

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 Garry 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	CAATACACAGAGGTGATCTTAC	<i>iv</i> -FP, <i>cbh2</i> oligo 1 rev
RG99	CATTAGCCTCAAGTAGAGCCTATTTCTC	<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]GCCTCAAGTAGAGCCTATTTCTCGCC	<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	GGAGTTGTTGTGCTTTTGGGCTTG	<i>iv</i> -FP, <i>xyn2</i> oligo 2 rev
RG131	[6-FAM]CCGTTATTCAGACAATGTATGTGCCGGG	<i>iv</i> -FP, <i>xyn2</i> oligo 3 fw
RG132	[6-FAM]GTTGTTGTGCTTTTGGGCTTGAGGG	<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	CCGAGAAGTTGATGACCTTGTT	<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

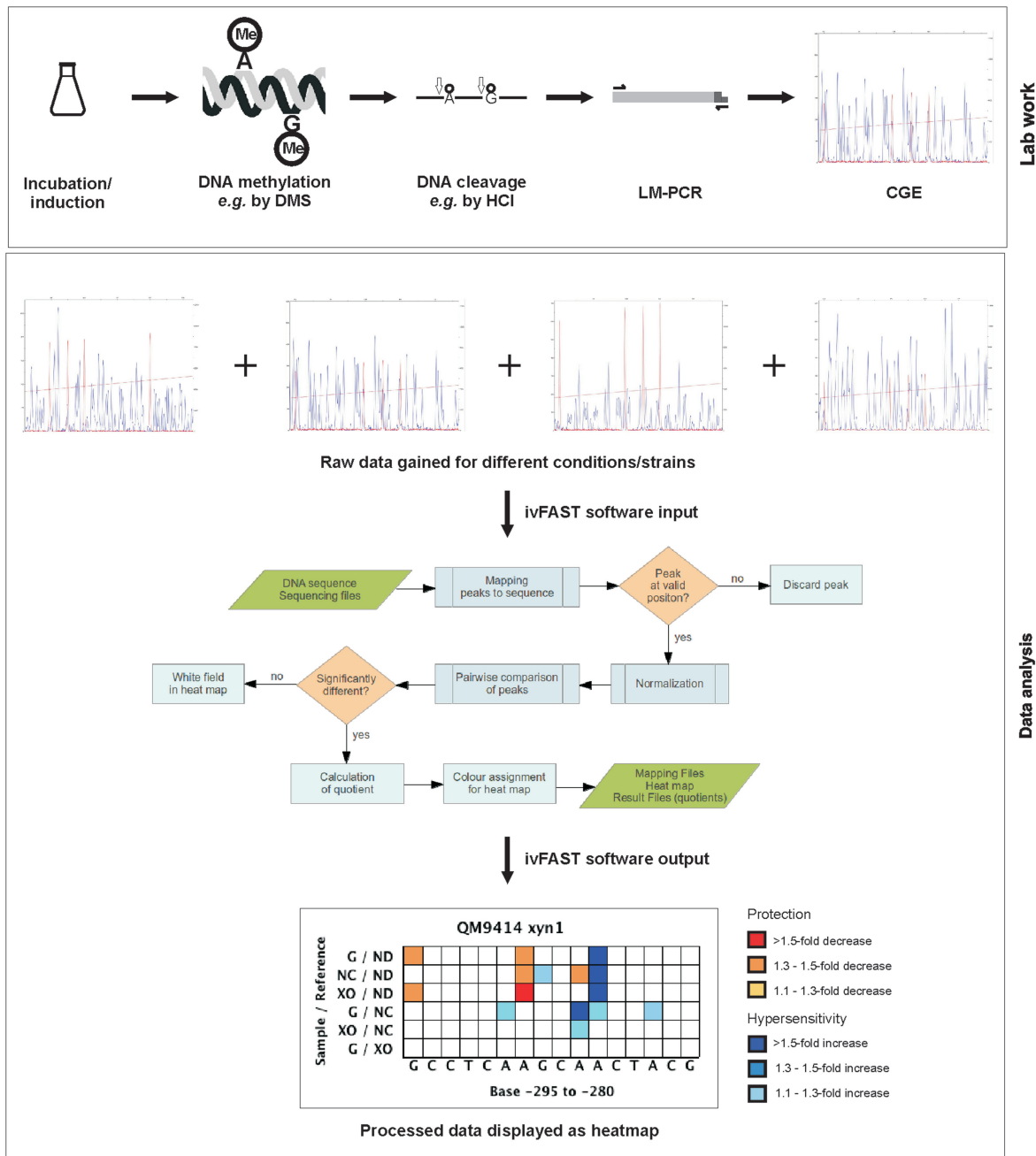


Figure 1. Schematic presentation of the workflow and generation of final data. The main steps of the software-based, high-throughput *in vivo* footprinting method comprise growing/incubating the microorganism under conditions to be investigated (e.g. inducing conditions), *in vivo* DNA methylation using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

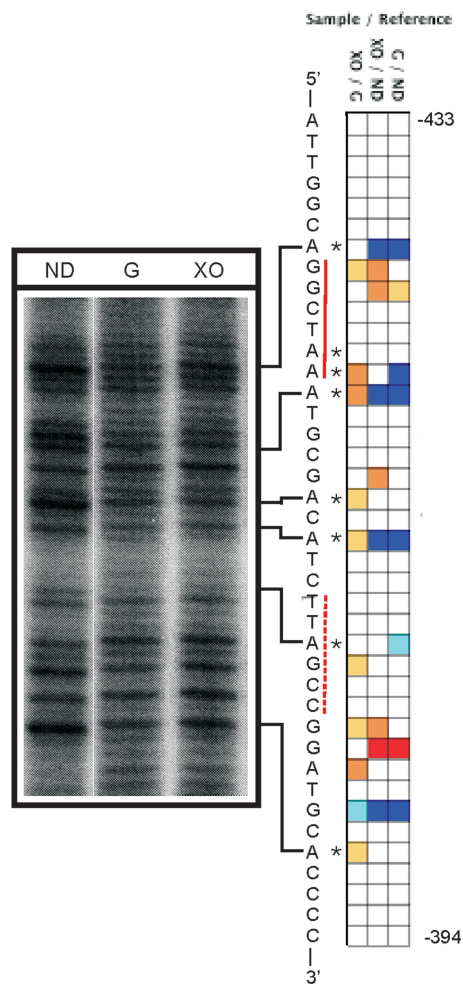


Figure 2. Comparison of the traditional *in vivo* footprinting to the newly developed method. *In vivo* footprinting analysis of the coding strand of a *xyn1* URR (–433- to –394-bp upstream from ATG) covering two Xyr1-binding sites, which are indicated by red lines (solid, site is located on the coding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

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 *xyn1* deletion strain (Figure 4b, $\Delta xyn1$ -G/ $\Delta xyn1$ -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

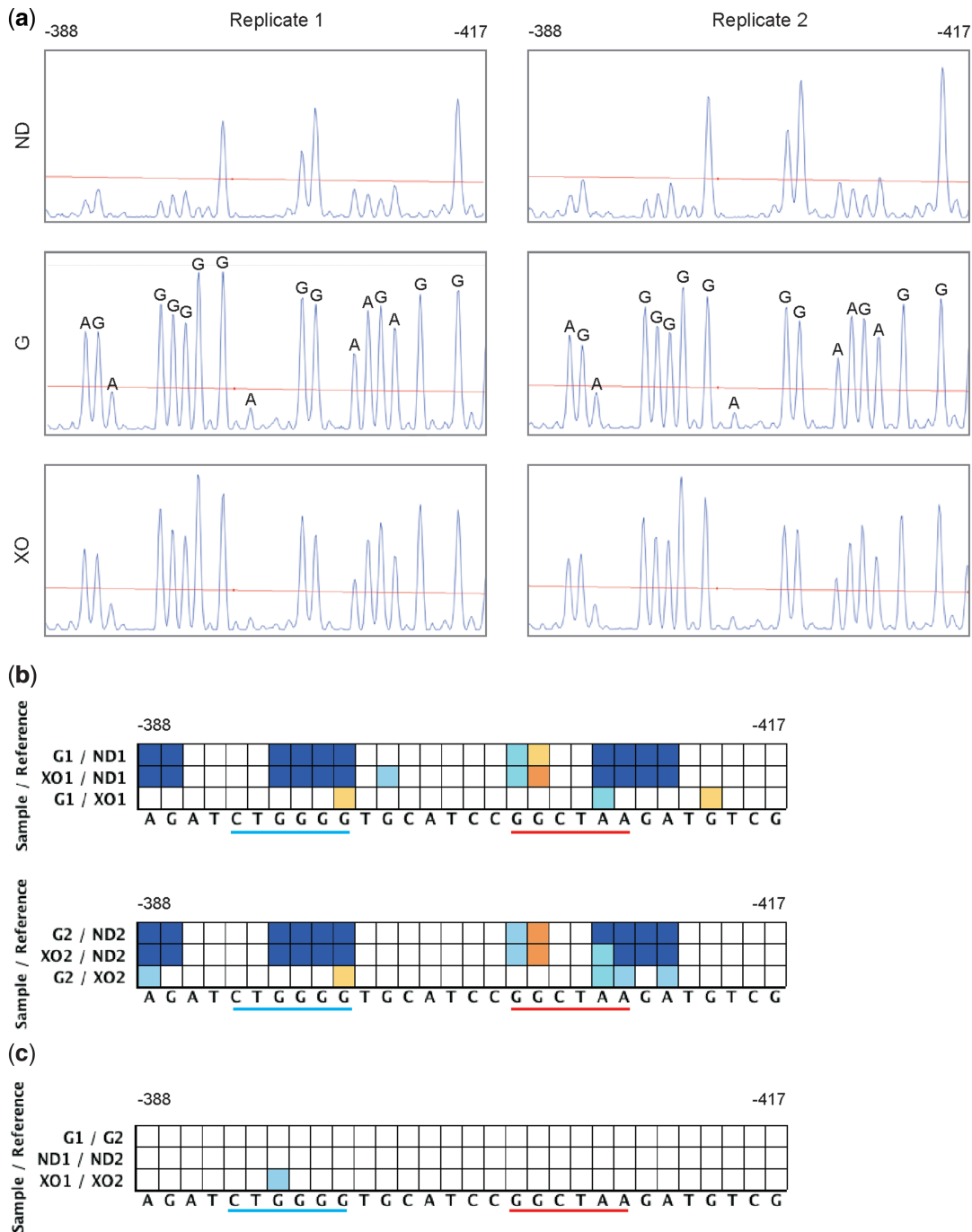


Figure 3. Comparison of two biological replicates analysed by the newly developed *in vivo* footprinting method. *In vivo* footprinting analysis of the non-coding strand of a *xynI* URR (–388- to –417-bp upstream from ATG) covering a Cre1-binding site (underlined in blue) and a Xyr1-binding site (underlined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

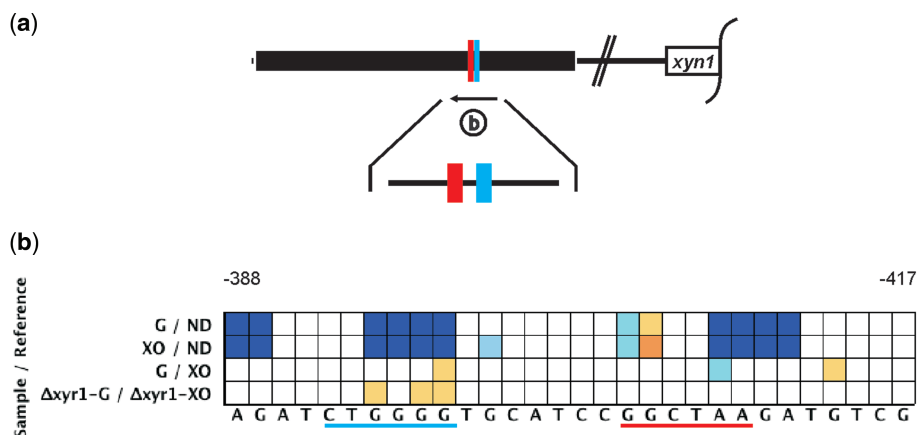


Figure 4. Comparison of *in vivo* footprinting results of a deletion and a parental strain. (a) *In vivo* footprinting analysis of a *xyn1* URR (–388- to –417-bp upstream from ATG) covering a Cre1-binding site (underlined in blue) and a Xyr1-binding site (underlined in red) was performed. (b) The non-coding 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).

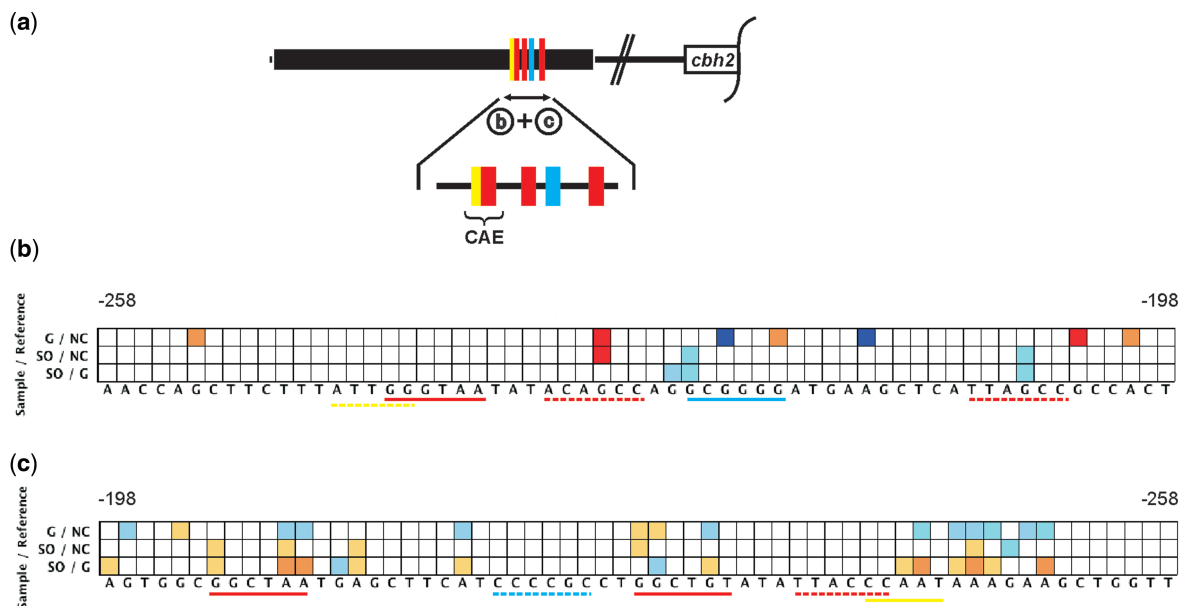


Figure 5. *In vivo* footprinting analysis of the *T. reesei* *cbh2* URR. (a) A *cbh2* URR covering the cellulase activating element (CAE) comprising a CCAAT-box (yellow) and a Xyr1-binding site (red), two additional Xyr1-binding sites and a Cre1-binding site (blue) was investigated (–258- to –198-bp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

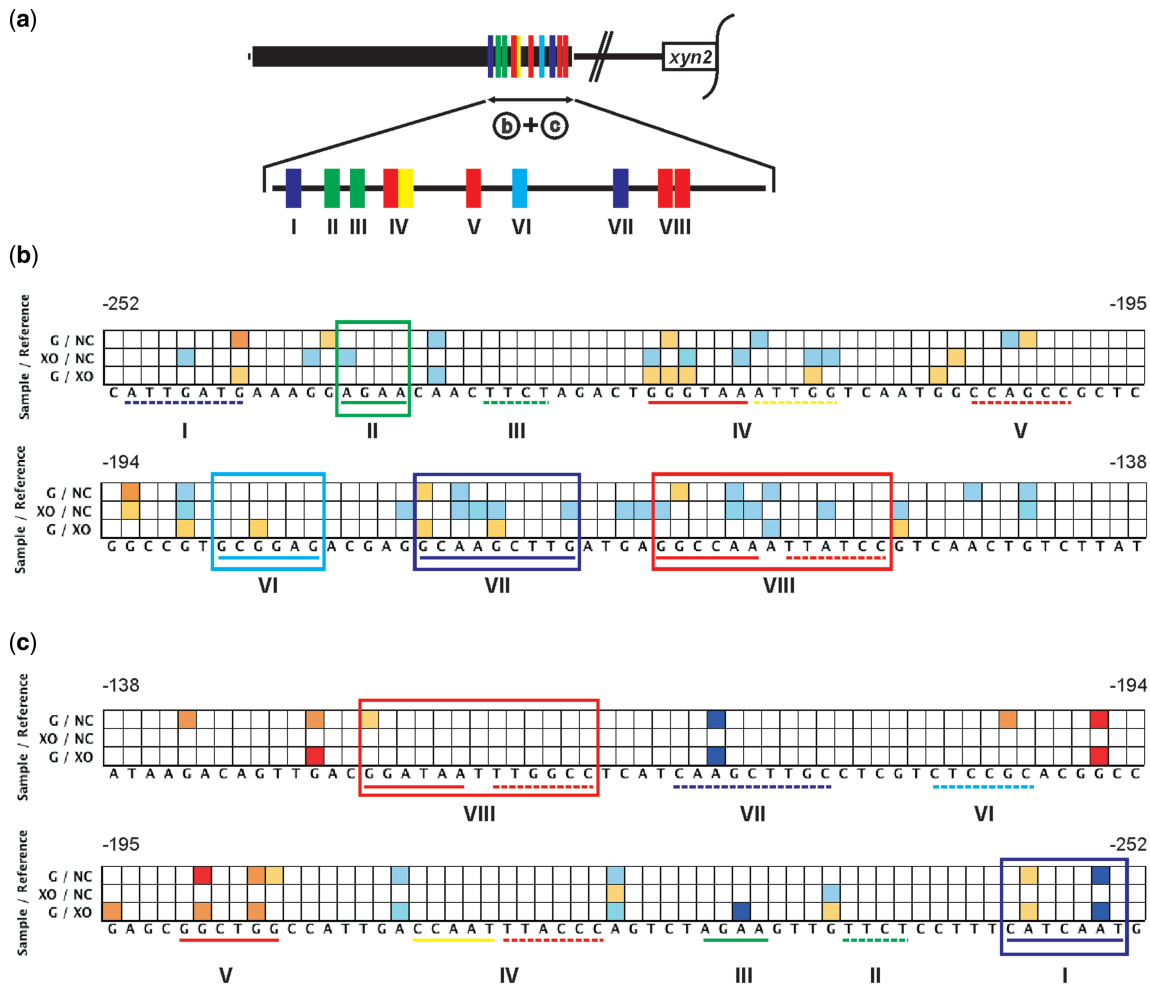


Figure 6. *In vivo* footprinting analysis of the *T. reesei* *xyn2* URR. (a) A *xyn2* URR showing a high number of cis elements [AGAA-box (green), CCAAT-box (yellow), Xyr1-binding site (red) and Cre1-binding site (blue)] was investigated (−252- to −138-bp upstream from ATG). The coding strand (b) and 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 Acel-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.