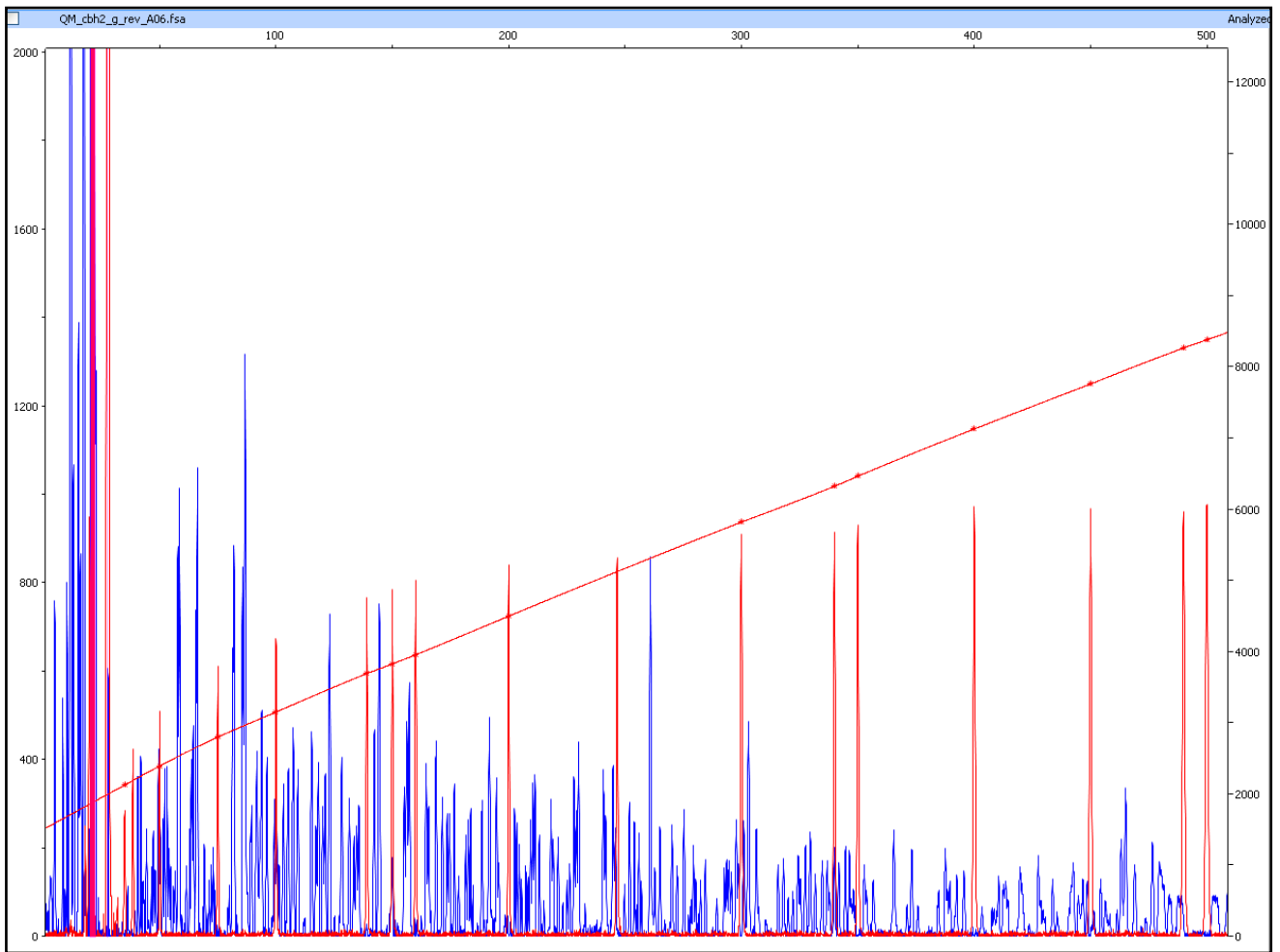
a Xyr1-binding site (red) (shown in lane SA) was investigated. 5'-UR, 5' upstream region; DA, regulatory region identified by promoter deletion analyses (49), EMSA, region involved in DNA-protein complex formation according to EMSA analyses (49); SA, functional binding sites identified by site mutagenesis (31); AF, fragment analyzed during the present study (indicated by a black arrow). (b) The coding strand (-263 to -148 bp upstream from ATG) of the *T. reesei* parental strain cultivated on glucose (G) or sophorose (SO) followed by DMS-induced *in vivo* methylation and naked DNA as a reference (ND) was analyzed. Grey arrows indicate the start and the end of the previously identified regulatory region. For settings of the gradual visualization see legend to Fig. S3. Additionally detected signals could be assigned to putative regulatory elements, *i.e.* three Xyr1-binding sites and one Cre1-binding sites (blue).

Supplementary figure S5. *In vivo* footprinting analysis of the *T. reesei xyn2* 5' upstream region. (a) A *xyn2* fragment covering a non-regulatory region (white bar) and a functional regulatory region (grey bar) bearing four already characterised, functional regulatory elements, *i.e.* one AGAA-box (green), two Xyr1-binding sites (red), and a CCAAT-box (yellow) (shown in lane SA) was investigated. 5'-UR, 5' upstream region; DA, regulatory region identified by promoter deletion analyses (41); SA, functional binding sites identified by site mutagenesis (30,46); AF, analyzed during the present study (indicated by a black arrow). (b) The coding strand (-260 to -82 bp upstream from ATG) of the *T. reesei* parental strain cultivated on glucose (G) or D-xylose (XO) followed by DMS-induced *in vivo* methylation and naked DNA as a reference (ND) was analyzed. Grey arrows indicate the start and the end of the previously identified regulatory region. For settings of the gradual visualization see legend to Fig. S3. Additionally detected signals could mostly be assigned to putative regulatory elements, *i.e.* two Xyr1 sites, a Cre1-binding site (blue). A putative TATA-box (black) yielded strong signals. The two newly identified motives (purple) could be detected, one of which bearing an unusual TCAAT-box, the other represents an octameric palindrome overlapping with an Ace1-site (48).

Supplementary references

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.
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.
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.
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.
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.

Fig. S1
a



b

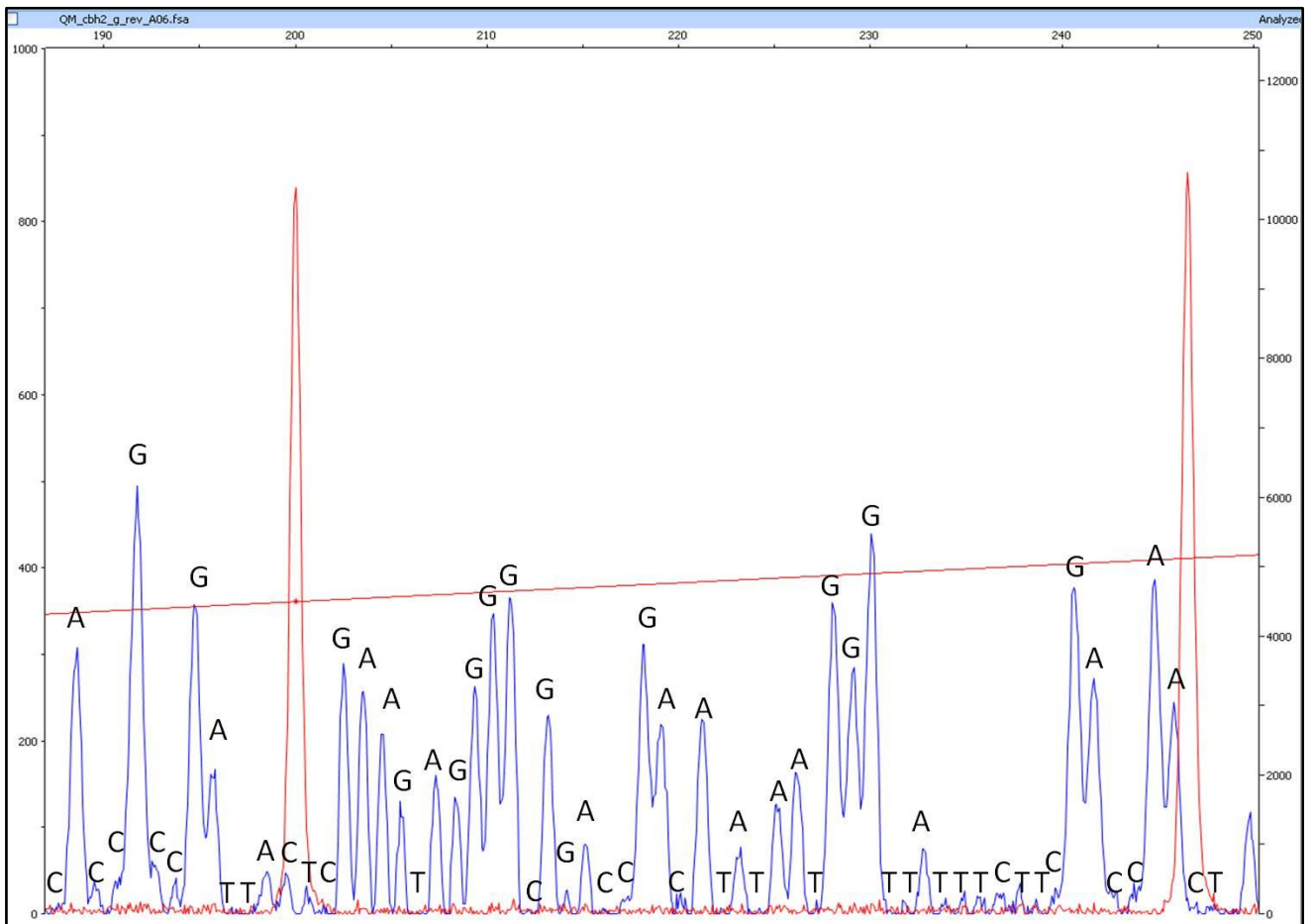
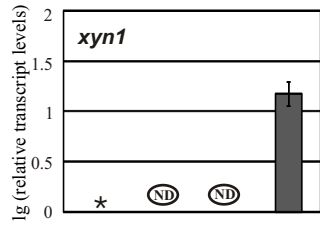
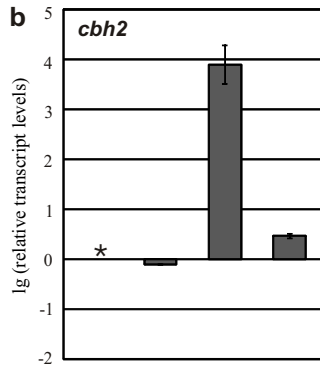


Fig. S2

a



b



c

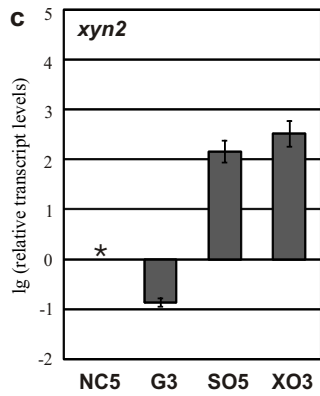


Fig. S3

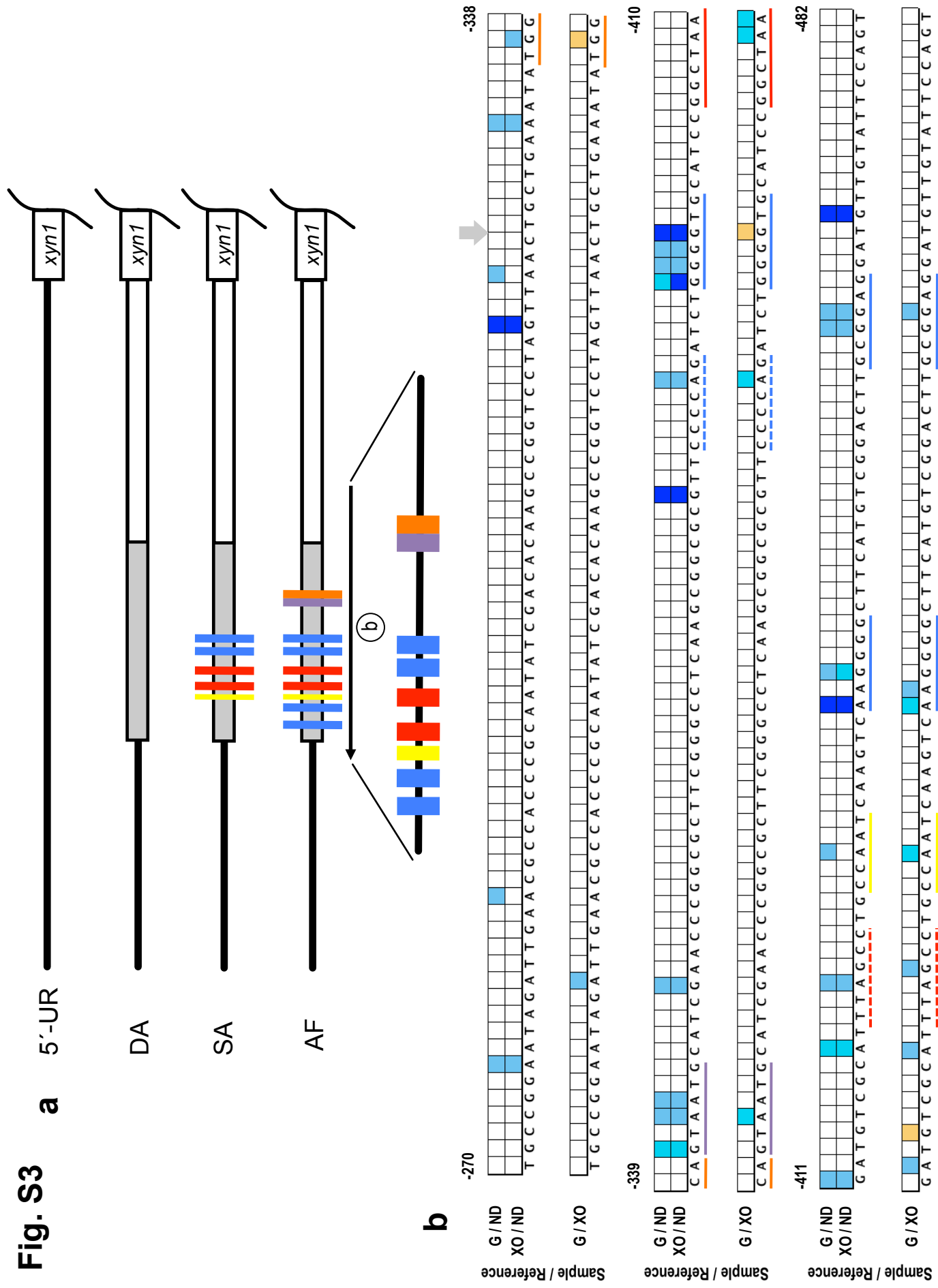
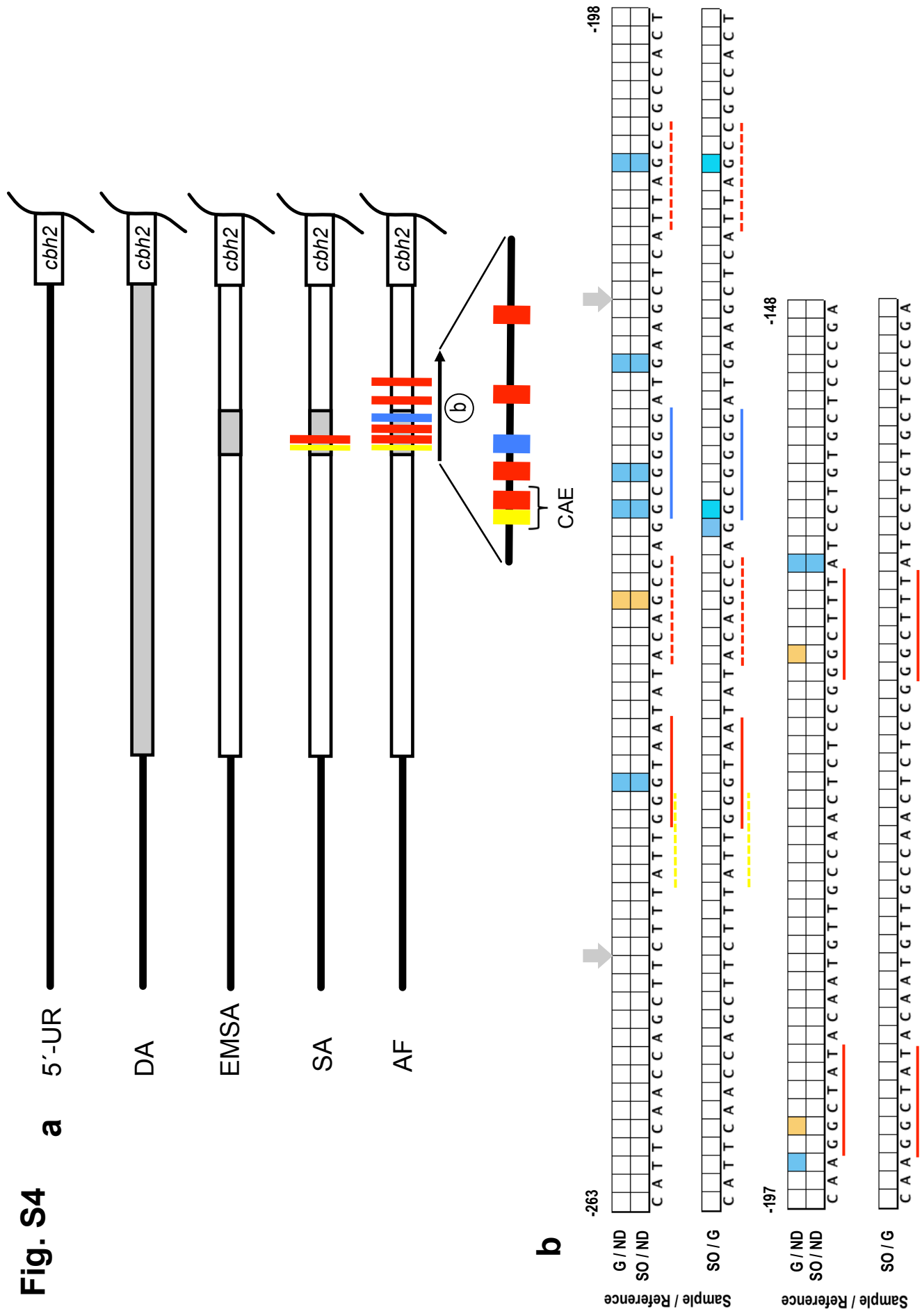


Fig. S4





TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

ivFAST

Contents

1	What is ivFAST?	2
2	How does ivFAST work?	2
3	How to use ivFAST	4
3.1	Running the program	4
3.2	File specifications	5
3.2.1	Input files	5
3.2.2	Output files	6
3.2.3	Configuration file	6
3.3	Manual adjustments	7
4	Credits	8

Copyright notice: This program is open source. It incorporates the JHeatChart library (<http://www.javaheatmap.com>), which is under LGPL licence.

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

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

the sample variance

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

and the confidence interval for the true mean μ_p

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

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 \bar{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 $\bar{x}_p(S)$ and $\bar{x}_p(R)$ can be said to be different. As criterion for this non-overlapping confidence intervals of the sample means

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

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

5. Output of the result

The values of the calculated quotients $\frac{\bar{x}_p(S)}{\bar{x}_p(R)}$ 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).

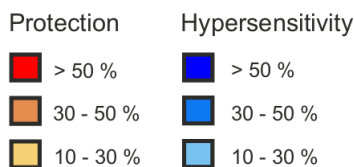


Figure 1: Legend heatmap

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 sub-directory 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 $\langle f|r \rangle$ has to contain either 'f', if the forward primer was used for the analysis, or 'r', 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 $\langle \text{CONDITION} \rangle$ 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 $\langle \text{REPLICATE} \rangle$ 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 $\langle \text{CONDITION} \rangle$, whose names only differ in the field $\langle \text{REPLICATE} \rangle$. Any characters except '-' can be used in this field.

The field $\langle a|p|s \rangle$ 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:

```

validBases=AG
primerLength=40
probability=0.95
lowerRange=0.1
middleRange=0.3
upperRange=0.5

```

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 delimiter. 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

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

(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)

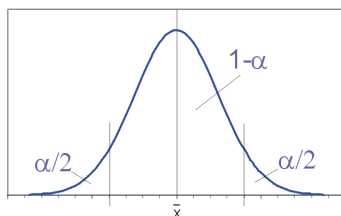


Figure 2: Confidence interval of \bar{x} (sample mean) with probability α . $1-\alpha$ and $\alpha/2$ denote the portion of the area included in the according region. (Source: Philipendula at the German language Wikipedia, GNU-FDL)

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 $\text{lowerRange} < \text{middleRange} < \text{upperRange}$.

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)	$S/R < 1 \div upperRange$
middle red	(230,140,80)	$1 \div upperRange \leq S/R < 1 \div middleRange$
light red	(245,208,115)	$1 \div middleRange \leq S/R < 1 \div lowerRange$
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 detect such possible mapping failures and produce a warning message on the command-line. But the program is not able to correct these issues automatically, since it cannot judge whether it is a real failure or not. Therefore the user has to correct the according size values in the sequencing 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.

Characterization of a nuclear receptor-like domain of the Xylanase regulator 1, a Gal4-like zinc binuclear cluster regulatory protein*

Thiago M. Mello-de-Sousa^{1#}, Rita Gorsche^{1#}, Birgit Jovanovic¹, Martina Marchetti-Deschmann¹, Marcio J. Poças-Fonseca², Robert L. Mach¹, and Astrid R. Mach-Aigner¹

¹From the Department of Biotechnology and Microbiology, Institute of Chemical Engineering, and the Division of Instrumental Analytical Chemistry, Institute of Chemical Technologies and Analytics, Vienna University of Technology, Vienna, Austria

²Department of Genetics and Morphology, Institute of Biological Sciences, University of Brasilia, Brasília-DF, Brazil

[#]These authors contributed equally to this article.

Running title: *Characterization of a nuclear receptor-like domain of Xyr1*

To whom correspondence should be addressed: Astrid R. Mach-Aigner, Department of Biotechnology and Microbiology, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Str. 1a, Vienna, Austria Tel.: +43 1 58801 166558, Fax: +43 1 58801 17299, E-mail: astrid.mach-aigner@tuwien.ac.at

Keywords: nuclear receptor-like domain; Gal-4 like, eukaryotic transactivator; Xylanase regulator 1; circular dichroism; *Trichoderma*

Background: A single point mutation in a Gal4-like transactivator leads to a glucose-blind (hemi)cellulase expressing phenotype in an industrially relevant fungus.

Results: The mutation in the transactivator leads to a loss of allosteric response to carbohydrates resulting in deregulation.

Conclusion: The mutation affects a nuclear receptor-like domain.

Significance: Understanding how signals are mediated is crucial for targeted transcription factor engineering.

SUMMARY

Engineering transcription factors is an interesting research target that gains increasing attention in case of industrially used organisms. With respect to sustainability, biomass-degrading saprophytic fungi such as *Trichoderma reesei* are promising industrial work horses because they exhibit a high secretory capacity of native and heterologously-expressed enzymes and compounds. A single point mutation in the main transactivator of xylanase and cellulase expression in *T. reesei*, Xyr1, led to a strongly deregulated and enhanced xylanase expression. Circular dichroism spectroscopy revealed a change in secondary structure caused by this mutation. According to

electrophoretic mobility shift assays and determination of the equilibrium binding constants, the DNA-binding affinity of the mutated Xyr1 was considerably reduced compared to the wild-type Xyr1. Both techniques were also used to investigate the allosteric response to carbohydrates (D-glucose and D-xylose), which are known to signal repressors or inducers of Xyr1 target genes. The mutated Xyr1 no longer exhibited a conformational change in response to these carbohydrates, indicating that the observed deregulation is not a simple matter of a change in DNA-binding of the transactivator. This was further reflected by a proteolytic digestion assay yielding carbohydrate-specific structural differences only for the wild-type Xyr1. Altogether, we postulate that the part of Xyr1, where the mutation is located, functions as a nuclear receptor-like domain that mediates carbohydrate signals and modulates the Xyr1 transactivating activity.

The filamentous ascomycete *Trichoderma reesei* (teleomorph *Hypocrea jecorina*, (1)) is a saprophyte that is industrially exploited for its ability to secrete vast amounts of cellulolytic and hemicellulolytic enzymes. Among those are the two main cellobiohydrolases CBHI and CBHII (EC 3.2.1.91) (2), and the two main endo- β -1,4-xylanases XYNI and XYNII (EC.3.2.1.8) (3).

Together with other enzymes they degrade cellulose and xylan to yield oligo- and monosaccharides like D-glucose, D-xylose, and the transglycosylation product sophorose among others (4). These in turn affect the expression of these cellulolytic and hemicellulolytic enzymes (reviewed by *e.g.* (5,6)).

The Gal4-like Zn₂Cys₆ binuclear cluster protein Xylanase regulator 1 (Xyr1) works as a wide domain activator for the production of all major cellulose- and hemicellulose-degrading enzymes. Accordingly, it was shown that a *xyr1* deletion mutant no longer expresses cellulases or xylanases (7,8). Furthermore, the transcript formation of the major cellulase-encoding genes *cbh1* and *cbh2*, encoding CBHI and CBHII respectively, directly correlates with the transcript pattern of *xyr1*, encoding their activator Xyr1 (9,10). In particular this becomes obvious as both the regulator and these target genes are induced by sophorose and repressed by glucose (10,11). In contrast, the xylanase-encoding genes *xyn1* and *xyn2* are induced by low D-xylose concentrations and repressed by glucose or high D-xylose concentrations (*e.g.* (12,13)), and their expression does not precisely correlate with *xyr1* transcript levels. This suggests that the expression of cellulases is strongly dependent on the concentration of available Xyr1, while the expression of xylanases also depends on additional mechanisms (10).

Derntl and co-workers reported an industrially used *T. reesei* mutant strain, which exhibits a strongly enhanced and de-regulated xylanase expressing phenotype (10). This strain bears a point mutation that results in an A824V transition in the so-called fungal transcription factor regulatory middle homology (FTFRMH³) region of Xyr1 (according to DELTA-BLAST, <http://www.ncbi.nlm.nih.gov/>). The introduction of this mutation into the parental strain and the reconstitution of the wild-type *xyr1* in the mutant strain demonstrated that the xylanase de-regulated phenotype is directly related to this mutation (see Fig. 1A). *In silico* secondary structure prediction of Xyr1 finds A824 in the middle of an α -helix, which suggests that the A824V mutation might possibly lead to a change in secondary structure, which could be responsible for the observed deregulation of xylanase expression and the generally enhanced cellulase expression (10).

In case of the orthologue of Xyr1 in *Aspergillus niger*, XlnR (14,15), mutants were investigated bearing the following changes: the

mutation V756F and the introduction of stop codons after residues L668 and G797. All modifications resulted in higher xylanolytic activity of the fungus as well as an increased expression of target genes of XlnR, such as *xlnB* (coding for endoxylanase B) even in repressing conditions using D-glucose as the sole carbon source (16). The region that contains these mutations is located between a putative coiled-coil region and a C-terminal transcriptional activation domain and is probably related to modulation of activity of XlnR by D-glucose (16).

Circular dichroism (CD) spectroscopy is very useful for the characterisation of the protein secondary structure (in % α -helix, β -sheets, turns), as well as structure changes resulting from DNA-protein, protein-protein or protein-ligand interactions (*e.g.* reviewed by (17,18)). Especially in the case of transcription factors this method is used to investigate the influence of DNA-binding on the protein and DNA structure, an example of which is the investigation into structural changes of the transcription factor NF- κ B induced by DNA-binding (19).

In this study, we use CD spectroscopy, electrophoretic mobility shift assay (EMSA), and proteolytic digestion assay to investigate the influence of the previously described A824V mutation on the Xyr1 secondary structure and its DNA-binding affinity. Finally, the response of Xyr1 variants to carbohydrates in form of conformational change mediating different signals, which affect (hemi)cellulase expression (D-glucose and D-xylose), are investigated concerning the putative action of Xyr1 as nuclear receptor (NR).

EXPERIMENTAL PROCEDURES

Fungal strains – Throughout this study Iogen-M4 (parental strain) (10,20), Iogen-M8 (mutant strain), which was obtained by UV mutagenesis of Iogen-M4 (10), Iogen-M4X, which is a derivative of Iogen-M4 bearing an introduced point mutation (A824V) in Xyr1 (10), and Iogen-M8X, which is a derivative of Iogen-M8 bearing a reconstituted wild-type Xyr1 (10). All strains were maintained on malt extract agar or potato-dextrose-agar.

Growth conditions – For determination of xylanolytic properties, growth on xylan plates was performed using Mandels-Andreotti (MA) medium (21) containing 0.2 % (w/v) azo-xylan (Megazym, Wicklow, Ireland) at 30 °C for 72 h followed by a 2 h incubation at 50 °C.

For carbon source replacement experiments, mycelia were pre-cultured in 1 L Erlenmeyer flasks on a rotary shaker (180 rpm) at 30 °C for 18 h in 300 mL of MA medium containing 105 mM of glycerol as the sole carbon source. A total of 10^9 conidia per litre (final concentration) was used as inoculum. Pre-grown mycelia were washed, and equal amounts were resuspended in 20 mL MA media containing 1 % (w/v) D-xylose or D-glucose, or 1.5 mM sophorose. Mycelia were also grown in MA media without any carbon source (control). Samples were taken after 6 hours of incubation from three biological replicates and were pooled before RNA extraction.

Analysis of transcript levels – Fungal mycelia were homogenized in 1 mL of peqGOLD TriFast DNA/RNA/protein purification system reagent (PEQLAB Biotechnologie, Erlangen, Germany) using a FastPrep(R)-24 cell disrupter (MP Biomedicals, Santa Ana, CA, USA). RNA was isolated according to the manufacturer's instructions, and the concentration was measured using the NanoDrop 1000 (Thermo Scientific, Waltham, US).

Synthesis of cDNA from mRNA was carried out using the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, USA) according to the manufacturer's instructions.

Quantitative PCRs were performed in a Mastercycler® ep realplex 2.2 system (Eppendorf, Hamburg, Germany). All reactions were performed in triplicate. The amplification mixture (final volume 25 µL) contained 12.5 µL 2 x iQ SYBR Green Mix (Bio-Rad, Hercules, USA), 100 nM forward and reverse primer and 2.5 µL cDNA (diluted 1:100). Primer sequences are provided in Table 1. Cycling conditions and control reactions were performed as described previously (22). Data normalization using *sar1* and *act* as reference genes and calculations were performed as published previously (22). All transcript levels are referred to the wild-type incubated without carbon source for 3 h.

Plasmid construction – A 2.945 bp-fragment comprising the T7 promoter, the *lac* operator, and the coding region of *xyl1* fused with a C-terminal six-histidine tag was chemically synthesized (GeneArt®, part of Life Technologies, Paisley, UK) with codon optimization for *Escherichia coli*. This expression cassette was cloned into the Novagen® pET28a vector (Merck, Darmstadt, Germany) using *BglII/NotI* restriction sites and yielding pTS1. Likewise, the mutated gene

version, *xyl1*_{A824V}, bearing the A824V transition, was chemically synthesized and cloned into pET28a, yielding pTS2. Insertion of the correct fragments into both expression vectors was confirmed by restriction profile analysis and automated sequencing (LGC Genomics, Berlin, Germany).

Protein expression and purification – 300 mL LB medium with D-glucose (1 % w/v) and kanamycin (50 µg/mL) were inoculated with *E. coli* BL21(DE3)pLysS (Promega, Madison, WI, USA) carrying the respective expression vectors. At OD₆₀₀ protein expression was induced by adding IPTG to a final concentration of 0.5 mM. The culture was incubated at 18 °C for 24 h. The cells were harvested by centrifugation and stored frozen at -20 °C overnight. Cells were then resuspended in 10 mL binding buffer (0.5 M NaCl, 20 mM Tris-HCl, 5 mM imidazole, pH 7.9) and sonicated using a Sonifier® 250 Cell Disruptor (Branson, Danbury, USA) (power 40 %, duty cycle 70 %, power for 30 s, pause for 30 s, 4 cycles). After centrifugation the protein was purified from the extract using Novagen® HisBind® resin (Merck) and a modified elution buffer (0.5 M NaCl, 20 mM Tris-HCl, 120 mM imidazole, pH 7.9) according to manufacturer's guidelines.

Circular dichroism spectroscopy – Before CD spectroscopy measurements PD-10 columns (GE Healthcare, Uppsala, Sweden) were used for buffer exchange to buffer A (50 mM Tris, 200 mM NaCl, 50 mM NaH₂PO₄, 10 % (v/v) glycerol, pH 7.5). The protein samples were centrifuged at 14000 x g for 10 min to remove aggregates and thereafter protein concentration was determined using Bio-Rad Protein Assay (Bio-Rad). 200 µl of a 200 - 250 nM protein solution were used for CD spectroscopy measurements. To study DNA-binding to the upstream regulatory region (URR) of *xyn1*, synthetic complementary oligonucleotides (Sigma-Aldrich, St. Louis, MO, USA) were annealed and used at a final concentration of 75, 150, 225, 300, 375, 450, 525, 600, 675, 750, 825, and 900 nM. Oligonucleotide sequences are provided in Table 1. To study the influence of D-glucose and D-xylose, the sugars were used at final concentrations (or sugar to protein ratios, respectively) of 40 (1:5), 200 (1:1) and 1000 nM (5:1). Measurements were carried out in 0.2 cm SUPRASIL® quartz cells (HellmaAnalytics, Müllheim, Germany) in a J-815 CD Spectrometer (Jasco, Tokyo, Japan) at 22 °C. CD spectra of the proteins were collected from 260 – 200 nm as average of 5 scans and baseline

subtracted in order to exclude buffer influences. Data are presented as the mean residue ellipticity $[\theta]$ in $\text{deg cm}^2 \text{dmol}^{-1}$, *i.e.* (millidegrees \times MRW)/(pathlength in mm \times concentration in mg/ml), where MRW (mean residue weight) is 109.69 Da. Prediction of protein secondary structure was done using the K2D algorithm (23). The equilibrium binding constants (Kd) of the DNA-protein complexes were determined by following the changes of $[\theta]_{222}$, using the non-linear least squares method to fit a curve based on Engel's equation (24) with $n=2$ to the measured data.

Proteolytic digestion assay – D-glucose or D-xylose were added to 20 μL of a 800 nM protein solution to yield a final sugar concentration of 20 μM (1:25 protein to sugar ratio) in a final volume of 21 μL and incubated at 22 $^\circ\text{C}$ for 10 min. Chymotrypsin was added to final concentrations (or protein to chymotrypsin ratios, respectively) of 32 (25:1), 16 (50:1), 8 (100:1) and 4 (200:1) nM in a final volume of 22 μL and incubated another 10 min at 22 $^\circ\text{C}$. The reaction was stopped by adding 6 μL of 4x Laemmli Sample Buffer (Bio-Rad) and kept at 95 $^\circ\text{C}$ for 10 min.

Electrophoretic mobility shift assay - Synthetic FAM-labeled oligonucleotides (Sigma-Aldrich) used for EMSA were annealed with their complementary oligonucleotides (Table 1). The protein-DNA binding assay and non-denaturing polyacrylamide gel electrophoresis were performed essentially as described previously (25). Binding was achieved by incubating 350 ng of the heterologously expressed Xyr1 or Xyr1_{A824V} with 15 ng of a labeled, double-stranded DNA fragment in buffer A (10 min, 22 $^\circ\text{C}$). Fluorescence and image analysis of the gels was carried out using a Typhoon 8600 variable mode imager (Amersham Bioscience, part of GE Healthcare, CT, USA).

RESULTS

A single point mutation in the zinc finger regulatory protein Xyr1 changes its secondary structure - In *T. reesei*, the expression of genes encoding (hemi)cellulases is regulated by the Gal4-like transactivator Xyr1 (7,8). Recently, it was demonstrated that a point mutation in Xyr1 (A824V) leads to a glucose blind, (hemi)cellulase expressing phenotype (10). As it can be inferred from RT-qPCR results (Fig. 1B), the transcript levels of the genes encoding the main endoxylanases (*xyn1* and *xyn2*) and the most prominent cellobiohydrolase (*cbh1*) are strongly elevated in the mutant strain (bearing

the A824V transition) compared to the parental strain, particularly under conditions that are considered to be non-inducing (no carbon source) or repressing (D-glucose). Generally, transcript formation in the mutant strain seems to be carbon source-independent or non-responsive (Fig. 1B). Derntl and co-workers reported that the domain bearing the mutation is a FTFRMH region (according to DELTA-BLAST, <http://www.ncbi.nlm.nih.gov/>) (10). More precisely, the mutation is located in the α -helices-rich, C-terminal part of the domain leading to structural changes of the protein. In order to see if a change in structural conformation is associated with the A824V mutation, the wild-type Xyr1 and the Xyr1_{A824V} were heterologously expressed in *E. coli*. After the expression of the correct proteins was confirmed by MS analysis, both wild-type Xyr1 and Xyr1_{A824V} were analyzed by CD spectroscopy. The spectrum of the negative ellipticity strongly differed between the wild-type Xyr1 and the Xyr1_{A824V} (Fig. 2) demonstrating that these proteins vary in their secondary structure. The use of the K2D algorithm (23) estimated an α -helix content of 35%, β -sheet of 17%, and random coil of 48% for the wild-type Xyr1. This correlates well with the *in silico* prediction made with SOPMA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=NPSA/npsa_sopma.html) (26), which indicated a 32.5% α -helix content. The mutation A824V resulted in a 46% decrease in mean residue ellipticity at 222 nm and a strong decrease of the fraction of residues involved in helical conformation, hence indicating a possible role of this region for a specific folding of Xyr1.

The A824V mutation changes DNA-binding properties of Xyr1 - To investigate whether the change in secondary structure of Xyr1 caused by the A824V mutation also affects its DNA-binding properties, we analyzed Xyr1-binding to the URR of the *xyn1* gene, which is a well-studied target gene within the Xyr1 regulon. For this purpose, CD spectroscopy analysis was used again. In the presence of the double-stranded DNA probe, we found indication for protein-DNA interactions for both proteins, reflected by changes in the spectra of the protein alone compared to protein with target DNA. These changes indicate a change in secondary structure content (Figure 3A, B). The observed reduction of the negative ellipticity due to potential DNA interaction resulted in a 21% loss of α -helical content for the wild-type Xyr1 and a 60% loss

for the Xyr1_{A824V} according to calculations using the K2D algorithm.

As CD spectroscopy analyses are not a direct proof for a protein-DNA complex formation, we performed EMSA analysis employing the same fragment of the upstream regulatory region (URR) of *xyn1* as fluorescently labelled probe. Using different amounts of wild-type Xyr1, two shifts in mobility could be observed at lower concentrations of the protein (Fig. 3C). This observation is in accordance with previous findings that reported binding of Xyr1 to the *xyn1* URR as monomer and as dimer (27) yielding two shifts in EMSA experiments. Application of increasing amounts of protein favours binding as a dimer in a concentration-dependent way (Fig. 3C, compare (27)). Using Xyr1_{A824V} yielded only a weak double shift (monomer and dimer) at the highest protein concentration (Fig. 3C) indicating a decrease in the protein-DNA complex formation due to the A824V mutation.

Determination of the equilibrium dissociation constant of the Xyr1 variants – At the first glance the decreased ability of Xyr1_{A824V} to form complexes with *xyn1* URR contradicts the superior xylanase formation observed in the *T. reesei* strain bearing this mutation. Therefore, the equilibrium dissociation constants of both wild-type Xyr1 and Xyr1_{A824V} were determined by CD spectroscopy following the changes of $[\theta]_{222}$. For this purpose, a constant amount of protein with varying amounts of target DNA was analysed by CD spectroscopy following the changes of the mean residue ellipticity at 222 nm. Then, a non-linear least squares method was used to fit a curve based on Engel's equation (24) to the measured data. Figure 4 depicts the curve fitting results of both Xyr1 variants. The higher K_d value of Xyr1_{A824V} (523 nM) (Fig. 4B) in comparison to the one of the wild-type Xyr1 (128 nM) (Fig. 4A) indicates a lower DNA-binding affinity for Xyr1_{A824V} and supports the former findings of this study. As this is a rather unexpected finding in case of a transactivator's mutation, which caused an enhanced deregulated xylanase production in the respective strain, we assume that the regulatory mechanism is not a simple matter of change in Xyr1 DNA-affinity.

Conformational response to carbohydrates is partly lost in Xyr1_{A824V} – As it is known that in eukaryotes the presence of a certain metabolic signal can cause conformational changes in a regulatory protein, this prompted us to investigate possible allosteric modulations of Xyr1 induced by certain carbohydrates.

Therefore, we analyzed the protein both alone and in presence of D-glucose (repression of Xyr1 target genes and *xyr1*) or D-xylose (concentration-dependent induction of Xyr1 target genes) by CD spectroscopy measurements. Both carbohydrates cause a change in the Xyr1 secondary structure, of which the extend directly correlates with the amount of carbohydrate present (Fig. 5A, B). Generally, the decrease in α -helix content was similar to that observed for Xyr1 interacting with DNA. Notably, the greater loss of α -helix structures occurred in the presence of D-xylose (up to 33% using the K2D algorithm) compared to D-glucose (22%) (also compare Fig. 5A and B). This suggests that Xyr1 assumes different conformations depending on the carbohydrate used. Again, as a complementing experiment EMSA analysis using the *xyn1* URR fragment as probe and the same ratios of protein to carbohydrate as for the CD spectroscopy analyses was performed. Interestingly, the induced changes in secondary structure do not interfere with the ability to form protein-DNA complexes as exactly the same shifts are obtained, regardless if carbohydrate is added and in which carbohydrate-to-protein ratio (data not shown).

The same assay was applied to the Xyr1_{A824V} and analyzed by CD spectroscopy. Compared to the wild-type Xyr1, in this case a higher ratio of carbohydrate to protein needed to be applied to achieve changes in secondary structure (Fig. 5C, D), while no different results were obtained for D-glucose and D-xylose (compare Fig. 5C and D). These findings support the carbon source-independent expression of (hemi)cellulases in the respective *T. reesei* mutant strain. However, here again, the observed changes in secondary structure at a high carbohydrate-to-protein ratio did not affect DNA-protein complex formation according to EMSA analysis (data not shown) pointing to other regulatory mechanism(s) than mere DNA-binding.

Proteolytic digestion reveals a carbohydrate-dependent change in Xyr1 protein structure – As described above the most striking finding during CD spectroscopy analyses of Xyr1 variants in presence of carbohydrates was a clear decrease in α -helix content of the wild-type Xyr1 in the presence of D-xylose. In order to further support the observed conformational change, proteolytic digestion of both Xyr1 variants was conducted in presence of D-glucose or D-xylose (molar ratio of protein to carbohydrate of 1:25) as an additional approach. To this end, decreasing

molar ratios of protein to chymotrypsin were applied. Strikingly, a different pattern was obtained solely for the wild-type Xyr1 comparing D-glucose and D-xylose using mild digestion conditions (the highest ratio of protein to chymotrypsin) (Fig. 6A). In contrast, the proteolytic digestion pattern of Xyr1_{A824V} did not differ between D-glucose and D-xylose regardless of the amount of chymotrypsin used (Fig. 6B). Altogether, these findings support the hypothesis that the wild-type Xyr1 undergoes conformational changes in response to different carbohydrates, which is – at least partly - lost in the Xyr1_{A824V} protein. This is in good accordance with the non-responsiveness of xylanase and cellulase expression to different carbon sources in the corresponding *T. reesei* mutant strain.

DISCUSSION

As mentioned above, a mutant of an industrially used *T. reesei* strain showed an unusual deregulated and enhanced xylanase expression profile that prompted transcriptional analyses of prominent xylanase- and cellulase-encoding genes and their common transactivator-encoding gene *xyr1* (10). In the present study we aimed to identify a mechanistic explanation for the observed phenotype by investigating the effect of the formerly identified A824V mutation in Xyr1 (10) on the protein structure. While we could generally observe a different secondary structure when comparing the CD spectrum of the wild-type Xyr1 and the Xyr1_{A824V} consistent with a considerable loss of helical content in the mutant Xyr1, we also found a reduced capability of Xyr1_{A824V} for protein–DNA complex formation. This finding was rather the opposite than what we expected because Xyr1 is the essential transactivator of xylanase expression, and the mutation led to increased xylanase expression. However, the lower DNA-binding affinity of Xyr1_{A824V} was supported by the comparison of the K_d of both the wild-type Xyr1 and the Xyr1_{A824V} for the corresponding part of the URR of *xyn1*. This let us assume that the DNA-binding affinity does not play the major role in the regulatory mechanism behind this phenotype. Of course, it remains open if DNA-binding affinity plays the major role in regulating expression of cellulase-encoding genes as they strictly follow the *xyr1* expression pattern, unlike the xylanase-encoding genes such as *xyn1*, the URR of which was used for investigations during this study. Nonetheless, an increase in expression of cellulase-encoding

genes was observed in the respective mutant strain, but the regulatory response to certain carbohydrates, in particular the induction by sophorose, remained the same (compare Fig. 1B). In contrast to that a strong deregulation (*i.e.* a general loss of repression and induction) was reported for the xylanase expression. This makes the xylanase regulon an interesting research target and prompted us to investigate a possible allosteric response of Xyr1 in an NR-like way. The modulation of the transactivating function of Zn₂Cys₆ binuclear cluster proteins in response to metabolic signals has been reported before. Wang and colleagues described such a phenomenon for the activating regulatory protein LEU3 of *Saccharomyces cerevisiae* (29,30), in which the protein undergoes a conformational change in the presence of isopropylmalate, a metabolic intermediate of the biosynthesis of leucine, and thus exposes its transcriptional activation domain (31). In case of GAL4, the member of Zn₂Cys₆ binuclear cluster proteins best studied so far, a masking of the transcriptional activation domain occurs through the regulator GAL80. In the presence of galactose a third protein, GAL3, forms a heterodimer with GAL80, thus freeing the GAL4 activation domain (reviewed by (32)). Besides a modulation of the activity of GAL4 through intermolecular interactions, there is evidence for a direct mechanism of inhibition by D-glucose. Stone & Sadowski showed that in the central region of GAL4 there is a domain responsive to D-glucose with an intramolecular role in inhibiting the transcriptional activity of the factor. In this model the hypothesis of a direct interaction of D-glucose or its metabolites triggering the conformational change in GAL4 is considered (33). In case of the closely related orthologue of Xyr1 in *A. niger*, XlnR, the proposed model for repression by D-glucose (provided by (16)) was speculated to occur through inter- or intramolecular interaction with the C-terminal region of the protein, which would result in an inactive state. It is noteworthy that Xyr1 has a high sequence identity in this region including the V756 residue of XlnR for which the mutation was described as mentioned in the introduction part. This amino acid is located at position 821 in Xyr1, close to the described mutation A824V.

Notably, we found that the allosteric response of Xyr1 to two carbohydrates, which are responsible for either induced or repressed gene expression in the parental strain, was lost in the Xyr1_{A824V}. These observations strongly suggest

that the part of the Xyr1 protein functions as a NR-like domain, which is in good accordance with the observed deregulation of xylanase expression. In our postulated model we assume that the mutation in the NR-like domain of Xyr1 leads to an altered presentation of the activation domain, independently of the carbohydrate signal. In contrast to that the intact NR-like domain in the wild-type Xyr1 would in the

simplest scenario directly interact with the carbohydrate and would according to the received signal reveal the transactivating domain (compare Fig. 7A, B). To study this more extensively, a crystallographic study of Xyr1 variants alone and in presence of the carbohydrates will be aspired.

REFERENCES

1. Kuhls, K., Lieckfeldt, E., Samuels, G. J., Kovacs, W., Meyer, W., Petrini, O., Gams, W., Borner, T., and Kubicek, C. P. (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**, 7755-7760
2. Teeri, T., Salovouri, I., Knowles, J. (1983) The molecular cloning of the major cellulase gene from *Trichoderma reesei*. *Biotechnology* **1**, 696-699
3. Törrönen, A., Harkki, A., and Rouvinen, J. (1994) Three-dimensional structure of endo-1,4-beta-xylanase II from *Trichoderma reesei*: two conformational states in the active site. *EMBO J* **13**, 2493-2501
4. Vaheri, M. P., Leisola, M., Kaupinnen, V. (1979) Transglycosylation products of the cellulase system of *Trichoderma reesei*. *Biotechnol Lett* **1**, 41-46
5. Mach, R. L., and Zeilinger, S. (2003) Regulation of gene expression in industrial fungi: *Trichoderma*. *Appl Microbiol Biotechnol* **60**, 515-522
6. 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
7. 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
8. 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*. *Eukaryotic Cell* **5**, 2128-2137
9. Portnoy, T., Margeot, A., Seidl-Seiboth, V., Le Crom, S., Ben Chaabane, F., Linke, R., Seiboth, B., and Kubicek, C. P. (2011) Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. *Eukaryotic Cell* **10**, 262-271
10. 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
11. 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
12. Mach, R. L., Strauss, J., Zeilinger, S., Schindler, M., and Kubicek, C. P. (1996) Carbon catabolite repression of xylanase I (*xynI*) gene expression in *Trichoderma reesei*. *Mol Microbiol* **21**, 1273-1281.
13. Mach-Aigner, A. R., Pucher, M. E., and Mach, R. L. (2010) D-Xylose as a repressor or inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*). *Appl Environ Microbiol* **76**, 1770-1776
14. Gielkens, M. M., Dekkers, E., Visser, J., and de Graaff, L. H. (1999) Two cellobiohydrolase-encoding genes from *Aspergillus niger* require D- xylose and the xylanolytic transcriptional activator XlnR for their expression. *Appl Environ Microbiol* **65**, 4340-4345.
15. Hasper, A. A., Visser, J., and de Graaff, L. H. (2000) The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression. *Mol Microbiol* **36**, 193-200.
16. Hasper, A. A., Trindade, L. M., van der Veen, D., van Ooyen, A. J., and de Graaff, L. H. (2004) Functional analysis of the transcriptional activator XlnR from *Aspergillus niger*. *Microbiology* **150**, 1367-1375
17. Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta* **1751**, 119-139
18. Ranjbar, B., and Gill, P. (2009) Circular dichroism techniques: biomolecular and nanostructural analyses - a review. *Chem Biol Drug Des* **74**, 101-120
19. Matthews, J. R., Nicholson, J., Jaffray, E., Kelly, S. M., Price, N. C., and Hay, R. T. (1995) Conformational changes induced by DNA binding of NF-kappa B. *Nuc Acids Res* **23**, 3393-3402

20. Hui, J., Lanthier, P., White, T., McHugh, S., Yaguchi, M., Roy, R., and Thibault, P. (2001) Characterization of cellobiohydrolase I (Cel7A) glycoforms from extracts of *Trichoderma reesei* using capillary isoelectric focusing and electrospray mass spectrometry. *J Chromatogr B* **752**, 349-368
21. Mandels, M. (1985) Applications of cellulases. *Biochem Soc Trans* **13**, 414-416
22. 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
23. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng* **6**, 383-390
24. Engel, G. (1974) Estimation of binding parameters of enzyme-ligand complex from fluorometric data by a curve fitting procedure: seryl-tRNA synthetase-tRNA Ser complex. *Anal Biochem* **61**, 184-191
25. Stangl, H., Gruber, F., and Kubicek, C. P. (1993) Characterization of the *Trichoderma reesei* *cbh2* promoter. *Curr Genet* **23**, 115-122.
26. Geourjon, C., and Deleage, G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci* **11**, 681-684
27. 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 *xynI*, encoding xylanase I, in *Hypocrea jecorina*. *Eukaryotic Cell* **5**, 447-456
28. Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J. C., Hochstrasser, D. F., and Appel, R. D. (1997) Detailed peptide characterization using PEPTIDEMASS - a World-Wide-Web-accessible tool. *Electrophoresis* **18**, 403-408
29. Friden, P., and Schimmel, P. (1987) LEU3 of *Saccharomyces cerevisiae* encodes a factor for control of RNA levels of a group of leucine-specific genes. *Mol Cell Biol* **7**, 2708-2717
30. Friden, P., and Schimmel, P. (1988) LEU3 of *Saccharomyces cerevisiae* activates multiple genes for branched-chain amino acid biosynthesis by binding to a common decanucleotide core sequence. *Mol Cell Biol* **8**, 2690-2697
31. Wang, D., Hu, Y., Zheng, F., Zhou, K., and Kohlhaw, G. B. (1997) Evidence that intramolecular interactions are involved in masking the activation domain of transcriptional activator Leu3p. *J Biol Chem* **272**, 19383-19392
32. Traven, A., Jelicic, B., and Sopta, M. (2006) Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep* **7**, 496-499
33. Stone, G., and Sadowski, I. (1993) GAL4 is regulated by a glucose-responsive functional domain. *EMBO J* **12**, 1375-1385
34. Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006) Pfam: clans, web tools and services. *Nuc Acids Res* **34**, D247-251

Acknowledgements – We thank Loreta Gudynaite-Savitch and John Tomashek for critical discussion of the experiments. We thank Theresa White for critically reading and commenting the manuscript.

FOOTNOTES

*This work was supported by two grants from the Austrian Science Fund (FWF): [P20192, P24851] given to RLM and ARMA, respectively, by Iogen Corp., by an Innovative Project (RAKI-MINT) from Vienna University of Technology to ARMA, by two doctoral programs of Vienna University of Technology (AB-Tec and CatMat), and by a fellowship from the Brazilian CNPq to TMMS.

¹Department of Biotechnology and Microbiology, Institute of Chemical Engineering, and the Division of Instrumental Analytical Chemistry, Institute of Chemical Technologies and Analytics, Vienna University of Technology, Vienna, Austria

²Department of Genetics and Morphology, Institute of Biological Sciences, University of Brasilia, Brasilia-DF, Brazil

³The abbreviations used are: CD, circular dichroism; EMSA, electrophoretic mobility shift assay; FTFRMH, fungal transcription factor regulatory middle homology region; MA, Mandels-Andreotti; NR, nuclear receptor; URR, upstream regulatory region;

FIGURE LEGENDS

FIGURE 1. Phenotype of an industrially used *T. reesei* mutant strain caused by a single point mutation in Xyr1. (A) Xylanolytic properties of the parental (boxed in yellow) and the mutant strain (boxed in red) as well as the parental strain with targeted introduction of the A824V mutation (Xyr1_{A824V}) and the mutant strain with restored wild-type Xyr1 (Xyr1_{V824A}). Grey arrow indicates strain generation by random UV-mutagenesis; black arrows indicate strain generation by targeted gene replacement. Strains were grown on plates containing 0.2 % (w/v) azo-xylan as carbon source. (B) Transcript levels of the *xyn1*, *xyn2* and *cbh1* genes of the parental (yellow bars) and the mutant strain (red bars). Both strains were pre-cultured on glycerol and thereafter transferred to MA media without carbon source (NC) or containing 1% (w/v) D-glucose (G) or D-xylose (34), or 1.5 mM sophorose (SO) and incubated for 6 h. Results are given as relative transcript ratios in logarithmic scale (lg). The values are means from three biological replicates. Error bars indicate standard deviations.

FIGURE 2. CD analyses of the wild-type Xyr1 and A824V mutant. Far-UV spectra (200 – 260 nm) of 300 nM Xyr1 (dashed line) and Xyr1_{A824V} (solid line) at 22 °C are given.

FIGURE 3. CD analyses of Xyr1 and Xyr1_{A824V} binding to target DNA. Far-UV spectra (200 – 260 nm) of 300 nM Xyr1 (A) or Xyr1_{A824V} (B) alone (dashed lines) and their respective protein-DNA complexes (solid lines) using a *xyn1* URR fragment (-430 to -396 bp from ATG) in a molar ratio of 1:1. Data were plotted after correction for the contribution of the DNA. (C) EMSA analysis of Xyr1 and Xyr1_{A824V} DNA-binding behavior using increasing amounts of protein (170 ng, 340 ng, 680 ng, and 1360 ng, respectively) and a fluorescent-labeled *xyn1* URR fragment. FP, free probe.

FIGURE 4. Determination of the equilibrium dissociation constants (K_d). Titrations were performed with 300 nM of Xyr1 (A) or Xyr1_{A824V} (B) and increasing amounts of DNA (*xyn1* URR fragment, -430 to -396 bp from ATG) ranging from 0 to 900 nM. The values are means of three independent experiments. Error bars indicate standard errors of the means.

FIGURE 5. CD analyses of Xyr1 and Xyr1_{A824V} in the presence of carbohydrates. Far-UV spectra (200 - 260 nm) of 250 nM Xyr1 (A, B) or 200 nM Xyr1_{A824V} (C, D) in presence of D-glucose (A, C) or D-xylose (B, D) applying molar ratios (protein to carbohydrate) of 1:0 (solid lines), 5:1 (short dashed lines), 1:1 (dotted lines), and 1:5 (long dashed lines). Data were plotted after correction for the contribution of the carbohydrates.

FIGURE 6. Analysis of the protein digestion patterns of Xyr1 and Xyr1_{A824V}. SDS-PAGE of the digestions of 800 nM Xyr1 (A) or Xyr1_{A824V} (B) in the presence of 20 μM D-glucose or D-xylose, respectively, using decreasing amounts of chymotrypsin (CT). The applied molar ratio of protein to CT was 25:1, 50:1, 100:1, and 200:1. PL, protein ladder for size estimation (kDa) of detected bands; red arrows indicate bands detected in a condition-specific way only in the wild-type Xyr1.

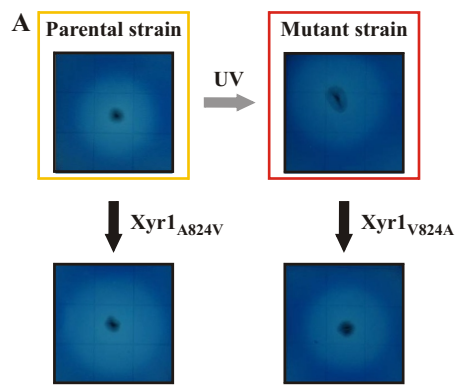
FIGURE 7. Schematic presentation of the allosteric response of Xyr1 variants. A simplified model of the secondary structure of the wild-type Xyr1 (A) and the Xyr1_{A824V} (B) without carbohydrates and their conformational change in the presence of D-xylose (X) or D-glucose (G), respectively, is presented. Yellow box, DNA-binding domain; red box, NR-like domain; red framed box, NR-like domain bearing the A824V mutation; blue box, transactivating domain.

TABLES

Table 1. Oligonucleotides used throughout the study

Name	Sequence (5' - 3')	Employment
act1f	TGAGAGCGGTGGTATCCACG	
act1r	GGTACCACCAGACATGACAATGTTG	
cbh1f	GATGATGACTACGCCAACATGCTG	
cbh1r	ACGGCACCGGGTGTGG	
xyn1f	CAGCTATTCGCCTTCCAACAC	
xyn1r	CAAAGTTGATGGGAGCAGAAG	qPCR
taqxyn2f	GGTCCAACCTCGGGCAACTTT	
taqxyn2r	CCGAGAAGTTGATGACCTTGTTT	
sar1fw	TGGATCGTCAACTGGTTCTACGA	
sar1rev	GCATGTGTAGCAACGTGGTCTTT	
Pxyn1f_FAM	[FAM]-TTGGCAGGCTAAATGCGACATCTTAGCCGGATGCA	EMSA
Pxyn1f	TTGGCAGGCTAAATGCGACATCTTAGCCGGATGCA	CD
Pxyn1r	TGCATCCGGCTAAGATGTCGCATTTAGCCTGCCAA	EMSA/CD

Fig. 1



B

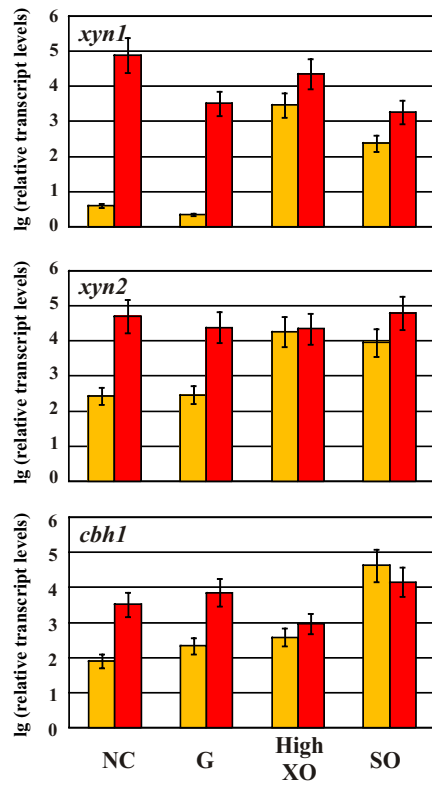


Fig. 2

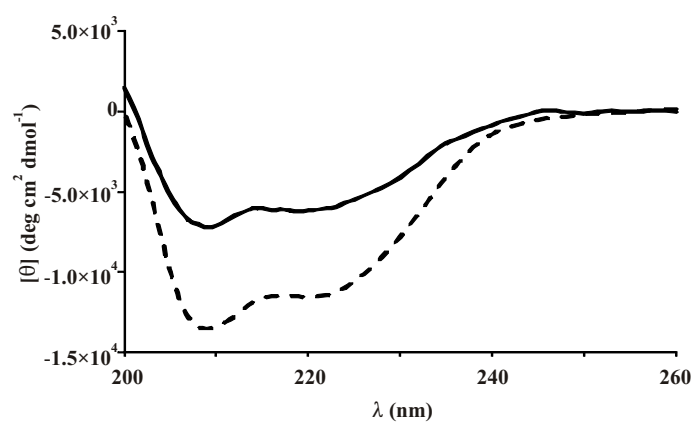


Fig. 3

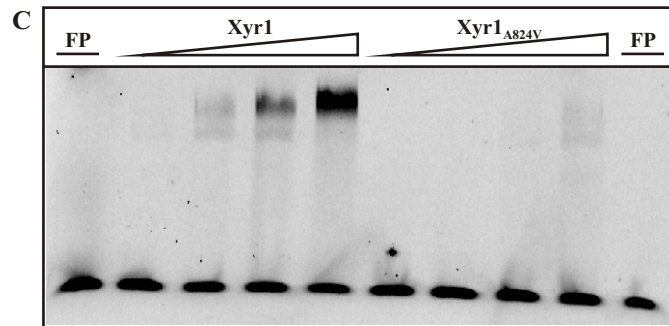
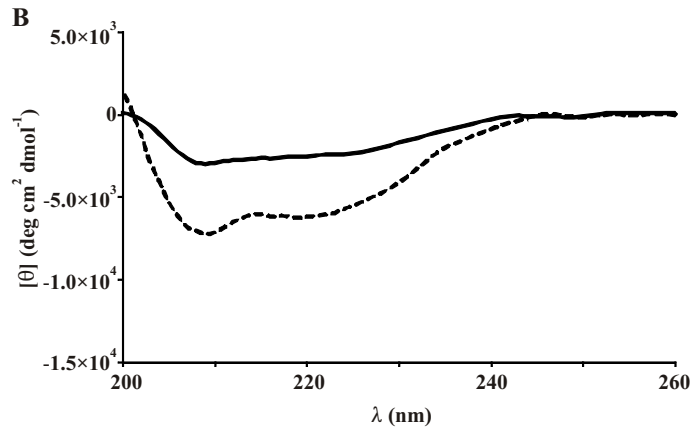
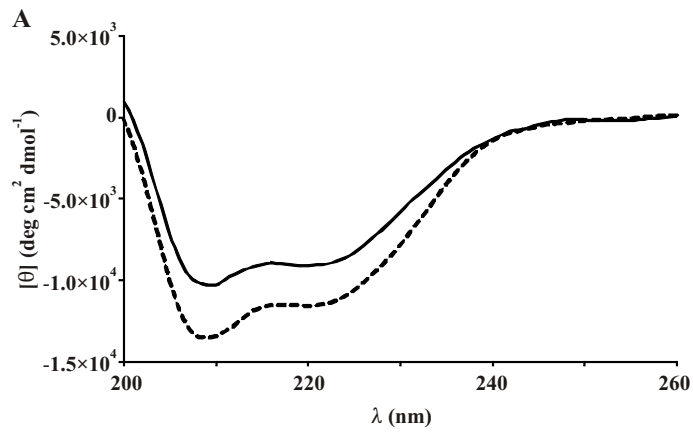


Fig. 4

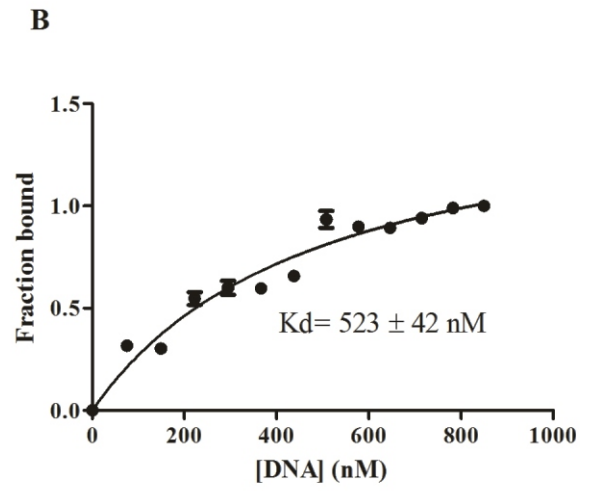
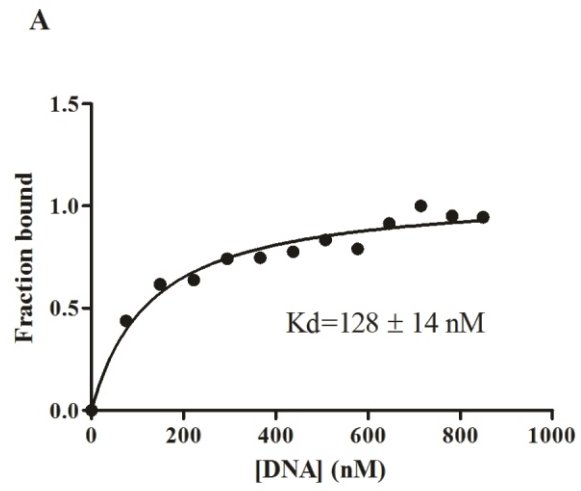


Fig. 5

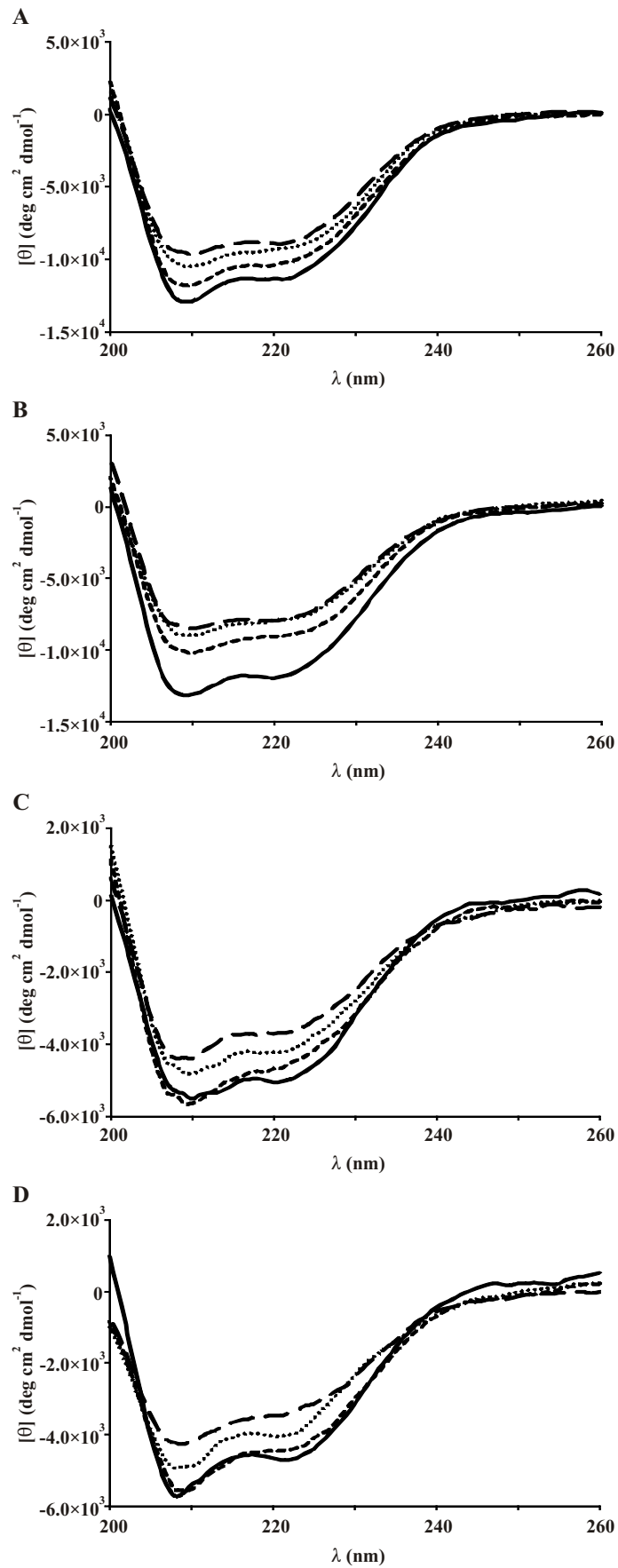
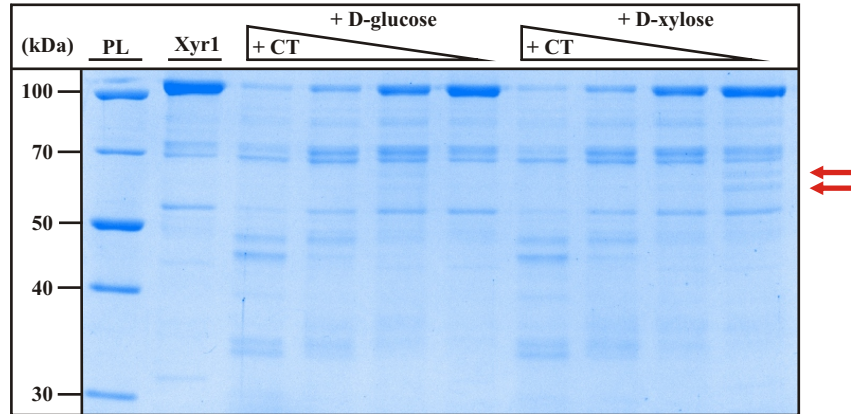


Fig. 6

A



B

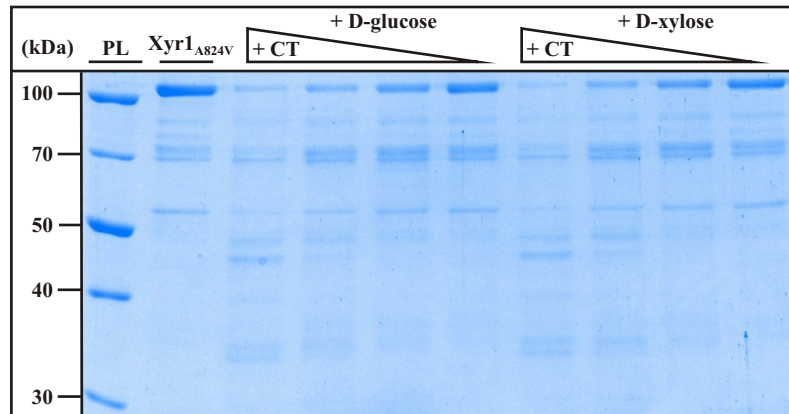
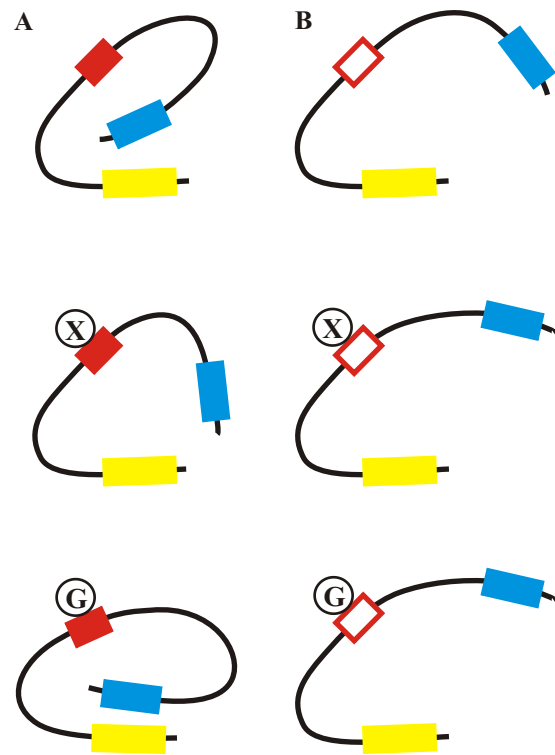


Fig. 7



PART II: *Trichoderma* as Cell Factory

Publication I: Synthesis of an antiviral drug precursor from chitin using a saprophyte as a whole-cell catalyst

(published in *Microbial Cell Factories*, 2011, 10: 102)

Publication II: *Trichoderma* as Cell Factories

(published in *Trichoderma: Biology and Applications*, edited by P Mukherjee, B Horwitz, U Singh, M Mukherjee, M Schmoll; *CABI*, Wallingford, UK 2013, pp 280-291)

RESEARCH

Open Access

Synthesis of an antiviral drug precursor from chitin using a saprophyte as a whole-cell catalyst

Matthias G Steiger^{1,4†}, Astrid R Mach-Aigner^{1†}, Rita Gorsche¹, Erwin E Rosenberg², Marko D Mihovilovic³ and Robert L Mach^{1*}

Abstract

Background: Recent incidents, such as the SARS and influenza epidemics, have highlighted the need for readily available antiviral drugs. One important precursor currently used for the production of Relenza, an antiviral product from GlaxoSmithKline, is N-acetylneuraminic acid (NeuNAc). This substance has a considerably high market price despite efforts to develop cost-reducing (biotechnological) production processes. *Hypocrea jecorina* (*Trichoderma reesei*) is a saprophyte noted for its abundant secretion of hydrolytic enzymes and its potential to degrade chitin to its monomer N-acetylglucosamine (GlcNAc). Chitin is considered the second most abundant biomass available on earth and therefore an attractive raw material.

Results: In this study, we introduced two enzymes from bacterial origin into *Hypocrea*, which convert GlcNAc into NeuNAc via N-acetylmannosamine. This enabled the fungus to produce NeuNAc from the cheap starting material chitin in liquid culture. Furthermore, we expressed the two recombinant enzymes as GST-fusion proteins and developed an enzyme assay for monitoring their enzymatic functionality. Finally, we demonstrated that *Hypocrea* does not metabolize NeuNAc and that no NeuNAc-uptake by the fungus occurs, which are important prerequisites for a potential production strategy.

Conclusions: This study is a proof of concept for the possibility to engineer in a filamentous fungus a bacterial enzyme cascade, which is fully functional. Furthermore, it provides the basis for the development of a process for NeuNAc production as well as a general prospective design for production processes that use saprophytes as whole-cell catalysts.

Background

NeuNAc is the most prevalent exponent of sialic acids [1]. In mammals, sialic acids are usually found as terminal residues of glycol conjugates on the outermost cell surface. As a result of their location and their negative carboxylate functionality, sialic acids play important roles in mediating cellular recognition and adhesion processes [2] and in the infection cycles of severe viral diseases, such as influenza viruses A and B [3]. In these cases, *de novo*-synthesized viral particles attach to their respective sialic acids at the cell surface. Neuraminidase (sialidase) activity is needed for the propagation of the virus in the host. Consequently,

sialic acid derivatives are successfully applied in the therapy of such virus-related diseases. One well-known product that functions as a neuraminidase inhibitor is Relenza. Its active pharmaceutical ingredient is Zanamivir, which is a direct derivative of the NeuNAc precursor [4].

Traditionally, NeuNAc is prepared through extraction from natural sources, such as bird nests, milk, or eggs [5], through the hydrolysis of colominic acid (a homopolymer of NeuNAc) in a culture broth of *Escherichia coli* K1 [6], or through chemical synthesis [7]. Methods for NeuNAc production have included a chemo-enzymatic process [8,9], a two-enzyme reaction process [10,11], a biotransformation process using *E. coli* [12], and an *E. coli* whole-cell system [13]. However, the requirement for ATP or an excess of pyruvate and the subsequent expensive downstream processing has kept the costs of NeuNAc production considerably high (current market price is \$100/g).

* Correspondence: rmach@mail.zserv.tuwien.ac.at

† Contributed equally

¹Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Str. 1a, A-1060 Wien, Austria

Full list of author information is available at the end of the article

Chitin is considered the second most abundant biomass available on earth [14]. The estimated annual biosynthesis of chitin is more than 10^{11} tons in marine waters alone [15]. Unlike cellulose, the other dominant biopolymer, chitin can serve as a source for both carbon and nitrogen (C:N = 8:1) [16]. This property suggests that chitin is an optimal resource for the production of NeuNAc (C:N = 11:1) because no additional nitrogen would need to be applied as it would be if glucose or cellulose were used as raw material. Chitin is found in the exoskeletons of arthropods, such as crustaceans (including crab, lobster, and shrimp) and insects (including ants and beetles), the cell walls of fungi, the radula of mollusks, and the beaks of cephalopods (including squid and octopi). This polymer is composed of β -(1,4)-linked units of the amino sugar N-acetylglucosamine (GlcNAc) that is currently produced using hydrolysis of deproteinized and demineralized crustacean shells [17]. Chitinolytic enzymes from fungi of the genus *Hypocrea* have been extensively studied for decades [18]. More recently, the chitinolytic enzyme system of *H. jecorina* has been studied using genome-wide analysis [19,20]. Unlike their bacterial counterparts (e.g., *Serratia marcescens* [21]), *Hypocrea* chitinolytic preparations have a high ratio of exochitinase to endochitinase activity and almost exclusively release monomeric GlcNAc from chitin [22], which is another advantageous aspect of chitin

compared to cellulose. Nevertheless, this raw material has not been adequately used. Therefore, the basic premise of this study was to exploit the potential of a saprophytic fungus to degrade the cheap biowaste chitin to its monomer GlcNAc and to further metabolize this product to NeuNAc.

Results and Discussion

Engineering a NeuNAc synthesis pathway into *Hypocrea*

The biosynthesis of NeuNAc begins with the formation of N-acetylmannosamine (ManNAc) from GlcNAc or UDP-N-acetylglucosamine (UDP-GlcNAc). In mammals, ManNAc is then phosphorylated to give ManNAc-6-phosphate (ManNAc-6P). The second step involves the condensation of either ManNAc (in bacteria) or ManNAc-6P (in mammals) with phosphoenolpyruvate (PEP) to give NeuNAc or NeuNAc-9P, respectively. In mammals, NeuNAc-9P is then dephosphorylated to generate NeuNAc (see Figure 1). *Hypocrea* naturally degrades chitin almost exclusively to GlcNAc [22]. Therefore, the challenge was to engineer a pathway to convert GlcNAc to NeuNAc via ManNAc, which would enable the use of *Hypocrea* as a whole-cell catalyst.

Lee and coworkers found that whole-cell extracts of several photobacteria could convert GlcNAc to ManNAc [13]. Among them, *Anabaena* sp. CH1 exhibited the highest GlcNAc 2-epimerase activity; consequently, they

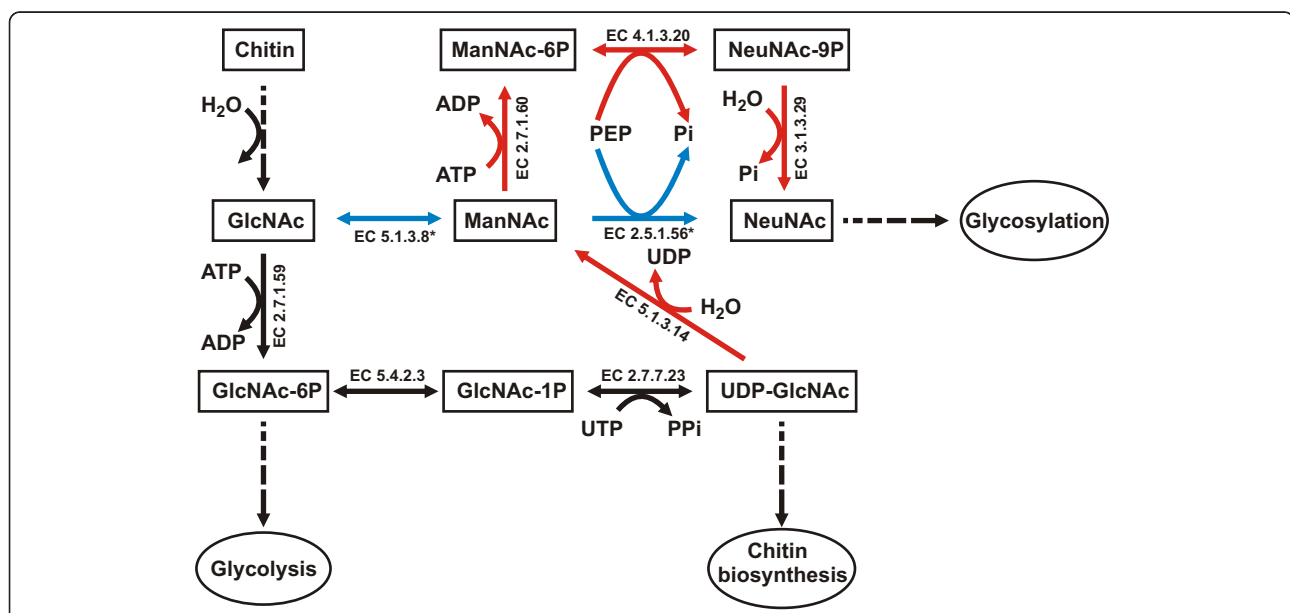


Figure 1 An illustration of the NeuNAc synthesis-related pathways in microorganisms. Boxes represent substances or metabolites, ellipses represent general biochemical processes, black solid lines represent enzymatic reactions in *H. jecorina*, black dashed lines represent pathways involving more than one enzymatic reaction in *H. jecorina*, red lines represent enzyme reactions not occurring in *H. jecorina*, and blue lines represent heterologously expressed bacterial enzymes. *EC numbers EC 5.1.3.8 (GlcNAc-2-epimerase) and EC 2.5.1.56 (NeuNAc synthase) refer to bacterial enzymes that were heterologously expressed using *Hypocrea* codon-optimized genes. EC 4.1.3.20, N-acetylneuraminic acid synthase; EC 3.1.3.29, N-acetylneuraminic acid phosphatase; EC 2.7.1.60, N-acetylmannosamine kinase; EC 5.1.3.14, UDP-N-acetylglucosamine 2-epimerase; EC 5.4.2.3 Phosphoacetylglucosamine mutase; EC 2.7.7.23, UDP-N-GlcNAc diphosphorylase; EC 2.7.1.59, N-acetylglucosamine kinase; GlcNAc, N-acetylglucosamine; ManNAc, N-acetylmannosamine; NeuNAc, N-acetylneuraminic acid; P, phosphate group.

cloned and characterized a gene encoding GlcNAc 2-epimerase from *Anabaena* sp. CH1 (E.C. 5.1.3.8), which was used in the present study as a *Hypocrea* codon-optimized gene. For the second step (the condensation of ManNAc to NeuNAc), the currently used enzyme-catalyzed processes use a lyase, which requires an excess of pyruvate. Use of this incurs high downstream processing costs. Therefore, we used the NeuNAc synthase (EC 2.5.1.56) from *Campylobacter jejuni* [23] in the *Hypocrea* process. This enzymatic step entails the use of PEP instead of pyruvate, which in the intended *in vivo* process is supplied by the fungus, thereby leading to an irreversible and more efficient reaction towards NeuNAc [24]. Moreover, the need for an excess of pyruvate becomes obsolete, and the resulting downstream process is significantly simplified. Similar to the GlcNAc 2-epimerase, the coding sequence for the NeuNAc synthase was codon-optimized for the usage in *Hypocrea*. The synthetic pathway is presented in Figure 1. The complete nucleotide sequences for both genes encoding the recombinant enzymes, *tbage* and *tneub*, are shown in additional file 1.

Metabolization or uptake of NeuNAc can not be observed in *Hypocrea*

As the ability of the fungus to metabolize NeuNAc is an important issue, a possible uptake of NeuNAc by *H. jecorina* was investigated. Therefore, the fungus was pre-grown on glycerol in liquid culture, and half of the mycelium was autoclaved and half of it was harvested. The dead and living mycelia were transferred to glycerol-containing medium to study growth conditions or to medium without a carbon source to study resting cell conditions. NeuNAc was added to both media, and cultures were incubated for 8 h. Supernatants from all conditions were analyzed after incubation for 0 and 8 h by HPLC after derivatization using 1,2-diamino-4,5-methylenedioxybenzene dihydrochloride (DMB). Similar amounts of NeuNAc were present under all conditions regardless of whether the fungus was alive or dead (Figure 2a). This result indicates that NeuNAc uptake does not occur in *H. jecorina*. As a positive control experiment we did a similar experiment but instead of NeuNAc, GlcNAc was added to the media. As can be inferred from Figure 2b GlcNAc was completely consumed after eight hours under both growth and resting cell conditions when the mycelium was viable.

Characterization of the recombinant *H. jecorina* strain

Recombinant *Hypocrea* strains were generated using protoplast transformation of *H. jecorina* QM9414. In the derived strains, the two *Hypocrea* codon-optimized genes (without GST-tag) were placed under the control of either the *H. jecorina* pyruvate kinase (*pki*) promoter, which is a strong constitutive promoter, or the *H. jecorina* xylanase 1 (*xyn1*) promoter, which is a strict shut-off system if an

inducer (e.g. D-xylose) is missing. Such a system was used to avoid interference of the introduced recombinant pathway with cell wall biosynthesis and consequently, biomass formation. However, when comparing both promoter systems the strong *pki* promoter did not lead to decreased growth, diminished cell integrity or other adverse effects (data not shown). Therefore, we used strains in which both genes were under the control of the *pki* promoter for further studies as we observed a remarkably higher NeuNAc formation.

Transcriptional analysis of the recombinant *H. jecorina* strains was done by RT-qPCR to compare expression of both inserted genes using *sar1* (SAR/ARF-type small GTPase) as a stable reference gene [25]. Furthermore, the copy numbers of both genes was measured by qPCR of genomic DNA using *pki* as a reference, which in the native *H. jecorina* genome is present as a single copy gene. Based on these analyses a strain (termed PEC/PSC1) was chosen for further investigations because it showed the highest equal expression of both inserted genes. This was confirmed by the finding that this strain bears two copies of each recombinant gene in the genome. These data were also supported using Southern blot analysis (data not shown).

GlcNAc 2-epimerase and NeuNAc synthase are fully functional as GST-fusion proteins

Both recombinant enzymes were heterologously expressed as glutathione S-transferase (GST) fusion proteins in *E. coli*; the affinity chromatography purified proteins were used to confirm that their enzymatic capability was not altered by the codon usage adaptation and to provide a positive control for the enzymatic assays later on.

To determine if the recombinant enzymes were functional, both GST fusion proteins were used in an enzymatic assay. The presence of GlcNAc and the formation of the intermediate product ManNAc and the final product NeuNAc were monitored using HPLC-MS analysis and results are presented in Figure 3. Application of the GST-fusion proteins of both enzymes in the *in vitro* assay led to the formation of ManNAc and NeuNAc demonstrating that the synthetic genes are expressed as functional proteins (Figure 3a1 and 3b1).

NeuNAc synthesis *in vitro* by recombinant *H. jecorina* strains

According to the GST-fusion proteins, cell-free extracts of the recombinant *H. jecorina* strain PEC/PSC1 were applied in the enzymatic assay. The formation of ManNAc (Figure 3a2) and NeuNAc could be detected (Figure 3b2). This demonstrates that both enzymes are also fully functionally expressed in the recombinant *H. jecorina* strain PEC/PSC1. Neither ManNAc nor NeuNAc was detected

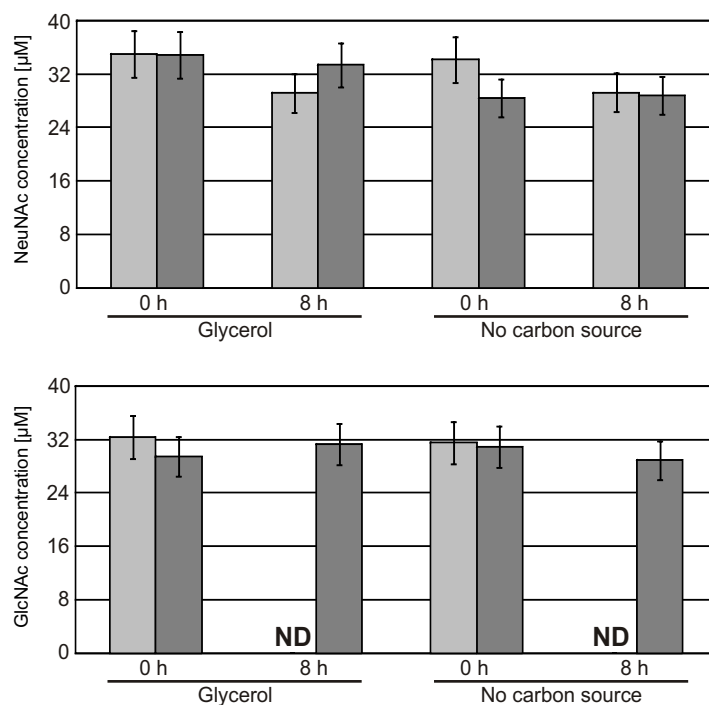


Figure 2 The analysis of possible metabolism of NeuNAC in *H. jecorina*. The parental strain was pre-cultured on glycerol, and half of the mycelia were autoclaved. Living (light grey) and dead (dark grey) mycelia were transferred to MA media containing glycerol or MA media that lacked a carbon source, and NeuNAC (a) or GlcNAc (b) (as positive control) was added to both media. Samples were collected at 0 and 8 h. For NeuNAC analysis supernatants were derivatized using DMB before analyses using HPLC. The presented values are the means of two biological duplicates that were derivatized in duplicate. Error bars indicate the standard deviations.

using cell-free extracts from the parental strain in the assay (Figure 3a3 and 3b3), indicating that these pathways are normally not active in *Hypocrea*.

To investigate the stability of NeuNAC in cell-free extracts of the recombinant strain, according cell-free extracts obtained from the cultivation in a bioreactor on chitin (*vide infra*) were spiked with NeuNAC and incubated for 24 h. As a control, a heat-inactivated cell-free extract was similarly treated. Using HPLC analysis after derivatization with DMB, similar amounts of NeuNAC were detected in both extract preparations (Figure 3c), suggesting that components of the cell-free extract do not actively degrade NeuNAC. In addition, a similar amount of NeuNAC was measured in a NeuNAC-spiked cell-free extract of the recombinant strain that was not incubated, assuming that the 24-h incubation period at 30°C did not decrease the NeuNAC levels. As a final control, a cell-free extract without NeuNAC was also analyzed after a 24-h incubation period and, as expected, showed a lower amount of NeuNAC, which could only have resulted from its formation during the cultivation on chitin. In summary, we did not observe degradation of NeuNAC by *H. jecorina*. These data suggest that NeuNAC is not metabolized by the recombinant *Hypocrea* strain.

NeuNAC synthesis *in vivo* by the recombinant *H. jecorina* strain

We next addressed whether the recombinant *H. jecorina* strain had the ability to produce NeuNAC *in vivo*. To test this, the strain was grown on GlcNAc in shake flasks and cultivated on colloidal chitin in a bioreactor. Data on the corresponding cultivation monitoring are provided in additional file 2. As a positive control, an enzyme assay using the GST fusion proteins was again performed and resulted in the detection of ManNAc using HPLC-MS analysis as shown in Figure 4a1. Notably, the intermediate ManNAc was detected in the recombinant strain, regardless of the carbon source (Figure 4a2 und Figure 4a4), whereas the parental strain did not form ManNAc (Figure 4a3). In the parental strain, only the first metabolite, GlcNAc, was detected, and it was present because it was either directly used as a carbon source or formed by degradation of the biopolymer chitin due to the native chitinolytic activity of the fungus. The synthesis of the product NeuNAC was analyzed using HPLC-MS/MS analysis (Figure 4b). As a positive control, the reaction products (ManNAc, NeuNAC) generated by the use of the GST fusion proteins in an enzymatic assay are shown (Figure 4b1). Importantly, the recombinant *H. jecorina* strain formed NeuNAC using either carbon source,

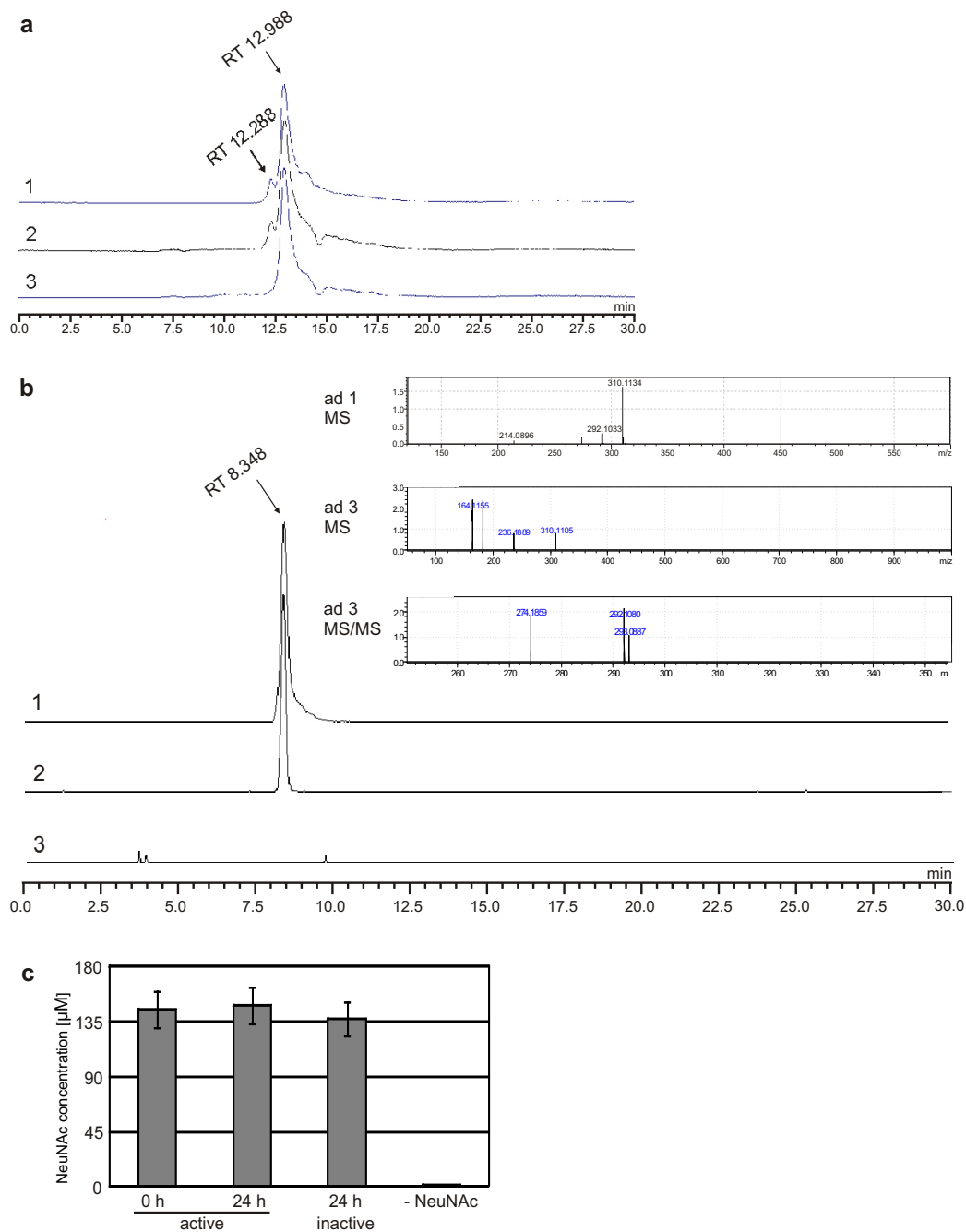


Figure 3 NeuNAc synthesis *in vitro* in an enzymatic assay. (a) EICs of the HPLC-MS analysis at 222.098 atomic mass units (amu) corresponding to the mass of the $[\text{GlcNAc}+\text{H}]^+$ ion and the $[\text{ManNAc}+\text{H}]^+$ ion. Retention times (RTs) of ManNAc (12.288 min) and GlcNAc (12.988 min) are indicated. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase. (2) Chromatogram of the *in vitro* assay using a cell-free extract of the PEC/PSC1 strain. (3) Chromatogram of the *in vitro* assay using a cell-free extract of the parental strain. (b) EICs at 310.1134 amu, corresponding to the mass of the $[\text{NeuNAc}+\text{H}]^+$ ion. RT of NeuNAc (8.348 min) is indicated. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase. (2) Chromatogram of the *in vitro* assay using a cell-free extract of the PEC/PSC1 strain showing an 8-fold amplification compared to (1). (3) Chromatogram of a cell-free extract of the parental strain showing a 250-fold amplification compared to (1). (ad 1 MS) and (ad 2 MS) are MS spectra of chromatograms 1 and 2, respectively, at a RT of 8.348 min. (c) Cell-free extracts (active) of the PEC/PSC1 strain obtained from cultivation on chitin were mixed with NeuNAc and incubated for 0 and 24 h. A heat-inactivated cell-free extract was similarly treated. An active cell-free extract without NeuNAc (-NeuNAc) was also incubated, DMB-derivatized and analyzed. Values are means of biological duplicates derivatized in duplicate. Error bars indicate standard deviations.

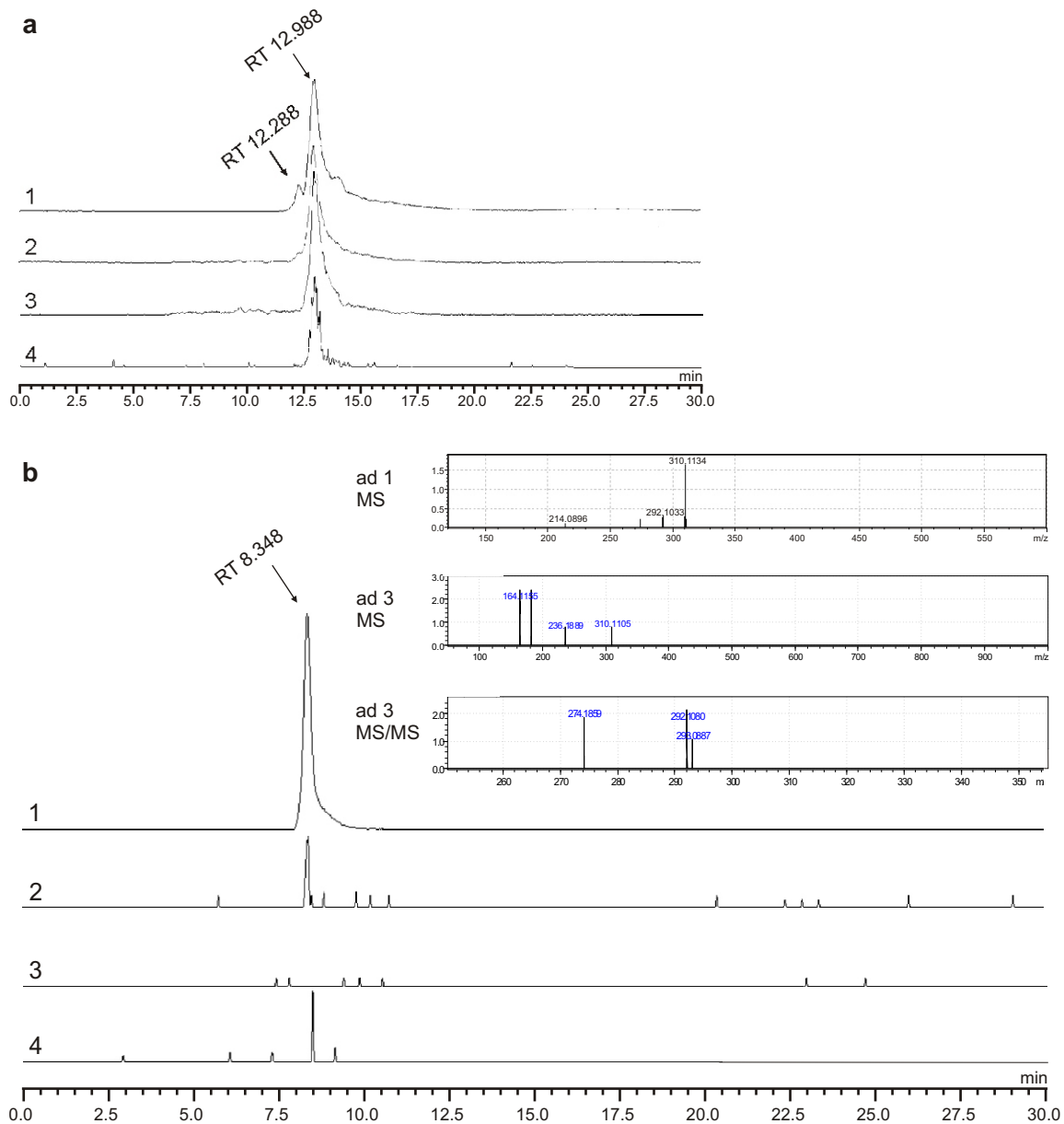


Figure 4 NeuNAc synthesis *in vivo* in the *H. jecorina* PEC/PSC1 strain. (a) EICs from the HPLC-MS analysis at 222.098 atomic mass units (amu) corresponding to the mass of the $[\text{GlcNAc}+\text{H}]^+$ and $[\text{ManNAc}+\text{H}]^+$ ions. Retention times (RTs) of ManAc (12.288 min) and GlcNAc (12.988 min) are indicated. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase. (2) Chromatogram of the cell-free extract of the PEC/PSC1 strain grown on GlcNAc. (3) Chromatogram of the cell-free extract of the parental strain grown on GlcNAc. (4) Chromatogram of the cell-free extract of the PEC/PSC1 strain grown on chitin. (b) EICs at 310.1134 amu corresponding to the mass of the $[\text{NeuNAc}+\text{H}]^+$ ion. RT of NeuNAc (8.348 min) is indicated. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase. (2) Chromatogram of the cell-free extract of the PEC/PSC1 strain grown on GlcNAc showing 25-fold amplification compared to (1). (3) Chromatogram of the cell-free extract of the parental strain showing a 250-fold amplification compared to (1). (4) Chromatogram of the cell-free extract of the PEC/PSC1 strain after cultivation on chitin for 90 h showing a 25-fold amplification compared to (1). (ad 1 MS) and (ad 4 MS) are the MS spectra of chromatograms 1 and 4, respectively, at RTs of 8.348 min. (ad 4 MS/MS) is the MS/MS spectrum of chromatogram 3 at a RT of 8.348 min.

GlcNAc or chitin (Figure 4b2 and 4b4), whereas in the parental strain no formation of NeuNAc was detected (Figure 4b3). This analysis allowed us to estimate that 13 μg NeuNAc per g mycelium (dry weight) was formed

in the recombinant strain. Thus, on its own, this would not be a competitive production process, but it does demonstrate the possibility for engineering a saprophyte and using it as a whole-cell catalyst that expresses a

bacterial enzyme cascade. This method has enormous potential considering its use of a cheap starting material and the relatively simple, inexpensive cultivation of a fungus.

Conclusions

Taken together, we successfully engineered *Hypocrea* in a way that this fungus now produces NeuNAc from the biopolymer chitin by employing its natural saprophytic activity in combination with the introduction of a bacterial enzyme cascade. Because human society will face severe bottlenecks in the supply of energy and in obtaining certain raw materials in the upcoming years, we hope that this study will highlight the potential advantages of biopolymers, such as chitin, and stimulate their efficient usage. Furthermore, we anticipate that such strategies will support efforts to create sustainable production processes.

Methods

Strains and cultivation conditions

The parental strain *H. jecorina* (*T. reesei* [26]) QM9414 (ATCC 26921) was maintained on malt extract (MEX) agar.

Mycelia for the enzymatic assay were cultivated in 3% (w/v) MEX medium using 10^8 conidia/L at 30°C.

Cultivation of *H. jecorina* on colloidal chitin was performed in a bench top bioreactor (Bioengineering, Wald, Switzerland) as previously described [27]. Briefly, 500 mL Mandels-Andreotti (MA) [28] medium containing 1% (w/v) colloidal chitin [29], 0.5% GlcNAc, and 0.1% (w/v) bacto peptone (Difco, Detroit, US) was inoculated with 10^8 conidia/L. Some drops glanapon (Becker, Wien, Austria) were added to the medium to avoid excessive foam formation. Cultivation was performed at 30°C temperature, pH 5, 0.3 vvm aeration rate, and 500 rpm agitation rate for 96 h. Each sample drawing was followed by a microscopic analysis for infection control. Culture supernatant and mycelia were separated by filtration through GF/F glass microfiber filters (Whatman, Brentford, UK). All strains (parental, recombinant) showed similar growth on rich media as well as MA medium.

Plasmid construction

The synthetic gene *tbage* (for sequence see additional file 1) is based on the protein sequence of *Anabaena sp.* CH1 GlcNAc-2-epimerase (GenBank: ABG57042) and was reverse translated into a nucleotide sequence using the GeneOptimizer[®] software (Geneart, Regensburg, Germany). The codon usage was optimized for *H. jecorina* (<http://www.kazusa.or.jp/codon>). The synthetic gene *tneub* (for sequence see additional file 1) was similarly obtained based on the protein sequence from *Campylobacter jejuni* NCTC11168 NeuNAc synthase (<http://old.genedb.org/genedb/cjejuni/index.jsp>, Cj1141).

The synthetic genes *tbage* and *tneub* were excised from the production plasmid using *XbaI/NsiI* digestion and inserted into pRLM_{ex}30 [30] to generate the plasmids pMS-PEC and pMS-PSC.

For the construction of pGEX-epi and pGEX-syn, the oligonucleotides GEXfw and GEXrev (Table 1) were used to introduce an *XbaI* and *NsiI* site into plasmid pGEX4T-2 (GE Healthcare, Chalfont St Giles, UK), yielding pGEX-MS. *tbage* and *tneub* were inserted into pGEX-MS via *XbaI/NsiI* digestion to yield the plasmids pGEX-epi and pGEX-syn.

Protoplast transformation

The protoplast transformation of *H. jecorina* was performed as described previously [31]. The plasmid pHylox2 (2 µg) [32], which confers hygromycin B resistance [30], and 4 µg of each plasmid pMS-PEC and pMS-PSC were co-transformed into the fungal genome.

DNA analysis

Fungal genomic DNA was isolated as described previously [31]. Southern hybridization and detection were performed using the DIG High Prime DNA Labeling and Detection Starter Kit II following the manufacturer's instructions (Roche, Basel, Switzerland).

Transcriptional analysis

RNA extraction, cDNA synthesis and qPCR analysis were performed as described elsewhere [25]. Primer sequences are given in Table 1.

Glutathione S-transferase (GST) fusion proteins

GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase were generated using plasmids pGEX-epi and pGEX-syn in *E. coli* BL21 (DE3). Purification of the proteins was performed using GSTrap[™]FF (GE Healthcare) according to standard procedures.

Table 1 Oligonucleotides used during this study

Name	Sequence (5'→3')	Usage
NANASfw	GTGGTGTGCAGGAGGACGAA	qPCR <i>tneub</i>
NANASrev	CAAGCACATCGCCAGTTCAAG	qPCR <i>tneub</i>
ManEfw	GCGATCTTGAGCCAGTTCTC	qPCR <i>tbage</i>
ManErev	GCTACTTCACCTGCCTCGAC	qPCR <i>tbage</i>
GEX-MSfw	AATTCCTTAGAGATATGCATC	Construction pGEX-MS
GEX-MSrev	TCGAGATGCATATCTCTAGAAGG	Construction pGEX-MS
pkifw R	CTGCGACACTCAGAACATGTACGT	qPCR <i>pki</i> cDNA
pkifw D	GCTCTGCTTGGAACTGATTGA	qPCR <i>pki</i> DNA
pkirev	GGTCTGGTCGTCCTTGATGCT	qPCR <i>pki</i>
sar1fw	TGGATCGTCAACTGGTTCTACGA	qPCR <i>sar1</i>
sar1rev	GCATGTGTAGCAACGTGGTCTTT	qPCR <i>sar1</i>

Enzymatic assay

Harvested mycelia were ground into fine powder and resuspended in 0.1 M Bicine buffer (pH 8) containing protease inhibitors (2 μ M leupeptin, 1 μ M pepstatin A, and 10 μ M PMSF) (0.3 g mycelia/mL). The suspension was sonicated using a Sonifier[®] 250 Cell Disruptor (Branson, Danbury, US) (power 40%, duty cycle 50%, power 20 sec, 40 sec pause, 10 cycles). Insoluble compounds were separated using centrifugation (10 min, 13000 g, 4°C). Enzymatic analysis was performed according to a previously described modified protocol [33]. The assay was performed in a total volume of 100 μ L containing 10 mM GlcNAc, 10 mM PEP, 12.5 mM MnCl₂, 100 mM Bicine buffer (pH 8) and 40 μ L cell-free extract. Reactions were incubated for 60 min at 37°C, terminated at 85°C for 10 min and analyzed using HPLC. As a positive control, 5 μ L of both GST fusion proteins were applied in place of the cell-free extracts.

The stability of NeuNAc in the cell-free extract was determined by adding NeuNAc (150 μ M) and incubating for 24 h at 30°C. After derivatization with DMB [34], the NeuNAc quantity was measured using HPLC.

Detection of NeuNAc synthesis *in vivo*

Harvested *H. jecorina* mycelia were ground into fine powder and resuspended in water (0.3 g mycelia/mL). The suspension was sonicated using a Sonifier[®] 250 Cell Disruptor (Branson) (power 70%, duty cycle 50%, power for 1 min, 1 min pause, 3 cycles). Insoluble compounds were separated using centrifugation (10 min, 13000 g, 4°C), and the supernatant was analyzed using HPLC-MS/MS.

NeuNAc and GlcNAc uptake

H. jecorina mycelia were pre-grown on MA containing 1% glycerol, transferred to MA medium containing 1% glycerol or no carbon source, spiked with 30 μ M NeuNAc or GlcNAc, respectively, and incubated for 8 h at 30°C. Autoclaved mycelia served as a negative control. After derivatization with DMB [34], the NeuNAc quantity was measured using HPLC.

HPLC and HPLC-MS/MS analysis

NeuNAc, ManNAc and GlcNAc formation was measured using LC-MS (IT-TOF-MS) (Shimadzu, Kyoto, Japan) with a Rezex[™] RHM-Monosaccharide H⁺-column (8%, 300 \times 7.8 mm) (Phenomenex, Torrance, USA). The mobile phase consisted of water with 0.1% (v/v) trifluoroacetic acid, the flow was 0.6 mL/min, the column temperature was 80°C, and the injected volume was 10 μ L. MS detection was performed in ESI+ mode, covering a scan range of 60–600 amu. The retention times were determined using pure standard substances. The identity of NeuNAc was confirmed by both, chromatographic

retention time and mass spectral signal, which are very well matched by authentic standards of NeuNAc. The better the mass accuracy obtained from exact mass determination by HR-MS, the lower is the number of possible isobaric candidates (e.g. [35]). In this case the mass accuracy is better than 2 ppm, leading to the number of candidates reduced to less than 10, with an even further reduction in the number of potential candidates because the isotopic pattern is also taken into account (what the software of the used IT-TOF-MS instrument does automatically).

DMB derivatives of NeuNAc were separated on a Kinetex RP C18 (Phenomenex) at 0.75 mL/min with a 40°C column temperature and a mobile phase of water: methanol:trifluoroacetic acid (74.25:25:0.75). A Shimadzu RF-20AXS fluorescence detector (excitation 373 nm, emission 448 nm) was used for detection.

Additional material

Additional file 1: Coding sequences of the synthetic genes *tbage* and *tneub*. Coding sequences of the synthetic genes *tbage* and *tneub*. The sequences are provided in FASTA format. The **XbaI** site is underlined, and the **NsiI** site is double-underlined. The start codon ATG and the stop codon TAA are presented in bold letters.

Additional file 2: Parameters of *H. jecorina* cultivation on chitin in a bioreactor. Oxygen consumption (pO₂; blue line), consumption of the intermediate N-acetylglucosamine (GlcNAc; red line), formation of the product N-acetylneuraminic acid (NeuNAc; pink line), and formation of biomass (given as dry weight, BDW; green line) of the *H. jecorina* PEC/PSC1 strain cultivated on chitin in a bioreactor for 96 h are displayed.

Acknowledgements

This study was supported by a grant from the Austrian Science Fund FWF (P21287) and a grant from the Vienna University of Technology ("DemoTech", Innovative Project), which are gratefully acknowledged. We thank Michael Schön for assistance with analytical work.

Author details

¹Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Str. 1a, A-1060 Wien, Austria. ²Instrumental Analytical Chemistry, Institute of Chemical Technologies and Analytics, Vienna University of Technology, Getreidemarkt 9/164, A-1060 Wien, Austria. ³Institute of Applied Synthetic Chemistry, Vienna University of Technology, Getreidemarkt 9/163, A-1060 Wien, Austria. ⁴Austrian Centre of Industrial Biotechnology, Muthgasse 107, 1190 Vienna, Austria.

Authors' contributions

MGS produced and characterized the recombinant strains and contributed to the manuscript. ARMA prepared the manuscript and contributed to the design of the study. RG established enzymatic assays. EER performed HPLC-MS/MS analyses; MDM suggested the target substance and supported analytics; RLM contributed to the design and coordination of the study. All authors critically read the manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 26 July 2011 Accepted: 5 December 2011
Published: 5 December 2011

References

- Schauer R, Kelm S, Reuter G, Roggentin P, Shaw L: **Biochemistry and Role of Sialic Acids**. In *Biology of the Sialic Acids*. Edited by: Rosenberg A. NY and London: Plenum Press; 1995:7-67.
- Varki A: **Sialic acids as ligands in recognition phenomena**. *Faseb J* 1997, **11(4)**:248-255.
- Herrler G, Hausmann J, Klenk HD: **Sialic acid as receptor determinant of ortho- and paramyxoviruses**. In *Biology of the Sialic Acids*. Edited by: Rosenberg A. NY and London: Plenum Press; 1995:315-331.
- Tremblay JF: **The other drug for avian flu**. *C&EN* 2006, **84(15)**:33-36.
- Koketsu M, Juneja LR, Kawanami H, Kim M, Yamamoto T: **Preparation of N-acetylneuraminic acid from delipidated egg yolk**. *Glycoconj J* 1992, **9(2)**:70-74.
- Maru I, Ohnishi J, Ohta Y, Tsukada Y: **Why is sialic acid attracting interest now? Complete enzymatic synthesis of sialic acid with N-acetylglucosamine 2-epimerase**. *J Biosci Bioeng* 2002, **93(3)**:258-265.
- de Ninno M: **The synthesis and glycosidation of N-acetyl-D-neuraminic acid**. *Synthesis* 1991, **8**:583-593.
- Blayer S, Woodley JM, Dawson MJ, Lilly MD: **Alkaline biocatalysis for the direct synthesis of N-acetyl-D-neuraminic acid (Neu5Ac) from N-acetyl-D-glucosamine (GlcNAc)**. *Biotechnol Bioeng* 1999, **66(2)**:131-136.
- Mahmoudian M, Noble D, Drake CS, Middleton RF, Montgomery DS, Piercey JE, Ramlakhan D, Todd M, Dawson MJ: **An efficient process for production of N-acetylneuraminic acid using N-acetylneuraminic acid aldolase**. *Enzyme Microb Technol* 1997, **20(5)**:393-400.
- Kragl U, Gygax D, Ghisalba O, Wandrey C: **Enzymatic two step synthesis of N-acetylneuraminic acid in the enzyme membrane reactor**. *Angew Chem Int Ed Engl* 1991, **30**:827-828.
- Maru I, Ohnishi J, Ohta Y, Tsukada Y: **Simple and large-scale production of N-acetylneuraminic acid from N-acetyl-D-glucosamine and pyruvate using N-acetyl-D-glucosamine 2-epimerase and N-acetylneuraminic acid lyase**. *Carbohydr Res* 1998, **306(4)**:575-578.
- Tabata K, Koizumi S, Endo T, Ozaki A: **Production of N-acetyl-D-neuraminic acid by coupling bacteria expressing N-acetyl-D-glucosamine 2-epimerase and N-acetyl-D-neuraminic acid synthetase**. *Enzyme Microb Technol* 2002, **30**:327-333.
- Lee YC, Chien HC, Hsu WH: **Production of N-acetyl-D-neuraminic acid by recombinant whole cells expressing *Anabaena* sp. CH1 N-acetyl-D-glucosamine 2-epimerase and *Escherichia coli* N-acetyl-D-neuraminic acid lyase**. *J Biotechnol* 2007, **129(3)**:453-460.
- Ballenweg S: **Roempp Online**. Thieme Chemistry, Stuttgart 2005.
- Li X, Roseman S: **The chitinolytic cascade in *Vibrios* is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase**. *Proc Natl Acad Sci USA* 2004, **101(2)**:627-631.
- Khoushab F, Yamabhai M: **Chitin research revisited**. *Mar Drugs* 2010, **8(7)**:1988-2012.
- Ferrer J, Paez G, Marmol Z, Ramones E, Garcia H, Forster C: **Acid hydrolysis of shrimp-shell wastes and the production of single cell protein from the hydrolysate**. *Bioresour Technol* 1996, **57**:55-60.
- Lorito M: **Chitinolytic enzymes and their genes**. In *Trichoderma & Gliocladium. Volume 2*. Edited by: Harman GE, Kubicek CP. London, UK: Taylor 1998:73-92.
- Seidl V, Huemer B, Seiboth B, Kubicek CP: **A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases**. *FEBS J* 2005, **272(22)**:5923-5939.
- Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, et al: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)**. *Nat Biotechnol* 2008, **26(5)**:553-560.
- Watanabe T, Kimura K, Sumiya T, Nikaidou N, Suzuki K, Suzuki M, Taiyoji M, Ferrer S, Regue M: **Genetic analysis of the chitinase system of *Serratia marcescens* 2170**. *J Bacteriol* 1997, **179(22)**:7111-7117.
- Donzelli BG, Ostroff G, Harman GE: **Enhanced enzymatic hydrolysis of langostino shell chitin with mixtures of enzymes from bacterial and fungal sources**. *Carbohydr Res* 2003, **338(18)**:1823-1833.
- Sundaram AK, Pitts L, Muhammad K, Wu J, Betenbaugh M, Woodard RW, Vann WF: **Characterization of N-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni***. *Biochem J* 2004, **383(Pt 1)**:83-89.
- Paccalet T, Bardor M, Rihouey C, Delmas F, Chevalier C, D'Aoust MA, Faye L, Vezina L, Gomord V, Lerouge P: **Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthesizing enzymes**. *Plant Biotechnol J* 2007, **5(1)**:16-25.
- Steiger MG, Mach RL, Mach-Aigner AR: **An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*)**. *J Biotechnol* 2010, **145(1)**:30-37.
- Kuhls K, Lieckfeldt E, Samuels GJ, Kovacs W, Meyer W, Petrini O, Gams W, Borner T, Kubicek CP: **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* 1996, **93(15)**:7755-7760.
- Stricker AR, Trefflinger P, Aro N, Penttilä M, Mach RL: **Role of Ace2 (Activator of Cellulases 2) within the *xyn2* transcriptosome of *Hypocrea jecorina***. *Fungal Genet Biol* 2008, **45(4)**:436-445.
- Mandels M: **Applications of cellulases**. *Biochem Soc Trans* 1985, **13(2)**:414-416.
- Roberts WK, Selitrennikoff CP: **Plant and Bacterial Chitinases Differ in Antifungal Activity**. *J Gen Microbiol* 1988, **134**:169-176.
- Mach RL, Schindler M, Kubicek CP: **Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expression signals**. *Curr Genet* 1994, **25(6)**:567-570.
- 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(1)**:71-76.
- Steiger MG, Vitikainen M, Uskonen P, Brunner K, Adam G, Pakula T, Penttilä M, Saloheimo M, Mach RL, 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(1)**:114-121.
- Vann WF, Tavarez JJ, Crowley J, Vimr E, Silver RP: **Purification and characterization of the *Escherichia coli* K1 *neuB* gene product N-acetylneuraminic acid synthetase**. *Glycobiology* 1997, **7(5)**:697-701.
- Nakamura M, Hara S, Yamaguchi M, Takemori Y, Ohkura Y: **1,2-Diamino-4,5-methylenedioxybenzene as a Highly Sensitive Fluorogenic Reagent for α -Keto Acids**. *Chem Pharm Bull (Tokyo)* 1987, **35(2)**:687-692.
- Holcapek M, Jirasko R, Lisa M: **Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules**. *J Chromatogr A* 2010, **1217(25)**:3908-3921.

doi:10.1186/1475-2859-10-102

Cite this article as: Steiger et al.: Synthesis of an antiviral drug precursor from chitin using a saprophyte as a whole-cell catalyst. *Microbial Cell Factories* 2011 **10**:102.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



16 *Trichoderma* as Cell Factories

Rita Gorsche, Astrid R. Mach-Aigner and Robert L. Mach*
Institute of Chemical Engineering, Vienna University of Technology, Austria

16.1 Introduction

The various strains of *Trichoderma reesei* (anamorph of *Hypocrea jecorina*) used in industry today are the results of a long history in the industrial production of hydrolysing enzymes such as cellulases and hemicellulases. Following years of strain improvement and mutations, the industrial strains of today reach protein production and secretion on a scale of 100 g/l, up to 60% of which comprises the major cellulase cellobiohydrolase I (CBHI) (Schuster and Schmoll, 2010). These excellent protein production properties are achieved through cultivation on relatively cheap and simple media, such as mixtures of cellulose and xylan in plant materials from agricultural waste (Mach and Zeilinger, 2003), or lactose, an industrial byproduct (Seiboth *et al.*, 2007). In contrast to most other industrially used expression hosts, *Trichoderma* does not require any expensive additives, such as vitamins, amino acids or other supplements, in the growth medium. As a result of these characteristics *T. reesei* constitutes an interesting host organism for the production of heterologous proteins and has been studied in that respect for some time.

A further advantage to using *T. reesei* for the expression of heterologous proteins is the

fact that years of research towards improving hydrolytic enzyme production has resulted in the development of all essential tools needed for the expression of heterologous proteins in this particular host organism. There are a number of useful mutant strains available, such as hypercellulolytic, cellulose-negative strains or even some partially protease-deficient strains (Mantyla, 1998). In addition, *T. reesei* can be transformed with a number of different selection markers, from auxotrophic markers, such as *pyr4* (Gruber *et al.*, 1990), to resistance to hygromycin (Mach *et al.*, 1994) or benomyl (Peterbauer *et al.*, 1992), or even to allow growth on acetamide as sole nitrogen source mediated through the *Aspergillus nidulans amdS* gene (Penttilä *et al.*, 1987), most of which are recyclable. Moreover, the use of a Cre/loxP based transformation system allows bidirectional positive selection via hygromycin resistance or loss of sensitivity to fluoroacetamide, as well as excision of the marker gene to enable multiple deletions and sequential transformations (Steiger *et al.*, 2011b). Finally, considerable efforts have been taken to investigate the regulation of gene expression, protein production and protein secretion (reviewed in Aro *et al.*, 2005; Kubicek *et al.*, 2009; Saloheimo and Pakula, 2012),

* E-mail: rmach@mail.zserv.tuwien.ac.at

the results of which are a basis for research into the improvement of heterologous protein production.

16.2 Attempts at Improving Heterologous Protein Production in *Trichoderma*

16.2.1 Promoters used for expression of heterologous proteins

Up to 60% of hydrolytic enzymes secreted by industrial *T. reesei* strains consist of CBHI (Schuster and Schmoll, 2010), which is expressed from the single copy gene *cbh1*, the promoter of which can be strongly induced by growth on plant material, cellulose and especially sophorose, making it one of the most frequently used promoters for protein production (Keränen and Penttilä, 1995). Unfortunately there are some drawbacks to using the *cbh1* promoter, such as CRE1-mediated repression on glucose containing media and co-regulation with most of the other hydrolases expressed by *T. reesei*, which might impair the purification of the target protein. There is also one case known in which, despite reasonable transcript levels, a heterologous protein could not be produced using the *cbh1* promoter (Schmoll *et al.*, 2010). In addition growth on *cbh1*-inducing media might also induce extracellular proteases (Keränen and Penttilä, 1995). Therefore a number of other promoters have been tested for their usefulness in heterologous protein production.

For the expression on media containing glucose promoters, regulating genes such as *pgk1*, encoding phosphoglycerate kinase (Vanhanen *et al.*, 1989), *pki*, encoding pyruvate kinase (Schindler *et al.*, 1993), *tef1*, encoding the translation elongation factor 1 α , some as yet undefined genes as well as a glucose derepressed version of the *cbh1* promoter have been tested for their protein production capacity. Of these, the promoter of a still unidentified gene 'cDNA1' proved most productive, yielding up to 50% of the total secreted protein on glucose (Nakari *et al.*, 1993; Nakari-Setälä and Penttilä, 1995). And just recently the use of the promoters of *pdv*, encoding

pyruvate decarboxylase, and *eno*, encoding enolase, was found to yield up to 80% of total protein secreted (Li *et al.*, 2012).

16.2.2 Protein glycosylation in *Trichoderma*

Besides obtaining high yields when producing heterologous proteins, it is crucial that these proteins are obtained in their active form. This is especially important for the production of antibodies or other pharmaceutically active products, where differences in glycosylation have a considerable effect on the therapeutic activity of a heterologously produced component. In order to reach the desired glycosylation patterns of a target product, a detailed understanding of the glycosylation pathways in *Trichoderma* is essential, albeit research in that direction is still in the early stages (reviewed by Nevalainen *et al.*, 2005). Only a few enzymes involved in the N-glycosylation pathway have been isolated and characterized, such as α -1-2-mannosidase, which can cleave four 1,2-linked mannose sugars from Man₉GlcNAc₂. This activity is indicative of Golgi-based mannosidases but the exact location of this enzyme in the cell has not been found yet (Maras *et al.*, 2000; Van Petegem *et al.*, 2001). Efforts towards reproducing mammalian glycan structures for proteins expressed in *Trichoderma* resulted in *in vitro* processing of the protein with mammalian enzymes (Maras *et al.*, 1997), as well as expression of mammalian genes, such as the human N-acetylglucosaminyltransferase I gene under the *T. reesei cbh1* promoter, which resulted in the formation of GlcNAcMan₅GlcNAc₂ (Maras *et al.*, 1999). In addition, overexpression of the *T. reesei* mannose-1-phosphate guanyltransferase has resulted in increased levels of GDP-mannose, thus securing the mannose supply for the formation of high-mannose mammalian-like glycan structures (Zakrzewska *et al.*, 2003b).

Studies on O-glycosylation in CBHI (Klarskov *et al.*, 1997; Maras *et al.*, 1997; Harrison *et al.*, 1998) showed that threonine and serine residues in the linker region between the catalytic and the substrate binding domains are connected to O-linked sugars

comprised mainly of one to three mannose units, as well as occasional glucose and galactose units (Harrison *et al.*, 1998, 2002). In addition, studies into the function of O-glycosylation in *T. reesei*, as well as in *Saccharomyces cerevisiae* and *Aspergillus niger var. awamori*, linked it to protein secretion, stability and localization to the cell wall (Kubicek *et al.*, 1987; Bourdineaud *et al.*, 1998; Zakrzewska *et al.*, 2003a; Goto *et al.*, 1999; Harty *et al.*, 2001), which makes further studies in this direction valuable for further improving the secretion of heterologous proteins.

16.2.3 Quality control, protein secretion and secretion stress

Before proteins can be secreted outside the cell, they have to undergo a process of folding, post-translational modification and quality control in the endoplasmic reticulum (ER), followed by either transport to the Golgi and final secretion or elimination by ER-associated degradation (ERAD). In the case of accumulation of unfolded proteins or proteins waiting for secretion in the ER, the unfolded protein response (UPR) is activated, which leads to transcriptional up-regulation of a number of genes encoding ER chaperones and foldases. In *T. reesei*, unfolded proteins in the ER bind to the chaperone BiP. Thereby, binding to the ER-associated IRE1 is abolished, which results in the induction of the UPR. This response in turn is mediated at the promoter level by the transcription factor HAC1 (reviewed by Nevalainen *et al.*, 2005; Saloheimo and Pakula, 2012).

Overexpression of IRE1 in *T. reesei* led to the induction of a number of genes involved in the protein processing pathways, such as the ER chaperone genes *bip1* and *lhs1*, the protein disulfide isomerase *pdi1*, the glycosylation pathway gene *pmi40* and the translocation channel gene *sec61*, and resulted in improved heterologous protein production in *S. cerevisiae* (Valkonen *et al.*, 2003a) and *A. niger var. awamori* (Valkonen *et al.*, 2003b). Nevertheless, in *T. reesei* strains expressing the *Phlebia radiata* laccase, the overexpression of *ire1* did not have any advantageous effects on the protein

production, even though the UPR was induced (Valkonen *et al.*, 2004). Similar results were obtained for overexpression of *hac1*, which led to increased production of a *Bacillus* α -amylase in *S. cerevisiae*, but had no effect on the expression of a heterologous laccase in *T. reesei* (Valkonen *et al.*, 2003a). Such varying effects of the overexpression of UPR-related genes might be dependent on the protein expressed or the strain used. In any case there are clearly regulatory mechanisms involved that could be important for an improvement of heterologous protein production.

In addition to the UPR, *Trichoderma* possesses a second way of dealing with large amounts of unfolded proteins in the ER, namely repression under secretion stress (RESS). It was shown that in cells treated with DTT, Brefeldin A or A23187, inhibiting protein folding and transport, transcript levels of extracellular genes, such as the major cellulase and xylanase genes, were down-regulated. Studies on the *cbh1* promoter with a *lacZ* reporter gene system showed, that this down-regulation takes place at the transcriptional level and is connected to the region upstream of -161 from the ATG triplet (Pakula *et al.*, 2003). Moreover, when comparing transcriptional patterns in a two-phase fermentation, where cells were first grown under repressing conditions and cellulose production was only induced after the exhaustion of glucose, the transcript levels of *pdi1* and *bip1* were found to follow the same temporal expression pattern as the cellulose-encoding genes (Collén *et al.*, 2005). Consequently, further studies are required to elucidate the molecular mechanisms involved in RESS and to find out whether it is connected to the UPR.

16.2.4 Hydrophobin fusion technology

Even though the high protein secretion capacity of *T. reesei* is decisive when choosing it as the target organism for heterologous protein production, that same secretion process can at times, especially in the case of more distantly related recombinant proteins, be detrimental to the protein yield. In addition, the presence of proteases in the culture supernatant can

further decrease protein stability. One solution to these obstacles would be an intracellular expression of the target proteins. This approach would, however, have serious disadvantages, such as an eventual toxicity of the heterologous protein to the host cell as well as a very problematic purification of the target protein. In order to overcome these complications, Mustalahti and co-workers introduced the use of hydrophobin fusions and protein body (PB) formation (Mustalahti *et al.*, 2011). To test this approach, green fluorescent protein (GFP) was expressed in *T. reesei* Rut-C30 (Montenecourt and Eveleigh, 1979) as a fusion protein to the class II hydrophobin HBF1 and the secretory signal peptide from CBHI, which was targeted to the ER and induced the formation of PB-like structures. A similar induction of PBs had already been observed using ZERA (a domain of maize γ -zein) as fusion partner (Torrent *et al.*, 2009). In addition to PB formation, the use of hydrophobin fusions constitutes a means for a rather simple and inexpensive purification of the target protein because the GFP-HBF1 fusion proteins were purified from the cell extract using a surfactant in an aqueous two-phase system, obtaining up to 62% of the target protein (Mustalahti *et al.*, 2011).

16.3 Recombinant Proteins Produced by *Trichoderma*

In industry, *Trichoderma* spp. are primarily used for the production of their homologous hydrolytic enzymes, in which case the secretory capacity of *Trichoderma* can be utilized to the fullest. In addition a number of heterologous enzymes, mostly from closely related donor organisms and expressed in either *T. reesei* or *Trichoderma longibrachiatum*, are produced efficiently so as to make them competitive (Table 16.1).

In addition to heterologous enzymes already produced on an industrial scale using *Trichoderma* as host organism, great efforts have been undertaken to express a number of other enzymes partly from more distantly related donor organisms, albeit still restricted to laboratory scale, with protein yields that need further optimization (Table 16.2).

Table 16.1. List of industrial enzymes heterologously produced in *T. reesei* or *T. longibrachiatum* (Amfep, 2009).

Enzyme	Donor organism
Alpha-Amylase	<i>Aspergillus</i> sp.
Aminopeptidase	<i>Aspergillus</i> sp.
Beta-Glucanase	<i>Trichoderma</i> sp.
Endo-1,4-beta Mannanase	<i>Trichoderma</i> sp.
Laccase	<i>Thielavia</i> sp.
Pectin methylesterase	<i>Aspergillus</i> sp.
Pectine lyase	<i>Aspergillus</i> sp.
Phospholipase A	<i>Aspergillus</i> sp.
Phospholipase A	<i>Thermomyces</i> sp.
Phospholipase B	<i>Aspergillus</i> sp.
Phytase	<i>Aspergillus</i> sp.
Phytase	<i>Buttiauxella</i> sp.
Polygalacturonase	<i>Aspergillus</i> sp.
Pullulanase	<i>Hormiconis</i> sp.
Xylanase	<i>Actinomadura</i> sp.

The first heterologous protein expressed in *T. reesei* was calf chymosin (Penttilä, 1998), which was expressed in *T. reesei* Rut-C30 under the *cbh1* promoter. The highest protein yields of more than 100 mg/l of active chymosin in shake-flask cultures were obtained through expression of a fusion protein of prochymosin with the CBHI core-linker region and optimization of the culture conditions (Harkki *et al.*, 1989; Uusitalo *et al.*, 1991). This work showed nicely how satisfactory protein yields are the result of integration of the target gene into favourable loci of the genome, such as the *cbh1* locus, as well as the use of the CBHI signal sequences to ascertain efficient translation, folding and secretion of the final product.

Not long after these promising results for the expression of a heterologous protein from a distantly related donor organism, first attempts to heterologously express an antibody in *Trichoderma* proved successful when Nyysönen and co-workers reported the expression of murine Fab fragments in *T. reesei* (Nyysönen *et al.*, 1993; Nyysönen and Keränen, 1995). These molecules consist of the light chain and the Fd part of the heavy chain, connected by a disulfide bridge. Expression of the gene encoding the light chain was achieved in Rut-C30, yielding 0.2 mg/l of product in shake-flask cultures. Subsequent introduction of the gene encoding the heavy

Table 16.2. Heterologous proteins expressed in *Trichoderma*.

Protein	Donor organism	Host strain	Protein (g/l)	Reference
Acid phosphatase	<i>Aspergillus niger</i>	ALKO2221	0.5 (shake flask)	Miettinen-Oinonen et al., 1997
Antibody Fab fragments	Murine	Rut-C30	0.15 (small-scale fermentor)	Nyssonen et al., 1993
Beta-Glucosidase	<i>Talaromyces emersonii</i>	Rut-C30	0.0027 (shake flask)	Murray et al., 2004
Chymosin	Calf	Rut-C30	0.1 (shake flask)	Harkki et al., 1989; Uusitalo et al., 1991
Cinnamoyl esterase EstA	<i>Piromyces equi</i>	Rut-C30	0.033 (shake flask)	Poidevin et al., 2009
Cutinase	<i>Coprinopsis cinerea</i>	D-00775 <i>cbh1</i> -neg	1.4 (small-scale fermentor)	Kontkanen et al., 2009
DewA	<i>Aspergillus nidulans</i>	QM9414	0.033 (small scale fermentor)	Schmoll et al., 2010
Endochitinase	<i>Trichoderma harzianum</i>	Rut-C30	0.13 (shake flask)	Margolles-Clark et al., 1996
Endopeptidase B	Barley	Rut-C30	0.05 (small-scale fermentor)	Saarelainen et al., 1997
Glucoamylase P	<i>Hormoconis resiniae</i>	ALKO2221	0.5 (shake flask)	Joutsjoki et al., 1993
Laccase	<i>Phlebia radiata</i>	Rut-C30	0.7 (shake flask)	Saloheimo and Niku-Paavola, 1991
Laccase	<i>Melanocarpus albomyces</i>	Rut-C30	0.02 (small-scale fermentor)	Kiiskinen et al., 2004
Steryl esterase	<i>Melanocarpus albomyces</i>	Rut-C30	0.23 (small-scale fermentor)	Kontkanen et al., 2006
Xylanase II	<i>Humicola grisea</i>	HEP1	0.076 (shake flask)	De Faria et al., 2002
Xyn VI	<i>Acrophialophora nainiana</i>	Rut-C30	0.5 (small-scale fermentor)	Salles et al., 2007
Xyn11A	<i>Nonomuraea flexuosa</i>	ALKO3620	0.172 (shake flask)	Paloheimo et al., 2003
			0.82 (small-scale fermentor)	

Adapted from Penttilä, 1998; Sharma et al., 2009; Peterson and Nevalainen, 2012.

Fd chain resulted in the secretion of fully functional antigen-binding Fab molecules on a scale of 1 mg/l in shake-flask culture. A further fusion of the CBHI core-linker region to the heavy Fd chain increased the protein yield even further, reaching 40 mg/l in shake-flask cultures and 150 mg/l in bioreactor cultivation. The decrease in immunoreactivity owing to the presence of the CBHI core-linker was reversed by cleavage of the CBHI part, releasing the fully functional Fab molecules (Nyyssönen *et al.*, 1993; Nyyssönen and Keränen, 1995)

Two further interesting examples of the introduction of heterologous proteins are the expression of endopeptidase B from barley under the control of the *cbh1* promoter (Saarelainen *et al.*, 1997) and cinnamoyl esterase from *Piromyces equi* (Poidevin *et al.*, 2009). On the one hand, the expression of endopeptidase B from barley proved once more that it is possible to express fully functional proteins from higher eukaryotes in *Trichoderma*. On the other hand, the expression of cinnamoyl esterase from *Piromyces equi* resulted in the production of EstA, a fully functional catalytic domain, capable of releasing ferulic acid, a valuable aromatic compound, from a variety of natural substrates such as wheat and maize bran.

16.4 *Trichoderma* as a Whole-cell Catalyst

In addition to using *T. reesei* solely for the production of proteins, recent efforts in metabolic engineering are aimed at the use of recombinant *T. reesei* strains for the production of valuable chemical components.

16.4.1 Ethylene

In 2010, researchers at the Agricultural University in Beijing reported the introduction of the ethylene-forming enzyme from *Pseudomonas syringae* into *T. reesei* QM9414. They obtained 14 stable transformants expressing the heterologous protein under the 3-phosphoglycerate kinase I (*pgk1*) promoter, the cellobiohydrolase I (*cbh1*) promoter,

as well as the glyceraldehyde-phosphate dehydrogenase (*gpd*) promoter from *A. nidulans*, all of which produced ethylene on a number of different carbon sources, such as cellulose, carboxymethyl cellulose (CMC) and wheat straw. The most promising ethylene producer was transformant C30-3 with an ethylene production of 4 ml/h/l of ethylene when cultivated on minimal medium with 2% wheat straw. These efforts showed nicely that it is possible to use agricultural wastes and a recombinant *T. reesei* strain to produce an important building block for the chemical industry (Chen *et al.*, 2010).

16.4.2 N-acetylneuraminic acid

One year later, Steiger and co-workers used a recombinant *T. reesei* strain to produce the antiviral drug precursor N-acetylneuraminic acid (NeuNAc) through cultivation on chitin (Steiger *et al.*, 2011a). NeuNAc belongs to the so-called sialic acids, a family of nine-carbon polyhydroxylated α -keto acids that are found as terminal residues of glycol-conjugates on the cell surface, where they are involved in mediating cellular recognition and cell adhesion processes (Varki, 1997; Angata and Varki, 2002; Tanner, 2005). Furthermore, they play an important role in the infection cycles of various viral diseases, for example influenza A and B (Herrler *et al.*, 1995; van der Merwe *et al.*, 1996; Brinkman-Van der Linden and Varki, 2000), which are dependent on a neuraminidase to cleave the bond between sialic acid and glycoprotein on the host cell surface. In order to prevent a further propagation of the virus, derivatives of sialic acids such as NeuNAc are used as neuraminidase inhibitors in antiviral pharmaceutical preparations (Steiger *et al.*, 2011a).

Traditional NeuNAc production processes, such as extraction from natural sources (Koketsu *et al.*, 1992), hydrolysis of colominic acid (Maru *et al.*, 2002), chemical synthesis (de Ninno, 1991), chemo-enzymatic (Mahmoudian *et al.*, 1997; Blayer *et al.*, 1999) and two-enzyme reaction processes (Kragl *et al.*, 1991; Maru *et al.*, 1998), as well as using *E. coli* in a whole-cell system (Tabata *et al.*, 2002; Lee *et al.*, 2007)

require the use of expensive additives, such as ATP, GlcNAc or an excess of pyruvate, leading to considerably high production costs for NeuNAc. The biosynthesis of NeuNAc usually starts out with N-acetylglucosamine (GlcNAc), which is converted into N-acetylmannosamine (ManNAc). Condensation of ManNAc with phosphoenolpyruvate (PEP) results in NeuNAc (Fig. 16.1; Tanner, 2005).

Chitin, one of the most prominent renewable resources (10^9 – 10^{11} tons/year), is composed of long chains of β -(1,4)-linked units of GlcNAc. In the case of *T. reesei*, genome-wide analysis of its chitinolytic enzyme system found 18 genes encoding proteins of the glycoside hydrolase family 18 (Perrakis *et al.*, 1994; Terwisscha van Scheltinga *et al.*, 1996; Robertus and Monzingo, 1999; Hollis *et al.*, 2000; Seidl *et al.*, 2005) and two proteins of glycoside hydrolase family 20 (Kubicek *et al.*, 2001). These results are indicative for a prominent chitinolytic potential of *T. reesei*, making it possible to utilize colloidal chitin as a cheap renewable carbon source for fermentation. In addition

the release of GlcNAc further enables the biosynthesis of NeuNAc with a recombinant *Trichoderma* strain.

In order to engineer *T. reesei* to produce NeuNAc, two bacterial enzymes, namely a GlcNAc-2-epimerase from *Anabaena sp.* CH1 and a NeuNAc synthase from *Campylobacter jejuni*, both in *T. reesei* codon-optimized form, were successfully introduced into *T. reesei* QM9414 (Fig. 16.2; Steiger *et al.*, 2011a).

Both enzymes were shown to be functional and the recombinant *T. reesei* strain was shown to produce NeuNAc using either GlcNAc or colloidal chitin as a carbon source on a scale of several μg NeuNAc per g mycelium (dry weight) (Steiger *et al.*, 2011a).

Even though these results do not constitute a competitive production process, it was nevertheless shown that it is possible to engineer a whole bacterial enzyme cascade into a saprophytic fungus and use the resulting recombinant strain as a whole-cell catalyst for the production of a valuable chemical building block for the pharmaceutical industry

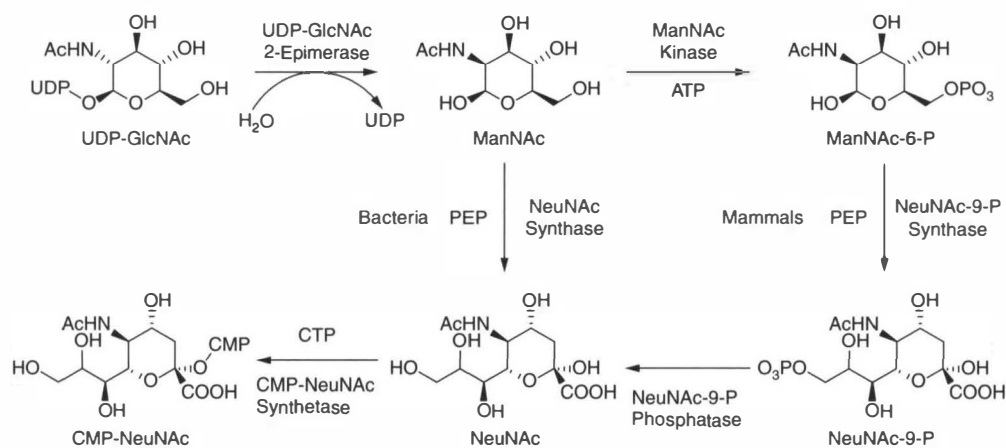


Fig. 16.1. The biosynthesis of NeuNAc.

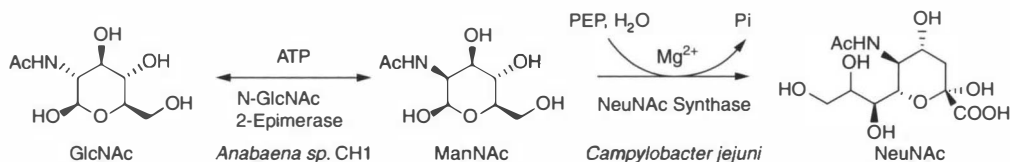


Fig. 16.2. The NeuNAc synthesis pathway engineered into *T. reesei*.

from an abundant starting material and using a simple and cheap cultivation method (Steiger *et al.*, 2011a).

16.5 Conclusion

Recombinant *Trichoderma* strains are successfully being used in industry for the production of homologous as well as heterologous proteins from closely related donor organisms. Research into relevant areas such as gene regulation, protein processing and secretion connected to improving the protein yield for industrial strains has resulted in the development of all the necessary tools to introduce foreign genes from more distantly related organisms into *Trichoderma* and recover the resulting proteins in a correctly folded and fully functional form (Nevalainen *et al.*, 2005; Schuster and Schmoll, 2010).

Moreover, recent efforts attempting to use *Trichoderma* as a whole-cell catalyst returned promising results for the production of valuable chemical building blocks using cheap, renewable resources as growth medium (Chen *et al.*, 2010; Steiger *et al.*, 2011a). One of the obvious benefits of using a saprophyte in a whole-cell approach is the fact that there is no need to add expensive co-factors for enzyme reactions, which as a consequence facilitates a simplified downstream processing. In addition, if a saprophyte is used as the production host, the use of cheap, renewable resources as growth media and the general lack of harsh chemicals in biotechnological applications support the general trend towards sustainable and environmentally friendly production processes. These characteristics, combined with the years of experience concerning cultivation conditions and improvement of product yields, make research into *Trichoderma*-based production systems a promising endeavour.

References

- Amfep (2009) *List of commercial enzymes*. Available at: <http://www.amfep.org/content/list-enzymes> (Accessed 5 November 2012).
- Angata, T. and Varki, A. (2002) Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chemical Reviews* 102, 439–469.
- Aro, N., Pakula, T. and Penttilä, M. (2005) Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiology Reviews* 29, 719–739.
- Blayer, S., Woodley, J.M., Dawson, M.J. and Lilly, M.D. (1999) Alkaline biocatalysis for the direct synthesis of N-acetyl-D-neuraminic acid (Neu5Ac) from N-acetyl-D-glucosamine (GlcNAc). *Biotechnology and Bioengineering* 66, 131–136.
- Bourdineaud, J.P., Van Der Vaart, J.M., Donzeau, M., De Sampaio, G., Verrips, C.T. and Lauquin, G.J. (1998) Pmt1 mannosyl transferase is involved in cell wall incorporation of several proteins in *Saccharomyces cerevisiae*. *Molecular Microbiology* 27, 85–98.
- Brinkman-Van Der Linden, E.C. and Varki, A. (2000) New aspects of siglec binding specificities, including the significance of fucosylation and of the sialyl-Tn epitope. Sialic acid-binding immunoglobulin superfamily lectins. *Journal of Biological Chemistry* 275, 8625–8632.
- Chen, X., Liang, Y., Hua, J., Tao, L., Qin, W. and Chen, S. (2010) Overexpression of bacterial ethylene-forming enzyme gene in *Trichoderma reesei* enhanced the production of ethylene. *International Journal of Biological Sciences* 6, 96–106.
- Collén, A., Saloheimo, M., Bailey, M., Penttilä, M. and Pakula, T.M. (2005) Protein production and induction of the unfolded protein response in *Trichoderma reesei* strain Rut-C30 and its transformant expressing endoglucanase I with a hydrophobic tag. *Biotechnology and Bioengineering* 89, 335–344.
- De Faria, F.P., Te'o, V.S.J., Bergquist, P.L., Azevedo, M.O. and Nevalainen, K.M.H. (2002) Expression and processing of a major xylanase (XYN2) from the thermophilic fungus *Humicola grisea* var. *thermoidea* in *Trichoderma reesei*. *Letters in Applied Microbiology* 34, 119–123.
- De Ninno, M. (1991) The synthesis and glycosidation of N-acetyl-D-neuraminic acid. *Synthesis* 8, 583–593.

- Goto, M., Tsukamoto, M., Kwon, I., Ekino, K. and Furukawa, K. (1999) Functional analysis of O-linked oligosaccharides in threonine/serine-rich region of *Aspergillus* glucoamylase by expression in mannosyltransferase-disruptants of yeast. *European Journal of Biochemistry / FEBS* 260, 596–602.
- 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. *Current Genetics* 18, 71–76.
- Harkki, A., Uusitalo, J., Bailey, M., Penttilä, M. and Knowles, J.K.C. (1989) A novel fungal expression system: secretion of active calf chymosin from the filamentous fungus *Trichoderma reesei*. *Bio/technology* 7, 596–603.
- Harrison, M.J., Nouwens, A.S., Jardine, D.R., Zachara, N.E., Gooley, A.A., Nevalainen, H. and Packer, N.H. (1998) Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. *European Journal of Biochemistry / FEBS* 256, 119–127.
- Harrison, M.J., Wathugala, I.M., Tenkanen, M., Packer, N.H. and Nevalainen, K.M. (2002) Glycosylation of acetylxyylan esterase from *Trichoderma reesei*. *Glycobiology* 12, 291–298.
- Harty, C., Strahl, S. and Romisch, K. (2001) O-mannosylation protects mutant alpha-factor precursor from endoplasmic reticulum-associated degradation. *Molecular Biology of the Cell* 12, 1093–1101.
- Herrler, G., Hausmann, J. and Klenk, H.D. (1995) Sialic acid as receptor determinant of ortho- and paromyxoviruses. In: Rosenberg, A. (ed.) *Biology of the Sialic Acids*. Plenum Press, NY and London.
- Hollis, T., Monzingo, A.F., Bortone, K., Ernst, S., Cox, R. and Robertus, J.D. (2000) The X-ray structure of a chitinase from the pathogenic fungus *Coccidioides immitis*. *Protein Science* 9, 544–551.
- Joutsjoki, V.V., Torkkeli, T.K. and Nevalainen, K.M. (1993) Transformation of *Trichoderma reesei* with the *Hormoconis resiniae* glucoamylase P (*gamP*) gene: production of a heterologous glucoamylase by *Trichoderma reesei*. *Current Genetics* 24, 223–228.
- Keränen, S. and Penttilä, M. (1995) Production of recombinant proteins in the filamentous fungus *Trichoderma reesei*. *Current Opinion in Biotechnology*, 6, 534–537.
- Kiiskinen, L.L., Kruus, K., Bailey, M., Ylösmäki, E., Siika-Aho, M. and Saloheimo, M. (2004) Expression of *Melanocarpus albomyces* laccase in *Trichoderma reesei* and characterization of the purified enzyme. *Microbiology* 150, 3065–3074.
- Klarskov, K., Piens, K., Stahlberg, J., Hoj, P.B., Beeumen, J.V. and Claeysens, M. (1997) Cellobiohydrolase I from *Trichoderma reesei*: identification of an active-site nucleophile and additional information on sequence including the glycosylation pattern of the core protein. *Carbohydrate Research* 304, 143–154.
- Koketsu, M., Juneja, L.R., Kawanami, H., Kim, M. and Yamamoto, T. (1992) Preparation of N-acetylneuraminic acid from delipidated egg yolk. *Glycoconjugate Journal* 9, 70–74.
- Kontkanen, H., Reinikainen, T. and Saloheimo, M. (2006) Cloning and expression of a *Melanocarpus albomyces* steryl esterase gene in *Pichia pastoris* and *Trichoderma reesei*. *Biotechnology and Bioengineering* 94, 407–415.
- Kontkanen, H., Westerholm-Parvinen, A., Saloheimo, M., Bailey, M., Rättö, M., Mattila, I., Mohsina, M., Kalkkinen, N., Nakari-Setälä, T. and Buchert, J. (2009) Novel *Coprinopsis cinerea* polyesterase that hydrolyzes cutin and suberin. *Applied and Environmental Microbiology* 75, 2148–2157.
- Kragl, U., Gygax, D., Ghisalba, O. and Wandrey, C. (1991) Enzymatic two step synthesis of N-acetylneuraminic acid in the enzyme membrane reactor. *Angewandte Chemie International Edition English* 30, 827–828.
- Kubicek, C.P., Panda, T., Schrefel-Kunar, G., Gruber, F. and Messner, R. (1987) O-linked but not N-linked glycosylation is necessary for the secretion of endoglucanases I and II by *Trichoderma reesei*. *Canadian Journal of Microbiology* 33, 698–703.
- Kubicek, C.P., Mach, R.L., Peterbauer, C.K. and Lorito, M. (2001) *Trichoderma*: from genes to biocontrol. *Journal of Plant Pathology* 83, 11–23.
- Kubicek, C.P., Mikus, M., Schuster, A., Schmoll, M. and Seiboth, B. (2009) Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnology for Biofuels* 2, 19.
- Lee, Y.C., Chien, H.C. and Hsu, W.H. (2007) Production of N-acetyl-D-neuraminic acid by recombinant whole cells expressing *Anabaena* sp. CH1 N-acetyl-D-glucosamine 2-epimerase and *Escherichia coli* N-acetyl-D-neuraminic acid lyase. *Journal of Biotechnology* 129, 453–460.
- Li, J., Wang, J., Wang, S., Xing, M., Yu, S. and Liu, G. (2012) Achieving efficient protein expression in *Trichoderma reesei* by using strong constitutive promoters. *Microbial Cell Factories* 11, 84.
- Mach, R.L. and Zeilinger, S. (2003) Regulation of gene expression in industrial fungi: *Trichoderma*. *Applied Microbiology and Biotechnology* 60, 515–522.
- Mach, R.L., Schindler, M. and Kubicek, C.P. (1994) Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expression signals. *Current Genetics* 25, 567–570.

- Mahmoudian, M., Noble, D., Drake, C.S., Middleton, R.F., Montgomery, D.S., Piercey, J.E., Ramlakhan, D., Todd, M. and Dawson, M.J. (1997) An efficient process for production of N-acetylneuraminic acid using N-acetylneuraminic acid aldolase. *Enzyme and Microbial Technology* 20, 393–400.
- Mäntylä, A., Paloheimo, M. and Suominen, P. (1998) Industrial mutants and recombinant strains of *Trichoderma reesei*. In: Harman, G.E. and Kubicek, C.P. (eds) *Trichoderma and Gliocladium*. Taylor & Francis Ltd, London.
- Maras, M., De Bruyn, A., Schraml, J., Herdewijn, P., Claeysens, M., Fiers, W. and Contreras, R. (1997) Structural characterization of N-linked oligosaccharides from cellobiohydrolase I secreted by the filamentous fungus *Trichoderma reesei* RUTC 30. *European Journal of Biochemistry / FEBS* 245, 617–625.
- Maras, M., De Bruyn, A., Vervecken, W., Uusitalo, J., Penttilä, M., Busson, R., Herdewijn, P. and Contreras, R. (1999) *In vivo* synthesis of complex N-glycans by expression of human N-acetylglucosaminyltransferase I in the filamentous fungus *Trichoderma reesei*. *FEBS Letters* 452, 365–370.
- Maras, M., Callewaert, N., Piens, K., Claeysens, M., Martinet, W., Dewaele, S., Contreras, H., Dewerte, I., Penttilä, M. and Contreras, R. (2000) Molecular cloning and enzymatic characterization of a *Trichoderma reesei* 1,2- α -D-mannosidase. *Journal of Biotechnology* 77, 255–263.
- Margolles-Clark, E., Hayes, C.K., Harman, G.E. and Penttilä, M. (1996) Improved production of *Trichoderma harzianum* endochitinase by expression in *Trichoderma reesei*. *Applied and Environmental Microbiology* 62, 2145–2151.
- Maru, I., Ohnishi, J., Ohta, Y. and Tsukada, Y. (1998) Simple and large-scale production of N-acetylneuraminic acid from N-acetyl-D-glucosamine and pyruvate using N-acetyl-D-glucosamine 2-epimerase and N-acetylneuraminase lyase. *Carbohydrate Research* 306, 575–578.
- Maru, I., Ohnishi, J., Ohta, Y. and Tsukada, Y. (2002) Why is sialic acid attracting interest now? Complete enzymatic synthesis of sialic acid with N-acetylglucosamine 2-epimerase. *Journal of Bioscience and Bioengineering* 93, 258–265.
- Miettinen-Oinonen, A., Torkkeli, T., Paloheimo, M. and Nevalainen, H. (1997) Overexpression of the *Aspergillus niger* pH 2.5 acid phosphatase gene in a heterologous host *Trichoderma reesei*. *Journal of Biotechnology* 58, 13–20.
- Montenecourt, B.S. and Eveleigh, D.E. (1979) Selective screening methods for the isolation of high yielding cellulase mutants of *Trichoderma reesei*. In: *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis*. American Chemical Society, Washington DC, pp. 289–301.
- Murray, P., Aro, N., Collins, C., Grassick, A., Penttilä, M., Saloheimo, M. and Tuohy, M. (2004) Expression in *Trichoderma reesei* and characterisation of a thermostable family 3 beta-glucosidase from the moderately thermophilic fungus *Talaromyces emersonii*. *Protein Expression and Purification* 38, 248–257.
- Mustalahti, E., Saloheimo, M. and Joensuu, J.J. (2011) Intracellular protein production in *Trichoderma reesei* (*Hypocrea jecorina*) with hydrophobin fusion technology. *Nature Biotechnology* 30, 262–268.
- Nakari, T., Alatalo, E. and Penttilä, M.E. (1993) Isolation of *Trichoderma reesei* genes highly expressed on glucose-containing media: characterization of the *tef1* gene encoding translation elongation factor 1 alpha. *Gene*, 136, 313–318.
- Nakari-Setälä, T. and Penttilä, M. (1995) Production of *Trichoderma reesei* cellulases on glucose-containing media. *Applied and Environmental Microbiology*, 61, 3650–3655.
- Nevalainen, H., Te'o, V., Penttilä, M.J. and Pakula, T. (2005) Heterologous gene expression in filamentous fungi: A holistic view. In: Dilip, K.A. and Randy, M.B. (eds) *Applied Mycology and Biotechnology*. Elsevier, the Netherlands.
- Nyyssönen, E. and Keränen, S. (1995) Multiple roles of the cellulase CBHI in enhancing production of fusion antibodies by the filamentous fungus *Trichoderma reesei*. *Current Genetics* 28, 71–79.
- Nyyssönen, E., Penttilä, M., Harkki, A., Saloheimo, A., Knowles, J.K. and Keränen, S. (1993) Efficient production of antibody fragments by the filamentous fungus *Trichoderma reesei*. *Bio/technology*, 11, 591–595.
- Pakula, T.M., Laxell, M., Huuskonen, A., Uusitalo, J., Saloheimo, M. and Penttilä, M. (2003) The effects of drugs inhibiting protein secretion in the filamentous fungus *Trichoderma reesei*. Evidence for down-regulation of genes that encode secreted proteins in the stressed cells. *Journal of Biological Chemistry* 278, 45011–45020.
- Paloheimo, M., Mäntylä, A., Kallio, J. and Suominen, P. (2003) High-yield production of a bacterial xylanase in the filamentous fungus *Trichoderma reesei* requires a carrier polypeptide with an intact domain structure. *Applied and Environmental Microbiology* 69, 7073–7082.
- Penttilä, M. (1998) Heterologous protein production in *Trichoderma*. In: Harman, G.E. and Kubicek, C.P. (eds) *Trichoderma and Gliocladium*. Taylor & Francis Ltd, London.

- Penttilä, M., Nevalainen, H., Ratto, M., Salminen, E. and Knowles, J. (1987) A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. *Gene* 61, 155–164.
- Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A.B., Chet, I., Wilson, K.S. and Vorgias, C.E. (1994) Crystal structure of a bacterial chitinase at 2.3 Å resolution. *Structure* 2, 1169–1180.
- Peterbauer, C.K., Heidenreich, E., Baker, R.T. and Kubicek, C.P. (1992) Effect of benomyl and benomyl resistance on cellulase formation by *Trichoderma reesei* and *Trichoderma harzianum*. *Canadian Journal of Microbiology* 38, 1292–1297.
- Peterson, R. and Nevalainen, H. (2012) *Trichoderma reesei* RUT-C30—thirty years of strain improvement. *Microbiology*, 158, 58–68.
- Poidevin, L., Levasseur, A., Paes, G., Navarro, D., Heiss-Blanquet, S., Asther, M. and Record, E. (2009) Heterologous production of the *Piromyces equi* cinnamoyl esterase in *Trichoderma reesei* for biotechnological applications. *Letters in Applied Microbiology* 49, 673–678.
- Robertus, J.D. and Monzingo, A.F. (1999) The structure and action of chitinases. *EXS* 87, 125–135.
- Saarelainen, R., Mantyla, A., Nevalainen, H. and Suominen, P. (1997) Expression of barley endopeptidase B in *Trichoderma reesei*. *Applied and Environmental Microbiology* 63, 4938–4940.
- Salles, B.C., Te'o, V.S., Gibbs, M.D., Bergquist, P.L., Filho, E.X., Ximenes, E.A. and Nevalainen, K.M. (2007) Identification of two novel xylanase-encoding genes (*xyn5* and *xyn6*) from *Acrophialophora nainiana* and heterologous expression of *xyn6* in *Trichoderma reesei*. *Biotechnology Letters* 29, 1195–1201.
- Saloheimo, M. and Niku-Paavola, M.-L. (1991) Heterologous production of a ligninolytic enzyme: expression of the *Phlebia radiata* laccase gene in *Trichoderma reesei*. *Bio/technology* 9, 987–990.
- Saloheimo, M. and Pakula, T.M. (2012) The cargo and the transport system: secreted proteins and protein secretion in *Trichoderma reesei* (*Hypocrea jecorina*). *Microbiology* 158, 46–57.
- Schindler, M., Mach, R.L., Vollenhofer, S.K., Hodits, R., Gruber, F., Visser, J., De Graaff, L. and Kubicek, C.P. (1993) Characterization of the pyruvate kinase-encoding gene (*pk1*) of *Trichoderma reesei*. *Gene* 130, 271–275.
- Schmoll, M., Seibel, C., Kotlowski, C., Wollert Genannt Vendt, F., Liebmann, B. and Kubicek, C.P. (2010) Recombinant production of an *Aspergillus nidulans* class I hydrophobin (DewA) in *Hypocrea jecorina* (*Trichoderma reesei*) is promoter-dependent. *Applied Microbiology and Biotechnology* 88, 95–103.
- Schuster, A. and Schmoll, M. (2010) Biology and biotechnology of *Trichoderma*. *Applied Microbiology and Biotechnology* 87, 787–799.
- Seiboth, B., Gamauf, C., Pail, M., Hartl, L. and Kubicek, C.P. (2007) The D-xylose reductase of *Hypocrea jecorina* is the major aldose reductase in pentose and D-galactose catabolism and necessary for beta-galactosidase and cellulase induction by lactose. *Molecular Microbiology* 66, 890–900.
- Seidl, V., Huemer, B., Seiboth, B. and Kubicek, C.P. (2005) A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS Journal* 272, 5923–5939.
- Sharma, R., Katoch, M., Srivastava, P. and Qazi, G. (2009) Approaches for refining heterologous protein production in filamentous fungi. *World Journal of Microbiology and Biotechnology* 25, 2083–2094.
- Steiger, M.G., Mach-Aigner, A.R., Gorsche, R., Rosenberg, E.E., Mihovilovic, M.D. and Mach, R.L. (2011a) Synthesis of an antiviral drug precursor from chitin using a saprophyte as a whole-cell catalyst. *Microbial Cell Factories*, 10, 102.
- Steiger, M.G., Vitikainen, M., Uskonen, P., Brunner, K., Adam, G., Pakula, T., Penttilä, M., Saloheimo, M., Mach, R.L. and Mach-Aigner, A.R. (2011b) Transformation system for *Hypocrea jecorina* (*Trichoderma reesei*) that favors homologous integration and employs reusable bidirectionally selectable markers. *Applied and Environmental Microbiology* 77, 114–121.
- Tabata, K., Koizumi, S., Endo, T. and Ozaki, A. (2002) Production of N-acetyl-D-neuraminic acid by coupling bacteria expressing N-acetyl-D-glucosamine 2-epimerase and N-acetyl-D-neuraminic acid synthetase. *Enzyme and Microbial Technology* 30, 327–333.
- Tanner, M.E. (2005) The enzymes of sialic acid biosynthesis. *Bioorganic Chemistry* 33, 216–228.
- Terwisscha Van Scheltinga, A.C., Hennig, M. and Dijkstra, B.W. (1996) The 1.8 Å resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. *Journal of Molecular Biology* 262, 243–257.
- Torrent, M., Llopart, B., Lasserre-Ramassamy, S., Llop-Tous, I., Bastida, M., Marzabal, P., Westerholm-Parvinen, A., Saloheimo, M., Heifetz, P.B. and Ludevid, M.D. (2009) Eukaryotic protein production in designed storage organelles. *BMC Biology* 7, 5.
- Uusitalo, J.M., Nevalainen, K.M., Harkki, A.M., Knowles, J.K. and Penttilä, M.E. (1991) Enzyme production by recombinant *Trichoderma reesei* strains. *Journal of Biotechnology* 17, 35–49.
- Valkonen, M., Penttilä, M. and Saloheimo, M. (2003a) Effects of inactivation and constitutive expression of the unfolded-protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology* 69, 2065–2072.

- Valkonen, M., Ward, M., Wang, H., Penttilä, M. and Saloheimo, M. (2003b) Improvement of foreign-protein production in *Aspergillus niger* var. *awamori* by constitutive induction of the unfolded-protein response. *Applied and Environmental Microbiology* 69, 6979–6986.
- Valkonen, M., Penttilä, M. and Saloheimo, M. (2004) The *ire1* and *ptc2* genes involved in the unfolded protein response pathway in the filamentous fungus *Trichoderma reesei*. *Molecular Genetics and Genomics* 272, 443–451.
- Van Der Merwe, P.A., Crocker, P.R., Vinson, M., Barclay, A.N., Schauer, R. and Kelm, S. (1996) Localization of the putative sialic acid-binding site on the immunoglobulin superfamily cell-surface molecule CD22. *Journal of Biological Chemistry* 271, 9273–9280.
- Van Petegem, F., Contreras, H., Contreras, R. and Van Beeumen, J. (2001) *Trichoderma reesei* alpha-1,2-mannosidase: structural basis for the cleavage of four consecutive mannose residues. *Journal of Molecular Biology* 312, 157–165.
- Vanhanen, S., Penttilä, M., Lehtovaara, P. and Knowles, J. (1989) Isolation and characterization of the 3-phosphoglycerate kinase gene (*pgk*) from the filamentous fungus *Trichoderma reesei*. *Current Genetics* 15, 181–186.
- Varki, A. (1997) Sialic acids as ligands in recognition phenomena. *FASEB Journal* 11, 248–255.
- Zakrzewska, A., Migdalski, A., Saloheimo, M., Penttilä, M.E., Palamarczyk, G. and Kruszewska, J.S. (2003a) cDNA encoding protein O-mannosyltransferase from the filamentous fungus *Trichoderma reesei*; functional equivalence to *Saccharomyces cerevisiae* PMT2. *Current Genetics* 43, 11–16.
- Zakrzewska, A., Palamarczyk, G., Krotkiewski, H., Zdebska, E., Saloheimo, M., Penttilä, M. and Kruszewska, J.S. (2003b) Overexpression of the gene encoding GTP:mannose-1-phosphate guanyltransferase, *mpg1*, increases cellular GDP-mannose levels and protein mannosylation in *Trichoderma reesei*. *Applied and Environmental Microbiology* 69, 4383–4389.

CURRICULUM VITAE

DI Rita Gorsche

Ausbildung:

- Seit 10/2009 Dissertation an der Technischen Universität Wien zum Thema
„*Trichoderma reesei*: new insights into gene regulation and potential use
as cell factory“
- 08/2011-12/2011 Auslandsaufenthalt zu Forschungszwecken bei IOGEN Corp, Ottawa,
Canada
- 10/2009 Abschluss des Diplomstudiums Technische Chemie mit Auszeichnung
- 10/2008-10/2009 Diplomarbeit an der Technischen Universität Wien zum Thema
„Einbringen eines N-Azetyl-Neuraminsäure Syntheseweges in
Trichoderma reesei“
- 10/2003-10/2009 Studium der Technischen Chemie (Spezialisierung: Biotechnologie und
Bioanalytik) an der Technischen Universität Wien
- 02/2009-06/2009 Auslandsaufenthalt an der Technischen Universität Perm, Russische
Föderation, im Rahmen des „Joint Study“ Programms der Technischen
Universität Wien
- 09/2002-06/2003 Auslandsaufenthalt an der EF International Language School, Santa
Barbara, USA
- 09/1994-06/2002 BG/BRG Mattersburg: Matura mit Auszeichnung bestanden

LIST OF PUBLICATIONS

Research articles (peer-reviewed)

Gorsche, R., Jovanovic, B., Gudynaite-Savitch, L., Mach, R. L., Mach-Aigner, A. R. (2013) A highly sensitive *in vivo* footprinting technique for condition-dependent identification of *cis* elements. *Nucleic Acids Res.* 1-12

Steiger, M.G., Mach-Aigner, A.R., **Gorsche, R.**, Rosenberg, E.E., Mihovilovic, M.D., Mach, R.L. (2011) Synthesis of an antiviral drug precursor from chitin using a saprophyte as a whole-cell catalyst. *Microb Cell Fact.* 10:102.

Review articles

Gorsche, R., Mach-Aigner, A. R., Mach, R. L. (2013) *Trichoderma* as cell factories. In: Mukherjee, P., Horwitz, B., *et al.* (eds) *Trichoderma: Biology and Applications*, CABI, Wallingford, UK; pp 280-291

Presentations

- talks

Gorsche, R., Derntl, C., Mach, R.L., Mach-Aigner, A.R. (2012) Improving *in vivo* footprinting – finding a more sensitive way to detect protein-DNA interactions. The 12th International *Trichoderma* & *Gliocladium* Workshop. Lincoln University, New Zealand; August 27-30, 2012

Gorsche, R., Mach-Aigner, A.R., Mach, R.L. (2010) Transcriptional regulation of cellulolytic and hemicellulolytic enzymes in *Trichoderma reesei*. Topical Issues of Rational Use of Natural Resources. St. Petersburg State Mining University, Russia; April 21-23, 2010

- Poster presentations

Gorsche, R., Mach, R.L., Mach-Aigner, A.R. (2012) High throughput *in vivo* footprinting: an improved method to detect protein-DNA interactions. ECFG11. Marburg, Germany; March 30 – April 2, 2012

Gorsche, R., Mach, R.L., Mach-Aigner A.R. (2011) A high throughput *in vivo* footprinting technique to detect protein-DNA interactions. Einsteins in the City. New York City College, USA; April 14-15, 2011