TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

zum Thema

Engineering an N-acetylneuraminic acid synthesis pathway into *Hypocrea* (*Trichoderma)*

(Einbringung eines N-Acetylneuraminsäure-Syntheseweges in *Hypocrea* (*Trichoderma*))

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Ao.Univ.Prof. Mag. Dr.rer.nat. Robert L. Mach
E166
Institut für Verfahrenstechnik, Umwelttechnik und Technische Biowissenschaften

eingereicht an der Technischen Universität Wien

Fakultät für Technische Chemie

von

DI Matthias Georg Steiger
Mtk.Nr. 0125435
Gentzgasse 159/4/13
A-1180 Wien

Diese Dissertation wurde begutachtet von

_____ _____

Ao.Univ.Prof. Mag. Dr.rer.nat. Robert L. Mach Univ.Prof. Dr. Peter J. Punt

Wien, am

_____
DI Matthias Georg Steiger

CONTENTS / *INHALTSVERZEICHNIS*

I DEUTSCHE KURZFASSUNG DER DISSERTATION

N-Acetyl-D-Neuraminsäure (NeuNAc) gehört zur funktionellen Klasse der Sialinsäuren und ist der biochemische Vorläufer für alle bekannten natürlich vorkommenden Sialinsäurederivate. Diese Karbonsäuren treten vor allem als terminale Reste an Glykokonjugaten in Proteinen auf. An dieser Stelle sind sie an diversen Interaktiosmechanismen beteiligt. Besondere Aufmerksamkeit haben künstliche Derivate von NeuNAc als Neuraminidaseinhibitoren erlangt, die zur Bekämpfung von viralen Krankheiten wie Influenza eingesetzt werden.

Die Herstellung von NeuNAc selbst ist teuer und kann einerseits durch Isolation aus natürlichen Rohstoffen (Eier, Milch), aber auch auf biochemischen Weg erfolgen. Dazu wurden enzymatische Verfahren entwickelt, die aber den Nachteil haben, dass der Schritt von N-Acetyl-D-Mannosamin hin zu NeuNAc mithilfe einer Aldolase und dem Kosubstrat Phosphoenolpyruvat zu bewerkstelligen ist. Diese Reaktion ist eine Gleichgewichtsrektion, die nur mit einem Überschuss an Pyruvat zu ausreichenden Mengen an NeuNAc führt. Zudem wurden bakterielle Ganzzellfermentationen entwickelt, die aber auf dem teuren Rohstoff N-Acetyl-D-Glukosamin angewiesen sind.

In dieser Arbeit wird daher eine neuer biochemischer Syntheseweg vorgestellt, der auf dem filamentösen Pilz *Hypocrea jecorina* (anamorph: *Trichoderma reesei*) basiert. Aufgrund der massiven Sekretion von hydrolytischen Enzymen besitzt dieser Pilz die Fähigkeit Biopolymere wie Cellulose aber auch Chitin in ihre Monomere abzubauen und zu metabolisieren. Da das Monomer von Chitin, N-Acetyl-D-Glukosamin, ein Substrat für die Synthese von NeuNAc darstellt und Chitin ein günstiger erneuerbarer Rohstoff ist, soll dieser Pilz zur Produktion von NeuNAc eingesetzt werden. *H. jecorina* weist aber keinen bekannten Stoffwechselweg zur Herstellung von NeuNAc auf. Deswegen soll ein künstlicher Stoffwechselweg durch heterologe Proteinexpression eingeführt werden. Zwei bakterielle Gene kodierend für eine N-Acetylglukosamin-2-epimerase und eine NeuNAc-Synthase werden in das Genom von *H. jecorina* eingeführt. Um die methodischen Voraussetzungen für diese Arbeit zu schaffen, wird in dieser Arbeit eine Transformationsstrategie für *H. jecorina* vorgestellt, die das zielgerichtete Einbringen oder Deletieren von Genen im Pilzgenom erlaubt und zudem Selektionsmarker rückgewinnt. Ein weiterer Aspekt behandelt die Messung von mRNA mittels quantitativer PCR, die ein unerlässliches Werkzeug zur genetischen Analyse darstellt. Für die Messung ist eine Normalisierung auf ein stabil transkribiertes Gen notwendig. In dieser Arbeit wird die Evaluation stabiler Gene für *H. jecorina* durchgeführt. Diese Dissertation belegt, dass ein künstlicher Stoffwechselweg zur Herstellung von NeuNAc in *H. jecorina* eingefügt werden kann und *in vivo* funktionstüchtig ist.

II SCIENTIFIC BACKGROUND OF THE PHD

1. Sialic Acids, *N*-Acetylneuraminic Acid (NeuNAc)

The sialic acids are a large family of nine carbon polyhydroxylated α-keto acids which play a wide variety of roles in nature (5). Most are derivatives of the common core structures *N*-acetylneuraminic acid (NeuNAc) or 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN) (Fig. 1).

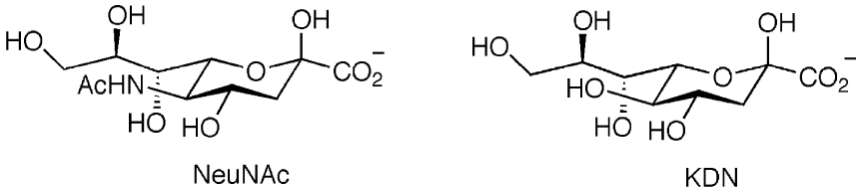


Figure 1: NeuNAc and 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid

NeuNAc is the most prevalent exponent of sialic acids (78). Currently, more than 50 derivatives of sialic acids are known to exist in nature. NeuNAc is believed to serve as a precursor of all these derivatives, since all biochemical pathways proceed via this substance (5). Commonly found modifications include the decoration of the hydroxyl groups with acetates, sulphates, phosphates, and methyl ethers. In addition, lactonization of the carboxylic acid and hydroxylation of the acetamido group of NeuNAc are observed (5).

2. Sialic Acids and their Function in Nature

In mammals, the sialic acids are usually found as terminal residues on the outermost cell surface as glycol-conjugates. As a result of their location, and their negative carboxylate functionality, sialic acids play important roles in mediating cellular recognition and adhesion processes (93). Some notable examples include the sialylation of neural cell adhesion molecules, which is responsible for the proper establishment of the vertebrate embryonic nervous system (16), and the necessity for recognition of sialic acids by the lectins of the inflammatory and immune response pathways (94). In addition, the extent of cell surface sialylation correlates with tumorigenesis and metastasis in certain cancers (28, 30). Although most bacteria do not produce sialic acids, several pathogenic strains can biosynthesize them and display them on their cell surface to mimic mammalian cells and evade the host's immune system (71). These include the neuroinvasive organisms *Escherichia coli* K1 and *Neisseria meningitidis* that cause meningitis (6, 22, 29), and *Campylobacter jejuni* that causes food-borne gastroenteritis (53).

3. Pharmaceutical Properties of NeuNAc

As mentioned *vide supra*, in biological systems sialic acids are mostly terminal components of glycoproteins presented at the respective cell surface. In such cases, sialic acids serve as recognition sites for respective endogenous and exogenous proteins, all of them belonging to the class of lectins (12, 90). Because of this paramount position of sialic acids in the cellular systems, they play a major role in infection cycles of important viral diseases, such as influenza viruses A and B (38). In such a case, *de novo* synthesized virus particles attach to respective sialic acids presented at the cell surface. For a propagation of the virus in the host, a neuraminidase (sialidase) activity is needed, which cleaves the linkage between the sialic acid and the glycoprotein. Therefore, sialic acid derivatives are nowadays successfully applied in the therapy of such virus related diseases. In

particular, they are used as neuraminidase inhibitors to prevent a further propagation of the virus. To date the best known preparations on the market are „Tamiflu“ (active pharmaceutical ingredient: “Oseltamivir”) by Hoffmann la Roche and „Relenza“ (active pharmaceutical ingredient: “Zanamivir”) by GlaxoSmithKline, both functioning as neuraminidase inhibitors, the latter one being a derivate of the precursor NeuNAc (http://www.gsk.com, (89)).

4. Biosynthesis of NeuNAc in Bacteria and Mammals

The biosynthesis of the most common sialic acid, NeuNAc, begins with the formation of *N*-acetylmannosamine (ManNAc) from UDP-*N*-acetylglucosamine (UDP-GlcNAc) (Fig. 2 from (86)).

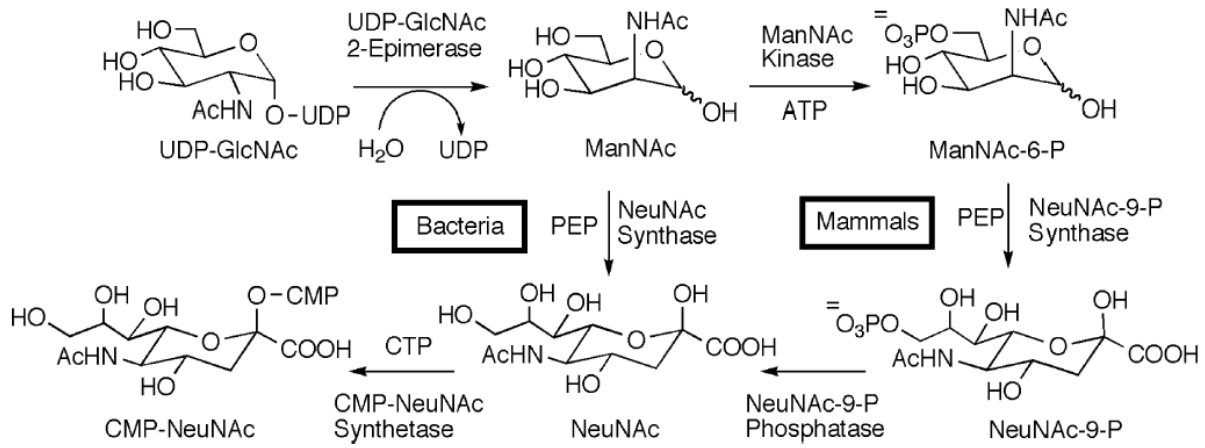


Figure 2: Biosynthesis of CMP-NeuNAc in bacteria and mammals (86)

In mammals, the ManNAc is then phosphorylated to give ManNAc-6-phosphate (ManNAc-6P). The second step involves the condensation of either ManNAc (in bacteria) or MacNAc-6P (in mammals) with phosphoenolpyruvate (PEP) to give NeuNAc or NeuNAc-9P, respectively. In mammals, NeuNAc-9P is then de-phosphorylated to generate NeuNAc. Finally, the activated form of the sialic acid, CMP-NeuNAc, is generated with the use of cytosine triphosphate (CTP). CMP-sialic acid is the activated substrate for all the sialyltransferases that incorporate the keto acid into glycoproteins and glycolipids.

In detail, biosynthesis of NeuNAc in mammalian cells is initiated and regulated by the synthesis of its precursor ManNAc (39, 47). Two enzymes have been identified to participate in the balance of the ManNAc pool. One is UDP-*N*-acetylglucosamine 2-epimerase (UDP-GlcNAc 2-epimerase, EC 5.1.3.14) that catalyzes the formation of ManNAc from UDP-GlcNAc (39, 86). The other is the *N*-acetylglucosamine 2-epimerase (GlcNAc 2-epimerase, EC 5.1.3.8) that catalyzes reversible the interconversion of *N*-acetylglucosamine (GlcNAc) and ManNAc (17). The GlcNAc 2-epimerase was previously known as a renin-binding protein. The tight GlcNAc 2-epimerase-renin complex is found when purified from porcine kidney (85), and the binding causes the block of renin activity. However, the studies of GlcNAc 2-epimerase gene knockout mice have proven that the GlcNAc 2-epimerase is not involved in renin-angiotensin system (79). Recently, the GlcNAc 2-epimerase has been found to catalyze the conversion of ManNAc to GlcNAc in human cells to serve as catabolic role in NeuNAc metabolism (55).

Several enzymatic steps involved in the pathway for the biosynthesis of NeuNAc in bacteria have been identified. In *Neisseria meningitidis*, NeuNAc is formed directly from ManNAc and PEP in the presence of Mn^{2+} (9). In *E. coli* K1, two pathways have been proposed for the formation of NeuNAc: i) condensation of ManNAc

and pyruvate by a NeuNAc lyase. Usually, the NeuNAc lyases are involved in the catabolism of sialic acids by catalyzing the cleavage of NeuNAc into ManNAc and pyruvate in a reversible reaction. However, at high concentrations of ManNAc and pyruvate, the equilibrium can be shifted to the synthesis of NeuNAc (59). ii) Condensation of ManNAc and PEP by a NeuNAc synthase, which directly is involved in NeuNAc synthesis (74, 92).

The first cloning of a prokaryotic GlcNAc 2-epimerase from the phototrophic cyanobacterium *Synechocystis* sp. PCC6803 was reported by (84). More recently, Lee and co-workers (52) found that whole-cell extracts of several photobacteria can convert GlcNAc to ManNAc. Among them, *Anabaena* sp. CH1 exhibited the highest GlcNAc 2-epimerase activity; consequently the authors cloned and characterised a gene encoding GlcNAc 2-epimerase from *Anabaena sp.* CH1.

5. NeuNAc in Filamentous Fungi

As described vide supra, NeuNAc is synthesized in higher eukaryotes and in gram-negative bacteria from the precursor GlcNAc. To the current knowledge in the kingdom of fungi such a pathway does not exist. This conclusion is based on *in silico* analyses of numerous fungal genomes (e.g. reported by (5), but also carried out in this PhD thesis), as these analyses did not reveal any sequences significantly similar to the enzymes known to be involved in the biosynthesis, activation, or transfer of sialic acids in bacteria and mammals. Nevertheless, there are reports on the presence of sialic acids in some pathogenic fungal cells (reviewed in (3)), such as *Candida albicans* (81), *Cryptococcus neoformans* (73), *Aspergillus fumigatus* (95), and *Sporothrix schenckii* (2). Although some of these claims are based only on binding of lectins, which can recognize sialic acids, others include more solid evidence, such as mass spectrometry (73, 95).

6. Biocatalyzed Production Strategies of NeuNAc

NeuNAc is traditionally prepared by extraction from natural sources, such as bird nest, milk or eggs (48), by the hydrolysis of colominic acid (a homopolymer of NeuNAc) in the culture broth of *E. coli* K1 (60), or by chemical synthesis (18). A chemo-enzymatic process for the production of NeuNAc from GlcNAc and pyruvate has also been reported (10, 58). An alkaline-catalyzed epimerization is performed for a conversion of GlcNAc to ManNAc. The ManNAc is then condensed with pyruvate by *E. coli* NeuNAc lyase to yield NeuNAc.

Since it has been known that epimerization of GlcNAc and ManNAc can be catalyzed by porcine kidney GlcNAc 2-epimerase (pGlcNAc 2-epimerase) (17), a two-enzyme reaction process has also been developed for the enzymatic production of NeuNAc (49, 59). In this process, the formation of ManNAc from GlcNAc is catalyzed by pGlcNAc 2-epimerase. Condensation of ManNAc with pyruvate to generate NeuNAc is similar to the one in the chemo-enzymatic process. The pGlcNAc 2-epimerase/NeuNAc lyase process has been proposed to be simpler and more efficient compared with the chemo-enzymatic method (Maru et al., 1998). However, the requirement of ATP for the activation of pGlcNAc 2-epimerase activity leads to high costs of the process.

Tabata and co-workers (84) described a biotransformation for the synthesis of NeuNAc using bacterial cells (*E. coli*) expressing GlcNAc 2-epimerase from *Synechocystis* and NeuNAc synthase from *E. coli* K1 without the isolation of the enzymes and/or ManNAc. The NeuNAc synthase irreversibly condenses ManNAc and PEP to NeuNAc; if PEP has to be bought and added, high costs will arise and moreover, it is not readily available as a substrate. In the microbial reaction, the authors could confirm the accumulation of NeuNAc without the addition of PEP, indicating that PEP could be supplied by the glycolytic activities of the involved bacterial expression

hosts. Recently, a whole-cell system based on the separate expression of GlcNAc 2-epimerase from *Anabaena sp.* CH1 and NeuNAc lyase from *E. coli* was reported (52). This system allowed the production of NeuNAc from GlcNAc and pyruvate by omitting the additional supplementation with ATP, such as needed in the enzymatic processes. Another system, which is similar to the process of Lee et al., reports the production of NeuNAc also with an optimized coupled cell system based on a temperature-induced expression rather based on IPTG as inducer (97). Recently, a review focusing on the biotechnological production of NeuNAc was published (87).

7. Engineering of the NeuNAc Pathway in different Eukaryotic Hosts

Recently, two remarkable reports on the engineering of a NeuNAc pathway in respective eukaryotic expression hosts, namely *Pichia pastoris* (36) and two plant models i) Bright Yellow 2 tobacco cells, and ii) *Medicago sativa* (64) were published. It is noteworthy that in both cases NeuNAc synthesis could be achieved, however, none of these studies focused on the production of NeuNAc or CMP-NeuNAc itself, but on the secretion of recombinant human glycoproteins with complex, fully, terminally sialylated N-glycans. Nonetheless, these studies impressively demonstrate that a NeuNAc - respectively CMP-NeuNAc - pathway can be engineered into none-producing hosts, such as fungi and plants.

Engineering the Yeast *Pichia pastoris* to Secrete Human Glycoproteins with fully, terminally Sialylated N-glycans

In order to improve the half-life and the therapeutic potency of human glycoproteins that are heterologously produced in *P. pastoris*, Hamilton and co-workers substantially reengineered its secretory pathway (36). In addition to several other steps leading to the construction of a human glycosylation pathway into a respective yeast strain (11, 15, 35), sialylation, the final step of human glycosylation, was accomplished. In particular the following steps were engineered: i) production of the N-glycosylated precursors terminating in β-1,4-galactose, ii) the biosynthetic capability to produce CMP-NeuNAc, iii) the introduction of a transporter to shuttle CMP–NeuNAc into the Golgi, and iv) the expression of a sialyltransferase to transfer sialic acid to terminal galactose on the nascent glycoprotein (36). It should be noted that the whole pathway was engineered based on mammalian genes. As a final proof of succeeding in the concept, the authors reported on cell lines secreting a terminally sialylated, complex, bi-antennary glycoprotein as exemplified by recombinant Erythropoietin (EPO).

Engineering of a Sialic Acid Synthesis Pathway in Transgenic Plants by Expression of Bacterial NeuNAc-synthesizing Enzymes

Paccalet and co-workers investigated the NeuNAc synthesis in plants by expressing a recombinant NeuNAc lyase and a NeuNAc synthase. Both the NeuNAc lyase from *E. coli* K1 and NeuB2 (encoding for a NeuNAc synthase) from *Campylobacter jejuni* (reviewed by (86)) were expressed in the cytosol of tobacco BY2 cells or *M. sativa* (alfalfa) plants. The NeuNAc lyase expressed in BY2 cells was able to cleave NeuNAc into ManNAc and pyruvate in a reversible reaction, but also formed NeuNAc in the presence of pyruvate and ManNAc. Furthermore, feeding experiments carried out in the presence of exogenous NeuNAc demonstrated that the enzyme was functional *in planta* (64). NeuB2 from *C. jejuni* expressed in tobacco BY2 cells was also demonstrated to be able to synthesize NeuNAc in the presence of ManNAc and PEP and expression of the same sequence in *M. sativa* (alfalfa) plants resulted in an accumulation of a functional enzyme (64). However, the activity of NeuB2 expressed in plants was relatively low, probably due to the fact that NeuB2 of *C. jejuni* was

recently shown to have a much lower capacity for NeuNAc synthesis than NeuB1, an isoform of the NeuNAc synthase (83). Furthermore, the expression of a functional CMP-NeuNAc synthase and CMP-NeuNAc transporter (45) in tobacco BY2 cells has also recently been described. Transportation of the CMP activated form of NeuNAc from the cytoplasm into Golgi vesicles could be demonstrated (61).

A combination of these data, strongly indicates that the introduction of the whole sialylation machinery in plants is feasible and may only be limited by the capacity to stably transform plants with multiple gene sequences (64), which in filamentous fungi - such *Trichoderma spp.* and *Aspergillus spp.* - will not be a restraining factor.

8. *Trichoderma* and its Chitinolytic Activity

The Biopolymer Chitin

After cellulose, chitin is the second most abundant renewable organic source in nature (8), with an estimated annual biosynthesis of 10^9 to 10^{11} tons. The major natural resources are the exoskeletons of arthropods, such as crustaceans (like crab, lobster, and shrimp), and insects (including ants, beetles), cell walls of fungi, the radula of mollusks, and the beaks of the cephalopods (including squid and octopi). The polymer is composed of ß-(1,4)-linked units of the amino sugar GlcNAc. For industrial applications this renewable resource is mainly extracted from crustacean shells, and can be processed into many derivatives. These derivatives are used for a number of commercial products, such as medical applications including the therapeutic potential for the treatment of a variety of diseases, [e.g. arthritis, (40, 62), inflammatory bowel disease, (76), and general inflammatory damage (46)], cosmetics, dietary supplements, agriculture, and water treatment (8, 20, 63). However, thitherto only a minute amount of this renewable natural resource is used in industrial and agricultural applications.

Biotechnological GlcNAc production, Microbial Chitinases

Commercial production of GlcNAc currently relies upon acid hydrolysis of de-proteinized and de-mineralized crustacean shells (24) Acid hydrolysis is relatively efficient, but involves strong acids (4 to 8 M HCl) at high temperatures and results in production of toxic wastes. Moreover, the extreme conditions used in the process may result in unwanted modifications of the hydrolysis products. Conversely, enzymatic degradation of e.g. crustacean shells is environmentally friendly, but it is more complex, since it involves both production of the enzyme and the digestion of the substrate. More importantly, enzymatic processes have been plagued by low yields and have resulted in incomplete conversion of chitin into its monomer (14, 70). However, there has been substantial, recent progress in studies of chitinases from both prokaryotic and eukaryotic sources. (23, 75, 96) Among these, the most widely examined enzyme sources for commercial chitin degradation are strains of *Serratia marcescens* (96). Nevertheless, its enzyme preparations do not completely convert chitin to GlcNAc, but instead accumulate a blend of both, monomers and chito-oligomers (mainly chitobiose) (1). The phenomenon is particularly evident when high concentrations of chitin (5 to 10 % (w/v)) are digested. Chitinolytic enzymes from fungi in the genus *Trichoderma* have also been extensively studied (54). Unlike their bacterial counterparts, *Trichoderma* chitinolytic preparations have a high ratio of exochitinase to endochitinase activity and release almost exclusively monomeric GlcNAc from chitin. Recently, the chitinolytic enzyme system of *T. reesei* has extensively been studied by genome-wide analysis (80). The authors identified 18 genes encoding proteins belonging to glycoside hydrolase family 18 (41, 67, 72, 88), and two members of glycoside hydrolase family 20 (51) in the respective genome, whereas no members of family 19 (27, 34, 37), primarily found in plants, could be detected. Such an amount of enzymes indicate for a high chitinolytic potential of

Trichoderma, which has also been proven for various species, such as *T. harzianum* (teleomorph *Hypocrea lixii*), *T. virens* (teleomorph *H. virens*), *T. asperellum,* and *T. atroviride* (teleomorph *H. atroviridis*), all known as potent mycoparasites. Nonetheless, Donzelli and co-workers (21) reported on the enhanced efficacy and the synergy of mixtures of enzymes from *T. atroviride*, *S. marcescens* and *Streptomyces albidoflavus* in hydrolyzing several forms of chitin (e.g. native chitin from langostino crab shells, colloidal chitin, and chitosan). Such findings warrant for a high potential of improvement of the chitinolytic abilities of *Trichoderma* strains applying respective genetic engineering strategies (i.e. heterologous expression of bacterial chitinases in *Trichoderma spp.* under chitin-inducible promoters such as p*nag1* (13, 56)).

III AIM OF THE PHD THESIS

1. Engineering an N-Acetylneuraminic acid synthesis pathway into *Hypocrea jecorina*

The synthesis of NeuNAc is costly. It is currently, if a biotechnological process is followed, carried out by a two-step enzyme catalyzed procedure (49, 59). The production costs are especially detrimentally influenced by the second step of biosynthesis, in which ManNAc is metabolized to NeuNAc via an equilibrium reaction. In this reaction step an excess of pyruvate has to be added to drive the equilibrium towards NeuNAc. Anyway, the removal of pyruvate from the reaction is a cost-intensive downstream processing step (98).

The aim of this PhD thesis is to optimise the synthesis of NeuNAc by the introduction of a whole-cell catalyzed process. In particular, it is planned to genetically engineer the filamentous ascomycete *Trichoderma* in such a way that it can synthesize NeuNAc using chitin (a cheap renewable biopolymer) as a carbon source. In comparison to the currently applied enzyme catalyzed process, one major advantage has to be mentioned. In contrast to the application of a lyase for the second synthesis step, which requires the use of an excess of pyruvate (vide supra), the newly designed in vivo synthesis would allow the application of NeuNAc synthase in the process. This enzymatic step implies the use of PEP instead of pyruvate (which in the intended in vivo process would be supplied by the organism), thereby leading to an irreversible process (64). Consequently, the insertion of an excess of pyruvate becomes obsolete and the resulting downstream process is significantly simplified.

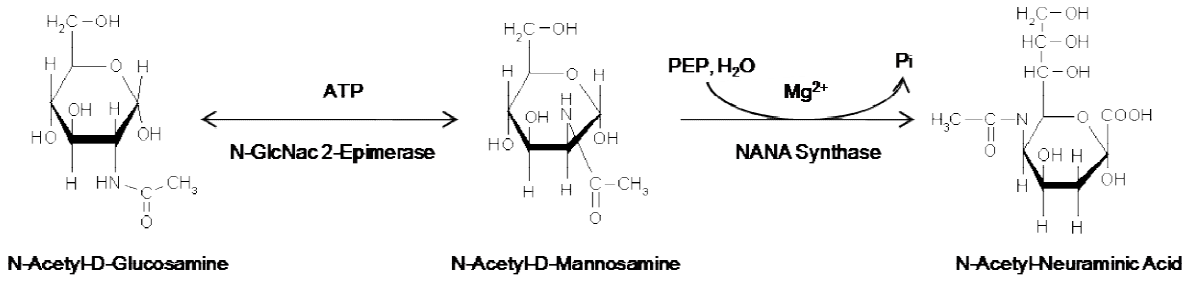


Figure 3: NeuNAc synthesis pathway to be engineered in *Trichoderma*

In addition, the application of *Trichoderma* as a whole-cell catalyst offers the advantage of using cheap renewable raw materials, such as chitin.

2. A transformation system for *Hypocrea* (*Trichoderma*) that favours homologous integration and that uses reusable bi-directionally selectable markers

In order to engineer a NeuNAc synthesis pathway into *Hypocrea jecorina*, it is necessary to have a versatile transformation system in hand. In particular, it could become necessary to perform serial transformations of the respective *Hypocrea* strains. Such a system should meet certain requirements: i) a high rate of homologous integration (gene replacement) events; ii) a recyclable marker for unrestricted rounds of knock-outs; and iii) a bi-directional positive selection system. Transformation systems which are known to work in *Hypocrea*, based on the markers *hph* (57), *amdS*(66), and *pyr4* (31), shall be combined to a new system. The advantage of the *amdS* marker is that it was reported to work as a bi-directional marker in *Aspergillus* (43). The recycling of the marker genes will be achieved by using the Cre/loxP system from bacteriophage P1 (7, 82). This strategy involves the

excision of DNA fragments flanked by loxP-sites catalysed by the expression of Cre recombinase. Although the use of the Cre/loxP system in an eukaryote (*Saccharomyces cerevisiae*) was described for the first time in 1987 (77), it was not used as a gene disruption tool in yeast until 1996 (33). Since then, the Cre/loxP system has also been used in filamentous fungi such as *Candida albicans*, *Aspergillus sp.*, *Epichloë festucae*, *Neotyphodium* sp., and *Cryptococcus neoformans* (19, 25, 26, 50, 65).

To increase the rate of homologous integration, a non homologous end-joining deficient (NHEJ) strain shall be constructed. Such a NHEJ deficient strain was previously reported for *Hypocrea jecorina* based on a gene deletion of *tku70* (YKU70) (32). Nevertheless, this system was constructed in a UV-mutant strain of *Hypocrea jecorina* TU6 (31), which cannot be used as a universally applicable recipient strain. Furthermore, it was shown for *Neurospora* that MUS-53 (a homolog of human Lig4) is required for NHEJ. The authors demonstrated that in contrast to *mus-51* and *mus-52* mutants, a *mus-53* mutant had a gene targeting efficiency of 100% even if the homologous DNA sequence was very short. *mus-53* was proposed as a highly efficient alternative to *mus-51* (YKU70) and *mus-52* (YKU80) (44). Therefore, a tmus53 deletion mutant shall be constructed for *Hypocrea jecorina* and tested on its ability to confer NHEJ deficiency.

3. An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*)

Measuring the mRNA level of a certain gene is one important tool for the characterisation of recombinant strains expressing heterologous proteins. The transcription of the corresponding genes is the first necessary step for the production of heterologous proteins and by quantification of the mRNA level it is possible to analyze the different expression patterns in different strains. To some extend the mRNA level correlates with the protein level and is therefore of special interest.

One efficient tool to measure the mRNA content is reverse transcription-quantitative PCR (RT-qPCR). Major advantages of this method are its high sensitivity, large dynamic range, and accurate quantification (42). However, an accurate and robust normalization system is needed when performing relative quantification of qPCR data. Normalizing to a stably expressed gene of the target organism, often called reference or housekeeping gene, is a powerful method for internal error prevention. In general, when applying mRNA quantification techniques, an error is caused by the multistage process required to extract, process (i.e. in vitro reverse transcription) and detect mRNA. A major challenge is to find a suitable reference gene. An important requirement for such a reference gene is its robust expression under all conditions applied in a certain experiment. Therefore, a careful evaluation is necessary to obtain one or even more suitable genes for normalization. Vandesompele and co-workers suggest the use of multiple reference genes rather than relying on a single one (91). However, it depends on the experimental setup how many reference genes have to be included in order to obtain consistent and reliable results. Especially for the setup of new experimental conditions, it has to be evaluated if a certain set of reference genes is suitable or not. Different approaches have been published, which allow the evaluation of multiple reference genes (4, 69, 91). The software geNorm ranks the genes according to the similarity of their expression profiles by a pairwise comparison (91). The authors calculated the average pairwise variation of a particular gene with all other control genes and denominated it M, the internal control gene-stability measure. Within this system, genes with the lowest M values have the most stable expression. Another calculation method is used by the program NormFinder (4), which uses a model-based approach for the estimation of expression variation. The advantage of this algorithm compared to geNorm is that it can deal with systematic differences in the data set like different tissues or strains.

The relative expression software tool (REST; version: 2008) developed by Pfaffl and co-workers allows the estimation of gene expression using qPCR amplification data (i.e. the threshold cycle values). This software enables the measurement of uncertainty in expression ratios by introducing randomization and bootstrapping techniques. Confidence intervals for expression levels allow measurement of not only the statistical significance of deviations but also of their likely magnitude even in the presence of outliers (68).

Therefore, potential reference genes shall be evaluated with the software tools GeNorm and NormFinder. The suitability of the reference genes shall then be evaluated in a correlation experiment between mRNA level and enzyme activity.

4. Structure of the PhD thesis

The above mentioned aims of the PhD thesis are elaborated in three autonomous manuscripts following this introduction (section IV, V and VI). The first paper, "Engineering an N-Acetylneuraminic acid synthesis pathway into H*ypocrea jecorina*", deals with the metabolic engineering of *H. jecorina* and how an intracellular pathway is engineered into this fungus leading to NeuNAc. The two other papers present methods which are needed for the realization of the first paper. Thus, the second paper, "A transformation system for *Hypocrea* (*Trichoderma*) that favours homologous integration and that uses reusable bi-directionally selectable markers", presents a transformation strategy for *H. jecorina* and the third paper, "An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*)", describes how quantitative PCR data shall be treated in order to obtain meaningful results.

References

1. **Aloise, P. A., M. Lumme, and C. A. Haynes.** 1996. *In* R. Muzzarelli (ed.), Chitin Enzymology, vol. 2, Grottamare, Italy.
2. **Alviano, C., M. Pereira, W. Souza, L. Oda, and L. Travassos.** 1982. Sialic acids are surface components of Sporothrix schenckii yeast forms. FEMS Microbiol Lett **15:**223-227.
3. **Alviano, C. S., L. R. Travassos, and R. Schauer.** 1999. Sialic acids in fungi: a minireview. Glycoconj J **16:**545-54.
4. **Andersen, C. L., J. L. Jensen, and T. F. Orntoft.** 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res **64:**5245-50.
5. **Angata, T., and A. Varki.** 2002. Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chem Rev **102:**439-69.
6. **Annunziato, P. W., L. F. Wright, W. F. Vann, and R. P. Silver.** 1995. Nucleotide sequence and genetic analysis of the *neuD* and *neuB* genes in region 2 of the polysialic acid gene cluster of *Escherichia coli* K1. J Bacteriol **177:**312-9.
7. **Austin, S., M. Ziese, and N. Sternberg.** 1981. A novel role for site-specific recombination in maintenance of bacterial replicons. Cell **25:**729-36.
8. **Ballenweg, S.** 2005. Roempp Online. Thieme Chemistry, Stuttgart.
9. **Blacklow, R. S., and L. Warren.** 1962. Biosynthesis of sialic acids by Neisseria meningitidis. J Biol Chem **237:**3520-6.
10. **Blayer, S., J. M. Woodley, M. J. Dawson, and M. D. Lilly.** 1999. Alkaline biocatalysis for the direct synthesis of N-acetyl-D-neuraminic acid (Neu5Ac) from N-acetyl-D-glucosamine (GlcNAc). Biotechnol Bioeng **66:**131-6.
11. **Bobrowicz, P., R. C. Davidson, H. Li, T. I. Potgieter, J. H. Nett, S. R. Hamilton, T. A. Stadheim, R. G. Miele, B. Bobrowicz, T. Mitchell, S. Rausch, E. Renfer, and S. Wildt.** 2004. Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. Glycobiology **14:**757-66.
12. **Brinkman-Van der Linden, E. C., and A. Varki.** 2000. New aspects of siglec binding specificities, including the significance of fucosylation and of the sialyl-Tn epitope. Sialic acid-binding immunoglobulin superfamily lectins. J Biol Chem **275:**8625-32.
13. **Brunner, K., C. K. Peterbauer, R. L. Mach, M. Lorito, S. Zeilinger, and C. P. Kubicek.** 2003. The Nag1 N-acetylglucosaminidase of Trichoderma atroviride is essential for chitinase induction by chitin and of major relevance to biocontrol. Curr Genet **43:**289-95.
14. **Carroad, P. A., and R. A. Tom.** 1978. Bioconversion of shellfisch chitin wastes: process conception and selection of microorganisms. J. Food Sci. **42:**1158-1161.
15. **Choi, B. K., P. Bobrowicz, R. C. Davidson, S. R. Hamilton, D. H. Kung, H. Li, R. G. Miele, J. H. Nett, S. Wildt, and T. U. Gerngross.** 2003. Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. Proc Natl Acad Sci U S A **100:**5022-7.
16. **Cunningham, B. A., S. Hoffman, U. Rutishauser, J. J. Hemperly, and G. M. Edelman.** 1983. Molecular topography of the neural cell adhesion molecule N-CAM: surface orientation and location of sialic acid-rich and binding regions. Proc Natl Acad Sci U S A **80:**3116-20.
17. **Datta, A.** 1970. Regulatory role of adenosine triphosphate on hog kidney N-acetyl-D-glucosamine 2-epimerase. Biochemistry **9:**3363-70.
18. **de Ninno, M.** 1991. The synthesis and glycosidation of N-acetyl-D-neuraminic acid. Synthesis **8:**583-593.

19. **Dennison, P. M., M. Ramsdale, C. L. Manson, and A. J. Brown.** 2005. Gene disruption in *Candida albicans* using a synthetic, codon-optimised Cre-*loxP* system. Fungal Genet Biol **42:**737-48.
20. **Di Martino, A., M. Sittinger, and M. V. Risbud.** 2005. Chitosan: a versatile biopolymer for orthopaedic tissue-engineering. Biomaterials **26:**5983-90.
21. **Donzelli, B. G., G. Ostroff, and G. E. Harman.** 2003. Enhanced enzymatic hydrolysis of langostino shell chitin with mixtures of enzymes from bacterial and fungal sources. Carbohydr Res **338:**1823-33.
22. **Edwards, U., A. Muller, S. Hammerschmidt, R. Gerardy-Schahn, and M. Frosch.** 1994. Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid capsule by *Neisseria meningitidis* serogroup B. Mol Microbiol **14:**141-9.
23. **Felse, P., and T. Panda.** 2000. Production of microbial chitinases - A revisit Bioprocess. Eng. **23:**127-134.
24. **Ferrer, J., G. Paez, Z. Marmol, E. Ramones, H. Garcia, and C. Forster.** 1996. Acid hydrolysis of shrimp-shell wastes and the production of single cell protein from the hydrolysate. Bioresour. Technol. **57:**55-60.
25. **Florea, S., K. Andreeva, C. Machado, P. M. Mirabito, and C. L. Schardl.** 2009. Elimination of marker genes from transformed filamentous fungi by unselected transient transfection with a Cre-expressing plasmid. Fungal Genet Biol **46:**721-30.
26. **Forment, J. V., D. Ramon, and A. P. MacCabe.** 2006. Consecutive gene deletions in *Aspergillus nidulans*: application of the Cre/*loxP* system. Curr Genet **50:**217-24.
27. **Fukamizo, T.** 2000. Chitinolytic enzymes: catalysis, substrate binding, and their application. Curr Protein Pept Sci **1:**105-24.
28. **Fukuda, M.** 1996. Possible roles of tumor-associated carbohydrate antigens. Cancer Res **56:**2237-44.
29. **Ganguli, S., G. Zapata, T. Wallis, C. Reid, G. Boulnois, W. F. Vann, and I. S. Roberts.** 1994. Molecular cloning and analysis of genes for sialic acid synthesis in *Neisseria meningitidis* group B and purification of the meningococcal CMP-NeuNAc synthetase enzyme. J Bacteriol **176:**4583-9.
30. **Gorelik, E., U. Galili, and A. Raz.** 2001. On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis. Cancer Metastasis Rev **20:**245-77.
31. **Gruber, F., J. Visser, C. P. Kubicek, and L. H. de Graaff.** 1990. Cloning of the *Trichoderma reesei pyrG* gene and its use as a homologous marker for a high-frequency transformation system. Curr Genet **18:**447-451.
32. **Guangtao, Z., L. Hartl, A. Schuster, S. Polak, M. Schmoll, T. Wang, V. Seidl, and B. Seiboth.** 2009. Gene targeting in a nonhomologous end joining deficient *Hypocrea jecorina*. J Biotechnol **139:**146-51.
33. **Güldener, U., S. Heck, T. Fielder, J. Beinhauer, and J. H. Hegemann.** 1996. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res **24:**2519-24.
34. **Hahn, M., M. Hennig, B. Schlesier, and W. Hohne.** 2000. Structure of jack bean chitinase. Acta Crystallogr D Biol Crystallogr **56:**1096-9.
35. **Hamilton, S. R., P. Bobrowicz, B. Bobrowicz, R. C. Davidson, H. Li, T. Mitchell, J. H. Nett, S. Rausch, T. A. Stadheim, H. Wischnewski, S. Wildt, and T. U. Gerngross.** 2003. Production of complex human glycoproteins in yeast. Science **301:**1244-6.
36. **Hamilton, S. R., R. C. Davidson, N. Sethuraman, J. H. Nett, Y. Jiang, S. Rios, P. Bobrowicz, T. A. Stadheim, H. Li, B. K. Choi, D. Hopkins, H. Wischnewski, J. Roser, T. Mitchell, R. R. Strawbridge, J. Hoopes, S. Wildt, and T. U. Gerngross.** 2006. Humanization of yeast to produce complex terminally sialylated glycoproteins. Science **313:**1441-3.
37. **Hart, P. J., H. D. Pfluger, A. F. Monzingo, T. Hollis, and J. D. Robertus.** 1995. The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 A resolution. J Mol Biol **248:**402-13.
38. **Herrler, G., J. Hausmann, and H. D. Klenk.** 1995. Sialic acid as receptor determinant of ortho-and paramyxoviruses, p. 315-331. *In* A. Rosenberg (ed.), Biology of the Sialic Acids. Plenum Press, NY and London.
39. **Hinderlich, S., R. Stasche, R. Zeitler, and W. Reutter.** 1997. A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Purification and characterization of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. J Biol Chem **272:**24313-8.
40. **Hochberg, M. C., and M. Dougados.** 2001. Pharmacological therapy of osteoarthritis. Best Pract Res Clin Rheumatol **15:**583-93.
41. **Hollis, T., A. F. Monzingo, K. Bortone, S. Ernst, R. Cox, and J. D. Robertus.** 2000. The X-ray structure of a chitinase from the pathogenic fungus *Coccidioides immitis*. Protein Sci **9:**544-51.
42. **Huggett, J., K. Dheda, S. Bustin, and A. Zumla.** 2005. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun **6:**279-84.
43. **Hynes, M. J.** 1977. Induction of the acetamidase of Aspergillus nidulans by acetate metabolism. J Bacteriol **131:**770-5.
44. **Ishibashi, K., K. Suzuki, Y. Ando, C. Takakura, and H. Inoue.** 2006. Nonhomologous chromosomal integration of foreign DNA is completely dependent on MUS-53 (human Lig4 homolog) in Neurospora. Proc Natl Acad Sci U S A **103:**14871-6.
45. **Ishida, N., N. Miura, S. Yoshioka, and M. Kawakita.** 1996. Molecular cloning and characterization of a novel isoform of the human UDP-galactose transporter, and of related complementary DNAs belonging to the nucleotide-sugar transporter gene family. J Biochem **120:**1074-8.
46. **Kamel, M., M. Hanafi, and M. Bassiouni.** 1991. Inhibition of elastase enzyme release from human polymorphonuclear leukocytes by N-acetyl-galactosamine and N-acetyl-glucosamine. Clin Exp Rheumatol **9:**17-21.
47. **Keppler, O. T., S. Hinderlich, J. Langner, R. Schwartz-Albiez, W. Reutter, and M. Pawlita.** 1999. UDP-GlcNAc 2-epimerase: a regulator of cell surface sialylation. Science **284:**1372-6.
48. **Koketsu, M., L. R. Juneja, H. Kawanami, M. Kim, and T. Yamamoto.** 1992. Preparation of N-acetylneuraminic acid from delipidated egg yolk. Glycoconj J **9:**70-4.
49. **Kragl, U., D. Gygax, O. Ghisalba, and C. Wandrey.** 1991. Enzymatic two step synthesis of N-acetylneuraminic acid in the enzyme membrane reactor. Angew Chem Int Ed Eng **30:**827-828.
50. **Krappmann, S., O. Bayram, and G. H. Braus.** 2005. Deletion and allelic exchange of the *Aspergillus fumigatus veA* locus via a novel recyclable marker module. Eukaryot Cell **4:**1298-307.
51. **Kubicek, C. P., R. L. Mach, C. K. Peterbauer, and M. Lorito.** 2001. *Trichoderma*: from genes to biocontrol. J Plant Pathol **83:**11-23.
52. **Lee, Y. C., H. C. Chien, and W. H. Hsu.** 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. J Biotechnol **129:**453-60.
53. **Linton, D., A. V. Karlyshev, P. G. Hitchen, H. R. Morris, A. Dell, N. A. Gregson, and B. W. Wren.** 2000. Multiple N-acetyl neuraminic acid synthetase (*neuB*) genes in *Campylobacter jejuni* identification and characterization of the gene involved in sialylation of lipo-oligosaccharide. Mol Microbiol **35:**1120-34.
54. **Lorito, M.** 1998. Chitinolytic enzymes and theire genes, p. 73-92. *In* G. E. Harman and C. P. Kubicek (ed.), Trichoderma & Gliocladium, vol. 2. Taylor & Francis, London, UK.
55. **Luchansky, S. J., K. J. Yarema, S. Takahashi, and C. R. Bertozzi.** 2003. GlcNAc 2-epimerase can serve a catabolic role in sialic acid metabolism. J Biol Chem **278:**8035-42.

56. **Mach, R. L., C. K. Peterbauer, K. Payer, S. Jaksits, S. L. Woo, S. Zeilinger, C. M. Kullnig, M. Lorito, and C. P. Kubicek.** 1999. Expression of two major chitinase genes of *Trichoderma atroviride* (*T. harzianum* P1) is triggered by different regulatory signals. Appl Environ Microbiol **65:**1858-63.
57. **Mach, R. L., M. Schindler, and C. P. Kubicek.** 1994. Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expression signals. Curr Genet **25:**567-70.
58. **Mahmoudian, M., D. Noble, C. S. Drake, R. F. Middleton, D. S. Montgomery, J. E. Piercey, D. Ramlakhan, M. Todd, and M. J. Dawson.** 1997. An efficient process for production of N-acetylneuraminic acid using N-acetylneuraminic acid aldolase. Enzyme Microb Technol **20:**393-400.
59. **Maru, I., J. Ohnishi, Y. Ohta, and Y. Tsukada.** 1998. Simple and large-scale production of N-acetylneuraminic acid from N-acetyl-D-glucosamine and pyruvate using N-acyl-D-glucosamine 2-epimerase and N-acetylneuraminate lyase. Carbohydr Res **306:**575-8.
60. **Maru, I., J. Ohnishi, Y. Ohta, and Y. Tsukada.** 2002. Why is sialic acid attracting interest now? Complete enzymatic synthesis of sialic acid with N-acylglucosamine 2-epimerase. J Biosci Bioeng **93:**258-65.
61. **Misaki, R., K. Fujiyama, and T. Seki.** 2006. Expression of human CMP-N-acetylneuraminic acid synthetase and CMP-sialic acid transporter in tobacco suspension-cultured cell. Biochem Biophys Res Commun **339:**1184-9.
62. **Muller-Fassbender, H., G. L. Bach, W. Haase, L. C. Rovati, and I. Setnikar.** 1994. Glucosamine sulfate compared to ibuprofen in osteoarthritis of the knee. Osteoarthritis Cartilage **2:**61-9.
63. **Muzzarelli, R. A.** 1999. Native, industrial and fossil chitins. EXS **87:**1-6.
64. **Paccalet, T., M. Bardor, C. Rihouey, F. Delmas, C. Chevalier, M. A. D'Aoust, L. Faye, L. Vezina, V. Gomord, and P. Lerouge.** 2007. Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthesizing enzymes. Plant Biotechnol J **5:**16-25.
65. **Patel, R. D., J. K. Lodge, and L. G. Baker.** 2010. Going green in *Cryptococcus neoformans*: the recycling of a selectable drug marker. Fungal Genet Biol **47:**191-8.
66. **Penttilä, M., H. Nevalainen, M. Ratto, E. Salminen, and J. Knowles.** 1987. A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. Gene **61:**155-64.
67. **Perrakis, A., I. Tews, Z. Dauter, A. B. Oppenheim, I. Chet, K. S. Wilson, and C. E. Vorgias.** 1994. Crystal structure of a bacterial chitinase at 2.3 A resolution. Structure **2:**1169-80.
68. **Pfaffl, M. W., G. W. Horgan, and L. Dempfle.** 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res **30:**e36.
69. **Pfaffl, M. W., A. Tichopad, C. Prgomet, and T. P. Neuvians.** 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol Lett **26:**509-15.
70. **Pichyangkura, R., S. Kudan, K. Kuttiyawong, M. Sukwattanasinitt, and S. Aiba.** 2002. Quantitative production of 2-acetamido-2-deoxy-D-glucose from crystalline chitin by bacterial chitinase. Carbohydr Res **337:**557-9.
71. **Preston, A., R. E. Mandrell, B. W. Gibson, and M. A. Apicella.** 1996. The lipooligosaccharides of pathogenic gram-negative bacteria. Crit Rev Microbiol **22:**139-80.
72. **Robertus, J. D., and A. F. Monzingo.** 1999. The structure and action of chitinases. EXS **87:**125-35.
73. **Rodrigues, M. L., S. Rozental, J. N. Couceiro, J. Angluster, C. S. Alviano, and L. R. Travassos.** 1997. Identification of N-acetylneuraminic acid and its 9-O-acetylated derivative on the cell surface of *Cryptococcus neoformans*: influence on fungal phagocytosis. Infect Immun **65:**4937-42.
74. **Rodriguez-Aparicio, L. B., M. A. Ferrero, and A. Reglero.** 1995. N-acetyl-D-neuraminic acid synthesis in *Escherichia coli* K1 occurs through condensation of N-acetyl-D-mannosamine and pyruvate. Biochem J **308 (Pt 2):**501-5.
75. **Saito, A., T. Fujii, T. Yoneyama, M. Redenbach, T. Ohno, T. Watanabe, and K. Miyashita.** 1999. High-multiplicity of chitinase genes in *Streptomyces coelicolor* A3(2). Biosci Biotechnol Biochem **63:**710-8.
76. **Salvatore, S., R. Heuschkel, S. Tomlin, S. E. Davies, S. Edwards, J. A. Walker-Smith, I. French, and S. H. Murch.** 2000. A pilot study of N-acetyl glucosamine, a nutritional substrate for glycosaminoglycan synthesis, in paediatric chronic inflammatory bowel disease. Aliment Pharmacol Ther **14:**1567-79.
77. **Sauer, B.** 1987. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol Cell Biol **7:**2087-96.
78. **Schauer, R., S. Kelm, G. Reuter, R. P., and L. Shaw.** 1995. Biochemistry and Role of Sialic Acids, p. 7-67. *In* A. Rosenberg (ed.), Biology of the Sialic Acids. Plenum Press, NY and London.
79. **Schmitz, C., M. Gotthardt, S. Hinderlich, J. R. Leheste, V. Gross, H. Vorum, E. I. Christensen, F. C. Luft, S. Takahashi, and T. E. Willnow.** 2000. Normal blood pressure and plasma renin activity in mice lacking the renin-binding protein, a cellular renin inhibitor. J Biol Chem **275:**15357-62.
80. **Seidl, V., B. Huemer, B. Seiboth, and C. P. Kubicek.** 2005. A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. FEBS J **272:**5923-39.
81. **Soares, R. M., A. S. R. M. de, D. S. Alviano, J. Angluster, C. S. Alviano, and L. R. Travassos.** 2000. Identification of sialic acids on the cell surface of *Candida albicans*. Biochim Biophys Acta **1474:**262-8.
82. **Sternberg, N., and D. Hamilton.** 1981. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. J Mol Biol **150:**467-86.
83. **Sundaram, A. K., L. Pitts, K. Muhammad, J. Wu, M. Betenbaugh, R. W. Woodard, and W. F. Vann.** 2004. Characterization of N-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. Biochem J **383:**83-9.
84. **Tabata, K., S. Koizumi, T. Endo, and A. Ozaki.** 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 Microb Techn **30:**327-333.
85. **Takahashi, S., T. Ohsawa, R. Miura, and Y. Miyake.** 1983. Purification and characterization of renin binding protein (RnBP) from porcine kidney. J Biochem **93:**1583-94.
86. **Tanner, M. E.** 2005. The enzymes of sialic acid biosynthesis. Bioorg Chem **33:**216-28.
87. **Tao, F., Y. Zhang, C. Ma, and P. Xu.** Biotechnological production and applications of N-acetyl-D: -neuraminic acid: current state and perspectives. Appl Microbiol Biotechnol.
88. **Terwisscha van Scheltinga, A. C., M. Hennig, and B. W. Dijkstra.** 1996. The 1.8 A resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. J Mol Biol **262:**243-57.
89. **Tremblay, J. F.** 2006. The other drug for avian flu. C&EN **84:**33-36.
90. **van der Merwe, P. A., P. R. Crocker, M. Vinson, A. N. Barclay, R. Schauer, and S. Kelm.** 1996. Localization of the putative sialic acid-binding site on the immunoglobulin superfamily cell-surface molecule CD22. J Biol Chem **271:**9273-80.

91. **Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol **3:**RESEARCH0034.
92. **Vann, W. F., J. J. Tavarez, J. Crowley, E. Vimr, and R. P. Silver.** 1997. Purification and characterization of the Escherichia coli K1 neuB gene product N-acetylneuraminic acid synthetase. Glycobiology **7:**697-701.
93. **Varki, A.** 1997. Sialic acids as ligands in recognition phenomena. Faseb J **11:**248-55.
94. **Vestweber, D., and J. E. Blanks.** 1999. Mechanisms that regulate the function of the selectins and their ligands. Physiol Rev **79:**181-213.
95. **Wasylnka, J. A., M. I. Simmer, and M. M. Moore.** 2001. Differences in sialic acid density in pathogenic and non-pathogenic *Aspergillus* species. Microbiology **147:**869-77.
96. **Watanabe, T., K. Kimura, T. Sumiya, N. Nikaidou, K. Suzuki, M. Suzuki, M. Taiyoji, S. Ferrer, and M. Regue.** 1997. Genetic analysis of the chitinase system of *Serratia marcescens* 2170. J Bacteriol **179:**7111-7.
97. **Zhang, Y., F. Tao, M. Du, C. Ma, J. Qiu, L. Gu, X. He, and P. Xu.** 2010. An efficient method for N-acetyl-D-neuraminic acid production using coupled bacterial cells with a safe temperature-induced system. Appl Microbiol Biotechnol **86:**481-9.
98. **Zimmermann, V., H. G. Hennemann, T. Daussmann, and U. Kragl.** 2007. Modelling the reaction course of N-acetylneuraminic acid synthesis from N-acetyl-D-glucosamine--new strategies for the optimisation of neuraminic acid synthesis. Appl Microbiol Biotechnol **76:**597-605.

ENGINEERING AN N-ACETYLNEURAMINIC ACID SYNTHESIS PATHWAY INTO *HYPOCREA* (*TRICHODERMA*)

MATTHIAS G. STEIGER, ASTRID R. MACH-AIGNER, RITA GORSCHE, ERWIN E. ROSENBERG, MARKO D. MIHOVILOVIC, ROBERT L. MACH

ENGINEERING AN N-ACETYLNEURAMINIC ACID SYNTHESIS PATHWAY INTO *HYPOCREA* (*TRICHODERMA*)

MATTHIAS G. STEIGER[1#], ASTRID R. MACH-AIGNER[1#], RITA GORSCHE[1], ERWIN E. ROSENBERG[2], MARKO D. MIHOVILOVIC[3], ROBERT L. MACH[1]

[1] Gene Technology and Applied Biochemistry Group, Institute of Chemical Engineering, TU Wien, Getreidemarkt 9/166/5/2, A-1060 Vienna, Austria
[2] Institute of Applied Synthetic Chemistry, TU Wien, Getreidemarkt 9/163, A-1060 Vienna, Austria
[3] Institute of Chemical Technologies and Analytics, TU Wien, Getreidemarkt 9/164 AC, A-1060 Vienna, Austria
# These authors contributed equally to this work.

ABSTRACT

N-acetylneuraminic acid (NeuNAc) belongs to the structural class of sialic acids. Derivatives of NeuNAc are used as neuraminidase inhibitors to treat viral infections like influenza. Therefore, the pharmaceutical industry is interested in a cheap source for NeuNAc, but its synthesis is costly (current market price: ~1000 €/kg). Currently, NeuNAc can be produced either by an *E. coli* process or enzymatic conversion. In both cases the substrate used is N-acetyl-D-glucosamine (GlcNAc), which itself is costly. We developed a whole-cell bio-catalysis process based on the fungus *Hypocrea*. *Hypocrea* is known for its high secretory capacity of hydrolytic enzymes and is able to utilize renewable polysaccharides, like cellulose or chitin, which commonly occur in nature. A metabolic pathway, consisting of two bacterial enzymes, is engineered into *Hypocrea* for the synthesis of NeuNAc. We will illustrate the properties of such recombinant *Hypocrea* strains to form NeuNAc using the renewable biopolymer chitin.

1. INTRODUCTION

NeuNAc is the most prevalent exponent of sialic acids and is an acidic α-keto sugar with a C9 backbone (28). 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 an important role in mediating cellular recognition and adhesion processes (34). Such a process can be found in the infectious cycle of influenza. Influenza is a highly infectious disease, which is caused by RNA viruses. For a propagation of the virus in the host, a neuraminidase (sialidase) activity is needed, which cleaves the linkage between a terminal sialic acid residue and the glycoprotein to release the virus from the host receptor (23). Therefore, sialic acid derivatives are nowadays successfully applied as neuraminidase inhibitors in the therapy of such virus related diseases. Currently, the best known products on the market are „Tamiflu“ (active pharmaceutical ingredient: “Oseltamivir”) by Hoffmann la Roche and „Relenza“ (active pharmaceutical ingredient: “Zanamivir”) by GlaxoSmithKline, both are neuraminidase inhibitors, and Zanamivir represents a direct derivate of the precursor NeuNAc (18, 32).

NeuNAc is traditionally prepared by extraction from natural sources, such as bird nest, milk or eggs (8), by the hydrolysis of colominic acid (a homopolymer of NeuNAc) in the culture broth of *E. coli* K1 (22), or by chemical synthesis (3). A chemoenzymatic process for the production of NeuNAc from GlcNAc and pyruvate has also been reported (1, 19). An alkaline-catalyzed epimerization is performed for a conversion of GlcNAc to N-acetyl-D-mannosamine (ManNAc). ManNAc is then condensed with pyruvate by *E. coli* NeuNAc lyase to yield NeuNAc. Two step enzymatic conversions were also reported converting GlcNAc to ManNAc with a GlcNAc-2-epimerase and further on two NeuNAc using NeuNAc aldolase (10, 15, 21, 36, 41). The drawback of this technology is the second step where an excess of pyruvate is needed to drive the reaction equilibrium towards NeuNAc. Moreover, the substrate GlcNAc is expensive (~1700 €/ kg GlcNAc suitable for cell culture, Sigma-Aldrich) and thus is not an optimal industrial precursor.

In this manuscript, we introduce a strategy, in which the renewable and highly abundant bio-polymer chitin is used as a substrate for the production of NeuNAc. As a production host, *Hypocrea jecorina* (anamorph: *Trichoderma reesei* (12)) was selected, because this filamentous fungus expresses a broad variety of chitinases (29) and is able to depolymerize chitin (7). *H. jecorina* is a saprophyte known for its high secretory capacity of hydrolytic enzymes. The enzymes are used in different industrial applications like pulp and paper production (2, 24, 38), food and feed industry (4, 14, 35), textile industry (9, 13, 25) and biofuels production (6).

H. jecorina is unable to produce NeuNAc. We insert a two step enzyme-catalytic pathway into this fungus, which consist of a GlcNAc-2-epimerase and a NeuNAc-synthase. GlcNAc-2-epimerase requires ATP as a cofactor and catalyzes the conversion of GlcNAc to ManNAc. ManNAc is further converted to NeuNAc using phosphoenolpyruvate (PEP) as co-substrate by

NeuNAc-synthase. Both substances, ATP and PEP, are supplied by the cellular metabolism of the fungus in the whole-cell catalytic process. Therefore, no excess of pyruvate is needed for this process facilitating the downstream processing. In addition, the application of *H. jecorina* offers the advantage of using the cheap and renewable raw material chitin.

2. MATERIALS AND METHODS

2.1. Strains and cultivation conditions

H. jecorina QM9414 (ATCC 26921) was used as parental strain throughout this study and maintained on malt extract agar (MEX).

Mycelia for the enzyme assay were obtained by cultivation of the strains in 1000-mL Erlenmeyer flasks with 200 mL 3% (w/v) MEX media. The flasks were inoculated with 10^8 conidia per liter. Incubation was performed for 40 h at 30 °C and 250 rpm. Cultivation of *H. jecorina* on colloidal chitin was done in 1000-mL Erlenmeyer flasks with 200 mL Mandels-Andreotti (MA) (20) media without the phosphate-citrate buffer containing 1% (w/v), but colloidal chitin, 0.1% (w/v) bacto peptone (Difco, Detroit, US) and inoculated with 10^8 conidia per liter. Incubation was performed for 90 h at 30 °C and 250 rpm. Colloidal chitin was prepared after a protocol previously published (27) using conc. HCl. instead of washing with water, the pH was adjusted with NaOH until neutral and the chitin suspension, containing NaCl, was dialyzed against water.

For the *in vivo* production of NeuNAc, direct cultivations of the fungi were performed in 1000-ml Erlenmeyer flasks with 250 ml MA medium containing 1% (w/v) GlcNAc and inoculated with 10^8 conidia per liter. Incubation was performed for 66 h at 30°C and 250 rpm.

2.2. Synthetic genes and plasmid construction

The synthetic gene *tbage* (supplementary data 2) was designed by using the protein sequence of *Anabaena sp.* CH1 GlcNAc-2-epimerase (GenBank: ABG57042) and by translating it into a DNA sequence with the software GeneOptimizer® (Geneart, Regensburg, Germany). Hereby, the codon usage was optimized for *H. jecorina* (http://www.kazusa.or.jp/codon). The synthetic gene *tneub* (supplementary data 2) was obtained in the same way, but the protein sequence was from *Campylobacter jejuni* NCTC11168 NeuNAc-synthase (http://old.genedb.org/genedb/cjejuni/index.jsp, Cj1141).

For the construction of plasmids pMS-PEC and pMS-PSC, the synthetic genes *tbage* and *tneub* were released from the production plasmid via *Xba*I/*Nsi*I digestion and inserted into $pRLM_{ex}30$ (17) replacing the *hph* cassette between *Xba*I and *Nsi*I sites.

For the construction of pGEX-epi and pGEX-syn, the plasmid pGEX4T-2 (GE Healthcare, Chalfont St Giles, UK) was digested with *EcoR*I and *Xho*I. A double strand linker, consisting of oligonucleotide GEXfw and GEXrev (Table 1), containing an *Xba*I and *Nsi*I site was ligated into the open pGEX4T-2 resulting in pGEX-MS. *tbage* and *tneub* were inserted into pGEX-MS via *Xba*I/*Nsi*I resulting in plasmids pGEX-epi and pGEX-syn.

Table 1 Oligonucleotide sequences

Name	Sequence (5´→3´)	Usage
NANASfw	GTGGTGTGCAGGA GGACGAA	qPCR *tneub*
NANASrev	CAAGCACATCGCC CAGTTCAAG	qPCR *tneub*
ManEfw	GCGATCTTGAGCC AGTTCTC	qPCR *tbage*
ManErev	GCTACTTCACCTG CCTCGAC	qPCR *tbage*
GEX-MSfw	AATTCCTTCTAGA GATATGCATC	Construction pGEX-MS
GEX-MSrev	TCGAGATGCATAT CTCTAGAAGG	Construction pGEX-MS
pkifw R	CTGCGACACTCAG AACATGTACGT	qPCR *pki* cDNA
pkifw D	GCTCTGCTTGGAA CCTGATTGA	qPCR *pki* DNA
pkirev	GGTCTGGTCGTCC TTGATGCT	qPCR *pki*
sar1fw	TGGATCGTCAA CTGGTTCTACGA	qPCR *sar1*
sar1rev	GCATGTGTAGCAA CGTGGTCTTT	qPCR *sar1*

2.3. Protoplast transformation of *H. jecorina*

The protoplast transformation of *H. jecorina* was carried out as described earlier (5). A total amount 10 µg of DNA was used in a transformation. In a co-transformation the plasmids pMS-PEC (4 µg) and pMS-PSC (4 µg) were transformed into the fungal genome together with the plasmid pHylox2 (2 µg) (paper in preparation, M.G.S., M. Vitikainen, K. Brunner, G. Adam, M. Saloheimo, R.L.M and A. R. M.) conferring hygromycin B resistance. Recombinant strains were selected for hygromycin B resistance (17).

2.4. RNA analysis

RNA extraction, reverse transcription and qPCR were carried out as described earlier (30). Oligonucleotide sequences are listed in Table 1. *sar1* was used as reference gene for RT-qPCR normalization. For *tbage* qPCR primers ManEfw and ManErev were used at an elongation temperature of 64 °C and 2 mM $MgCl_2$ were added to the qPCR amplification mixture. For *tneub* qPCR primers NANAfw and NANArev were used at an elongation temperature of 64 °C. For *pki* qPCR primers pkifwR and pkirev were used at an elongation temperature of 64 °C. Data analysis of qPCR data was carried out using REST 2008 (26).

2.5. DNA analysis

Fungal genomic DNA was isolated as described previously (5). Southern hybridization and detection was carried out with the DIG High Prime DNA Labeling and Detection Starter Kit II using the standard procedure (Roche, Basel, Switzerland). qPCR of genomic DNA for quantification of gene copies was done using ~50 ng genomic DNA as template. The same primers were used for *tbage* and *tneub* as metioned for RNA analysis. As reference, the gene *pki* was used with the primers pkifwD and pkirev at an elongation temperature of 64 °C. Data analysis of qPCR data was carried out using REST 2008 (26).

2.6. Gluthathione S-transferase (GST) fusion proteins

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

2.7. Enzyme assay using cell free extracts

Harvested mycelia were ground to a fine powder under liquid nitrogen and resuspended in a 0.1 M bicine buffer (pH 8) containing protease inhibitors (2 µM leupeptin, 1 µM pepstatin A, 10 µM PMSF) (0.3 g mycelia/1 mL buffer). The suspension was sonicated on a Sonifier® 250 Cell Disruptor (Branson, Danbury, US) (power 40%, duty cycle 50%, power for 20 sec, 40 sec pause, 10 cycles) and the insoluble compounds separated via centrifugation (10 min, 13000 x g, 4 °C). Enzymatic analysis was carried out following a modified protocol reported earlier (33). The assay for GlcNAc-2-epimerase contained 10 mM GlcNAc, 0.2 mM ATP, 100 mM bicine buffer (pH 8) and 10-40 µL cell free extract in a total volume of 100 µL. The assay for NeuNAc-synthase contained 10 mM ManNAc, 10 mM PEP, 12.5 mM $MnCl_2$, 100 mM bicine buffer (pH 8) and 10-40 µL cell free extract in a total volume of 100 µL. The combination assay for GlcNAc-2-epimerase and NeuNAc-synthase contained 10 mM GlcNAc, 10 mM PEP, 12.5 mM $MnCl_2$, 100 mM bicine buffer (pH 8) and 40 µL cell free extract in a total volume of 100 µL. All reactions were incubated for 60 min at 37 °C, terminated via heat inactivation at 85 °C for 10 min and then analyzed by HPLC. As a positive control 5 µL of the GST-fusion proteins, GST-epi or GST-syn, were applied instead of the cell free extracts.

2.8. Chitinase assay

The assay measuring the release of GlcNAc from chitin contained 3.5% (w/v) colloidal crab-shell chitin or crude crab-shell chitin in a 30 mM phosphate buffer (pH 4.7) and 5, 10 or 50 µL of culture supernatant (from a 90 h cultivation of recombinant *H. jecorina* strain PEC/PSC1 on colloidal chitin). The total volume was adjusted to 1.5 mL. The reaction was incubated at 37 °C for 20 h and stopped via heat inactivation at 90 °C for 10 min. GlcNAc formation was measured via HPLC using pure standards for the calibration.

2.9. NeuNAc detection in cell free extracts

Harvested *H. jecorina* mycelia were ground to a fine powder under liquid nitrogen and resuspended water (0.3 g mycelia/1 mL H_2O). The suspension was sonicated on a Sonifier® 250 Cell Disruptor (Branson) (power 70%, duty cycle 50%, power for 1 min, 1 min pause, 3 cycles) and the insoluble compounds were separated via centrifugation (10 min, 13000 x g, 4 °C). The supernatant was filtered through a 0.45 µm filter and further analyzed via HPLC-MS/MS.

2.10 HPLC-MS/MS analysis

The formation of NeuNAc and ManNAc in the enzyme assay and in the cell free extracts was measured by means of LC-MS (IT-TOF-MS) (Shimadzu, Kyoto, Japan) with a Rezex™ RHM-Monosaccharide H^+-column (8%, 300 x 7.8 mm) (Phenomenex, Torrance, USA). The mobile phase consisted of water with 0.1% (v/v) trifluoroacetic acid and the flow was set to 0.6 mL/min. The column temperature was 80 °C and the injected volume 10 µL. MS detection was carried out in ESI+ mode covering a scan range of 60-600 amu.

3. RESULTS

3.1 *In silico* analysis of the NeuNAc pathway in *H. jecorina*

Until now, no evidence in literature can be found about the formation of sialic acids in *Hypocrea*. First, we had a look on the known metabolic pathways *in silico*, which lead to NeuNAc and whether those pathways exist in eukaryotes like filamentous fungi (www.expasy.ch/cgi-bin/search-biochem-index, Biochemical Pathways, Roche). Figure 1 depicts the currently known enzyme-catalyzed pathways starting from the polymer chitin.

Comparing those pathways with *H. jecorina* genome, we found several genes annotated in the genome which are necessary for the catabolisation of chitin. The first step from the polymer chitin to its monomer GlcNAc is catalyzed by chitinases (3.2.1.14) (29). Table 2 lists the protein candidates annotated in *H. jecorina* for this pathway. Furthermore, the activity of a hexokinase (EC 2.7.1.1), GlcNAc-6-phosphate deacetylase (EC 3.5.1.25) and glucosamine-6-phosphate deaminase (EC 3.5.99.6) are necessary till fructose-6-phosphate can enter glycolysis. For all those enzymatic steps at least one candidate gene can be found in the annotated *H. jecorina* genome. Genes

necessary for the synthesis of chitin can also be found including phosphoacetylglucosamine mutase (EC 5.4.2.3), UDP-N-GlcNAc diphosphorylase (EC 2.7.7.23) and chitin synthase (EC 2.4.1.16). But no genes are annotated for the synthesis of ManNAc (EC 5.1.3.8 in bacteria, EC 5.1.3.4 in mammals) nor for the synthesis of NeuNAc in *H. jecorina* (EC 2.5.1.6. in bacteria, EC 2.7.1.60, EC 2.5.1.57, EC 3.1.3.29 in mammals).

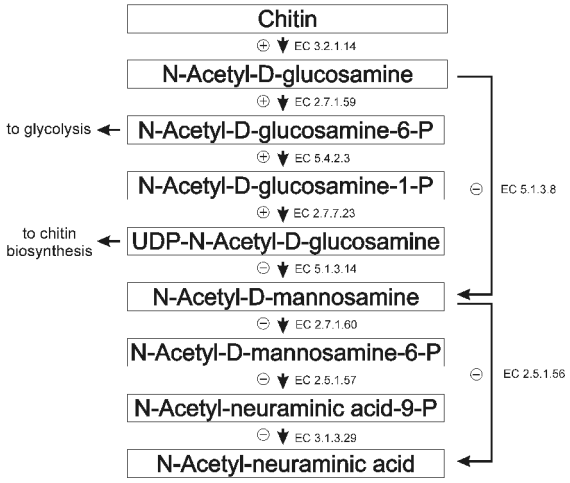


Figure 1 Flow chart of the metabolic pathways leading to the formation of NeuNAc starting from the polymer chitin. Intermediate metabolites are presented in boxes. Arrows indicate enzyme-catalyzed reactions. Next to the arrow the EC number of the catalyzing enzyme is presented. Circled + (plus) indicate that annotated candidate genes for this enzyme activity can be found in *H. jecorina* genome annotation (Table 2) (http://genome.jgi-psf.org/Trire2/Trire2.home.html). Circled – (minus) indicate that no genes for this enzyme activity can be found in *H. jecorina* genome annotation.

3.2 Gene cluster for the catabolisation of GlcNAc in *H. jecorina*

During the *in silico* analysis of the catabolic pathways of GlcNAc, it was found, that three candidate genes (estExt_GeneWisePlus.C_140427, estExt_GeneWisePlus.C_140421, estExt_Genewise1.C_140432) encoding for a hexokinase, GlcNAc-6-phosphate deacetylase and glucosamine-6-phosphate deaminase are located in close vicinity in the genome of *H. jecorina* (scaffold 14: 714385-729984). For *Candida albicans* it was shown by Yamada and co-workers, that six genes necessary for the catabolisation of GlcNAc are organized in a cluster (39, 40). Three of the genes (*nag1*, *nag2*, *nag5*) encode for the same activity as mentioned above in *H. jecorina*. The other three (*nag3*, *nag4*, *nag6*) were found to be required for the uptake of extracellular hexoses not only GlcNAc, but also glucose and galactose (40). With a tblastn alignment search it was not possible to find homologs of those genes in vicinity to the cluster in *H. jecorina*. A similar cluster was also found in *Fusarium graminearum* (*F. graminearum* genome: supercontig 5: 1999956-2009033, Broad Institute), *Neurospora crassa* (*N. crassa* genome: supercontig6: 553870-570540, Broad Institute), and *Aspergillus fumigatus* (*A. fumigatus* genome: chromosome 8: 878656-892484, Broad Institute). Therefore, it seems to be a conserved cluster among filamentous fungi, which is involved in the catabolisation of GlcNAc. The hexokinase (EC 2.7.1.1) annotated in the genome of *H. jecorina* with the protein ID 79677 could be further specified as a GlcNAc-kinase (EC 2.7.1.59) as it is annotated and characterized for *C. albicans* (39). Furthermore, a gene (estExt_GeneWisePlus.C_140419), next to the GlcNAc-6-phosphate deacetylase (estExt_GeneWisePlus.C_140421), could also belong to the cluster, because in *N. crassa* it is annotated as a β-N-acetylglucosaminidase (*N. crassa* OR74A (NC10): Supercontig 6: 560844-564980). This gene cluster outlines interesting targets for further genetic engineering of a NeuNAc production strain.

3.3 Construction of expression vectors

We chose a two enzyme strategy for the production of NeuNAc in *Hypocrea* using a GlcNAc 2-epimerase (EC 5.1.3.8) and a NeuNAc synthase (EC 2.5.1.99). For the GlcNAc 2-epimerase, we took the protein sequence of *Anabaena sp.* CH1 (GenBank: ABG57042) and for the NeuNAc-synthase the sequence of *Campylobacter jejuni* NCTC11168 (http://old.genedb.org/genedb/cjejuni/index.jsp, Cj1141). The protein sequences were translated into DNA sequences using the software GeneOptimizer® (Geneart) and the codon usage of *H. jecorina* (http://www.kazusa.or.jp/codon). The two synthetic genes were named *tbage* and *tneub* (sequences are given in the supplementary data). The coding sequences were synthesized by Geneart and inserted into the plasmid $pRLM_{ex}30$ replacing the *hph* gene. Thus, both genes are under the control of the constitutive *pki* promotor of *H. jecorina* and its *cbh2* terminator.

3.4 Insertion of the GlcNAc 2-Epimerase and the NeuNAc synthase into *H. jecorina*

In order to generate *H. jecorina* strains, which have the ability to produce NeuNAc, we transformed the parental strain QM9414 with the plasmids pMS-PEC (carrying the *tbage* gene), pMS-PSC (carrying the *tneub* gene) and pMS-Hylox2 (carrying the marker gene *hph* between two loxP-sites) in a co-transformation experiment. Fifteen hygromycin B resistant strains where screened on DNA level (PCR) for the integration of *tbage* and *tneub*. Two strains PEC/PSC1 and PEC/PSC10 contained both genes. In both strains, the transcription of the genes were analyzed by RT-qPCR after cultivation on glycerol for 20 h. *sar1* was used as reference gene, because this gene exhibits a stable transcription profile (30). The ratio of the transcription levels are presented in Table 3. The PEC/PSC1 strain shows a higher transcription (~2 fold) of *tbage* than

Table 2 Candidate genes annotated in the *H. jecorina* genome involved in the metabolic pathway of chitin and GlcNAc

EC No.	Name	Protein identities
EC 3.2.1.14	chitinase	2735, 43873, 53949, 62645, 62704, 66041, 68347, 72339, 80833, 81598, 104401, 110317, 119859, 123354, 124043
EC 2.7.1.1	hexokinase	56129, 73665, 79677
EC 3.5.1.25	GlcNAc-6-phosphate deacetylase	79671
EC 3.5.99.6	glucosamine-6-phosphate deaminase	49898
EC 5.4.2.3	phosphoacetylglucosamine mutase	80994
EC 2.7.7.23	UDP-N-GlcNAc diphosphorylase	79568
EC 2.4.1.16	chitin synthase	51492, 55341, 58188, 71563, 112271, 122172

Table 3 Comparison of gene transcription and gene copy numbers between two recombinant *H. jecorina* strains

Gene name	Transcript ratio PEC/PSC1 / PEC/PSC10 median [95% C.I.]	Copy number ratio PEC/PSC1 / PEC/PSC10 median [95% C.I.]
tbage	2.021 [1.589-2.836]	1.810 [1.376-2.585]
tneub	0.479 [0.385-0.622]	0.400 [0.320-0.492]

Table 4 Chitinase activity formed during cultivation of *H. jecorina* PEC/PSC1 strain on 1% chitin

Substrate	Chitinase activity [a] [mU/mL]
crab-shell chitin	2,7 ± 0,5
colloidal crab-shell chitin	25,0 ± 0,9

[a] 1U: release of 1µmol GlcNac/min at 37 °C

PEC/PSC10. In return, PEC/PSC10 has a higher transcription level (~2 fold) of *tneub* than PEC/PSC1. These different transcription ratios can be explained by different copy numbers of *tneub* and *tbage* integrated into the genome. The ratio of the copy numbers was measured by qPCR of genomic DNA using *pki* as reference because only one copy of *pki* is present in the *H. jecorina* genome. The same ratios between PEC/PSC1 and PEC/PSC10 can be measured for the genomic DNA of *tneub* and *tbage* (Table 3) compared to the transcript formation. The same ratio of transcripts and gene dose indicates that every gene copy is transcribed with the same efficiency (Table 3).

Gene copy numbers can also be estimated from a southern blot analysis, for which a restriction enzyme was used that cut only once in the *tbage* or *tneub* gene (supplementary data 1). In case of PEC/PSC1 strain, 4 different sized bands can be found for a *tbage* and a *tneub* southern blot (supplementary data 1). This indicates that 2 copies of each gene are present in the genome. In case of PEC/PSC10 strain, the southern for *tbage* shows 3 bands and 4 bands for *tneub* (supplementary data 1). In case of *tbage*, 1 to 2 copies could have integrated into the genome. In case of *tneub*, one band has the size of the plasmid (5445 bp) and the signal is more intense. This indicates that the plasmid has integrated in tandem and the copy number can be estimated from the intensity of the plasmid-sized band. Another way to determine the copy numbers in PEC/PSC 10 is to use the copy number ratios of PEC/PSC1 to PEC/PSC10 (Table 3) measured by qPCR. Assuming that PEC/PSC1 has 2 copies of *tneub* it results in 4 copy numbers for PEC/PSC10 strain.

3.5 Heterologous protein expression of GlcNAc-2-epimerase and NeuNAc synthase in *H. jecorina*

Because the recombinant strain PEC/PSC1 has 2 copies of each heterologous gene, this strain was further analyzed. After cultivation of *H. jecorina* strain PEC/PSC1 on malt extract agar, the cell free extract was tested for the presence of GlcNAc-2-epimerase and NeuNAc-synthase with an enzyme assay. The conversion of the substrates PEP and GlcNAc towards ManNAc and NeuNAc was measured applying the cell free extract as biocatalyst. The corresponding chromatograms obtained after HPLC analysis of the enzyme assays are presented in Figure 2. As a positive control, GST-fusion proteins of GlcNAc-2-epimerase (*tbage*) and NeuNAc-synthase (*tneub*) were expressed in *E. coli* and used in the enzyme assay. Using them, the formation of ManNAc and NeuNAc demonstrates that the synthetic genes *tbage* and *tneub* are expressed as functional proteins (Fig. 2a1 and Fig. 2b1). Applying the cell-free extracts of the recombinant strain PEC/PSC1, the formation of ManNAc (Fig. 2a2) and NeuNAc can be detected (Fig. 2b2). Those results demonstrate that both heterologous proteins are expressed in the strain PEC/PSC1 and are fully functional.

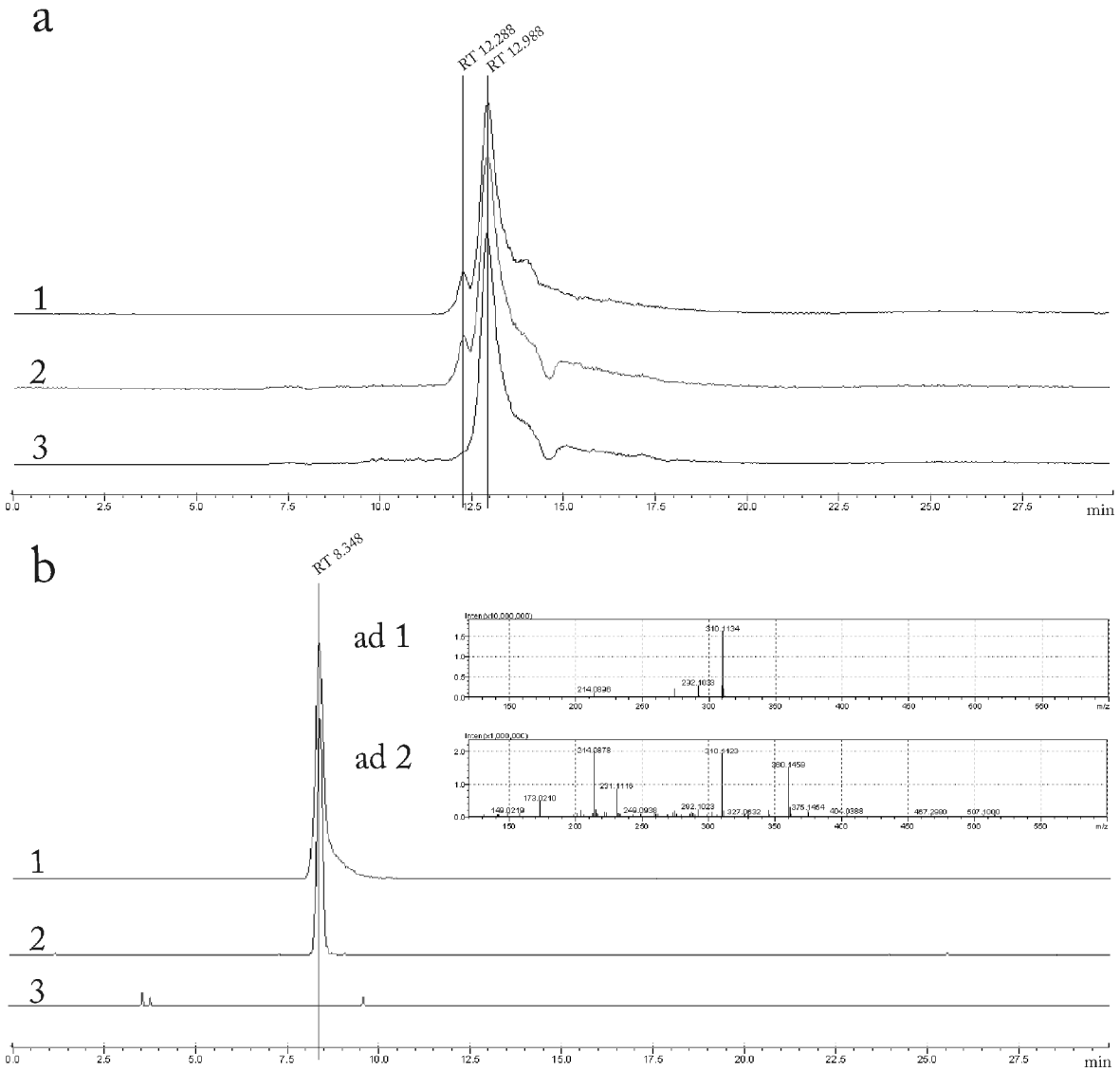


Figure 2 Formation of NeuNAc *in vitro* with heterologously expressed *H. jecorina* proteins of strain PEC/PSC1 in the presence of GlcNAc, ATP and PEP. (a) The extracted ion chromatograms (EIC) of the HPLC-MS analysis at a mass of 222.098 atomic mass units (amu) are shown. This mass corresponds to the mass of the ions of $[GlcNAc+H]^+$ as well as $[ManNAc+H]^+$ formed in the ESI ionisation. The retention time (RT) of GlcNAc (12.988 min) and ManNAc (12.288) were determined with pure standards of both substances and indicated as vertical lines in the chromatograms. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase expressed in *E. coli*. (2) Chromatogram of the *in vitro* assay with the cell free extract of the PEC/PSC1 strain. (3) Chromatogram of the *in vitro* assay with the cell free extract of the QM9414 strain. (b) The EICs at a mass of 310.1134 amu are shown, which corresponds to the mass of the ion of $[NeuNAc+H]^+$. The retention time (RT) of NeuNAc (8.348 min) was determined with a pure standard and is shown as vertical line in the chromatograms. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase expressed in *E. coli*. (2) Chromatogram of the *in vitro* assay with the cell free extract of the PEC/PSC1 strain. This chromatogram is amplified 10 times compared to chromatogram 1. (3) Chromatogram of the *in vitro* assay with the cell free extract of the QM9414 strain. This chromatogram is amplified 1000 times compared to chromatogram 1. (ad 1) MS spectrum of chromatogram 1 at RT of 8.348 min. (ad 2) MS spectrum of chromatogram 2 at RT of 8.348 min.

Neither NeuNAc nor ManNAc is formed using the cell free extract of the parental strain QM9414 (Fig. 2a3 and Fig. 2b3). This indicates that no GlcNAc-2-epimerase activity is present in *H. jecorina*. To test whether a NeuNAc-synthase activity is present in *H. jecorina*, ManNAc and PEP were used as substrates in an additional enzyme assay. Also in this assay, no NeuNAc was detectable using the cell free extract of QM9414 (data not shown).

3.6 Growth of *H. jecorina* on colloidal chitin and GlcNAc production

We cultivated *H. jecorina* PEC/PSC1 on colloidal chitin as sole carbon source in order to hydrolyze chitin into its monomer GlcNAc. The cultivation was monitored by measuring the chitinase activity with a p-nitrophenyl-β-N,N-diacetylchitobiose substrate (data not shown). After 90 h of cultivation the chitinase activity reached a maximum and the supernatant was tested for GlcNAc formation on crude and colloidal chitin. Table 4 presents the data.

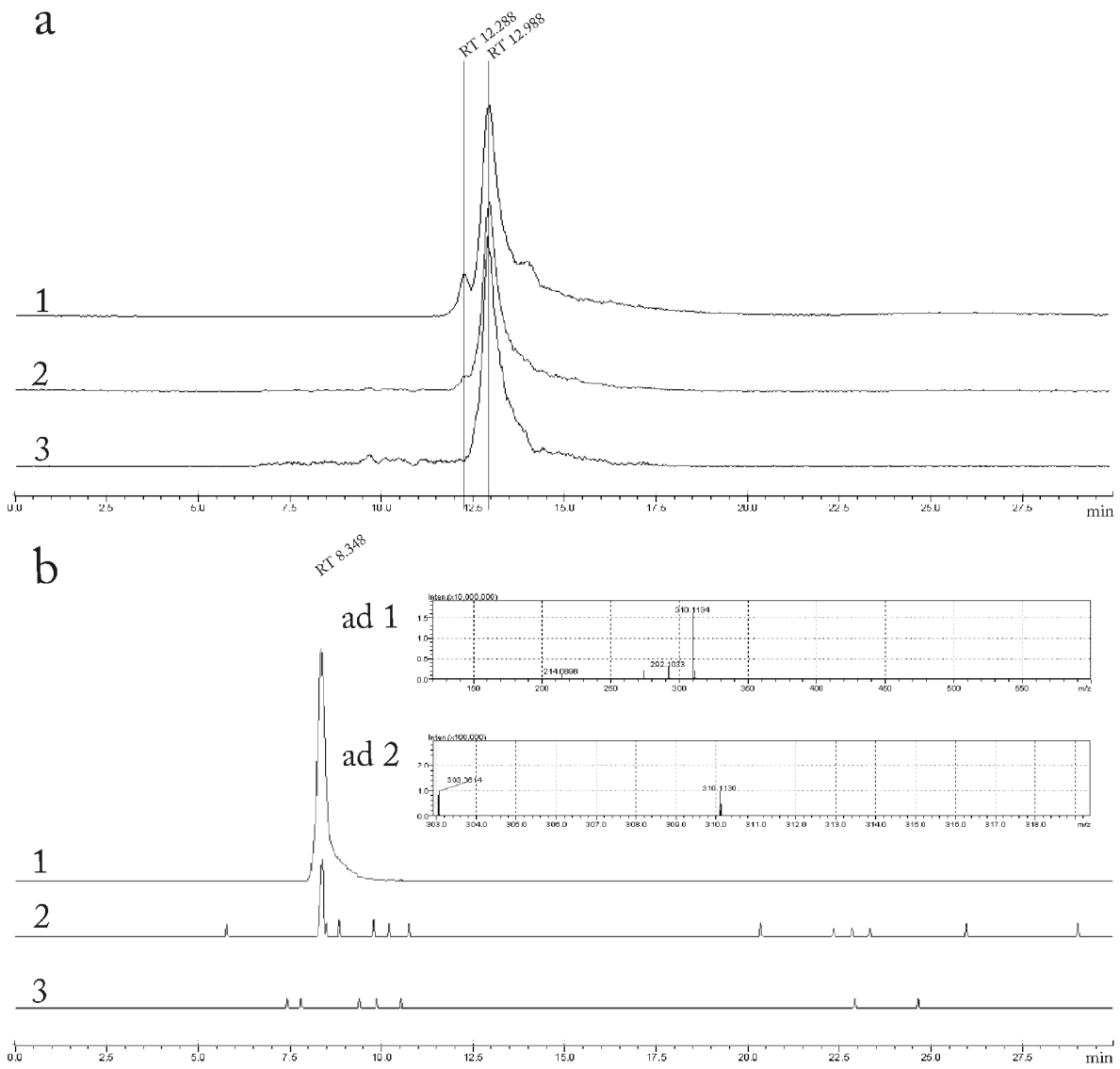


Figure 3 Formation of NeuNAc *in vivo* in *H. jecorina* recombinant strain PEC/PSC1 after cultivation on GlcNAc for 66 h. (a) EICs of the HPLC-MS analysis at a mass of 222.098 atomic mass units (amu) are shown. This mass corresponds to the mass of the ions of $[GlcNAc+H]^+$ as well as $[ManNAc +H]^+$ formed in the ESI ionisation. The RT of GlcNAc (12.988 min) and ManNAc (12.288) were determined with pure standards of both substances and indicated as vertical lines in the chromatograms. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase expressed in *E.coli* (positive control). (2) Chromatogram of the cell free extract of the PEC/PSC1 strain. (3) Chromatogram of the cell free extract of the QM9414 strain (negative control). (b) The EICs at a mass of 310.1134 amu are shown, which corresponds to the mass of the ion of $[NeuNAc+H]^+$. The retention time (RT) of NeuNAc (8.348 min) was determined with a pure standard and is shown as vertical line in the chromatograms. (1) Chromatogram of the *in vitro* assay using GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase expressed in *E. coli*. (2) Chromatogram of the cell free extract of the PEC/PSC1 strain. This chromatogram is amplified 100 times compared to chromatogram 1. (3) Chromatogram of the cell free extract of the QM9414 strain. This chromatogram is amplified 1000 times compared to chromatogram 1. (ad 1) MS spectrum of chromatogram 1 at RT of 8.348 min. (ad 2) MS spectrum of chromatogram 2 at RT of 8.348 min.

Ten times more GlcNAc can be released from colloidal crab-shell chitin than from crude crab-shell chitin. The released GlcNAc can be used as a carbon source for the cultivation of the recombinant PEC/PSC1 strain to produce NeuNAc.

3.7 *In vivo* formation of NeuNAc in *H. jecorina*

The next experiment shall elucidate whether the two heterologously expressed enzymes are functional *in vivo* and are able to process their substrates to form NeuNAc. Therefore, the recombinant strain PEC/PSC1 was cultivated on GlcNAc for 66 h. As a negative control, the parental strain QM9414 was co-cultivated. The harvested mycelia were decomposed and investigated for the formation of NeuNAc. The corresponding chromatograms are depicted in Figure 3. GST fusion proteins of GlcNAc-2-epimerase and NeuNAc synthase expressed in *E. coli* were used as positive control for ManNAc and NeuNAc formation (Fig. 3a1 and Fig. 3b1). The recombinant strain PEC/PSC1 forms ManNAc and NeuNAc (1 µg NeuNAc per g mycelia) (Fig. 3a2 and Fig. 3b2). These results demonstrate that NeuNAc can be produced in *H. jecorina* co-expressing two bacterial enzymes GlcNAc-2-

epimerase and NeuNAc-synthase. The parental strain QM9414 shows no formation of NeuNAc or ManNAc (Fig. 3a3 and Fig. 3b3).

4. DISCUSSION

In this paper, the engineering of an intracellular metabolic pathway into *H. jecorina* is presented. To the best of our knowledge, this is the first work where an intracellular two-step enzym cascade is introduced into a filamentous fungus. *H. jecorina* itself is not able to synthesize NeuNAc (Fig. 3b3), but the essential intermediate metabolite GlcNAc is formed by *H. jecorina*. This substance is obtained by depolymerization of the renewable resource chitin (Table 4). *H. jecorina* expresses a high variety of chitinases (reviewed by Seidl et al.(29)) and is able to metabolize and to grow on the monomer GlcNAc. The polymer chitin can also be synthesized by *H. jecorina*, because it is an essential component of the fungal cell wall (11). Therefore, candidate genes encoding pathways for the biosynthesis and the catabolisation of chitin can be found in *H. jecorina* (Table 2). The biosynthesis of NeuNAc starts from intermediates (GlcNAc, UDP-GlcNAc) of the chitin pathways (Fig. 1), which are readily available in *H. jecorina*. But, no candidate genes can be found in the *H. jecorina* genome, which have a similarity to genes encoding UDP-GlcNAc-2-epimerase, ManNAc kinase, NeuNAc-9-phosphate synthase or NeuNAc-9-phosphatase (Fig. 1). The pathway consisting of those four enzyme steps is found in mammals. An alternative pathway, found in bacteria, requires the enzyme activity of a GlcNAc-2- epimerase and a NeuNAc-synthase (Fig. 1). Also in this case, no candidate genes can be found in the genome of *H. jecorina*. In *A. fumigates*, NeuNAc is present on the surface of the conidia (37), but no formation of NeuNAc was detectable on the conidia of the *H. jecorina* parental strain QM9414 (M.G.S., A.R.M., R.L.M. unpublished data). Furthermore, cell free extracts of QM9414 were investigated for the detection of NeuNAc and ManNAc (Fig. 3a3 and Fig. 3b3) or enzymatic activity of GlcNAc-2-epimerase and NeuNAc-synthase activity (Fig. 2a3 and Fig. 2b3). Neither any enzyme activity nor traces of NeuNAc and ManNAc were detectable. Thus, we conclude that *H. jecorina* is natively not able to form ManNAc or NeuNAc.

Therefore, the necessary enzyme activities had to be engineered into *H. jecorina* in order to produce NeuNAc. The first enzyme in the cascade, a GlcNAc-2-epimerase, was taken from *Anabaena sp*. CH1. This enzyme is well characterized and requires a low concentration of ATP (20 µM) as cofactor for its maximal activity (15). The second enzyme, a NeuNAc-synthase, was chosen from *C. jejuni*, which is also well described in literature (16, 31). The codon-usage of both genes was adapted to the one of *H. jecorina* in order to optimize the expression of the bacterial genes in the fungal host. For the expression system, the strong and constitutive promotor of *pki* was chosen, which was already successfully applied for the expression of another bacterial gene (*hph*) (17) in *H. jecorina*.

The obtained *H. jecorina* recombinant strain PEC/PSC1, carrying both genes, showed GlcNAc-2-epimerase activity and NeuNAc-synthase activity (Fig. 2a2 and 2b2). In a further experiment, we tested the ability of the fungus to produce NeuNAc also *in vivo*. Therefore, we cultivated the fungus on the renewable biopolymer chitin, which cause the degradation to its monomer GlcNAc (Table 4). By cultivating the recombinant strain PEC/PSC1 on GlcNAc, we were able to show that NeuNAc is formed in the fungal cell (Fig. 3b2). Although the total amount of NeuNAc formed in the strain PEC/PSC1 is low (1 µg NeuNAc per g mycelia), this work presents for the first time a strategy, how an intracellular enzyme-cascade consisting of heterologously expressed genes can be realized in a filamentous fungus. Further engineering of the strains might help to increase the yields of NeuNAc. Especially, the metabolic pathways of GlcNAc catabolisation can be a promising target for its further engineering, because reducing the flux of GlcNAc towards glycolysis might increase the flux towards NeuNAc production. As the genes of the GlcNAc pathway seem to be organized in a gene cluster as it was found for *C. albicans* (39, 40), targeted engineering of this cluster is conceivable. The genes could be put under the control of weaker promoters or partially silenced. Further analysis of this gene cluster could also reveal the function of unknown genes located in the vicinity and their possible involvement in the metabolization of GlcNAc.

By introducing a two step enzyme cascade into *H. jecorina*, the fungus gains the ability to form NeuNAc. This work demonstrates such a high-value fine chemical can be produced from a renewable feedstock, like chitin, using the whole-cell catalyst *H. jecorina*. Not only chitin but a plethora of other renewable carbon sources like cellulose and hemicelluloses can be used by this organism. This highlights the potential of a saprophytic organism, like *H. jecorina*, to be used as a production host for chemicals.

ACKNOWLEDGMENTS

This study was supported by a grant from the Austrian Fonds zur Förderung Wissenschaftlicher Forschung (P21287) and by a grant from the Vienna University of Technology, “DemoTech – Innovative Project”, both of them given to R.L.M., which is gratefully acknowledged.

REFERENCES

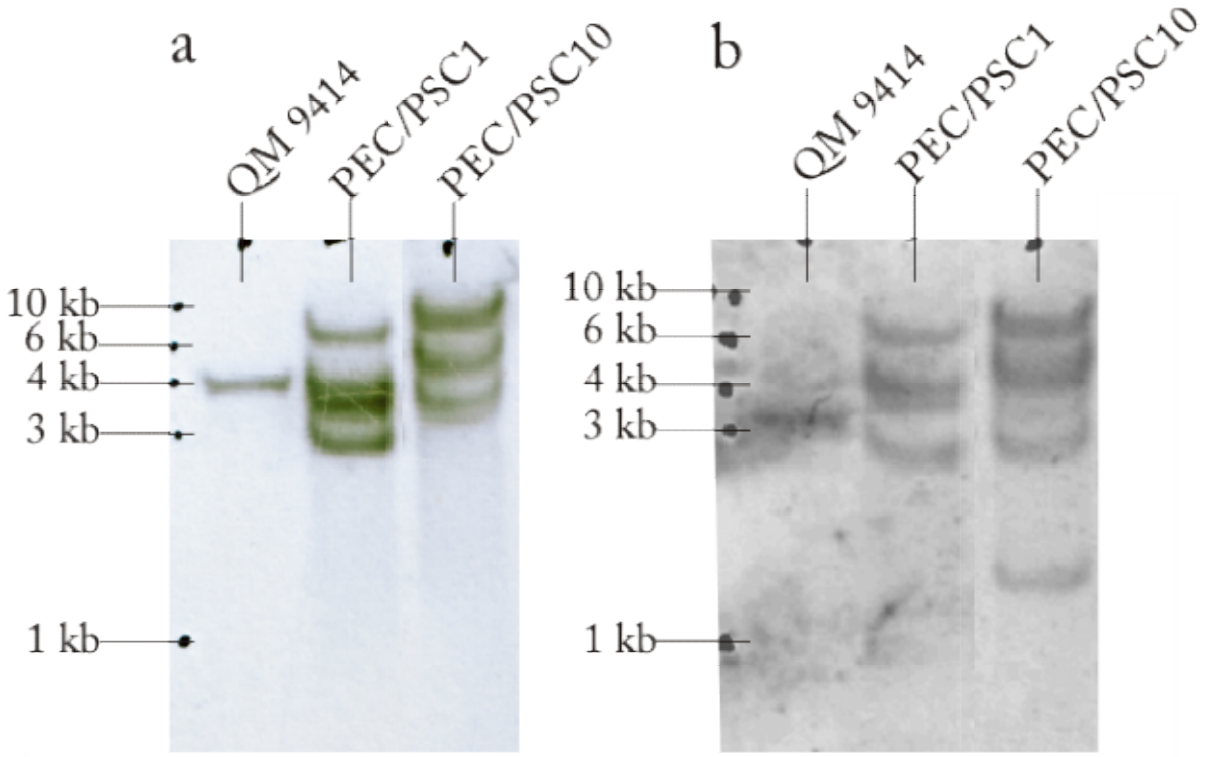
1. **Blayer, S., J. M. Woodley, M. J. Dawson, and M. D. Lilly.** 1999. Alkaline biocatalysis for the direct synthesis of N-acetyl-D-neuraminic acid (Neu5Ac) from N-acetyl-D-glucosamine (GlcNAc). Biotechnol Bioeng **66:**131-6.

2. **Buchert, J., T. Oksanen, J. Pere, M. Siika-aho, A. Suurnäkki, and L. Viikari.** 1998. Applications of *Trichoderma reesei* enzymes in the pulp and paper industry, p. 343-357. *In* G. E. Harman and C. P. Kubicek (ed.), *Trichoderma* & *Gliocladium*, vol. 2. Taylor & Francis Ltd., London, UK.
3. **de Ninno, M.** 1991. The synthesis and glycosidation of N-acetyl-D-neuraminic acid. Synthesis **8:**583-593.
4. **Galante, Y. M., Monteverdi, r., Inama, S., Caldini, c., De Conti, A., Lavelli V., Bonomi, F.** 1993. New applications of enzymes in wine making and olive oil production. Italian Biochem. Soc. trans. **4:**34.
5. **Gruber, F., J. Visser, C. P. Kubicek, and L. H. de Graaff.** 1990. The development of a heterologous transformation system for the cellulolytic fungus Trichoderma reesei based on a pyrG-negative mutant strain. Curr Genet **18:**71-6.
6. **Himmel, M. E., S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, and T. D. Foust.** 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science **315:**804-7.
7. **Klein, D., and D. E. Eveleigh.** 1998. Ecology of Trichoderma, p. 57-74. *In* G. E. Harman, Kubicek, C.P. (ed.), *Trichoderma* & *Gliocladium*, vol. 1. Taylor & Francis, London.
8. **Koketsu, M., L. R. Juneja, H. Kawanami, M. Kim, and T. Yamamoto.** 1992. Preparation of N-acetylneuraminic acid from delipidated egg yolk. Glycoconj J **9:**70-4.
9. **Koo, H., Ueda, M., Wakida, T., Yoshimura, Y., Igarashi, T.** 1994. Cellulase treatment of cotton fabrics. Textile Res. J. **64:**70-74.
10. **Kragl, U., D. Gygax, O. Ghisalba, and C. Wandrey.** 1991. Enzymatic two step synthesis of N-acetylneuraminic acid in the enzyme membrane reactor. Angew Chem Int Ed Eng **30:**827-828.
11. **Kubicek-Pranz, E. M.** 1998. Nutrition, cellular structure and basic metabolic patways in Trichoderma and Gliocladium, p. 95-119. *In* G. E. Harman, Kubicek, C.P. (ed.), *Trichoderma* & *Gliocladium*, vol. 1. Taylor & Francis, London.
12. **Kuhls, K., E. Lieckfeldt, G. J. Samuels, W. Kovacs, W. Meyer, O. Petrini, W. Gams, T. Borner, and C. P. Kubicek.** 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:**7755-60.
13. **Kumar, A., Lepola, M., Purtell, C.** 1994. Enyme finishing of man-made cellulosic fabrics. Textile Chemist and Colorist **26:**25-28.
14. **Lanzarini, G., Pifferi, P.G.** 1989. Enzymes in the fruit juice industry., p. 189-222. *In* C. Cantarelli, Lanzarini, G. (ed.), Biotechnology Applications in Beverage Production. Elsevier Science, London.
15. **Lee, Y. C., H. C. Chien, and W. H. Hsu.** 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. J Biotechnol **129:**453-60.
16. **Linton, D., A. V. Karlyshev, P. G. Hitchen, H. R. Morris, A. Dell, N. A. Gregson, and B. W. Wren.** 2000. Multiple N-acetyl neuraminic acid synthetase (*neuB*) genes in *Campylobacter jejuni* identification and characterization of the gene involved in sialylation of lipo-oligosaccharide. Mol Microbiol **35:**1120-34.
17. **Mach, R. L., M. Schindler, and C. P. Kubicek.** 1994. Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expression signals. Curr Genet **25:**567-70.
18. **Magano, J.** 2009. Synthetic approaches to the neuraminidase inhibitors zanamivir (Relenza) and oseltamivir phosphate (Tamiflu) for the treatment of influenza. Chem Rev **109:**4398-438.
19. **Mahmoudian, M., D. Noble, C. S. Drake, R. F. Middleton, D. S. Montgomery, J. E. Piercey, D. Ramlakhan, M. Todd, and M. J. Dawson.** 1997. An efficient process for production of N-acetylneuraminic acid using N-acetylneuraminic acid aldolase. Enzyme Microb Technol **20:**393-400.
20. **Mandels, M.** 1985. Applications of cellulases. Biochem Soc Trans **13:**414-6.
21. **Maru, I., J. Ohnishi, Y. Ohta, and Y. Tsukada.** 1998. Simple and large-scale production of N-acetylneuraminic acid from N-acetyl-D-glucosamine and pyruvate using N-acyl-D-glucosamine 2-epimerase and N-acetylneuraminate lyase. Carbohydr Res **306:**575-8.
22. **Maru, I., J. Ohnishi, Y. Ohta, and Y. Tsukada.** 2002. Why is sialic acid attracting interest now? Complete enzymatic synthesis of sialic acid with N-acylglucosamine 2-epimerase. J Biosci Bioeng **93:**258-65.
23. **Monto, A. S.** 2003. The role of antivirals in the control of influenza. Vaccine **21:**1796-800.
24. **Noé P., C. J., Mora F., Comtat J.** 1986. Action of enzymes in chemical pulp and fibers. PartII: Enzymatic beating. J. Wood Chem. Technol. **6:**167-184.
25. **Pedersen, G. P., Screws, G.A., Cereoni, D.A.** 1992. Biopolishing of cellulosic fabrics. Can. Text. J. **Dec:**31-35.
26. **Pfaffl, M. W., G. W. Horgan, and L. Dempfle.** 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res **30:**e36.
27. **Roberts, W. K., and C. P. Selitrennikoff.** 1988. Plant and Bacterial Chitinases Differ in Antifungal Activity. J Gen Microbiol **134:**169-176.
28. **Schauer, R., S. Kelm, G. Reuter, R. P., and L. Shaw.** 1995. Biochemistry and Role of Sialic Acids, p. 7-67. *In* A. Rosenberg (ed.), Biology of the Sialic Acids. Plenum Press, NY and London.
29. **Seidl, V., B. Huemer, B. Seiboth, and C. P. Kubicek.** 2005. A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. FEBS J **272:**5923-39.
30. **Steiger, M. G., R. L. Mach, and A. R. Mach-Aigner.** 2010. An accurate normalization strategy for RT-qPCR in Hypocrea jecorina (Trichoderma reesei). J Biotechnol **145:**30-7.
31. **Sundaram, A. K., L. Pitts, K. Muhammad, J. Wu, M. Betenbaugh, R. W. Woodard, and W. F. Vann.** 2004. Characterization of N-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. Biochem J **383:**83-9.
32. **Tremblay, J. F.** 2006. The other drug for avian flu. C&EN **84:**33-36.
33. **Vann, W. F., J. J. Tavarez, J. Crowley, E. Vimr, and R. P. Silver.** 1997. Purification and characterization of the Escherichia coli K1 neuB gene product N-acetylneuraminic acid synthetase. Glycobiology **7:**697-701.
34. **Varki, A.** 1997. Sialic acids as ligands in recognition phenomena. Faseb J **11:**248-55.
35. **Walsh, G. A., R. F. Power, and D. R. Headon.** 1993. Enzymes in the animal-feed industry. Trends Biotechnol **11:**424-30.
36. **Wang, T. H., Y. Y. Chen, H. H. Pan, F. P. Wang, C. H. Cheng, and W. C. Lee.** 2009. Production of N-acetyl-D-neuraminic acid using two sequential enzymes overexpressed as double-tagged fusion proteins. BMC Biotechnol **9:**63.
37. **Warwas, M. L., J. N. Watson, A. J. Bennet, and M. M. Moore.** 2007. Structure and role of sialic acids on the surface of Aspergillus fumigatus conidiospores. Glycobiology **17:**401-10.
38. **Welt, T., Dinus, R.J.** 1995. Enzymatic deinking - a review. Progress in Paper Recycling **4:**36-47.
39. **Yamada-Okabe, T., Y. Sakamori, T. Mio, and H. Yamada-Okabe.** 2001. Identification and characterization of the genes for N-acetylglucosamine kinase and N-acetylglucosamine-phosphate deacetylase in the pathogenic fungus Candida albicans. Eur J Biochem **268:**2498-505.
40. **Yamada-Okabe, T., and H. Yamada-Okabe.** 2002. Characterization of the CaNAG3, CaNAG4, and CaNAG6 genes of the pathogenic fungus Candida albicans: possible involvement of these genes in the susceptibilities of cytotoxic agents. FEMS Microbiol Lett **212:**15-21.
41. **Zhang, Y., F. Tao, M. Du, C. Ma, J. Qiu, L. Gu, X. He, and P. Xu.** 2010. An efficient method for N-acetyl-D-neuraminic acid production using coupled bacterial cells with a safe temperature-induced system. Appl Microbiol Biotechnol **86:**481-9.

Supplementary data 1

Southern blot of *H. jecorina* strains QM9414, PEC/PSC1, PEC/PSC10.

Genomic DNA was digested with *Sal*I and separated on a 0.9% agarose gel. (a) Detection of *tbage* was achieved with a DIG-labeled probe released from plasmid pMS-PEC by a *Sac*I digestion. The probe has a size of 1457 bp and contains a part of the *cbh2* terminator. (b) Detection of *tneub* was achieved with a DIG-labeled probe released from plasmid pMS-PSC by a *BssH*II digestion. The probe has a size of 1298 bp and contains a part of the *pki* promotor.



Supplementary data 2

Coding sequences of the synthetic genes *tbage* and *tneub*.

The sequences are given in FASTA format. The *Xba*I-*Nsi*I fragment introduced into pRLM$_{ex}$30 resulting in the plasmids pMS-PEC and pMS-PSC is shown. *Xba*I-site is highlighted by underlined letters and *Nsi*I-site by double-underlined letters. The start codon ATG and the stop codon TAA are given in bold letters.

>*tbage*

TCTAGA**ATG**GGCAAGAACCTCCAGGCCCTGGCCCAGCTCTACAAGAACGCCCTCCTCAACGACGTC
CTGCCCTTCTGGGAGAACCACAGCCTCGACAGCGAGGGCGGCTACTTCACCTGCCTCGACCGCCAG
GGCAAGGTCTACGACACCGACAAGTTCATCTGGCTCCAGAACCGCCAGGTCTGGACCTTCAGCATG
CTCTGCAACCAGCTGGAGAAGCGCGAGAACTGGCTCAAGATCGCCCGCAACGGCGCCAAGTTCCT
CGCCCAGCACGGCCGCGACGACGAGGGCAACTGGTACTTTGCCCTGACCCGCGGCGGCGAGCCTCT
GGTCCAGCCCTACAACATCTTCAGCGACTGCTTCGCCGCCATGGCCTTCAGCCAGTACGCCCTCGCC
AGCGGCGAGGAGTGGGCCAAGGACGTCGCCATGCAGGCCTACAACAACGTCCTCCGCCGCAAGGA
CAACCCCAAGGGCAAGTACACCAAGACCTACCCCGGCACCCGCCCCATGAAGGCCCTGGCTGTCCC
CATGATCCTCGCCAACCTCACCCTGGAGATGGAGTGGCTCCTCCCCCAGGAGACCCTGGAGAACGT
CCTCGCCGCCACCGTCCAGGAGGTCATGGGCGACTTCCTCGACCAGGAGCAGGGCCTCATGTACGA
GAACGTCGCCCCCGACGGCAGCCACATCGACTGCTTCGAGGGCCGCCTCATCAACCCCGGCCACGG
CATCGAGGCCATGTGGTTCATCATGGACATCGCCCGCCGCAAGAACGACAGCAAGACCATCAACC
AGGCCGTCGACGTCGTCCTCAACATCCTCAACTTCGCCTGGGACAACGAGTACGGCGGCCTCTACT
ACTTCATGGACGCCGCCGGCCACCCCCCCCCAGCAGCTGGAGTGGGACCAGAAGCTCTGGTGGGTCC
ACCTGGAGAGCCTCGTCGCCCTCGCCATGGGCTACCGCCTCACCGGCCGCGACGCCTGCTGGGCCT
GGTATCAGAAGATGCACGACTACAGCTGGCAGCACTTCGCCGACCCTGAGTACGGCGAGTGGTTCG
GCTACCTCAACCGCCGAGGCGAGGTCCTCCTCAACCTCAAGGGCGGCAAGTGGAAGGGCTGCTTCC
ACGTCCCCCGCGCCATGTACCTCTGCTGGCAGCAGTTCGAGGCCCTCAGC**TAA**ATGCAT

>*tneub*

TCTAGA**ATG**CAGATCAAGATCGACAAGCTCACCATCAGCCAGAAGAACCCCCTCATCATCCCCGA
GATCGGCATCAACCACAACGGCAGCCTGGAGATCGCCAAGCTCATGGTCGACGCCGCCAAGCGAG
CCGGCGCCAAGATCATCAAGCACCAGACCCACATCGTCGAGGACGAGATGAGCCAGGAGGCCAAG
AACGTCATCCCCGGCAACGCCAACATCAGCATCTACGAGATCATGGAGCAGTGCGCCCTCAACTAC
AAGGACGAGCTGGCCCTCAAGGAGTACGTCGAGAAGCAGGGCCTCGTCTACCTCAGCACCCCCTTC
AGCCGCGCCGCCGCCAACCGCCTGGAGGACATGGGCGTCAGCGCCTACAAGATCGGCAGCGGCGA
GTGCAACAACTACCCCCTGATCAAGCACATCGCCCAGTTCAAGAAGCCCATGATCATCAGCACCGG
CATGAACAGCATCGAGAGCATCAAGCCCACCGTCAAGATCCTCCGCGACTACGAGATCCCCTTCGT
CCTCCTGCACACCACCAACCTCTACCCCACCCCCAGCCACCTCGTCCGCCTCCAGGCCATGCTGGA
GCTGTACAAGGAGTTCAACTGCCTCTACGGCCTCAGCGACCACACGACGAACAACCTCGCCTGCAT
CGGCGCCATCGCCCTCGGCGCCAGCGTCCTGGAGCGCCACTTCACCGACACCATGGACCGCAAGGG
CCCCGACATCGTCTGCAGCATGGACGAGAGCACCCTCAAGGACCTCATCAACCAGACCCAGGAGA
TGGTCCTCCTCCGCGGCGACAACAACAAGAACCCCCTGAAGGAGGAGCAGGTCACCATCGACTTC
GCCTTCGCCAGCGTCGTCAGCATCAAGGACATCAAGAAGGGCGAGATCCTCAGCATGGACAACAT
CTGGGTCAAGCGCCCCAGCAAGGGCGGCATCAGCGCCAAGGACTTCGAGGCCATCCTCGGCAAGC
GCGCCAAGAAGGACATCAAGAACAACATCCAGCTCACCTGGGACGACTTCGAG**TAA**ATGCAT

A TRANSFORMATION SYSTEM FOR *HYPOCREA* (*TRICHODERMA*) THAT FAVOURS HOMOLOGOUS INTEGRATION AND THAT USES REUSABLE BI-DIRECTIONALLY SELECTABLE MARKERS

MATTHIAS G. STEIGER[1], MARIKA VITIKAINEN[2], KURT BRUNNER[1], GERHARD ADAM[3], MARKKU SALOHEIMO[2], ROBERT L. MACH[1] AND ASTRID R. MACH-AIGNER[1]

A TRANSFORMATION SYSTEM FOR *HYPOCREA* (*TRICHODERMA*) THAT FAVOURS HOMOLOGOUS INTEGRATION AND THAT USES REUSABLE BI-DIRECTIONALLY SELECTABLE MARKERS

MATTHIAS G. STEIGER[1], MARIKA VITIKAINEN[2], KURT BRUNNER[1], GERHARD ADAM[3], MARKKU SALOHEIMO[2], ROBERT L. MACH[1] AND ASTRID R. MACH-AIGNER[1]

[1] Gene Technology, Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, TU Wien, Getreidemarkt 9/166/5/2, A-1060 Wien, Austria
[2] VTT Technical Research Centre, Biotechnology, FIN-02044 VTT, Espoo, Finland
[3] Department für Angewandte Genetik und Zellbiologie, BOKU Wien, Muthgasse 18, A-1190 Wien, Austria

ABSTRACT

Hypocrea jecorina **is an industrially important filamentous fungus due to its effective production of hydrolytic enzymes. Recently, it has received increasing interest because of its ability to convert lignocellulosic biomass to monomeric sugars, which can be converted to biofuels or platform chemicals. Genetic engineering of strains is a highly important means of meeting the requirements of tailor-made applications. Therefore, we report the development of a transformation system that combines highly efficient gene targeting (by using a *tmus53* (human *LIG4*-homolog) deletion strain), marker recycling (with a Cre/loxP-based excision system), and bi-directional positive selection (by combining resistance to hygromycin B and loss of sensitivity to fluoroacetamide). Furthermore, the bi-directional marker *pyr4* is exploited to remove the *cre* gene again. Implementation of components of this system in *H. atroviridis* and *Fusarium* has highlighted it as a universal tool for filamentous fungi.**

1. INTRODUCTION

Hypocrea jecorina (anamorph: *Trichoderma reesei* (22)) is a saprophyte noted for its ability to abundantly secrete native hydrolytic enzymes. These enzymes are used in various industrial applications, such as pulp and paper production (2, 29, 40), the food and feed industries (8, 24, 39), and the textile industry (19, 23, 31). Recently, these enzymes have garnered attention because of their ability to turn lignocellulosic biomass into biofuel (13, 15, 34). Polymeric biomass compounds can be degraded into readily fermentable sugar monomers such as glucose and xylose. Strain improvement is an important focus of research because it can increase productivity and broaden the operating range. Optimal strain design by genetic engineering requires a potent transformation system, and such a system must meet certain requirements: i) a high rate of homologous integration (gene replacement) events; ii) a recyclable marker for unrestricted rounds of knock-outs; and iii) a bi-directional positive selection system.

Until now, no transformation system existed that fulfilled all the criteria mentioned above. Nevertheless, in the past, efforts have been made to meet some of these requirements. In *Neurospora*, *mus-51* and *mus-52* (homologous to human *KU70* and *KU80*) have been deleted to block nonhomologous end-joining (NHEJ), and this led to 100% of transformants exhibiting integration at the homologous site (Ninomiya et al 2004). In other filamentous fungi, highly efficient gene targeting has been achieved by following the same strategy (reviewed by (21)). More recently, MUS-53 (a homolog of human Lig4) was shown to be required for NHEJ in *Neurospora*. The authors demonstrated that in contrast to *mus-51* and *mus-52* mutants, a *mus-53* mutant had a gene targeting efficiency of 100% even if the homologous DNA sequence was very short. Therefore, *mus-53* was proposed as a highly efficient alternative to *mus-51* and *mus-52* (17).

In addition to gene targeting, an efficient marker recycling system is also very important for a powerful transformation system. One proposed strategy is the use of the Cre/loxP recombination system adapted from bacteriophage P1 (1, 37). This strategy involves the excision of DNA fragments flanked by loxP-sites catalysed by the expression of Cre recombinase. Although the use of the Cre/loxP system in a eukaryote (*Saccharomyces cerevisiae*) was described for the first time in 1987 (35), it was not used as a gene disruption tool in yeast until 1996 (12). Since then, the Cre/loxP system has also been used in filamentous fungi such as *Candida albicans*, *Aspergillus sp.*, *Epichloë festucae*, *Neotyphodium* sp., and *Cryptococcus neoformans* (5-7, 20, 30).

In this study, we report the deletion of the *Hypocrea LIG4* homolog *tmus53* to create a NHEJ-deficient *Hypocrea* strain. Furthermore, we developed a marker recycling system for *Hypocrea* based on the Cre/loxP system. The genomically integrated Cre recombinase can be induced in a controlled way, which makes the excision of the marker regulatable. An additional advantage of the marker system is that it offers the possibility of bi-

directional positive selection (resistance to hygromycin B (marker insertion) or loss of sensitivity to fluoroacetamide (marker removal)). Moreover, we constructed expression cassettes using a donor and an acceptor vector in conjunction with a Cre recombinase assay, and this allows a completely restriction enzyme-free cloning procedure. Altogether, this transformation system offers the possibility of serial targeted genome manipulation with absolute marker recycling in *Hypocrea* as well as in other filamentous fungi.

2. MATERIALS AND METHODS

2.1 Strains and cultivation conditions

Hypocrea jecorina QM6a (ATCC 13631) was used as a parental strain throughout this study and maintained on malt extract agar. The recombinant strains QM6aΔ*tmus53*(P^+), QM6aΔ*tmus53*(H^+,A^+,loxP), QM6aΔ*tmus53*Δ*pyr4*(loxP), and QM6aΔ*tmus53*(loxP) were maintained on malt extract agar, which was supplemented with 5 mM uridine in the case of *pyr4* deletion strains. QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP) and QM6aΔ*tmus53*Δ*pyr4env1*$^{PAS-}$(H^+) were maintained on Mandels-Andreotti (MA) medium (28) containing 1% glucose and 5 mM uridine. An overview of all strains used in this study and their cultivation/selection media is given in Table 1.

2.2 PCR

Unless noted otherwise, 50 µL PCRs contained 1.25 U GoTaq® Flexi DNA Polymerase (Promega, Wisconsin, US), 2 mM $MgCl_2$, 1x Green GoTaq® Flexi Buffer, 0.2 mM dNTPs, 0.1 µM forward and reverse primer, 1 µl template DNA (10 ng of plasmid DNA, 300 ng of genomic DNA) and nuclease-free water. For PCR screening experiments, the PCR volume was scaled down to 15 µL. The standard PCR consisted of an initial denaturation of 2 min at 95 °C followed by 35 cycles of 30 sec at 95 °C, 30 sec at 60 °C and 1 min/kb at 73 °C. All PCRs were performed in a UNO II (Biometra GmbH, Göttingen, Germany). All primers used in this work are listed in Table 2.

2.3 Plasmid construction

To delete the *tmus53* gene, a deletion cassette (pRS426-dmus) was constructed in the pRS426 plasmid (3) by yeast recombination cloning (4). The deletion cassette contained the phosphinothricin resistance (*bar*) marker flanked by 1.5-kb fragments from up- and downstream of *tmus53*. The flanking region fragments were amplified by PCR with the primers 5forwardmus and 5reversemus or 3forwardmus and 3reversemus. The *bar* marker gene was amplified with the primers barforward and barreverse. The template for the flanking region fragments was QM6a genomic DNA and pTJK1 (18) for the *bar* gene.

For the construction of pMS-HALS, pUG6 (12) was digested with *Bgl*II and blunt ends were generated with Klenow polymerase. Subsequently, the linearised plasmid was cut with *Sac*I. The plasmid $pRLM_{ex30}$ (26) was digested with *Xho*I, blunt-ended by Klenow polymerase and then cut with *Sac*I. The 2.6-kb fragment containing the *hph* gene (hygromycin B phosphotransferase-encoding) under control of the *pki* promoter and the *cbh2* terminator thereby obtained was cloned in between two loxP sites of the pUG6 backbone resulting in the plasmid pHylox2. With the primers amdSXbaDAMf and amdSXbaXhor, a 2.7-kb fragment containing the *amdS* gene was amplified from the plasmid pamdS (33). The resulting PCR product was *Xba*I-digested and ligated into the pHylox2 plasmid, which was opened by partial digestion using *Xba*I. The obtained plasmids were screened for the *amdS* integration upstream of the *pki* promoter and named pMS-HAL. With the primers sacBSalfw and sacBSalrev, the *sacB* gene (encoding the *Bacillus subtilis* levansucrase) was amplified using the plasmid pDNR-Dual (Clontech, Mountain View, CA) as a template, thereby creating *Sal*I restriction sites flanking the amplicon. The obtained PCR product was inserted into the plasmid pMS-HAL via the *Sal*I restriction site, resulting in the final plasmid pMS-HALS.

For the construction of the plasmid pMS-5loxP3-mus, a splicing by overlapping extension (SOE) PCR approach was used to connect the 5´flanking region and the 3´flanking region of the target gene. With a primer, a loxP site was inserted in between the two flanking regions. In the first PCR, the primers 5flmusfw and 5flmusrev were used for the first fragment, and the primers 3flmusfw and 3flmusrev were used for the second fragment. Genomic DNA of QM6a was used as the template. In the second PCR, both fragments were combined using the primers 5flmusfw and 3flmusrev. The template for the second PCR consisted of 300 ng of fragment 1 and 2 from the first PCR. The resulting PCR product was ligated into the pGEM®-T Vector (Promega) using the standard protocol resulting in plasmid pMS-5loxP3-tmus.

Plasmid pMS-loxPΔtmus53 was obtained by Cre-mediated recombination of plasmid pMS-HALS and pMS-5loxP3-tmus. The Cre recombinase assay from the Creator™ DNA cloning kit (Clontech) was carried out according to the manufacturer's instructions. pMS–HALS was used as the donor vector and pMS-5loxP3-tmus as the acceptor vector. A 0.5-µL aliquot of the Cre recombinase assay was transformed into *E. coli* Supercharge EZ10 Electrocompetent Cells (Clontech) and selected on plates containing LB medium supplemented with 100 µg/mL hygromycin B, 100 µg/mL ampicillin, and 70 mg/mL sucrose. Of the screened colonies, 98% contained the correct pMS-loxPΔtmus53 plasmid (11.123 kb).

For the construction of plasmid pMS-creΔpyr4, a fragment of the *H. jecorina xyn1* promoter was

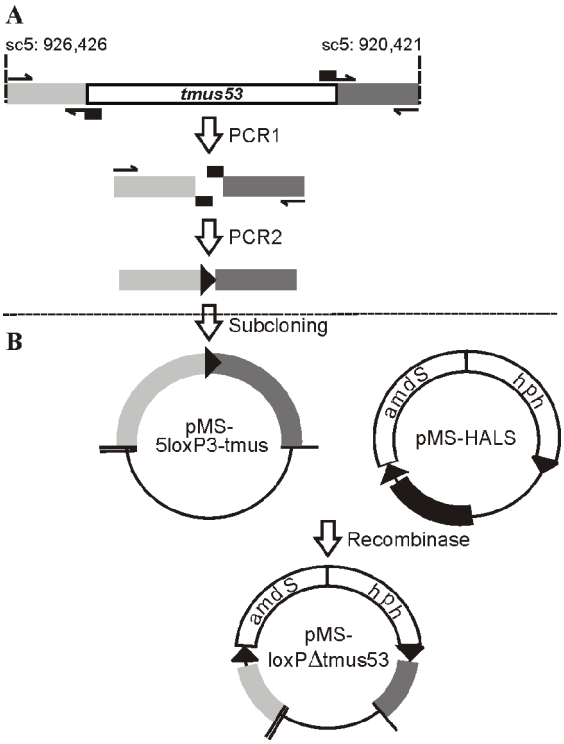


Figure 1 Construction of the vector bearing the loxP-sites and the bi-directional marker system. (A) Amplification of the *tmus53* 5'-flank (light grey) and 3'-flank (dark grey) introducing a loxP site (black triangle) in between. The first PCRs yielded 1.4- and 1.3-kb products of the 5' and 3'-flanks including a loxP site on their 3'-end or 5'-end, respectively (PCR1). A second PCR using the outer primers yielded an amplicon consisting of the 5'-flank of *tmus53*, a loxP site, and the 3'-flank of *tmus53* (PCR2). Primer sequences are given in Table 2. Thin black arrows indicate primers; thin black arrows with a black box indicate primers introducing a loxP site; sc means scaffold and indicates genomic position. (B) Subcloning and Cre recombinase reaction. The obtained PCR product was subcloned into pGEM®-T to obtain pMS-5loxP3-tmus, which was used as the acceptor vector in the subsequent Cre recombinase reaction. pMS-HALS, bearing two loxP sites flanking the *amdS* and *hph* genes as markers for fungal transformation and *sacB* as a killer gene (black segment), was used as the donor vector. The Cre-recombinase reaction yielded pMS-loxPΔtmus53, bearing the *tmus53* 5'-flank, a loxP site, *amdS*, *hph*, a loxP site, and the *tmus53* 3'-flank. Short black lines indicate *Not*I-sites; white arrows indicate enzymatic reactions.

amplified from genomic DNA by PCR with the primers pxyn1fw and pxyn1rev. The fragment was subcloned into the pGEM®-T vector (Promega). The Cre recombinase of bacteriophage P1 was amplified with the primers crefw and crerev and cloned into pGEM®-T (Promega). Both plasmids were digested with *Sac*I and *Xho*I and both fragments were ligated into alkaline phosphatase treated, *Sac*I-digested pGEM®-5Zf(+) (Promega) resulting in plasmid pKBxyn1-cre. With a SOE PCR strategy, the plasmid pMS-creΔpyr4 was constructed using the plasmid pKBxyn1-cre as the template. In the first PCR, the 5'- and 3'-flanking regions of the *pyr4* gene were amplified. For the 5'-flanking region, primers 5'FLANKdpyrfw and 5'Flankdpyrrev were used, and primers 3'FLANKdpyrfw and 3'FLANKdpyrrev were used for the respective 3'-flanking region. In both cases,

Table 1 Overview on the phenotype of H. jecorina mutant strains

strain/genotype	Media containing [X indicates growth]						
	Phosphinothricin	Hygromycine B	5-Fluorotic acid	Fluoroacetamide	Acetamide (minimal medium)	medium without uridine	genotype stable only on glucose
QM6a (ATCC 13631)				X		X	
QM6aΔ*tmus53*(P^+)	X			X		X	
QM6aΔ*tmus53*(H^+,A^+,loxP)		X			X	X	
QM6aΔ*tmus53*Δ*pyr4*(H^+, A^+,loxP)		X	X		X		X
QM6aΔ*tmus53*Δ*pyr4*(loxP)			X	X			X
QM6aΔ*tmus53*Δ*pyr4env1*$^{PAS-}$(H^+)		X					
QM6aΔ*tmus53*(loxP)				X		X	

genomic DNA of QM6a was used as the template. The 5'-flanking region is 1.5 kb and the 3'-flanking region is 1.4 kb. In the second PCR, those flanking regions were attached to the *pxyn1::cre* construct. Thereafter, pKBxyn1-cre was digested with *Pst*I and *Mlu*I, and the resulting 1.601-kb *Pst*I/*Mlu*I-fragment was purified via gel electrophoresis. In the second PCR, the amplified 5'-flanking region and the *Pst*I/*Mlu*I-fragment were used as templates together with the primers 5'FLANKdpyrfw and pxyn5'Flankdpyrrev to generate a 1.977-kb fragment. This was subcloned into the pGEM®-T vector (Promega) resulting in the plasmid pWD1. A further PCR fused the second part of the *pxyn1::cre* cassette to the *pyr4* 3'-flanking region, using the 3'-flanking region and the *Pst*I/*Mlu*I-fragment as templates and the primers cre3´FLANKdpyrfw and 3´FLANKdpyrrev in a PCR. This PCR product was digested with *Eco*RI and *Mlu*I and ligated into the plasmid pWD1, which was previously cut with *Eco*RI and *Mlu*I. The resulting plasmid was named pMS-creΔpyr4.

All the primers used in plasmid construction are listed in Table 2.

2.4 Protoplast transformation of Hypocrea

The protoplast transformation of *Hypocrea* was carried out as described earlier ((32) for the QM6aΔ*tmus53*(P^+) transformation and (10) for all other transformations). The amount of DNA used was normally in the range of 1 to 5 µg. The selection medium used depended on the transformation system. QM6a was used as the recipient strain for the generation of QM6aΔ*tmus53*(P^+). In the case of QM6aΔ*tmus53*(P^+), the transformation medium described in (32) was supplemented with 200 µg/ml

Table 2 Primer used during this study

Name	Sequence (5´→3´)	Usage
amdSXbaDAMf	ATTTCTAGATCTACGCCAGGACCG	construction pMS-HAL
amdSxbaXhor	ATTATCTAGACTCGAGCTGGAAACGCAACCCTGAAGG	construction pMS-HAL
SacBSal fw	ATAGTCGACGACGTCCACATATACCTGCCGTT	construction pMS-HALS
SacBSal rev	TATGTCGACGACGTCAATGCCAATAGGATATCGG	construction pMS-HALS
5forwardmus	GTAACGCCAGGGTTTTCCCAGTCACGACGAAGCTTCCGTATTTC AGCAGTAACCC	construction pRS426-dmus
5reversemus	ATCCACTTAACGTTACTGAAATCTCCAACTCGCGCTAGACTGAA GATTG	construction pRS426-dmus
3forwardmus	CTCCTTCAATATCATCTTCTGTCTCCGACTTTGCCTGAGAATGAG GAGG	construction pRS426-dmus
3reversemus	GCGGATAACAATTTCACACAGGAAACAGCAAGCTTGCCTTCCA TACTGATTTGCC	construction pRS426-dmus
barforward	GTCGGAGACAGAAGATGATATTGAAGGAGC	construction pRS426-dmus
barreverse	GATTTCAGTAACGTTAAGTGGATCCCGGTGACGGATCAGATCTC GGTGACG	construction pRS426-dmus
5flmusfw	TATGCGGCCGCAAGCTTCCGTATTTCAGCAGTAACC	construction pMS-5loxP3-tmus
5flmusrev	ATAACTTCGTATAGCATACATTATACGAAGTTATAAAGGACCTT GACAGAACGGAGTA	construction pMS-5loxP3-tmus
3flmusfw	ATAACTTCGTATAATGTATGCTATACGAAGTTATCAAGGAGCAT ATGCAAGGGTATCTC	construction pMS-5loxP3-tmus
3flmusrev	TATGCGGCCGCTCATTGTCATGACGCTACAGAAGCT	construction pMS-5loxP3-tmus
pxyn1fw	GCAAATGAGCTCAAGCAACTACG	construction pKBxyn1-cre
pxyn1rev	CCTCGAGGATGTTATTTGTGCGTGTTTTCCTTG	construction pKBxyn1-cre
crefw	CCTCGAGATGTCCAATTTACTGACCGTACACC	construction pKBxyn1-cre
crerev	CCTGAGCTCCTAATCGCCATCTTCCAGCAG	construction pKBxyn1-cre
5´FLANKdpyrfw	TATAGCGGCCGCGTTCTCCAAGGCGTCAAGCAT	construction pMS-creΔpyr4
5`FLANKdpyrrev	GGTCGACCTGCAGGCGGCCGTGAATGATACACACAAGTCTGCC AGAT	construction pMS-creΔpyr4
3´FLANKdpyrfw	GATGGCGATTAGGAGCTCCCAATTCGAGGACCGCAAGTTTGG	construction pMS-creΔpyr4
3´FLANKdpyrrev	ATATACGCGTGCTCTGAAGATGTTACGCTGCATAC	construction pMS-creΔpyr4
pxyn5´Flankdpyrrev	GCAGAATTCTTCCACATTCAAAACCTCTT	construction pMS-creΔpyr4
cre3´FLANKdpyrfw	GAAGAATTCTGCATATATAAAGCCATGGAAGAAG	construction pMS-creΔpyr4
5amdsrev	TTCAGGATGCTCTCCAGCGGG	control PCR loxP cassette integration/sequencing
Cbh2term fw	TAGATCAAGCTGGTAGATTCCAATTACTCC	control PCR loxP cassette/integration
3cre fw	ATGCAAGCTGGTGGCTGGACC	control PCR *cre* integration/sequencing
3 loc pyr fw	GTAGGTAGGTTAGGTAGGTAGTTTAGGTA	control PCR *pyr4* locus/sequencing
3loc pyr rev	TTGCTTCCTTCGATGGCACTCG	control PCR *pyr4* locus
fwxyncre	AAACTCCATGAGATTGCAGATGCG	control PCR *cre* integration
revxyncre	CTAATCGCCATCTTCCAGCAGG	control PCR *cre* integration/sequencing
5loc pyr fw	GTAGGTAGGTTAGGTAGGTAGTTTAGGTA	control PCR *pyr4* locus sequencing
5-loc mus fw	AATGGTCCTGCTTTGAGATGCTGG	control PCR *tmus53* locus sequencing
3 loc mus rev	TGTGATTCAGGCGATCGGTGC	control PCR *tmus53* locus sequencing
In 5 Fl mus fw	ACTAATCCCAAAAAGGTGGCGTGC	control PCR *tmus53* locus sequencing
In 3Fl mus rev	GTTTGTGAGGGTCATACAGAGGAAGC	control PCR *tmus53* locus sequencing
In 5Fl cre fw	CTTATCCCCATCTTTTCTCTCTCTCTGC	control PCR *pyr4* locus sequencing
In 3Fl cre rev	TTGACAATGTGCGCCCAGTCG	control PCR *pyr4* locus
5loc env fw	GCTGGACTTTGCCGCTAACACG	control PCR *env1* locus sequencing
3loc env rev	AAATCAAGCGGCTGCACGAGG	control PCR *env1* locus
phospho fw	GTGACGGATCAGATCTCGGTGACG	probe southern *bar*
phospho rev	TTTAGTCGTCCAGGCGGTGAGC	probe southern *bar*
hph(delta)f	GCCTCACCTGCTGATTTCTC	probe southern *hph.*
hph(delta)r	GATGTTGGCGACCTCGTATT	probe southern *hph*
amdsfw	CCGGGATCAATGAGGAGAATGAGG	probe southern *amdS*
amdsrev	ACGAATCCCAACGATCGCACC	probe southern *amdS*

phosphinothricin prepared from Basta herbicide. Phosphinothricin concentrations up to 800 µg/ml were tested on transformation medium plates. The DNA used for the transformation was the deletion cassette from plasmid pRS426-dmus. QM6aΔ*tmus53*(P^+) served as a recipient strain to obtain QM6aΔ*tmus53*(H^+,A^+,loxP) and QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP). To generate strain QM6aΔ*tmus53* (H^+,A^+,loxP), selection was carried out on selection medium for *amdS* as previously described (32), and this was additionally supplemented with 100 µg/mL hygromycin B. The 8.114-kb *Not*I-fragment from pMS-loxPΔtmus53 was used for the transformation. In the case of QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP), the same medium as for QM6aΔ*tmus53*(H^+,A^+,loxP) was used and was additionally supplemented with 5 mM uridine. Again, QM6aΔ*tmus53*(P^+) served as the recipient strain. The DNA used for this transformation consisted of an 8.114-kb *Not*I-fragment from pMS-loxPΔtmus53 and a 4.516-kb *Not*I/*Mlu*I-fragment from pMS-creΔpyr4. The fragments from pMS-loxPΔtmus53 and pMS-creΔpyr4 were used at a ratio of 1:6.

QM6aΔ*tmus53*Δ*pyr4*(loxP) was used as the recipient strain for generating QM6aΔ*tmus53*Δ*pyr4env1*$^{PAS-}$(H^+). This strain was selected on the same medium as QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP). The DNA used was a linear *Acc65*I/*BamH*I-fragment from vector pDELENV (36) containing the *hph* gene under control of the *pki* promoter.

2.5 Excision of the loxP-marker cassette

To excise the loxP cassette containing the *amdS* and *hph* genes, the strain QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP) was cultivated on an MA plate containing 1% (w/v) oat spelt xylan (Sigma Aldrich, St. Louise, MO) as the carbon source and 5 mM uridine. Conidia were harvested after four days and plated on an MA agar plate containing 0.5 µL/ml Igepal CA-360 (Sigma Aldrich), 1.5 mg/mL fluoroacetamide, 5 mM uridine, and 1 % (w/v) D-xylose. The phenotype of the fungal isolates was checked by plating them on a malt extract agar plate containing 100 µg/mL hygromycin B and 5 mM uridine. No growth indicated a successful excision of the loxP-marker cassette. If the fungus still had the ability to grow on hygromycin B-containing media, the single-spore purification was repeated.

2.6 DNA analysis by Southern blot

Fungal genomic DNA was isolated as described previously (9). Southern hybridisation and detection was carried out with the DIG High Prime DNA Labeling and Detection Starter Kit II using the standard procedure (Roche, Switzerland). To integrate the *bar* gene at the *tmus53* locus, the genomic DNA of QM6a and QM6aΔ*tmus53*(P^+) was digested with *Sal*I. The probe for the *bar* gene was amplified with the primers phosphofw and phosphorev using pRS426-dmus as the template. To test the homologous integration of *cre* at the *pyr4* locus, genomic DNA from QM6a, QM6aΔ*tmus53*(P^+), QM6aΔ*tmus53*(loxP), QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP), and QM6aΔ*tmus53*Δ*pyr4*(loxP) was digested with *Hind*III and *Nar*I. A 1.409-kb *Nsi*I-fragment from pKBxyn1-cre was used as a specific probe for *cre* and the *xyn1* promoter. The integration of the loxP-marker cassette into the *tmus53* locus was checked with an *amdS*-specific probe amplified with the primers amdsfw and amdsrev from pMS-HALS. The same blot used for *cre* (digested with *Hind*III/*Nar*I) was used for this Southern.

3. RESULTS

3.1 Construction of deletion cassettes by an acceptor/donor cloning technique

For efficient knock-out and gene replacement events in *Hypocrea*, it is necessary to produce cassettes with long (~ 500 bp – 1500 bp) 5' and 3' regions flanking the desired integration locus (11). Because of the considerable length of the flanking regions, it is difficult to find suitable restriction enzymes for the subcloning procedure. In this study, we modified the commercially available Creator™ DNA Cloning system (Clontech) to create an efficient system for the generation of deletion cassettes. As an example, the strategy used to create the deletion cassette for *tmus53* is pictured in Fig. 1. In the first step (Fig. 1A), a fragment containing the 5'- and 3'-flanking region was obtained via SOE PCR. With the inner primer pair, a loxP-site was introduced between the 5'- and 3'-flanking regions. This amplicon was subcloned into a vector system for PCR products such as the pGEM®-T system. The next step (Fig. 1B) was an *in vitro* Cre recombinase reaction, in which the acceptor vector recombined at the loxP-site with the donor vector. The recombination reaction was transformed into *E. coli* and the transformants were selected on hygromycin B and sucrose, because the loxP-marker cassette contains the *hph* gene and the donor vector carries the killer gene *sacB*, which is activated on sucrose. It is noteworthy that the *pki::hph*-construct is fully functional in *E. coli*, allowing the use of such vectors as shuttle systems between *E. coli* and filamentous fungi.
With this system, gene replacement cassettes can be obtained without using restriction enzymes. The donor vector, pMS-HALS, can be universally used for the construction of any other knock-out cassette.

3.2 Deletion of the tmus53 gene and generating an NHEJ-deficient Hypocrea strain

A BLAST search of the genome sequence of *H. jecorina* QM6a available at DOE Joint Genome Institute (http://genome.jgi-psf.org/Trire2/Trire2.home.html) identified a gene (protein id 58509) similar to human DNA ligase IV, and this gene was named *tmus53*. To delete *tmus53*, a deletion cassette containing the phosphinothricin resistance (*bar*) marker was used (Fig. 2A). This marker was used for the first time to screen *Hypocrea* transformants. The correct transformation of both loci was confirmed by means of Southern blot (data not shown) and PCR. The positions of the primers for the control PCR are given in Fig. 3 (C, D). The corresponding PCR amplicons were sequenced, and the results are given in Supplemental Material (see Sup. 1D). The QM6aΔ*tmus53*(P^+) strain showed no alteration in growth rate, conidiation or cellulase activity compared to the parental strain QM6a (data not shown).
In order to study the effect of the *tmus53* deletion on the efficiency of homologous recombination in *Hypocrea*, we used QM6aΔ*tmus53*(P^+) as a host strain to delete several genes. Table 3 lists the frequencies of homologous integration events obtained using nine deletion constructs. The deletion cassettes contained the *hph* gene as a marker flanked by 1 kb-fragments up- and

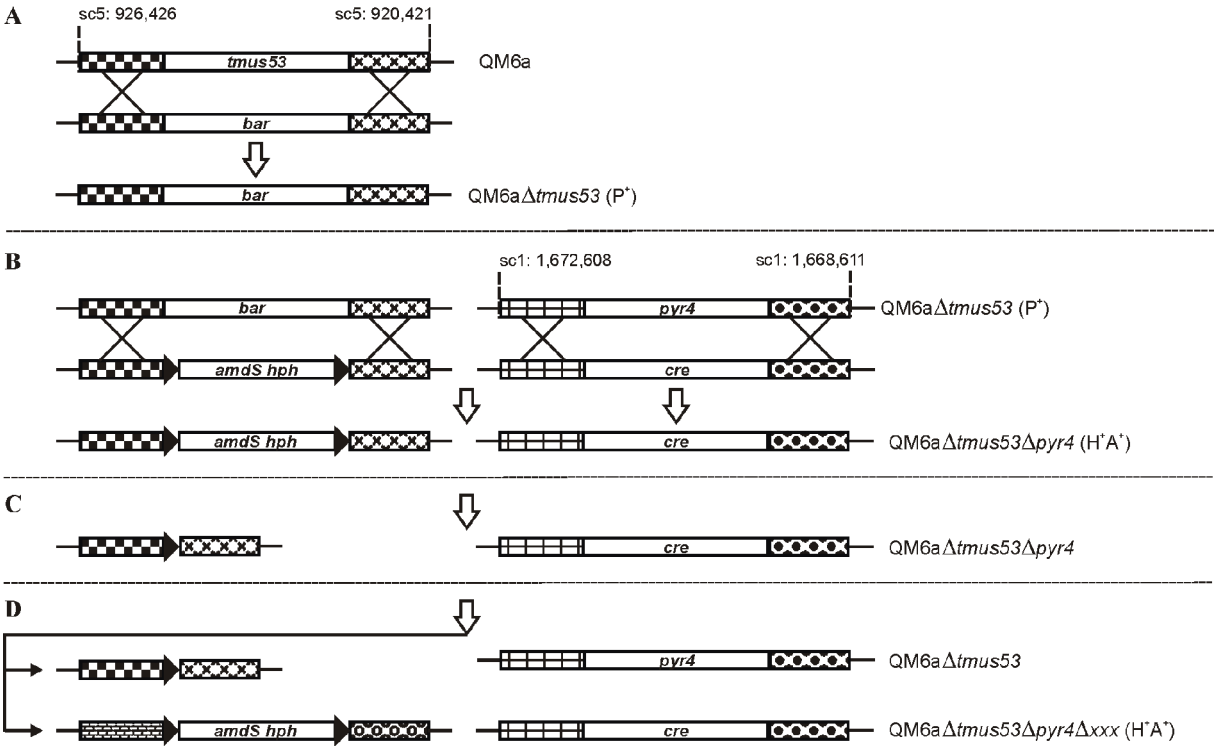


Figure 2 Schematic drawing of strain construction and marker reuse. (A) Construction of the *H. jecorina tmus53* deletion strain. The *tmus53* gene was replaced by the *bar* gene conferring phosphinothricin resistance (P^+) to yield QM6aΔ*tmus53.* The boxes containing squares or crosses indicate 5'-flanks or 3'-flanks of the *tmus53* gene, respectively; sc means scaffold and indicates genomic position; white arrows indicate a fungal transformation step. (B) Deletion of the *pyr4* gene and replacement of phosphinothricin resistance by a bi-directional marker system. In a co-transformation, the *bar* gene was replaced by the loxP-*amdS-hph*-cassette conferring hygromycin B resistance (H^+) and the ability to use acetamide as nitrogen source (A^+), and the *pyr4* gene was replaced by the *cre* gene encoding Cre-recombinase to yield QM6aΔ*tmus53*Δ*pyr4.* The boxes containing a grid pattern or points indicate the 5'-flank or the 3'-flank of the *pyr4* gene, respectively. (C) Marker excision. Growing the QM6aΔ*tmus53*Δ*pyr4* strains on xylan (1% w/v) for three days promoted excision of the marker genes (*amdS, hph*) via the flanking loxP-sites. (D) Two options for further manipulations: First, back-transformation of the *pyr4* gene replacing the *cre* gene to obtain a marker-free QM6aΔ*tmus53*; second, another round of any gene (*xxx*) deletion (in the present study the PAS domain of *env1*) to obtain QM6aΔ*tmus53*Δ*pyr4*Δ*xxx*, which is hygromycin B resistant (H^+) and able to use acetamide as a nitrogen source (A^+). The boxes containing a brick pattern or circles indicate 5'-flanks or 3'-flanks of the *env1* PAS domain, respectively; black arrows indicate the two options.

downstream of the intended deletion site. In general, the frequency of homologous integration was high (70 - 90%), and 100% efficiency was obtained in two transformations. However, in two cases the frequency was below 50%. We have not observed any correlation between the frequency of homologous integration and the size of the region to be replaced (Tab. 3).

Table 3 Frequencies of homologous integration events in the QM6aΔ*tmus53*(P^+) strain

Knock-out construct	Size of deleted DNA region (bp)	Homologous integration / No. of transformants screened	Homologous Integration rate (%)
KO1	934	8/9	88.9
KO2	1948	8/11	72.7
KO3	1075	4/15	26.7
KO4	1973	7/15	46.7
KO5	1716	15/15	100
KO6	5134	14/14	100
KO7	2227	12/13	92.3
KO8	2268	16/18	88.9
KO9	2202	12/13	92.3

3.3 Introducing a Cre/loxP-mediated marker recycling system into the tmus53 deletion strain

To use a QM6aΔ*tmus53*(P^+) strain efficiently, for instance for multiple sequential gene deletions, it is necessary to have enough markers available. Unfortunately, the markers available for *Hypocrea* are limited. Furthermore, the interference of markers with cellular processes points to the need for marker recycling. Therefore, we implemented a marker recycling system based on the Cre/loxP system. In a co-transformation, two loci of QM6aΔ*tmus53*(P^+) were changed (Fig. 2B). First, at the *tmus53*-site, a loxP-marker cassette was used to replace the *bar* gene, which confers phosphinothricin resistance. The cassette consists of two loxP-sites flanking the *amdS* and *hph* genes, the latter under control of the *Hypocrea pki* promoter. Both genes can be used as markers in *H. jecorina* (26, 32). Second, the *pyr4* gene was replaced with a cassette containing Cre recombinase under control of the *xyn1* promoter (Fig. 2B).

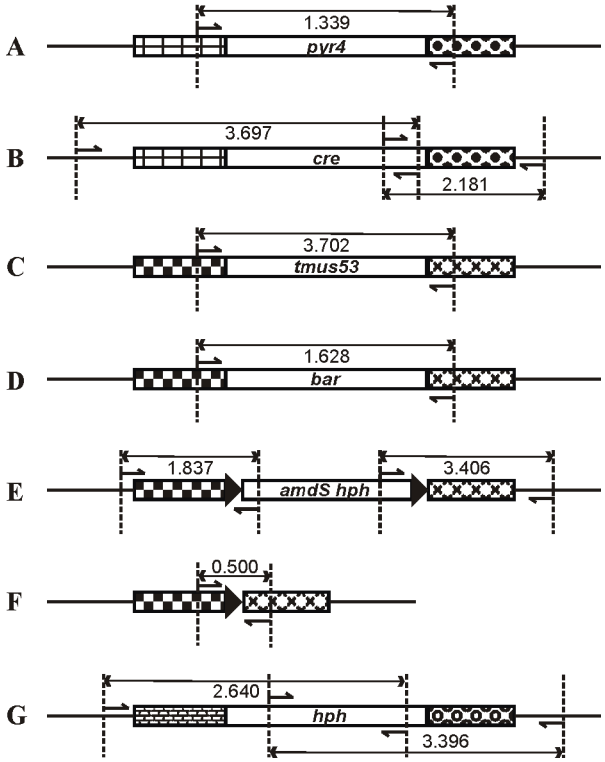


Figure 3 Schematic drawing of the primer positions for analytic PCR of the loci. Small black arrows indicate the locations of primers that were used for analyses of transformants and parental strains. The size of the expected amplicon is given between the dashed lines in kb. The boxes containing a grid pattern or points indicate the 5'-flank or the 3'-flank of the *pyr4* gene, respectively (A, B). The boxes containing squares or crosses indicate the 5'-flank or the 3'-flank of the *tmus53* gene, respectively (C – F). The boxes containing a brick pattern or circles indicate the 5'-flank or the 3'-flank of the *env1* PAS domain, respectively (G). The corresponding sequences are given in the Supplemental data (Sup. 1A – G).

The resulting strain QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP) has two functional deletions, *tmus53* and *pyr4*. The *pyr4* deletion causes uridine auxotrophy, which itself can be used as a transformation marker (9). Furthermore, this strain is able to express a Cre recombinase in a controlled way, because the driving *xyn1* promoter can be induced by D-xylose and xylan. On glucose, this promoter is shut off due to a double double-lock mechanism (25, 27). The correct transformation of both loci was confirmed by means of Southern blot (data not shown) and PCR. The positions of the primers for the control PCR are given in Fig. 3 (A, B, C, E). The corresponding PCR amplicons were sequenced, and the results are given in Supplemental Materials (see Sup. 1 A, B, C, E).

3.4 Marker recycling by induction of Cre recombinase expression

The loxP-marker cassette containing the marker genes *amdS* and *hph* can be looped out of the chromosomal DNA by the action of Cre recombinase. By cultivating the strain QM6aΔ*tmus53*Δ*pyr4*(H^+,A^+,loxP) on medium containing xylan, the *xyn1* promoter is induced, which leads to Cre recombinase expression. The resulting strain contains only one copy of the loxP-site at the *tmus53* locus, is sensitive to hygromycin B and is acetamidase-negative. Thus, screening for the excision event is performed in a fluoroacetamide background because fluoroacetamide has a slight toxic effect on strains containing the *amdS* gene and reduces their growth. This situation is depicted in Fig 2C. The strain obtained after the excision was named QM6aΔ*tmus53*Δ*pyr4*(loxP). The phenotype can be checked on plates containing hygromycin B. After two subsequent single-spore purification steps, approximately 70% of the transformed strains had successfully excised the loxP-marker cassette and did not show growth on hygromycin B-containing plates. For further verification, the excision was examined with PCR for the *tmus53* locus (Fig. 3E, F). The sequence results showed one remaining loxP-site at the locus (Sup. 1E, F). To test whether the excision was due to Cre recombinase activity, a QM6aΔ*tmus53*(H^+,A^+,loxP) strain containing no *cre* gene was used as a control in the excision experiment. In this case, it was not possible to obtain a strain that had lost its hygromycin B resistance. Therefore, we propose that *H. jecorina* has no native Cre recombinase activity and the activity is due to the heterologously expressed Cre recombinase. The strain QM6aΔ*tmus53*Δ*pyr4*(loxP) can be used as a recipient strain for multiple sequential gene deletions, because it still contains the *cre* gene and is free of marker genes.

3.5 Demonstration of marker reuse by deletion of the PAS domain of env1 from the QM6aΔtmus53Δpyr4(loxP) genome

We used the strain QM6aΔ*tmus53*Δ*pyr4*(loxP) as a recipient strain for the deletion of the PAS domain of *env1* (36) to demonstrate successful reuse of the hygromycin B marker (following the second option proposed in Fig. 2D). The deletion cassette for the PAS domain contains the *hph* gene under control of the *pki* promoter. The resulting deletion strains were characterised by means of PCR. The corresponding primer positions are indicated in Fig. 3G, and the sequences can be found in the Supplementary data (Sup. 1G). The transformants showed poor sporulation, as described earlier (36).

3.6 Demonstration of genotype recovery by pyr4 re-transformation into the QM6aΔtmus53Δpyr4(loxP) genome

The Cre recombinase is a heterologously expressed gene, and we already demonstrated in this paper that it is functional in *H. jecorina*. In order to construct a genetically stable recombinant strain, it is necessary to remove the *cre* gene from the genome. Therefore, we initially directed the *cre* gene to the *pyr4* locus. In a final re-transformation step with the *pyr4* gene, it was now possible to replace the *cre* gene and use the *pyr4* gene as a

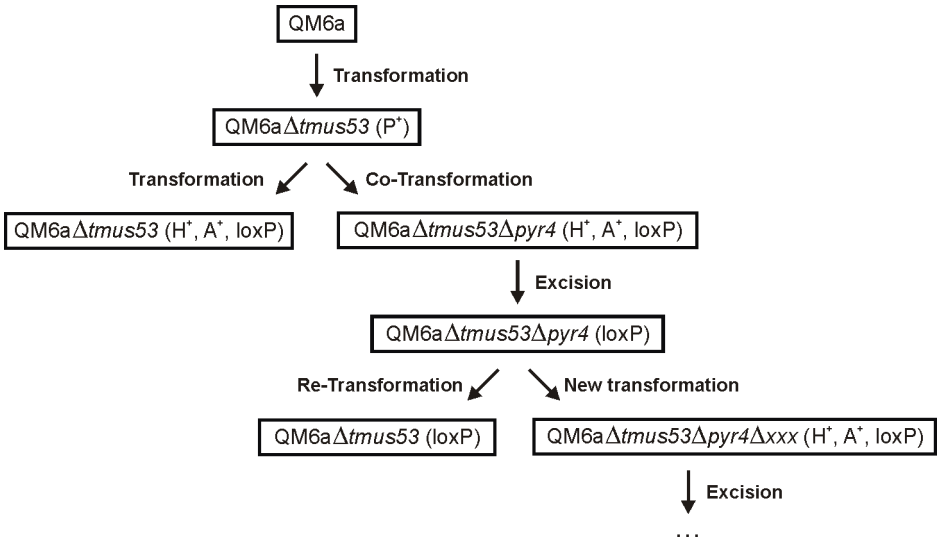


Figure 4 Flow chart of fungal transformation steps and the corresponding strains obtained from *H. jecorina* QM6a. Strain genotypes are given in boxes and their respective properties are given in brackets. H^+, hygromycin B resistance; A^+, ability to use acetamide as a nitrogen source; loxP, loxP-site(s); *xxx*, any gene to be deleted from the genome.

marker. This we demonstrated with the successful re-transformation of QM6aΔ*tmus53*Δ*pyr4*(loxP) with *pyr4* (following the first option proposed in Fig. 2D). The correct transformation was confirmed by means of Southern blot (data not shown) and PCR. The positions of the primers for the control PCR are given in Fig. 3A. The resulting strain (QM6aΔ*tmus53*(loxP)) has lost its uridine auxotrophy and does not contain the Cre recombinase. Thus, it can be assumed that this strain is genetically stable with respect to its loxP sites.

4. DISCUSSION

In this paper, we have presented a Cre/loxP-based donor/acceptor cloning system that is a promising strategy for the creation of deletion cassettes. This system permits the construction of cassettes without the need for restriction digests.

Furthermore, we describe the deletion of *tmus53* from the *H. jecorina* genome. As previously described for *N. crassa*, the disruption of the *lig4* homologue *mus53* leads to an NHEJ-deficient strain (17). We demonstrated that an *H. jecorina* deletion strain has an elevated level of homologous integration events, in some cases up to 100% (Table 3), during the transformation procedure. Recently, the construction of another NHEJ-deficient strain of *H. jecorina* was reported (11), where they deleted the *N.crassa* ortholog *tmus51*. Although, they described an elevated level of homologous recombination events (63-96%), this system cannot be directly compared to the results presented in this paper, because they used a *Hypocrea jecorina* UV-mutant strain (TU-6, (9)) as recipient strain and not the commonly used isolate QM6a.

Using the *tmus53* deletion strain, we developed a marker recycling system for *Hypocrea*. The whole strategy is summarised in Fig. 4. With the strain QM6aΔ*tmus53*Δ*pyr4*(loxP), it is possible to perform sequential gene replacements without running out of available markers. We integrated the *cre* gene into the genome of *Hypocrea* and thereby replaced the *pyr4* gene. The *cre* gene was expressed under the tightly controllable *xyn1* promoter. This system has several advantages: for the excision of the loxP-marker cassette, it is not necessary to perform another transformation. Simple cultivation on xylan induces the excision and leads to the removal of the marker genes. This is timesaving when performing multiple sequential gene deletions. Furthermore, the *xyn1* promoter provides good control over this system, because on glucose this promoter is shut off via a double double-lock mechanism (25, 27). This *xyn1* regulated Cre/loxP-system also works in *H. atroviridis* (38), and it was successfully transferred into *Fusarium* and led to the excision of the loxP-maker cassette in this host as well (K. Brunner, G. Adam, unpublished data). These findings point to the universal applicability of this system in filamentous fungi.

The *cre* gene can be easily removed from the genome by re-transformation with the *pyr4* gene. Hereby, the *pyr4* gene itself can be used as a marker, because uridine auxotrophy is generated in a *pyr4* deletion strain. The elimination of the Cre recombinase is necessary because otherwise undesired recombination events could genetically destabilise the fungus. For the transformation and excision, we used a two marker-system based on the *hph* and *amdS* genes. Hygromycin B is a potent marker for filamentous fungi that is often used in *Hypocrea*. *amdS* has the advantage that it can be used as a bi-directional marker (16), which

facilitates screening in a loxP excision experiment. Growth on fluoroacetamide is reduced in *amdS*-positive *Hypocrea* strains but is not completely eliminated (M. G. Steiger, A. R. Mach-Aigner, unpublished data). However, medium optimisation might help to increase the effect of fluoroacetamide selection in the transformation procedure. In addition to gene deletions, targeted gene integration can also be performed with this system; this can be achieved by integration of the desired gene between flanking regions and loxP sites in the acceptor vector. Then, it will be stably integrated into the locus but not excised by Cre recombinase. The advantages of this system over a previously reported transformation system based on a *pyr4* blaster cassette (14) are its universal applicability. It can also be used in other *Trichoderma* species and even other filamentous fungi. Additionally, the excision of the marker cassette can be easily controlled and is very efficient.

In sum, we present a strategy to genetically engineer *Hypocrea* in a targeted way and to recycle the markers used during the transformation process. This strategy enables the creation of recombinant strains without the disturbing effects of markers on their physiology. Furthermore, it is possible to perform sequential engineering steps without a shortage of markers.

ACKNOWLEDGMENTS

This study was supported by two grants from the Austrian Fonds zur Förderung Wissenschaftlicher Forschung (P21287, P20192), which is gratefully acknowledged. We thank Satu Hilditch for construction of pRS426-dmus and Wolfgang Däuble for the construction of pWD1. G. A. partially financed and hosted K. B. at BOKU Wien (FWF, SFP F-37).

REFERENCES

1. **Austin, S., M. Ziese, and N. Sternberg.** 1981. A novel role for site-specific recombination in maintenance of bacterial replicons. Cell **25:**729-36.

2. **Buchert, J., T. Oksanen, J. Pere, M. Siika-aho, A. Suurnäkki, and L. Viikari.** 1998. Applications of *Trichoderma reesei* enzymes in the pulp and paper industry, p. 343-357. *In* G. E. Harman and C. P. Kubicek (ed.), *Trichoderma* & *Gliocladium*, vol. 2. Taylor & Francis Ltd., London, UK.

3. **Christianson, T. W., R. S. Sikorski, M. Dante, J. H. Shero, and P. Hieter.** 1992. Multifunctional yeast high-copy-number shuttle vectors. Gene **110:**119-22.

4. **Colot, H. V., G. Park, G. E. Turner, C. Ringelberg, C. M. Crew, L. Litvinkova, R. L. Weiss, K. A. Borkovich, and J. C. Dunlap.** 2006. A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. Proc Natl Acad Sci U S A **103:**10352-7.

5. **Dennison, P. M., M. Ramsdale, C. L. Manson, and A. J. Brown.** 2005. Gene disruption in *Candida albicans* using a synthetic, codon-optimised Cre-*loxP* system. Fungal Genet Biol **42:**737-48.

6. **Florea, S., K. Andreeva, C. Machado, P. M. Mirabito, and C. L. Schardl.** 2009. Elimination of marker genes from transformed filamentous fungi by unselected transient transfection with a Cre-expressing plasmid. Fungal Genet Biol **46:**721-30.

7. **Forment, J. V., D. Ramon, and A. P. MacCabe.** 2006. Consecutive gene deletions in *Aspergillus nidulans*: application of the Cre/*loxP* system. Curr Genet **50:**217-24.

8. **Galante, Y. M., Monteverdi, r., Inama, S., Caldini, c., De Conti, A., Lavelli V., Bonomi, F.** 1993. New applications of enzymes in wine making and olive oil production. Italian Biochem. Soc. trans. **4:**34.

9. **Gruber, F., J. Visser, C. P. Kubicek, and L. H. de Graaff.** 1990. Cloning of the *Trichoderma reesei pyrG* gene and its use as a homologous marker for a high-frequency transformation system. Curr Genet **18:**447-451.

10. **Gruber, F., J. Visser, C. P. Kubicek, and L. H. de Graaff.** 1990. The development of a heterologous transformation system for the cellulolytic fungus Trichoderma reesei based on a pyrG-negative mutant strain. Curr Genet **18:**71-6.

11. **Guangtao, Z., L. Hartl, A. Schuster, S. Polak, M. Schmoll, T. Wang, V. Seidl, and B. Seiboth.** 2009. Gene targeting in a nonhomologous end joining deficient *Hypocrea jecorina*. J Biotechnol **139:**146-51.

12. **Güldener, U., S. Heck, T. Fielder, J. Beinhauer, and J. H. Hegemann.** 1996. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res **24:**2519-24.

13. **Hahn-Hägerdal, B., M. Galbe, M. F. Gorwa-Grauslund, G. Liden, and G. Zacchi.** 2006. Bio-ethanol--the fuel of tomorrow from the residues of today. Trends Biotechnol **24:**549-56.

14. **Hartl, L., and B. Seiboth.** 2005. Sequential gene deletions in Hypocrea jecorina using a single blaster cassette. Curr Genet **48:**204-11.

15. **Himmel, M. E., S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, and T. D. Foust.** 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science **315:**804-7.

16. **Hynes, M. J., and J. A. Pateman.** 1970. The genetic analysis of regulation of amidase synthesis in *Aspergillus nidulans*. I. Mutants able to utilize acrylamide. Mol Gen Genet **108:**97-106.

17. **Ishibashi, K., K. Suzuki, Y. Ando, C. Takakura, and H. Inoue.** 2006. Nonhomologous chromosomal integration of foreign DNA is completely dependent on MUS-53 (human Lig4 homolog) in Neurospora. Proc Natl Acad Sci U S A **103:**14871-6.

18. **Jones, C. A., S. E. Greer-Phillips, and K. A. Borkovich.** 2007. The response regulator RRG-1 functions upstream of a mitogen-activated protein kinase pathway impacting asexual development, female fertility, osmotic stress, and fungicide resistance in *Neurospora crassa*. Mol Biol Cell **18:**2123-36.

19. **Koo, H., Ueda, M., Wakida, T., Yoshimura, Y., Igarashi, T.** 1994. Cellulase treatment of cotton fabrics. Textile Res. J. **64:**70-74.

20. **Krappmann, S., O. Bayram, and G. H. Braus.** 2005. Deletion and allelic exchange of the *Aspergillus fumigatus veA* locus via a novel recyclable marker module. Eukaryot Cell **4:**1298-307.

21. **Kück, U., and B. Hoff.** 2010. New tools for the genetic manipulation of filamentous fungi. Appl Microbiol Biotechnol **86:**51-62.

22. **Kuhls, K., E. Lieckfeldt, G. J. Samuels, W. Kovacs, W. Meyer, O. Petrini, W. Gams, T. Borner, and C. P. Kubicek.** 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:**7755-60.

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

24. **Lanzarini, G., Pifferi, P.G.** 1989. Enzymes in the fruit juice industry., p. 189-222. *In* C. Cantarelli, Lanzarini, G. (ed.), Biotechnology Applications in Beverage Production. Elsevier Science, London.

25. **Mach-Aigner, A. R., M. E. Pucher, M. G. Steiger, G. E. Bauer, S. J. Preis, and R. L. Mach.** 2008. Transcriptional

regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in Hypocrea jecorina. Appl Environ Microbiol **74:**6554-62.

26. **Mach, R. L., M. Schindler, and C. P. Kubicek.** 1994. Transformation of *Trichoderma reesei* based on hygromycin B resistance using homologous expression signals. Curr Genet **25:**567-70.

27. **Mach, R. L., J. Strauss, S. Zeilinger, M. Schindler, and C. P. Kubicek.** 1996. Carbon catabolite repression of xylanase I (*xyn1*) gene expression in *Trichoderma reesei*. Mol Microbiol **21:**1273-81.

28. **Mandels, M.** 1985. Applications of cellulases. Biochem Soc Trans **13:**414-6.

29. **Noé P., C. J., Mora F., Comtat J.** 1986. Action of enzymes in chemical pulp and fibers. PartII: Enzymatic beating. J. Wood Chem. Technol. **6:**167-184.

30. **Patel, R. D., J. K. Lodge, and L. G. Baker.** 2010. Going green in *Cryptococcus neoformans*: the recycling of a selectable drug marker. Fungal Genet Biol **47:**191-8.

31. **Pedersen, G. P., Screws, G.A., Cereoni, D.A.** 1992. Biopolishing of cellulosic fabrics. Can. Text. J. **Dec:**31-35.

32. **Penttilä, M., H. Nevalainen, M. Ratto, E. Salminen, and J. Knowles.** 1987. A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. Gene **61:**155-64.

33. **Peterbauer, C. K., K. Brunner, R. L. Mach, and C. P. Kubicek.** 2002. Identification of the N-acetyl-D-glucosamine-inducible element in the promoter of the *Trichoderma atroviride nag1* gene encoding N-acetyl-glucosaminidase. Mol Genet Genomics **267:**162-70.

34. **Ragauskas, A. J., C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, Jr., J. P. Hallett, D. J. Leak, C. L. Liotta, J. R. Mielenz, R. Murphy, R. Templer, and T. Tschaplinski.** 2006. The path forward for biofuels and biomaterials. Science **311:**484-9.

35. **Sauer, B.** 1987. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol Cell Biol **7:**2087-96.

36. **Schmoll, M., L. Franchi, and C. P. Kubicek.** 2005. Envoy, a PAS/LOV domain protein of *Hypocrea jecorina* (Anamorph *Trichoderma reesei*), modulates cellulase gene transcription in response to light. Eukaryot Cell **4:**1998-2007.

37. **Sternberg, N., and D. Hamilton.** 1981. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. J Mol Biol **150:**467-86.

38. **Tisch, D.** 2007. Identification of pathogen inducible genes of *Trichoderma atroviride* and development of a *cre*-lox mediated marker rescue system. University of Technology, Vienna.

39. **Walsh, G. A., R. F. Power, and D. R. Headon.** 1993. Enzymes in the animal-feed industry. Trends Biotechnol **11:**424-30.

40. **Welt, T., Dinus, R.J.** 1995. Enzymatic deinking - a review. Progress in Paper Recycling **4:**36-47.

Supplementary data 1

Sequences of the loci according to Fig. 3
The PCR product that was verified by size by gel electrophoresis is given in bold letters, forward primer are given in underlined letters, reverse primer are given in double underlined letters, sequenced fragments of the PCR products are given in capitals. sc means scaffold position in *H. jecorina* QM6a genome, sc* means relative positions of loci in mutant strains compared to the native *H. jecorina* QM6a genome.

A) native* pyr4 *locus

sc1:1674000_acgccacggcaaaaggctcgagcattttagcgcgctcatcagggccctgtccctttcttctgcttacagtactttggacgcatcgcgacgaacg
gcaatgcagggttcagttgctgcatactgtatcaaggtagtagttatcatcaaggagagaaacgctcgacatataagcttgtacgtgcatgtgttagcatcggacctttt
ttgacgcaaggccagccaggttgggtaagacaacgacaagaggcaggtcagcacttgattcatacggttaccaccagacggtgattcacatatacggggtaagga
gaagcgctcgtcccgcggctggctcgcgcatttgcagcggaagatctcgagatagtatctcattcagtaataatactcgtacgagtttgaggggcgcgaggctaa
gctgcaagcgggcgcggcaacactacctgtatgtattaataaagcagcagagagagagagagccctggactgccggcaatcctgagaaggggggatcgaccg
gaccatctgcacgcagcagggctgctgctgccaacatggagagactaaagatgggattggtggtacatagcaagacgcttggatgcttgttgacgggctgctccaa
taagcaagctcacaagcagtgctttgggggtccgcacggacggtcgagagggctcggacggctgatattgcagacctggacgagtcagttgtgcaggtaggcac
tcgtatgtgtgtctttgtgcttgatcaaccttgagtaagtattattagggtgatggcttaggtaggtaggttaggtaggtagtttaggtacagggaggtgaggaggggag
acgtccagctgccatatgcgccacttgcagcctcaactctgcaaaagccagccccgatcacggcatgtcgtacatactgtaaggtagtacctagtagtagaacgaat
gctaaaggtgcagagcaactaacttgagcaaaaccccccctcccccccatcgagattgattgccctctacgctgaaaatagtctcgccagccggcgacgcttagccg
tggcgctaccgcgacatggcgcttgttagcctgcggtccaggttgatgcgaagcatgtcgcagggcgggtctcgtgcgtgccgacagccagcagagagacagaa
gagagcgagagagcgagagagcgagagagagagagagagagaaagacgcctcgggggtttggctggtggcttcgatgcttgtccgtgtcaagggttcacgaa
agacgggacggctgagttggtgactggcggcggtggcatcgaggctggagtccaggcacatgcatggcgcaaggcatgtcaggccatgtcgtgtcatgccccg
gcacagctgaagggaatactactattgagagctgttctccaaggcgtcaagcatgtgagcaaaaactcgagttgtctagactcgactcgacttgcggctgcagtaca
acaatacagtactactcgtacacgtatgtacagcaaggagcaggtggatcagcaaagttgcatgataatggactggaccggcaacgctcaaagttcgccttgcagct
ccggcatgaagggcagacccccgccattattagcgatgactagccgtggaggccactttggctttagtgatttaaaccccgagattcatgctcgatgctctacaggta
tgtatgtactagctcagctcagccttagtaggtaggtattgaagtctggtggtgcgaggaggacctgcaggtgggagttttgggggaaaaagatgatgctcaccaac
taaaataaatagtcaggccacacacacaacctactgagcagaaccaccacaaccagtgaagagctactagtagtagaaatggagggtaaactgaacatcagtcaa
aagacgtcgagacgcagatactactagatacctatcccagcgacggaaaatggacaaggtgtcgccgccgccaagctctcagatgcgctatcgtgcatgatgcct
ctcatcaacaaacggcgccagtgacgggctagccgtgggattgactcgagctcgcctcttctttgtgcttttctcctgtcgacatcccggcttgcgcttggacctcgcc
gggacgtgcaaacgaacacatcactttcaaagagagtgtttgatgctcacgctcggatgctccagccccccccttgcttgcttgtccatcattgaggctcgaaggggca
gttgctccccattccctctgctccaacaggaactgtaaccgaaaaaaaagccatcctccttcttcctcttcctcttccgcctcctcctcttcctcctcctactaatacctact
cctcctctcctcatcgactagactgacccccccggttgggcccctcgtcccgtctccaacagagcaccaccagacaaagacccctgcccgcgcgaatccagaccc
ccccagcaattccgggcctcgttgatcctcctctactgtagttgtacatacatacctaccgactgcattgcattggtacagctgcaggcacttccaggcacggccacca
aattgcagcggcccttgcttgcttgcttggttcgagaactaggatctgtgtcttttgccttgccttgtcttgtctgggttcctgctcgtctgcggcaatcggaacgccgcca
gtgcggtgccaagcaaaccagccaaggtaggtacctaccactaggcttcttttctcgttgtctcactctctctttcctctttgtcctct**cttatccccatcttttctctctct**
ctctgctcctttcctaaccacttccctacctttctcttttttcctttttcttgtcatctccatcttggctgacgaaaaaggtctgactgggtaggtattatctggcagac
ttgtgtgtatcattcaccctatttctgcttcatagtacatgtactgtacctgaacggctcaaccgctatttacgactcttatttttttgtggcgttggtcacgtttgc
cagctgttgtccgtctttctagggctcctcaaacttgacctgaccgagctccctttctggacccggtgggcttcacttccagctgctgagcgacctgagccga
acatcctcagtccttgtccagcgcaattcatttttctttccttttctttttttttattccttttctttacttttattctctcttttttctcctcttcctcttcttctttctcctcc
tcctccatatcctcactctcgtctccctcattactaccctctcggctcctcaggtccaccaaccctcccgcacccaaacctctgccgctgaaacccattcggtg
gtcgccgttttttttttttttttttttctcacccccaaagtcgcaatatcgggtatcgccgccggcattgaatcgccttctccgctagcatcgactactgctgctctg
ctctcgttgccagcgctgctccctagaattttgaccaggggacgagcccgacattaaagcaactccctcgcctcgagacgactcggatcgcacgaaattct
cccaatcgccgacagttcctactcctcttcctcccgcacggctgtcgcgcttccaacgtcattcgcacagcagaattgtgccatctctctctttttttttccccccc
tctaaaccgccacaacggcaccctaagggttaaactatccaaccagccgcagcctcagcctctctcagcctcatcagccatggcaccacacccgacgctc
aaggccaccttcgcggccaggagcgagacggcgacgcacccgctgacggcttacctgttcaagctcatggacctcaaggcgtccaacctgtgcctgagc
gccgacgtgccgacagcgcgcgagctgctgtacctggccgacaagattggcccgtcgattgtcgtgctcaagacgcactacgacatggtctcgggctggg
acttccacccggagacgggcacgggagcccagctggcgtcgctggcgcgcaagcacggcttcctcatcttcgaggaccgcaagtttggcgacattggcc
acaccgtcgagctgcagtacacgggcgggtcggcgcgcatcat**cgactgggcgcacattgtcaa**cgtcaacatggtgcccggcaaggcgtcggtggcctc
gctggcccagggcgccaagcgctggctcgagcgctacccctgcgaggtcaagacgtccgtcaccgtcggcacgcccaccatggactcgtttgacgacgacgcc
gactccagggacgccgagcccgccggcgccgtcaacggcatgggctccattggcgtcctggacaagcccatctactcgaaccggtccggcgacggccgcaag
ggcagcatcgtctccatcaccaccgtcacccagcagtacgagtccgtctcctcgccccggttaacaaaggccatcgccgagggcgacgagtcgctcttcccgggc
atcgaggaggcgccgctgagccgcggcctcctgatcctcgcccaaatgtccagccagggcaacttcatgaacaaggagtacacgcaggcctgcgtcgaggccg
cccgggagcacaaggactttgtcatgggcttcatctcgcaggagacgctcaacaccgagcccgacgatgccttttatccacatgacgcccggctgccagctgcccc
ccgaagacgaggaccagcagaccaacggatcggtcggtggagacggccagggccagcagtacaacacgccgcacaagctgattggcatcgccggcagcga
cattgccattgtgggccggggcatcctcaaggcctcagaccccgtagaggaggcagagcggtaccgatcagcagcgtggaaagcctacaccgagaggctgctg
cgataggggagggaagggaagaaagaagtaaagaaaggcatttagcaagaagggggaaaagggagggaggacaaacggagctgagaaagagctcttgtcc
aaagcccggcatcatagaatgcagctgtatttaggcgacctcttttccatcttgtcgattttgttatgacgtaccagttgggatgatggatgattgtaccccagctgcga
ttgatgtgtatctttgcatgcaacaacacgcgatggcggaggcgaactgcacattggaaggttcatatatggtcctgacatatctggtggatctggaagcatggaattg
tatttttgatttggcatttgcttttgcgcgtggagggaacatatcaccctcgggcatttttcatttggtaggatggtttggatgcagttgtcgacgatatcagcttccatatat

acggaattgttcaattgagtattacctgatggctatttgtgtcgatcctttgtgacattgacatgtgatggcgcctgtacctctaaatgtatgcagcgtaacatcttcagagc
acggcggcatgttctcccctctacaatattcagataagcagtacttgtgctgtgagaaacccattgatcgctggtggcaaccaaggcgatatgtctccaggagaggaa
attgcacgatcattgatcgtttgtgctactcgtttcgtgtccaagcgaacagagtattaatatataccattatatcatatgcatcgtggagtcgttactatgcagcctaacct
aaggtacctacaaatcggcatcccggagttaacgcggaggctctcatcgccgggcgtctccaccaattcgacatatcaatgctaatcgggcgcaaaggctccaacg
cggatgctggtcgcattgaggttccgcattccttttcctccttccgaccctctttcttacatctcactctcacatcccgtttgccggcccgctccccaccgagacacgaaa
cggacggcaacaatcaaccgtccccccggaacagcgacctgacctcccctcccccccatatacgattacgagctgaacgagacgagagagagagagagagagaga
agaagaagacagagaagaggcatggccgacattgaggtggcagtcaacgtacgggccgagcggggcccgactaagaagcggagacggatcgtcatctcgtg
caccgagtgccatcgaaggaagcaaaaggtcaggcttctttcctaccaagaaacaaagtgaagcttcacgatgctgcatgggcattcgcacaacagccctcgcctc
cgcctcaaggtggatgggaaataacatgaagatgatcacaccagcatgctgacaccttactgttcatatccgcaaatcagtgcgatcgcgagctcccctgcggcaa
ctgccggtcgagaaacaaggaagcccactgcgcctacgaggccggcgcgcccaccgcccgcgagcacaacgagacgcacgccggcctctacaactcctcctc
ctcctcctcctccgccgccgcccgcaatgaccgccacggctacggcagcaacagcggcagcggcaccccgacgtccagggccgagccgctctcgtccatggc
gccgtcgtgggggtacgcccagacaggcgccgcctcgacgctgagcttcctgcagaagattgagcgggcgagcacctcggacggcgacgacgggctaggcg
caggacacgaaggggacgaagcggccgcgggggcggcgtcgtcgtcgcactacc_sc1:1667400

B)* cre *at the* pyr4 *locus

sc*1:1674000_acgccacggcaaaaggctcgagcattttagcgcgctcatcagggccctgtccctttcttctgcttacagtactttggacgcatcgcgacgaac
ggcaatgcagggttcagttgctgcatactgtatcaaggtagtagttatcatcaaggagagaaacgctcgacatataagcttgtacgtgcatgtgttagcatcggacctttt
tttgacgcaaggccagccaggttgggtaagacaacgacaagaggcaggtcagcacttgattcatacggttaccaccagacggtgattcacatatacggggtaagg
agaagcgctcgtcccgcggctggctcgcgcatttgcagcggaagatctcgagatagtatctcattcagtaataatactcgtacgagtttgaggggcgcgaggcta
agctgcaagcgggcgcggcaacactacctgtatgtattaataaagcagcagagagagagagagccctggactgccggcaatcctgagaaggggggatcgaccg
gaccatctgcacgcagcagggctgctgctgccaacatggagagactaaagatgggattggtggtacatagcaagacgcttggatgcttgttgacgggctgctccaa
taagcaagctcacaagcagtgctttgggggtccgcacggacggtcgagagggctcggacggctgatattgcagacctggacgagtcagttgtgcaggtaggcac
tcgtatgtgtgtctttgtgcttgatcaaccttgagtaagtattattagggtgatggcttag**gtaggtaggttaggtaggtagtttaggta**cagggaGGTGAGG
AGGGGAGACGTCCAGCTGCCATATGCGCCACTTGCAGCCTCAACTCTGCAAAAGCCAGCCCC
GATCACGGCATGTCGTACATACTGTAAGGTAGTACCTAGTAGTAGAACGAATGCTAAAGGTG
CAGAGCAACTAACTTGAGCAAAACCCCCCCTCCCCCCATCGAGATTGATTGCCCTCTACGCT
GAAAATAGTCTCGCCAGCCGGCGACGCTTAGCCGTGGCGCTACCGCGACATGGCGCTTGTTA
GCCTGCGGTCCAGGTTGATGCGAAGCATGTCGCAGGGCGGGTCTCGTGCGTGCCGACAGCC
AGCAGAGAGACAGAAGAGAGCGAGAGAGCGAGAGAGCGAGAGAGAGAGAGAGAGAGAAAG
ACGCCTCGGGGGTTTGGCTGGTGGCTTCGATGCTTGTCCGTGTCAAGGGTTCACGAAAGACG
GGACGGCTGAGTTGGTGACTGGCGGCGGTGGCATCGAGGCTGGAGTCCAGGCACATGCATG
GCGCAAGGCATGTCAGGCCATGTCGTGTCATGCCCCGGCACAGCTGAAGGGAATACTACTAT
TGAGAGCTGTTCTCCAAGGCGTCAAGCATGTGAGCAAAAACTCGAGTTGTCTAGACTCGACT
CGACTTGCGGCTGCAGTACAACAATACAGTACTACTCGTACACGTATGTACAGCAAGGAGCA
GGTGGATCAGCAAAGTTGCATGATAATGGACTGGACCGGCAACGCTCAAAGTTCGCCTTGCA
GCTCCGGCATGAAGGGCAGACCCCCGCCATTATTAGCGATGACTAGCCGTGGAGGCCACTTT
GGCTTTAGTGATTTAAACCCCGAGATTCATGCTCGATGCTCTACAGGTATGTATGTACTAGCT
CAGCTCAGCCTTAGTAGGTAGGTATTGAAGTCTGGTGGTGCGAGGAGGACCTGCAGGTGGG
AGTTTTGGGGGAAAAAGATGATGCTCACCAACTAAAATAAATAGTCAGGCCACACACACAAC
CTACTGAGCAGAACCACCACAACCAGTGAAGAGCTACTAGTAGTAGAAATGGAGGGTAAACT
GAACATCAGTCAAAAGACGTCGAGACGCAGATACTACTAGATACCTATCCCAGCGACGGAAA
ATGGACAAGGTGTCGCCGCCGCCAAGCTCTCAGATGCGCTATCGTGCATGATGCCTCTCATC
AACAAACGGCGCCAGTGACGGGCTAGCCGTGGGATTGACTCGAGCTCGCCTCTTCTTTGTGC
TTTTCTCCTGTCGACATCCCGGCTTGCGCTTGGACCTCGCCGGGACGTGcaaacgaacacatcactttcaa
agagagtgtttgatgctcacgctcggatgctccagccccccttgcttgcttgtccatcattgaggctcgaaggggcagttgctccccattccctctgctccaa
caggaactgtaaccgaaaaaaaagccatcctccttcttcctcttcctctccgcctcctcctcttcctcctcctactaatacctactcctcctctcctcatcgact
agactgacccccccggttgggcccctcgtcccgtctccaacagagcaccaccagacaaagacccctgcccgcgcgaatccagacccccccagcaattcc
gggcctcgttgatcctcctctactgtagttgtacatacatacctaccgactgcattgcattggtacagctgcaggcacttccaggcacggccaccaaattgc
agcggcccttgcttgcttgcttggttcgagaactaggatctgtgtcttttgccttgccttgtcttgtctgggttcctgctcgtctgcggcaatcggaacgccgcc
agtgcggtgccaagcaaaccagccaaggtaggtacctaccactaggcttcttttctcgttgtctcactctctctttttcctctttgtcctctcttatccccatctttt
ctctctctctctgctcctttcctaacCACTTCCCTACCTTTCTCTTTTTTCCTTTTCTTGTCATCTCCATCTTGGC
TGACGAAAAAGGTCTGACTGGGTAGGTATTATCTGGCAGACTTGTGTGTATCATTCACGGCC
GCCTGCAGGTCGACCATATGGGAGAGCTCAAGCAACTACGTAAAACTCCATGAGATTGCAGA
TGCGGCCCACTGGAATACAACATCCTCCGCAAGTCCGACATGAAGCCCCTTGACTTGATTGG
CAGGCTAAATGCGACATCTTAGCCGGATGCACCCCAGATCTGGGGAACGCGCCGCTTGAGGC
CCGAAGCGCCGGGTTCGATGCATTACTGCCATATTTCAGCAGTTAACTAGGACCGGCTTGTG
TCGATATTGCGGGTGGCGTTCAATCTATTCCGGCACTCCTATGCCGTTTGATCCGATACCTGG
AGGGCGTGCTTTAGGCAAAATGCCAAGCTTCGAGGATACTGTACGAGCCGCTTTCAACCTCA
CTTGATGATGTCTGAGTTTCATCAAGAGAATTGAAGTCAAAGCTCAAATCATGATGTGAAGAG

GTTTTGAATGTGGAAGAATTCTGCATATATAAAGCCATGGAAGAAGACGTAAAACTGAGACA
GCAAGCTCAACTGCATAGTATCGACTTCAAGGAAAACACGCACAAATAACATCTCGAGATGT
CCAATTTACTGACCGTACACCAAAATTTGCCTGCATTACCGGTCGATGCAACGAGTGATGAG
GTTCGCAAGAACCTGATGGACATGTTCAGGGATCGCCAGGCGTTTTCTGAGCATACCTGGAA
AATGCTTCTGTCCGTTTGCCGGTCGTGGGCGGCATGGTGCAAGTTGAATAACCGGAAATGGT
TTCCCGCAGAACCTGAAGATGTTCGCGATTATCTTCTATATCTTCAGGCGCGCGGTCTGGCA
GTAAAAACTATCCAGCAACATTTGGGCCAGCTAAACATGCTTCATCGTCGGTCCGGGCTGCC
ACGACCAAGTGACAGCAATGCTGTTTCACTGGTTATGCGGCGGATCCGAAAAGAAAACGTTG
ATGCCGGTGAACGTGCAAAACAGGCTCTAGCGTTCGAACGCACTGATTTCGACCAGGTTCGT
TCACTCATGGAAAATAGCGATCGCTGCCAGGATATACGTAATCTGGCATTTCTGGGGATTGC
TTATAACACCCTGTTACGTATAGCCGAAATTGCCAGGATCAGGgttaaagatatctcacgtactgacggtgggag
aatgttaatccatattggcagaacgaaaacgctggttagcaccgcaggtgtagagaaggcacttagcctgggggtaactaaactggtcgagcgatggatt
tccgtctctggtgtagctgatgatccgaataactacctgttttgccgggtcagaaaaaatggtgttgccgcgccatctgccaccagccagctatcaactcgc
gccctggaagggatttttgaagcaactcatcgattgatttacggcgctaaggatgactctggtcagagatacctggcctggtctggacacagtgcccgtgtc
ggagccgcgcgagatatggcccgcgctggagtttcaataccggagatc<u>atgcaagctggtggctggacc</u>aatgtaaatattgtcatgaACTATAT
CCGTAACCTGGATAGTGAAACAGGGGCAATGGTGCG<u>CCTGCTGGAAGATGGCGATTAG</u>GAG
CTCCCAATTCGAGGACCGCAAGTTTGGCGACATTGGCCACACCGTCGAGCTGCAGTACACGG
GCGGGTCGGCGCGCATCATCGACTGGGCGCACATTGTCAACGTCAACATGGTGCCCGGCAA
GGCGTCGGTGGCCTCGCTGGCCCAGGGCGCCAAGCGCTGGCTCGAGCGCTACCCCTGCGAG
GTCAAGACGTCCGTCACCGTCGGCACGCCCACCATGGACTCGTTTGACGACGACGCCGACTC
CAGGGACGCCGAGCCCGCCGGCGCCGTCAACGGCATGGGCTCCATTGGCGTCCTGGACAAG
CCCATCTACTCGAACCGGTCCGGCGACGGCCGCAAGGGCAGCATCGTCTCCATCACCACCGT
CACCCAGCAGTACGAGTCCGTCTCCTCGCCCCGGTTAACAAAGGCCATCGCCGAGGGCGACG
AGTCGCTCTTCCCGGGCATCGAGGAGGCGCCGCTGAGCCGCGGCCTCCTGATCCTCGCCCAA
ATGTCCAGCCAGGGCAACTTCATGAACAAGGAGTACACGCAGGCCTGCGTCGAGGCCGCCC
GGGAGCACAAGGACTTTGTCATGGGCTTCATCTCGCAGGAGACGCTCAACACCGAGCCCGAC
GATGCCTTTATCCACATGACGCCCGGCTGCCAGCTGCCCCCCGAAGACGAGGACCAGCAGAC
CAACGGATCGGTCGGTGGAGACGGCCAGGGCCAGCAGTACAACACGCCGCACAAGCTGATT
GGCATCGCCGGCAGCGACATTGCCATTGTGGGCCGGGGCATCCTCAAGGCCTCAGACCCCGT
AGAGGAGGCAGAGCGGTACCGATCAGCAGCGTGGAAAGCCTACACCGAGAGGCTGCTGCGA
TAGGGGAGGGAAGGGAAGAAAGAAGTAAAGAAAGGCATTTAGCAAGAAGGGGGAAAAGGGA
GGGAGGACAAACGGAGCTGAGAAAGAGCTCTTGTCCAAAGCCCGGCATCATAGAATGCAGCT
GTATTTAGGCGACCTCTTTTTCCATCTTGTCGATTTTTGTTATGACGTACCAGTTGGGATGAT
GGATGATTGTACCCCAGCTGCGATTGATGTGTATCTTTGCATGCAACAACACGCGATGGCGG
AGGCGAACTGCACATTGGAAGGTTCATATATGGTCCTGACATATCTGGTGGATCTGGAAGCA
TGGAATTGTATTTTTGATTTGGCATTTGCTTTTGCGCGTGGAGGGAACATATCACCCTCGGGCATT
TTTCATTTGGTAGGATGGTTTGGATGCAGTTGTCGACGATATCAGCTTCCATATATACGGAATTGTT
CAATTGAGTATTACCTGATGGCTATTTGTGTCGATCCTTTGTGACATTGACATGTGATGGCGCCTGT
ACCTCTAAATGTATGCAGCGTAACATCTTCAGAGCACGGCGGCATGTTCTCCCCTCTACAATATTCA
GATAAGCAGTACTTGTGCTGTGAGAAACCCATTGATCGCTGGTGGCAACCAAGGCGATATGTCTCC
AGGAGAGGAAATTGCACGATCATTGATCGTTTGTGCTACTCGTTTCGTGTCCAAGCGAACAGAGTA
TTAATATATACCATTATATCATATGCATCGTGGAGTCGTTACTATGCAGCCTAACCTAAGGTACCTA
CAAATCGGCATCCCGGAGTTAACGCGGAGGCTCTCATCGCCGGGCGTCTCCACCAATTCGACATAT
CAATGCTAATCGGGCGCAAAGGCTCCAACGCGGATGCTGGTCGCATTGAGGTTCCGCATTCCTTTT
CCTCCTTCCGACCCTCTTTCTTACATCTCACTCTCACATCCCGTTTGCCGGCCCGCTCCCCACCGAGA
CACGAAACGGACGGCAACAATCAACCGTCCCCCCGGAACAGCGACCTGACCTCCCCTCCCCCCATA
TACGATTACGAGCTGAACGAGACGAGAGAGAGAGAGAGAGAGAGAAGAAGAAGACAGAGAAGAGG
CATGGCCGACATTGAGGTGGCAGTCAACGTACGGGCCGAGCGGGGCCCGACTAAGAAGCGGAGAC
GGATCG**tcatctcgtgcac<u>cgagtgccatcgaaggaagcaa</u>**aaggtcaggcttctttcctaccaagaaacaaagtgaagcttcacgatgctgcatgggcat
tcgcacaacagccctcgcctccgcctcaaggtggatgggaaataacatgaagatgatcacaccagcatgctgacaccttactgttcatatccgcaaatcagtgcgat
cgcgagctcccctgcggcaactgccggtcgagaaacaaggaagcccactgcgcctacgaggccggcgcgcccaccgcccgcgagcacaacgagacgcacg
ccggcctctacaactcctcctcctcctcctcctccgccgccgcccgcaatgaccgccacggctacggcagcaacagcggcagcggcaccccgacgtccagggc
cgagccgctctcgtccatggcgccgtcgtgggggtacgcccagacaggcgccgcctcgacgctgagcttcctgcagaagattgagcgggcgagcacctcggac
ggcgacgacgggctaggcgcaggacacgaaggggacgaagcggccgcgggggcggcgtcgtcgtcgcactacc_sc*1:1667400

C) native* tmus53 *locus

sc5:927000_caaaacctccttgttgtgcagttggggcagatgaagcgtcgtgacgactgagcccagggggtcattcggcggtgatttttctccagagccgtgcc
agatgtgactcgcgacccgggaatgcgccgacagcgttcgaacgtgaggatgaggccgcccgtaggaactggaacagagacgctcaagcagctccctttcgag

aacccagaccagtgtgaccaacttcttgacgtgaactggagaaatgtcttcagacaagggccgaacccggacgttgaatctgcatgttgagttgagatgcatgacga
ggccctttgtcaaaacgccgaggatccagtctgtcggtgactctatttcgttgtccacaatggtcctgctttgagatgctggactgcgaacccgtacgccaaaccattc
atacatctgcgggctgtctctctttgcatatgtagtgactgaaggctcaacatgccaccattcgtcatccgcagtgaggtactgacgagccttccgcgggtcatatgct
gctgcatatggttgctgatcagaggagtcttcacagattgccacaggttgtccgtatttcagcagtaacccccctgacctcattcagtatggtctcgtacggatcctttgca
tcggcgggacatgcccacggcagctcttcacgacgtccatggtattcgttgaaggtagagtcgtaattgaatcccggcagctcctgtgccacgagaaattcaacctc
gaaactgtaggtagtttgcgggagacggtcgtcatattctggacgcatggtgatgtgttgttacgcgcgggaatacttggacgagatgctgtcaagtcgaggtctttaa
cgattgtgagagctaaacgattgtagagtatacgtctgcggtcggagagtcgtcgccttttcctggtgactgtagtaaaatatgcctgcggggaacgatgtctgtgctg
agtcgaagggagaacggtagaatatagagaaaagtctctagagcctatcaatcgtctcggggtgctccctttacaaaagaatgactgttgatggttgtgtggaagga
acacgttgcctgctgttcacctggcagatgaactccttcctcttgactgcatgttgacactggcgacctgtttctttgagaaggccagtctcggaatctgggacaagga
ctgagtctggttacagccagtgacaggttaccttgcgttgccttgacgggcgagttgcactgaccccgagcgtcatgatgccacgagcaactttgtcgcggcctccat
cctcaccatatcaagccgcgggtgtgctgagatggcaacatactgtacgtcctcggatgagccaggcaaagaagtgacacgtatcatggaattgaaattcagctgct
gaatggtcaacgtcgacaaccttctaggtcagcttctgatggccaaaaaagcacgagcagctgaagcggcatctcccctctgataccaacagcgcggcgtgtatcg
ggcattcatgctgtgccatgtatcatgtagcatccggtatgtggcattgattttacccgtgcgattgtctcaagcagagctcttgagcccccagagatgcccgatggag
caatgaatgcgaaagacgaagctcaagttgcttggctttggagatccgcaggtaacagcaaacaagtcgaggctggccgcggttacataagcagatgcaagctag
gcagattgctttgtag**<u>actaatcccaaaaaggtggcgtgc</u>aggcagcagctccgggcccacaatttctggatagctgtacctacctactactagtaaggcaa**
tgacttcacggcaactttggccttgacccttgaatcagcagtacttaaagctgcatcatcacagcagcgacgcgattccttcaccgtctcgttcattcaattg
ctttgtctggtcctgtcacatttcccgtacaattcgagcccttctggaagaattgacagtgccaatctcctcagcggcgatactccgttctgtcaaggtcctttc
gttcactggaccatcatcaccactaacaaacaagccaatcttcagtctagcgcgacatgagcagatggacgcccaaggaagaaaaagccccgacgaa
gacgcgctgagggaggatgacctgcaatatggcacaagtggcatgacggaggctgaaatcgctgaaaagtaagaaatctatctctaccttatatctcaa
aacatcagcagcaggcgttaacccagcagccaaggtatcccaataagccgaggaaccactccaagacgctcaagttcgcagagctgttcaaggagctg
ttcaacccccctcaacgacaacaagaagcagccttctggcccatcaattcccgcaagcaaaaggggccctcatggactgacaaagctatcgcctcaggag
cagaggcggcacatcatcgagaggttcatcaccagatggagaaaggaagtcggcaatgacttttaccctgctttgcgactcatcctgccggacaaggatc
gagatcgcggcgtctacggcctgaaggagaatgccatcgggaagcttttggtgaaattgatgaagattgacaagaactcggaagatggctacaatctctt
gcactggaagctgccaggccaaacaaccggcgcgcgtctggcgggtgactttgcgggccggtgcttcgaggtcatctcgaaacggcccatgcgcaccca
agtcggagacatgaccattgcagaggtcaatcatcaactggataagctggctgcctcgagcggagaagctgaaaacctccgcatctttgaaacctttac
aatcgcatgaatgcggaagagcttatgtggctcatccgcatcattctgaagcagatgaaggtgggcgctaccgagaagacgttcctatacctgtggcacc
ccgatgccgaggctcttttcagcgtgtcatccagtctcaggcgcgtatgctgggagctttacgaccccgagtttcgactcgaacaggatgaggcgggcatt
gcagttatgcagtgcttccagccacagcttgcccagttccagatgccggcctcgttccaaaagatgattgagctgctacggccgacggaagatgacgccg
agttttggatcgaggagaagctcgacggcgagcgcatgcagatgcacatgatggaggaccctaaccaccccgggggtcggagattccgctggtggtcga
gaaaggccaaggactacacctatctctatggcgacggcttactggatgagaactccagcttgacgaggcatctcaagaacgcctttgcacccggcgtgcg
taacctgattctcgatggcgagatgatcacctgggacatgggcgtggacaagattgtccccttggtactctcaagtcggctgccatcgcagaacaaagca
acaagtccgaatccaacgcggccggccaccgcccgctgtttcgcgtgttcgacatactgtatctcaacgacaagcagctgacgcagtacaccctgcgcga
tcgccgaaatgcactgcaaaaggcaatcaggccggtccatcgccgactggaaatccacgaatacacgcccgccacctcagccgatgccatcgagcccct
tctgcgcgaggtcgtcgccaacgcgtcagaaggcctcgtgttgaagaacccacggtccatgtatcgcctcaacagccgcaacgacgactggcttaaggta
aagcccgagtacatgtccgagttcggcgagtcgctcgactgcgtcatcatcggcggctactatggctcgggacgacgaggagggatgctgtcgagcttcc
tgtgcggcctccgcgtcagcccgaaccacatccaggccggcgccaacccagaaaagtgcttcagcttcttcaaagtcggcggcggcttccgcacagagg
actacgccgagatcaggcaccgaacagacggcaagtggatggagtgggatcccaaaaacccgccgtccgagtacattgagctgggcggcggcgaggc
caagcagtacgaacggcccgacatgtggattcggccaaaggactccgtcgtggtctctgtcaaggcggccagcgtcggtccgtcggatcagttcgcccgg
ggcttcaccctgcgattccctcggtttcgaagactacgactcgaccgcagctgggacacggccctgtctctagaggagtttcaagagctcaagcagagggt
agaatcggaactcaaagcgaccgagatgacagtcgaagaccgcaggaggcgaaacccaaagcgaatcaagaaagagctcgtcatcgcgggcgacga
cacagctcctcccctattcgaaggcaacaagacgaagctcttcgaaggtctcgaattctgcgtcctatccgaatcgctcaagcccctacaaaaagaccaaa
ccacagctcgaagccatcatcaaagaaaacggaggctcaatatcgcaaagagccgccgcaggcacgagaatgatcctcatagccgacaaaagagtcg
tcaaagtcgccagcctcatcaaagacggcaacgcagacattatccggccccgctggatcagggactgtctggagcagcccgacggcggattcttgctccc
gtacgagccggggcggcatatattccacgcgacggacgccctggcagaagtcgctgcgcagaacacggatcggtacggggatagctacgcgcgcgatg
tggatgttgaggagttgcgcgagttgatgagggacatgccgaagctcgaggacaggcagacgtttgacaagatgctgtttctggaacaactggagcagc
atggcaaggacgtcggttccctgcgaaactttatgttttcgcgctgcagggtgctgctttacctcgtcgacggcgcaaaggagagcgcacgggataaatgg
gagaggtacgtccgctacgggggcggttccgtcagccgagatgcagacgacgactcggtgacgcatattgttgtgctgggagacgatgtgatggaggtt
ggcgacgtgggggacaagttgcggtcgaagctgagttcgaggaggaagatgccgaggttggtgagcgaggggtgggtggaagagagttggcgggag
aggacgcttttggacgaggagcggtatgctgtctgatttgcctgagaatgaggaggattctgaactggacaggatccggtcggtctagttccgacttcaatt
gagaggtctttgtaatgcaaaatagtcattttgataggagcatctatatttctaaacaaggagcatatgcaagggtatctcgacaagccttcaccaacccca
cgcctcaaagtatgcagagttaccgcatcatcaccgaccatcgtcagcccagtcaccgcgtcttcctcttggccttgttcttttttccggatattcagcaccc<u>gc</u>
<u>ttcctctgtatgaccctcacaaac</u>caggcgtaaatcatttcgcgcatgaggatgagcaccgtagcaacagccgtcggaacagcacaagcccagctatgacgaatt
cgtgttagcattcgtccttttgctttctctgcaaatagttcagaaaacacttacaagtaccatatcccgttgttaatctcgggaaagccctccggcggaaagttttgcccaa
agaaacccgtgataaacgtcaggggcaggaagatgatggtcaccgttgtcagttgcttcatgctctcattctggtacgccgagattgtgttgaaaatcaggttgatcaa
gttctcgctcgactgcttggactgctgcagcgcctccgtaatgatgatgcagtggtccagcacgtcgccgaggtacgtgtgtgtcatgggcgtgacgatgacgccg
ctggccgggttctccagctcgttcatggcctggtcgtgcgtcaggttcgtgcggtggtcgcgcagcacgttgaccaggttgtcgatggggttgaggaagcgcagca
tcttgttgatttcgctgatgcagatgtacagggtcttgcactgcttgatgttgggcgatgtcagcacgtccagctccaggtcgccgagcacgtcgttgtacgcggccgt
caggggcacggccaggtcgacgatggcgtcgatgatggcctgcaccagcagtgaggcgtcgcacgactcgcgcagcaccgtgcccgggttgctcaggcgcgt
cacaatcggacgctccacgtcgccggcagacacctcaaagaaggagatgacggtattgtcggacgttaggaaaatggacacctgctcggccgacactgccaggt

tgagcggcgtcagaatcgagttggcctccatgtactcggtgcggatctcgttgcccgaggcgtggtatcgctgcagcgagcgcatcatggacgtctcctgaagccc
cagcgtctcggacagctccgtcatgcacttggtgggcgcacgcttctcggggtcgagcggggtgttgttggtggccttgtctgcggctttgctaccaaagctgcgga
accagctcttgaacttggaggacgactttcccgagaggatgctgctcgtgtccgatgaggaatcgtcgtcgtccacgagatggaccagcttctgtagcgtcatgaca
atgaaggcatggctcggataccttgacagaggcaaatcagtatggaaggcataaccaaagaattggccgtgtgttaacccaccagtcagccttggtcctgtttctaa
tgttcatgatgtcctccaacgccagcttatgcagccccttgtgcgcaccgatcgcctgaatcacatcccaactgagcccattgacgttgatccaccgacactttgccca
ctcctccttgggctccgtcagaaacgacacaaacgagtcgttgtcaaagtgcttcttcaccatattatcttttgaaaaatctacgaccgtgatttggcaatccgcactcag
cgtcggcatcgaagcgtggccgccatcagggagatggggatcaaagcccggctcagaccccggttgccagcctggcctctccgtgtgcgtggcatcgaagtcga
agtcgtcgttgtcatcaatggtgaggaaagtctggacgcgagcgctcttgcgacggcgcagcggcgaaaagggctcgcgatggacggccgtgtcggagcgcga
gaggcgccgtacgtcggcctggggaggcgtattgttgttggcgttctcagtgtcgacgctgagatcgtagcggactcggttcagggatggccgtgcgtcggtagg
agagacggcgctgttcccttcttcgtggacgggatccatggtgtctgtgaggcctgtcgagacgagatcgagcgaggagatggcttcgagccagcgcaatcacgc
taacagttgcgctctgctttgcgcgtcgctctcgacagccgcggcgaagatgctggtgagcaatgacagccaatgctggagccgtcggttggttgagcttccgaac
cagcagtcgccagatttcgatctcacaatggtaacagatacgtcacctcgtcagattcaacggcttcggtcgtcttccatccaagtcgtgtcaggtcccaagagagttt
ccgtcgccaagaatcaaagataatatgatcgcattcacggcatcgtcatccaaaacatttcgccatgtttcttgaaacagagcatgagagagggaggagagaaaaaa
gattgatagataagcccatggtataattagaattggaggagcgtcttattgcatgttcgcttcagaacggatcgcagctgccgccccttagcctgggccttagccgca
gctgttagggaacaggttccctgaaatacattcccgtcagcatgacgtgatacatggcgatgatcaccgcgacccggccaatcacgccgcaaatcagaacatgatg
ggtcaattggctggagttgctccacgctggactgatatcgtcttttctacagtatgtactgtatgtgctcgtgataccagacatgcatctgatcgtcaagaatggatagtat
gtacttgtctcgtcatgatcaacgtctagac_sc5:919000

D)* bar *cassette at the* tmus53 *locus

sc*5:927000_caaaacctccttgttgtgcagttggggcagatgaagcgtcgtgacgactgagcccaggggtcattcggcggtgatttttctccagagccgtgc
cagatgtgactcgcgacccgggaatgcgccgacagcgttcgaacgtgaggatgaggccgcccgtaggaactggaacagagacgctcaagcagctccctttcga
gaacccagaccagtgtgaccaacttcttgacgtgaactggagaaatgtcttcagacaagggccgaacccggacgttgaatctgcatgttgagttgagatgcatgacg
aggccctttgtcaaaacgccgaggatccagtctgtcggtgactctatttcgttgtccacaatggtcctgctttgagatgctggactgcgaacccgtacgccaaaccatt
catacatctgcgggctgtctctctttgcatatgtagtgactgaaggctcaacatgccaccattcgtcatccgcagtgaggtactgacgagccttccgcgggtcatatgct
gctgcatatggttgctgatcagaggagtcttcacagattgccacaggttgtccgtatttcagcagtaacccccctgacctcattcagtatggtctcgtacggatcctttgca
tcggcgggacatgcccacggcagctcttcacgacgtccatggtattcgttgaaggtagagtcgtaattgaatcccggcagctcctgtgccacgagaaattcaacctc
gaaactgtaggtagtttgcgggagacggtcgtcatattctggacgcatggtgatgtgttgttacgcgcgggaatacttggacgagatgctgtcaagtcgaggtctttaa
cgattgtgagagctaaacgattgtagagtatacgtctgcggtcggagagtcgtcgccttttcctggtgactgtagtaaaatatgcctgcggggaacgatgtctgtgctg
agtcgaagggagaacggtagaatatagagaaaagtctctagagcctatcaatcgtctcggggtgctcccttacaaaagaatgactgttgatggttgtgtggaagga
acacgttgcctgctgttcacctggcagatgaactccttcctcttgactgcatgttgacactggcgacctgtttctttgagaaggccagtctcggaatctgggacaagga
ctgagtctggttacagccagtgacaggttaccttgcgttgccttgacgggcgagttgcactgaccccgagcgtcatgatgccacgagcaactttgtcgcggcctccat
cctcaccatatcaagccgcgggtgtgctgagatggcaacatactgtacgtcctcggatgagccaggcaaagaagtgacacgtatcatggaattgaaattcagctgct
gaatggtcaacgtcgacaaccttctaggtcagcttctgatggccaaaaaagcacgagcagctgaagcggcatctcccctctgataccaacagcgcggcgtgtatcg
ggcattcatgctgtgccatgtatcatgtagcatccggtatgtggcattgattttacccgtgcgattgtctcaagcagagctcttgagcccccagagatgcccgatggag
caatgaatgcgaaagacgaagctcaagttgcttggctttggagatccgcaggtaacagcaaacaagtcgaggctggccgcggttacataagcagatgcaagctag
gcagattgctttgtag**actaatcccaaaaaggtggcgtgcag**GCAGCAGCTCCGGGCCCACAATTTCTGGATAGCTGTA
CCTACCTACTACTAGTAAGGCAATGACTTCACGGCAACTTTGGCCTTGACCCTTGAATCAGCA
GTACTTAAAGCTGCATCATCACAGCAGCGACGCGATTCCTTCACCGTCTCGTTCATTCAATTG
CTTTGTCTGGTCCTGTCACATTTCCCGTACAATTCGAGCCCTTCTGGAAGAATTGACAGTGCC
AATCTCCTCAGCGGCGATACTCCGTTCTGTCAAGGTCCTTTCGTTCACTGGACCATCATCACC
ACTAACAAACAAGCCAATCTTCAGTCTAGCGCGAGTTGGAGATTTCAGTAACGTTAAGTGGA
TCCCGGTGACGGATCAGATCTCGGTGACGGGCAGGACCGGACGGGGCGGTACCGGCAGGCT
GAAGTCCAGCTGCCAGAAACCCACGTCATGCCAGTTCCCGTGCTTGAAGCCGGCCGCCCGCA
GCATGCCGCGGGGGGCATATCCGAGCGCCTCGTGCATGCGCACGCTCGGGTCGTTGGGCAG
CCCGATGACAGCGACCACGCTCTTGAAGCCCTGTGCCTCCAGGGACTTCAGCAGGTGGGTGT
AGAGCGTGGAGCCCAGTCCCGTCCGCTGGTGGCGGGGGGAGACGTACACGGTCGACTCGGC
CGTCCAGTCGTAGGCGTTGCGTGCCTTCCAGGGGCCCGCGTAGGCGATGCCGGCGACCTCG
CCGTCCACCTCGGCGACGAGCCAGGGATAGCGCTCCCGCAGACGGACGAGGTCGTCCGTCC
ACTCCTGCGGTTCCTGCGGCTCGGTACGGAAGTTGACCGTGCTTGTCTCGATGTAGTGGTTG
ACGATGGTGCAGACCGCCGGCATGTCCGCCTCGGTGGCACGGCGGATGTCGGCCGGGCGTC
GTTCTGGGCTCATGGTTACTTCCTAATCGAAGCTTTGCTTGGGTAGAATAGGTAAGTCAGATT
GAATCTGAAATAAAGGGAGGAAGGGCGAACTTAAGAAGGTATGACCGGGTCGTTCACTTACC
TTGCTTGACAAACGCACCAAGTTATCGTGCACCAAGCAGCAGATGATAATAATGTCCTCGTTC
CTGTCTGCTAATAAGAGTCACACTTCGAGCGCCGCCGCTACTGCTACAAGTGGGGCTGATCT
GACCAGTTGCCTAAATGAACCATCTTGTCAAACGACACAAATTTTGTGCTCACCGCCTGGACG
ACTAAACCAAAATAGGCATTCATTGTTGACCTCCACTAGCTCCAGCCAAGCCCAAAAAGTGCT
CCTTCAATATCATCTTCTGTCTCCGACTTTGCCTGAGAATGAGGAGGATTCTGAACTGGACAG
GATCCGGTCGGTCTAGTTCCGACTTCAATTGAGAGGTCTTTGTAATGCAAAATAGTCATTTTG

ATAGGAGCATCTATATTTCTAAACAAGGAGCATATGCAAGGGTATCTCGACAAGCCTTCACCA
ACCCCACGCCTCAAAGTATGCAGAGTTACCGCATCATCACCGACCATCGTCAGCCCAGTCAC
CGCGTCTTCCTCTTGGCCTtgttctttttccggatattcagcaccc<u>gcttcctctgtatgaccctcacaaac</u>caggcgtaaatcatttcgcgca
tgaggatgagcaccgtagcaacagccgtcggaacagcacaagcccagctatgacgaattcgtgttagcattcgtccttttgctttctctgcaaatagttcagaaaaca
cttacaagtaccatatcccgttgttaatctcgggaaagccctccggcggaaagttttgcccaaagaaacccgtgataaacgtcaggggcaggaagatgatggtcacc
gttgtcagttgcttcatgctctcattctggtacgccgagattgtgttgaaaatcaggttgatcaagttctcgctcgactgcttggactgctgcagcgcctccgtaatgatg
atgcagtggtccagcacgtcgccgaggtacgtgtgtgtcatgggcgtgacgatgacgccgctggccgggttctccagctcgttcatggcctggtcgtgcgtcaggt
tcgtgcggtggtcgcgcagcacgttgaccaggttgtcgatggggttgaggaagcgcagcatcttgttgatttcgctgatgcagatgtacagggtcttgcactgcttga
tgttgggcgatgtcagcacgtccagctccaggtcgccgagcacgtcgttgtacgcggccgtcaggggcacggccaggtcgacgatggcgtcgatgatggcctgc
accagcagtgaggcgtcgcacgactcgcgcagcaccgtgcccgggttgctcaggcgcgtcacaatcggacgctccacgtcgccggcagacacctcaaagaag
gagatgacggtattgtcggacgttaggaaaatggacacctgctcggccgacactgccaggttgagcggcgtcagaatcgagttggcctccatgtactcggtgcgg
atctcgttgcccgaggcgtggtatcgctgcagcgagcgcatcatggacgtctcctgaagccccagcgtctcggacagctccgtcatgcacttggtgggcgcacgct
tctcggggtcgagcggggtgttgttggtggccttgtctgcggctttgctaccaaagctgcggaaccagctcttgaacttggaggacgactttcccgagaggatgctg
ctcgtgtccgatgaggaatcgtcgtcgtccacgagatggaccagcttctgtagcgtcatgacaatgaaggcatggctcggataccttgacagaggcaaatcagtat
ggaaggcataaccaaagaattggccgtgtgttaacccaccagtcagccttggtcctgtttctaatgttcatgatgtcctccaacgccagcttatgcagccccttgtgcg
caccgatcgcctgaatcacatcccaactgagcccattgacgttgatccaccgacactttgcccactcctccttgggctccgtcagaaacgacacaaacgagtcgttgt
caaagtgcttcttcaccatattatcttttgaaaaatctacgaccgtgatttggcaatccgcactcagcgtcggcatcgaagcgtggccgccatcagggagatggggat
caaagcccggctcagaccccggttgccagcctggcctctccgtgtgcgtggcatcgaagtcgaagtcgtcgttgtcatcaatggtgaggaaagtctggacgcgag
cgctcttgcgacggcgcagcggcgaaaagggctcgcgatggacggccgtgtcggagcgcgagaggcgccgtacgtcggcctggggaggcgtattgttgttgg
cgttctcagtgtcgacgctgagatcgtagcggactcggttcagggatggccgtgcgtcggtaggagagacggcgctgttcccttcttcgtggacgggatccatggt
gtctgtgaggcctgtcgagacgagatcgagcgaggagatggcttcgagccagcgcaatcacgctaacagttgcgctctgctttgcgcgtcgctctcgacagccgc
ggcgaagatgctggtgagcaatgacagccaatgctggagccgtcggttggttgagcttccgaaccagcagtcgccagatttcgatctcacaatggtaacagatacg
tcacctcgtcagattcaacggcttcggtcgtcttccatccaagtcgtgtcaggtcccaagagagtttccgtcgccaagaatcaaagataatatgatcgcattcacggca
tcgtcatccaaaacatttcgccatgtttcttgaaacagagcatgagagagggaggagagaaaaaagattgatagataagcccatggtataattagaattggaggagc
gtcttattgcatgttcgcttcagaacggatcgcagctgccgccccttagcctgggccttagccgcagctgttagggaacaggttccctgaaatacattcccgtcagc
atgacgtgatacatggcgatgatcaccgcgacccggccaatcacgccgcaaatcagaacatgatgggtcaattggctggagttgctccacgctggactgatatcgt
cttttctacagtatgtactgtatgtgctcgtgataccagacatgcatctgatcgtcaagaatggatagtatgtacttgtctcgtcatgatcaacgtctagac_sc*5:919
000

E) loxP-marker cassette at the* tmus53 *locus

sc*5:927000_caaaacctccttgttgtgcagttggggcagatgaagcgtcgtgacgactgagcccaggggtcattcggcggtgatttttctccagagccgtgc
cagatgtgactcgcgacccgggaatgcgccgacagcgttcgaacgtgaggatgaggccgcccgtaggaactggaacagagacgctcaagcagctccctttcga
gaacccagaccagtgtgaccaacttcttgacgtgaactggagaaatgtcttcagacaagggccgaacccggacgttgaatctgcatgttgagttgagatgcatgacg
aggccctttgtcaaaacgccgaggatccagtctgtcggtgactctatttcgttgtccac**<u>aatggtcctgctttgagatgctgg</u>actgcgaaccCGTACGCC**
AAACCATTCATACATCTGCGGGCTGTCTCTCTTTGCATATGTAGTGACTGAAGGCTCAACATG
CCACCATTCGTCATCCGCAGTGAGGTACTGACGAGCCTTCCGCGGGTCATATGCTGCTGCAT
ATGGTTGCTGATCAGAGGAGTCTTCACAGATTGCCACAGGTTGTCCGTATTTCAGCAGTAAC
CCCCTGACCTCATTCAGTATGGTCTCGTACGGATCCTTTGCATCGGCGGGACATGCCCACGG
CAGCTCTTCACGACGTCCATGGTATTCGTTGAAGGTAGAGTCGTAATTGAATCCCGGCAGCT
CCTGTGCCACGAGAAATTCAACCTCGAAACTGTAGGTAGTTTGCGGGAGACGGTCGTCATAT
TCTGGACGCATGGTGATGTGTTGTTACGCGCGGGAATACTTGGACGAGATGCTGTCAAGTCG
AGGTCTTTAACGATTGTGAGAGCTAAACGATTGTAGAGTATACGTCTGCGGTCGGAGAGTCG
TCGCCTTTCCTGGTGACTGTAGTAAAATATGCCTGCGGGGAACGATGTCTGTGCTGAGTCGA
AGGGAGAACGGTAGAATATAGAGAAAAGTCTCTAGAGCCTATCAATCGTCTCGGGGTGCTCC
CTTTACAAAAGAATGACTGTTGATGGTTGTGTGGAAGGAACACGTTGCCTGCTGTTCACCTG
GCAGATGAACTCCTTCCTCTTGACTGCATGTTGACACTGGCGACCTGTTTCTTTGAGAAGGCC
AGTCTCGGAATCTGGGACAAGGACTGAGTCTGGTTACAGCCAGTGACAGGTTACCTTGCGTT
GCCTTGACGGGCGAGTTGCACTGACCCCGAGCGTCATGATGCCACGAGCAACTTTGTCGCGG
CCTCCATCCTCACCATATCAAGCCGCGGGTGTGCTGAGATGGCAACATACTGTACGTCCTCG
GATGAGCCAGGCAAAGAAGTGACACGTATCATGGAATTGAAATTCAGCTGCTGAATGGTCAA
CGTCGACAACCTTCTAGGTCAGCTTCTGATGGCCAAAAAAGCACGAGCAGCTGAAGCGGCAT
CTCCCCTCTGATACCAACAGCGCGGCGTGTATCGGGCATTCATGCTGTGCCATGTATCATGT
AGCATCCGGTATGTGGCATTGATTTTACCCGTGCGATTGTCTCAAGCAGAGCTCTTGAGCCC
CCAGAGATGCCCGATGGAGCAATGAATGCGAAAGACGAAGCTCAAGTTGCTTGGCTTTGGAG
ATCCGCAGGTAACAGCAAACAAGTCGAGGCTGGCCGCGGTTACATAAGCAGATGcaagctaggcag
attgctttgtagactaatcccaaaaaggtggcgtgcaggcagcagctccgggccCACAATTTCTGGATAGCTGTACCTACCTA
CTACTAGTAAGGCAATGACTTCACGGCAACTTTGGCCTTGACCCTTGAATCAGCAGTACTTAA
AGCTGCATCATCACAGCAGCGACGCGATTCCTTCACCGTCTCGTTCATTCAATTGCTTTGTCT

GGTCCTGTCACATTTCCCGTACAATTCGAGCCCTTCTGGAAGAATTGACAGTGCCAATCTCCT
CAGCGGCGATACTCCGTTCTGTCAAGGTCCTTTATAACTTCGTATAATGTATGCTATACGAAG
TTATTAGGTCTAGATCTACGCCAGGACCGAGCAAGCCCAGATGAGAACCGACGCAGATTTCC
TTGGCACCTGTTGCTTCAGCTGAATCCTGGCAATACGAGATACCTGCTTTGAATATTTTGAAT
AGCTCG<u>CCCGCTGGAGAGCATCCTGAA</u>tgcaagtaacaaccgtagaggctgacacggcaggtgttgctagggagcgtcgtgttctac
aaggccagacgtcttcgcggttgatatatatgtatgtttgactgcaggctgctcagcgacgacagtcaagttcgccctcgctgcttgtgcaataatcgcagtggggaa
gccacaccgtgactcccatctttcagtaaagctctgttggtgtttatcagcaatacacgtaatttaaactcgttagcatggggctgatagcttaattaccgtttaccagtgc
cgcggttctgcagctttccttggcccgtaaaattcggcgaagccagccaatcaccagctaggcaccagctaaaccctataattagtctcttatcaacaccatccgctcc
cccgggatcaatgaggagaatgaggggatgcggggctaaagaagcctacataaccctcatgccaactcccagtttacactcgtcgagccaacatcctgactataa
gctaacacagaatgcctcaatcctgggaagaactggccgctgataagcgcgcccgcctcgcaaaaaccatccctgatgaatggaaagtccagacgctgcctgcg
gaagacagcgttattgatttcccaaagaaatcggggatcctttcagaggccgaactgaagatcacagaggcctccgctgcagatcttgtgtccaagctggcggccg
gagagttgacctcggtggaagttacgctagcattctgtaaacgggcagcaatcgcccagcagttagtagggtcccctctacctctcagggagatgtaacaacgcca
ccttatgggactatcaagctgacgctggcttctgtgcagacaaactgcgcccacgagttcttccctgacgccgctctcgcgcaggcaagggaactcgatgaatacta
cgcaaagcacaagagacccgttggtccactccatggcctccccatctctctcaaagaccagcttcgagtcaaggtacaccgttgcccctaagtcgttagatgtccctt
tttgtcagctaacatatgccaccagggctacgaaacatcaatgggctacatctcatggctaaacaagtacgacgaaggggactcggttctgacaaccatgctccgca
aagccggtgccgtcttctacgtcaagacctctgtcccgcagaccctgatggtctgcgagacagtcaacaacatcatcgggcgcaccgtcaacccacgcaacaaga
actggtcgtgcggcggcagttctggtggtgagggtgcgatcgttgggattcgtggtggcgtcatcggtgtaggaacggatatcggtggctcgattcgagtgccggc
cgcgttcaacttcctgtacggtctaaggccgagtcatgggcggctgccgtatgcaaagatggcgaacagcatggagggtcaggagacggtgcacagcgttgtcg
ggccgattacgcactctgttgagggtgagtccttcgcctcttccttcttttcctgctctataccaggcctccactgtcctccttttcttgctttttatactatatacgagaccgg
cagtcactgatgaagtatgttagacctccgcctcttcaccaaatccgtcctcggtcaggagccatggaaatacgactccaaggtcatccccatgccctggcgccagt
ccgagtcggacattattgcctccaagatcaagaacggcgggctcaatatcggctactacaacttcgacggcaatgtccttccacaccctcctatcctgcgcggcgtg
gaaaccaccgtcgccgcactcgccaaagccggtcacaccgtgaccccgtggacgccatacaagcacgatttcggccacgatctcatctcccatatctacgcggct
gacggcagcgccgacgtaatgcgcgatatcagtgcatccggcgagccggcgattccaaatatcaaagacctactgaacccgaacatcaaagctgttaacatgaac
gagctctgggacacgcatctccagaagtggaattaccagatggagtaccttgagaaatggcgggaggctgaagaaaaggccgggaaggaactggacgccatca
tcgcgccgattacgcctaccgctgcggtacggcatgaccagttccggtactatgggtatgcctctgtgatcaacctgctggatttcacgagcgtggttgttccggttac
ctttgcggataagaacatcgataagaagaatgagagtttcaaggcggttagtgagcttgatgccctcgtgcaggaagagtatgatccggaggcgtaccatggggca
ccggttgcagtgcaggttatcggacggagactcagtgaagagaggacgttggcgattgcagaggaagtggggaagttgctgggaaatgtggtgactccatagcta
ataagtgtcagatagcaatttgcacaagaaatcaataccagcaactgtaaataagcgctgaagtgaccatgccatgctacgaaagagcagaaaaaaacctgccgta
gaaccgaagagatatgacacgcttccatctctcaaaggaagaatcccttcagggttgcgtttccagctcgagtctagagatctcgagataacggtgagactagcggc
cggtcccttatccagctgtccacgttggcctgccctcagttagcgctcaactcatgccctcactggcgaggcgagggcaaggatggaggggcagcatcgcctgag
ttggagcaaagcgccgccatgggagcagcgaaccaacggagggatgccgtgctttgtcgtggctgctgtggccaatccgggcccttggttggctcacagagcgtt
gctgtgagaccatgagctattattgctaggtacagtatagagagaggagagagagagagagagtgtgagagggaaaaggtgaggttgaagtgaggtagtaagtg
aaaaaaaaaaaaaaatcaacactgacggctgccgtctgccaccctcctccacccagccacctgcacactcagcgcgcagcatcacctaatcttggctcgccttccg
cagctcaggttgttttttttttttctctctcctcgtcgaagccgcccttgttcccttatttatttccctctcctaccttgtctgcctttggtccatctgccccttttgtctgcatctctttt
gcacgcatcgccttatcgtcgtctctttttttcactcacgggagcttgacgaagacctgactcgtga**<u>gcctcacctgctgatttctc</u>tccccccctcccGACCGG**
CTTGACTTTTGTTTCTCCTCCAGTACCTTATCGCGAAGCCGGAAGAACCCTCTTAACCTCTAG
AATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACA
GCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTA
GGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTA
TGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAAT
TCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTG
CCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGC
GGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACA
CTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGA
TGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAG
GACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGA
CAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACG
AGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTAC
TTCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCAT
TGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGC
AGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGC
CCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAAC
CGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAATGCATGGCTTTCGTGACCGGGCTTCA
AACAATGATGTGCGATGGTGTGGTTCCCGGTTGGCGGAGTCTTTGTCTACTTTGGTTGTCTGT
CGCAGGTCGGTAGACCGCAAATGAGCAACTGATGGATTGTTGCCAGCGATACTATAATTCAC
ATGGATGGTCTTTGCGATCAGTAGCTAGTGAGAGAGAGAGAACATCTATCCACAATGTCGAG
TGTCTATTAGACATACTCCGAGAATAAAGTCAACTGTGTCTGTGATCTAAAGATCGATTCGGC
AGTCGAGTAGCGTATAACAACTCCGAGTACCAGCAAAAGCACGTCGTGACAGGAGCAGGCTT
TGCCAACTGCGCAACCTTGCTTGAATGAGGATACACGGGGTGCAACATGGCTGTACTGATCC

ATCGCAACCAAAATTTCTGTTTATAGATCAAGCTGGTAGATTCCAATTACTCCACCTCTTGCG
CTTCTCCATGACATGTAAGTGCACGTAGGAAACCATACCCAAATTGCCTACAGCTGCGGAGC
ATGAGCCTATGGCGATCAGTCTGGTCATGTTAACCAGCCTGTGCTCTGACGTTAATGCAGAA
TAGAAAGCCGCGGTTGCAATGCAAATGATGATGCCTTTGCAGAAATGGCTTGCTCGCTGACT
GATACCAGTAACACCTTTGCTTGGCCGTCTAGCGCTGTTGATTGTATTCATCACAACCTCGTC
TCCCTCCTTTGGGTTGAGCTCTCGAGAACCCTTAATATAACTTCGTATAATGTATGCTATACG
AAGTTATCAAGGAGCATATGCAAGGGTATCTCGACAAGCCTTCACCAACCCCACGCCTCAAA
GTATGCAGAGTTACCGCATCATCACCGACCATCGTCAGCCCAGTCACCGCGTCTTCCTCTTG
GCCTTGTTCTTTttccggatattcagcacccgcttcctctgtatgaccctcacaaaccaggcgtaaatcatttcgcgcatgaggatgagcaccgt
agcaacagccgtcggaacagcacaagcccagctatgacgaattcgtgttagcattcgtccttttgctttctctgcaaatagttCAGAAAACACTT
ACAAGTACCATATCCCGTTGTTAATCTCGGGAAAGCCCTCCGGCGGAAAGTTTTGCCCAAAG
AAACCCGTGATAAACGTCAGGGGCAGGAAGATGATGGTCACCGTTGTCAGTTGCTTCATGCT
CTCATTCTGGTACGCCGAGATTGTGTTGAAAATCAGGTTGATCAAGTTCTCGCTCGACTGCTT
GGACTGCTGCAGCGCCTCCGTAATGATGATGCAGTGGTCCAGCACGTCGCCGAGGTACGTGT
GTGTCATGGGCGTGACGATGACGCCGCTGGCCGGGTTCTCCAGCTCGTTCATGGCCTGGTCG
TGCGTCAGGTTCGTGCGGTGGTCGCGCAGCACGTTGACCAGGTTGTCGATGGGGTTGAGGA
AGCGCAGCATCTTGTTGATTTCGCTGATGCAGATGTACAGGGTCTTGCACTGCTTGATGTTG
GGCGATGTCAGCACGTCCAGCTCCAGGTCGCCGAGCACGTCGTTGTACGCGGCCGTCAGGG
GCACGGCCAGGTCGACGATGGCGTCGATGATGGCCTGCACCAGCAGTGAGGCGTCGCACGA
CTCGCGCAGCACCGTGCCCGGGTTGCTCAGGCGCGTCACAATCGGACGCTCCACGTCGCCG
GCAGACACCTCAAAGAAGGAGATGACGGTATTGTCGGACGTTAGGAAAATGGACACCTGCTC
GGCCGACACTGCCAGGTTGAGCGGCGTCAGAATCGAGTTGGCCTCCATGTACTCGGTGCGGA
TCTCGTTGCCCGAGGCGTGGTATCGCTGCAGCGAGCGCATCATGGACGTCTCCTGAAGCCCC
AGCGTCTCGGACAGCTCCGTCATGCACTTGGTGGGCGCACGCTTCTCGGGGTCGAGCGGGG
TGTTGTTGGTGGCCTTGTCTGCGGCTTTGCTACCAAAGCTGCGGAACCAGCTCTTGAACTTG
GAGGACGACTTTCCCGAGAGGATGCTGCTCGTGTCCGATGAGGAATCGTCGTCGTCCACGAG
ATGGACCAGCTTCTGTAGCGTCATGACAATGAAGGCATGGCTCGGATACCTTTGACAGAGGC
AAATCAgtatggaaggcataaccaaagaattggccgtgtgttaacccaccagtcagccttggtcctgtttctaatgttcatgatgtcctccaacgccag
cttatgcagcccctgtgcgcaccgatcgcctgaatcacatcccaactgagcccattgacgttgatccaccgacactttgcccactcctccttgggctccgtcaga
aacgacacaaacgagtcgttgtcaaagtgcttcttcaccatattatcttttgaaaaatctacgaccgtgatttggcaatccgcactcagcgtcggcatcgaagcgtggc
cgccatcagggagatggggatcaaagcccggctcagaccccggttgccagcctggcctctccgtgtgcgtggcatcgaagtcgaagtcgtcgttgtcatcaatggt
gaggaaagtctggacgcgagcgctcttgcgacggcgcagcggcgaaaagggctcgcgatggacggccgtgtcggagcgcgagaggcgccgtacgtcggcct
ggggaggcgtattgttgttggcgttctcagtgtcgacgctgagatcgtagcggactcggttcagggatggccgtgcgtcggtaggagagacggcgctgttcccttct
tcgtggacgggatccatggtgtctgtgaggcctgtcgagacgagatcgagcgaggagatggcttcgagccagcgcaatcacgctaacagttgcgctctgctttgcg
cgtcgctctcgacagccgcggcgaaagatgctggtgagcaatgacagccaatgctggagccgtcggttggttgagcttccgaaccagcagtcgccagatttcgatct
cacaatggtaacagatacgtcacctcgtcagattcaacggcttcggtcgtcttccatccaagtcgtgtcaggtcccaagagagtttccgtcgccaagaatcaaagata
atatgatcgcattcacggcatcgtcatccaaaacatttcgccatgtttcttgaaacagagcatgagagagggaggagagaaaaaagattgatagataagcccatggt
ataattagaattggaggagcgtcttattgcatgttcgcttcagaacggatcgcagctgccgccccttagcctgggccttagccgcagctgttagggaacaggttccc
tgaaatacattcccgtcagcatgacgtgatacatggcgatgatcaccgcgacccggccaatcacgccgcaaatcagaacatgatgggtcaattggctggagttgctc
cacgctggactgatatcgtcttttctacagtatgtactgtatgtgctcgtgataccagacatgcatctgatcgtcaagaatggatagtatgtacttgtctcgtcatgatcaac
gtctagac_sc*5:919000

F) loxP-site at the* tmus53 *locus

sc*5:927000_caaaacctccttgttgtgcagttggggcagatgaagcgtcgtgacgactgagcccaggggtcattcggcggtgatttttctccagagccgtgc
cagatgtgactcgcgacccgggaatgcgccgacagcgttcgaacgtgaggatgaggccgcccgtaggaactggaacagagacgctcaagcagctccctttcga
gaacccagaccagtgtgaccaacttcttgacgtgaactggagaaatgtcttcagacaagggccgaacccggacgttgaatctgcatgttgagttgagatgcatgacg
aggccctttgtcaaaacgccgaggatccagtctgtcggtgactctatttcgttgtccacaatggtcctgctttgagatgctggactgcgaacccgtacgccaaaccatt
catacatctgcgggctgtctctctttgcatatgtagtgactgaaggctcaacatgccaccattcgtcatccgcagtgaggtactgacgagccttccgcgggtcatatgct
gctgcatatggttgctgatcagaggagtcttcacagattgccacaggttgtccgtatttcagcagtaacccccctgacctcattcagtatggtctcgtacggatcctttgca
tcggcgggacatgcccacggcagctcttcacgacgtccatggtattcgttgaaggtagagtcgtaattgaatcccggcagctcctgtgccacgagaaattcaacctc
gaaactgtaggtagtttgcgggagacggtcgtcatattctggacgcatggtgatgtgttgttacgcgcgggaatacttggacgagatgctgtcaagtcgaggtctttaa
cgattgtgagagctaaacgattgtagagtatacgtctgcggtcggagagtcgtcgcctttcctggtgactgtagtaaaatatgcctgcggggaacgatgtctgtgctg
agtcgaagggagaacggtagaatatagagaaaagtctctagagcctatcaatcgtctcggggtgctccctttacaaaagaatgactgttgatggttgtgtggaagga
acacgttgcctgctgttcacctggcagatgaactccttcctcttgactgcatgttgacactggcgacctgtttctttgagaaggccagtctcggaatctgggacaagga
ctgagtctggttacagccagtgacaggttaccttgcgttgccttgacgggcgagttgcactgaccccgagcgtcatgatgccacgagcaactttgtcgcggcctccat
cctcaccatatcaagccgcgggtgtgctgagatggcaacatactgtacgtcctcggatgagccaggcaaagaagtgacacgtatcatggaattgaaattcagctgct
gaatggtcaacgtcgacaaccttctaggtcagcttctgatggccaaaaaagcacgagcagctgaagcggcatctcccctctgataccaacagcgcggcgtgtatcg
ggcattcatgctgtgccatgtatcatgtagcatccggtatgtggcattgattttacccgtgcgattgtctcaagcagagctcttgagcccccagagatgcccgatggag

caatgaatgcgaaagacgaagctcaagttgcttggctttggagatccgcaggtaacagcaaacaagtcgaggctggccgcggttacataagcagatgcaagctag
gcagattgctttgtag**ACTAATCCCAAAAAGGTGGCGTGC**AGGCAGCAGCTCCGGGCCCACAATTTCTG
GATAGCTGTACCTACCTACTACTAGTAAGGCAATGACTTCACGGCAACTTTGGCCTTGACCCT
TGAATCAGCAGTACTTAAAGCTGCATCATCACAGCAGCGACGCGATTCCTTCACCGTCTCGTT
CATTCAATTGCTTTGTCTGGTCCTGTCACATTTCCCGTACAATTCGAGCCCTTCTGGAAGAAT
TGACAGTGCCAATCTCCTCAGCGGCGATACTCCGTTCTGTCAAGGTCCTTTATAACTTCGTAT
AATGTATGCTATACGAAGTTACAAGGAGCATATGCAAGGGTATCTCGACAAGCCTTCACCAA
CCCCACGCCTCAAAGTATGCAGAGTTACCGCATCATCACCGACCATCGTCAGCCCAGTCACC
GCGTCTTCCTCTTGGCCTTGTTCTTTTTCCGGATATTCAGCACCC<u>**GCTTCCTCTGTATGACCC**
TCACAAAC</u>caggcgtaaatcatttcgcgcatgaggatgagcaccgtagcaacagccgtcggaacagcacaagcccagctatgacgaattcgtgttagcatt
cgtccttttgctttctctgcaaatagttcagaaaacacttacaagtaccatatcccgttgttaatctcgggaaagccctccggcggaaagttttgcccaaagaaacccgt
gataaacgtcaggggcaggaagatgatggtcaccgttgtcagttgcttcatgctctcattctggtacgccgagattgtgttgaaaatcaggttgatcaagttctcgctcg
actgcttggactgctgcagcgcctccgtaatgatgatgcagtggtccagcacgtcgccgaggtacgtgtgtgtcatgggcgtgacgatgacgccgctggccgggtt
ctccagctcgttcatggcctggtcgtgcgtcaggttcgtgcggtggtcgcgcagcacgttgaccaggttgtcgatggggttgaggaagcgcagcatcttgttgatttc
gctgatgcagatgtacagggtcttgcactgcttgatgttgggcgatgtcagcacgtccagctccaggtcgccgagcacgtcgttgtacgcggccgtcaggggcacg
gccaggtcgacgatggcgtcgatgatggcctgcaccagcagtgaggcgtcgcacgactcgcgcagcaccgtgcccgggttgctcaggcgcgtcacaatcggac
gctccacgtcgccggcagacacctcaaagaaggagatgacggtattgtcggacgttaggaaaatggacacctgctcggccgacactgccaggttgagcggcgtc
agaatcgagttggcctccatgtactcggtgcggatctcgttgcccgaggcgtggtatcgctgcagcgagcgcatcatggacgtctcctgaagccccagcgtctcgg
acagctccgtcatgcacttggtgggcgcacgcttctcggggtcgagcggggtgttgttggtggccttgtctgcggctttgctaccaaagctgcggaaccagctcttg
aacttggaggacgactttcccgagaggatgctgctcgtgtccgatgaggaatcgtcgtcgtccacgagatggaccagcttctgtagcgtcatgacaatgaaggcat
ggctcggataccttgacagaggcaaatcagtatggaaggcataaccaaagaattggccgtgtgttaacccaccagtcagccttggtcctgtttctaatgttcatgatgt
cctccaacgccagcttatgcagccccttgtgcgcaccgatcgcctgaatcacatcccaactgagcccattgacgttgatccaccgacactttgcccactcctccttgg
gctccgtcagaaacgacacaaacgagtcgttgtcaaagtgcttcttcaccatattatcttttgaaaaatctacgaccgtgatttggcaatccgcactcagcgtcggcatc
gaagcgtggccgccatcagggagatggggatcaaagcccggctcagaccccggttgccagcctggcctctccgtgtgcgtggcatcgaagtcgaagtcgtcgtt
gtcatcaatggtgaggaaagtctggacgcgagcgctcttgcgacggcgcagcggcgaaaagggctcgcgatggacggccgtgtcggagcgcgagaggcgcc
gtacgtcggcctggggaggcgtattgttgttggcgttctcagtgtcgacgctgagatcgtagcggactcggttcagggatggccgtgcgtcggtaggagagacgg
cgctgttcccttcttcgtggacgggatccatggtgtctgtgaggcctgtcgagacgagatcgagcgaggagatggcttcgagccagcgcaatcacgctaacagttg
cgctctgctttgcgcgtcgctctcgacagccgcggcgaagatgctggtgagcaatgacagccaatgctggagccgtcggttggttgagcttccgaaccagcagtc
gccagatttcgatctcacaatggtaacagatacgtcacctcgtcagattcaacggcttcggtcgtcttccatccaagtcgtgtcaggtcccaagagagtttccgtcgcc
aagaatcaaagataatatgatcgcattcacggcatcgtcatccaaaacatttcgccatgtttcttgaaacagagcatgagagagggaggagagaaaaaagattgata
gataagcccatggtataattagaattggaggagcgtcttattgcatgttcgcttcagaacggatcgcagctgccgccccttagcctgggccttagccgcagctgtta
gggaacaggttccctgaaatacattcccgtcagcatgacgtgatacatggcgatgatcaccgcgacccggccaatcacgccgcaaatcagaacatgatgggtcaat
tggctggagttgctccacgctggactgatatcgtcttttctacagtatgtactgtatgtgctcgtgataccagacatgcatctgatcgtcaagaatggatagtatgtacttg
tctcgtcatgatcaacgtctagac_sc*5:919000

F) hph-marker cassette at the* env1 *locus deleting the PAS domain

sc*26:99000_cttcagctcttgatccggcacgacaagcgtcgagtacggagatggtgcacaaaggacgccggcagctgagaccagcttgaagctatcaatct
acaggtatattacctgacgcagaggtgcagcatccgaattggctacccacctatgcaaaatgtaaaagatccggatggcttcgattcattccgatggtcccatacagc
aggtagagtaccatgagtgatctgaatcatctgagatgcccgatccacgagactcacacacgagttccacccaaggcacggcgcttcccgaaccccttgctgcaat
ctcactctgcatcatctccgtccacgacagaggcacaactcgatggccgtacttgcatactagtcatagcgccgacccttttgtgctacactatgaccatggacaaaaac
cagatgcttgtgcctaggtagttgtacagcactgcgtgttaccaaatgtacttccggctgagacatgcagaggcagctgacgttgctcaggccaggccgaggtgac
aggagatgcatcggagccccagatgctgcaagcctccgtcagccatcatgctgctactactactactactacatgtgcttggcaagcctcagcgacgatgcctta
tcccaggccctggacggagttgtttgtccgccacggcgccaactcgacgctctgttgccagtcggacccctgcgagctggctggcctcggacaattgcacagacg
ctggaaggggggaagaccacgaaaggtagtagtagtggcactctgtgtgctccgcggcatagcacgccttgtctcgaccaggtacgtggatcccgatgctgcgagc
tgcctttgctattttcaattgtagctgctcatggcttgagatgcagcccatcaccacagggagcgtgcgggaggtctttcgaaggcttcagcatcacacaagcagtcaa
agcctagtcgctgcaaacagcctgatacaaagctactactatgtacgatgattgctgcccagcaaacgcctctttctgctctgctgtgtgtctgttccgggtcagcctga
gacagatcgacgaatcatgactattactactgggttgatgctggccaagaagactgccttattattgctccgtgcggatcagtagcacaaatcgacaatactagcagct
gtcttttcaagatgagtgctgcgtgggcaatcaggcaatgttgcccaagccggatcccggcttgcaaactgcggccaacatacgattattggcattgcccacggcgc
aacgccgcattcctgcgccctcgtccgcgactcaccatcgcgtgtccacaagtctagaaccacgaggtccccttctctcccgtgctgcgggtttccaacacactgact
gaggcgtcatggccgtcgggacaccagctgcccgaagcagtcaaagacatggacatggacatggacccagctaatgtacagtcctggatgtatgtacctgccgg
gctgcatgtgggcattgaccgaccggcatcagaagagactgggattctcggatcgcctttcatcagctccagcaatccagccgatcgtcgtccggtcccaatctgcc
gccatcgggccaatcagcagggtccgcgtcgacccctggg**gctggactttgccgctaacacg**atct**GGCGTGTGTCTGTCGCCGCCCG**
CGATAGTGCGCTAACACGCGACTAACACCATGTTCGCCCTCGATGACGGCCAGAGCCTCCCG
CGGCTTCGATCTGAGATCGCTGCCCTGTATCGATGATGTCCAGCCTTCTGCCAGCCTTTCCAG
CCCTCTTGCCCGGCCTCCGCTCTGCTGCCCCTGCCTGGAATCTGCACCTGCTAGAGCCCTTTT

CTGCCAACTCAAATCGCGCCATGGCTTGCGTCAGAATCAAACGCCCCCCCCTCGTGTCCATG
TGTCGTGTCCGAAAGGCCCAGAGAAAAGGCCCCGTTGAGATCCTATCCGGCGCCAGCAAGGA
GCAGGAGCAAACAGCAGGGAGCTGCTATCTGGGACAGTCAAGTTCTTTGCCTCTGCTTATGC
TTGCTTACCTAGGTATCCTTTTGCACGCAGCCGCTCAGATTCGTCCCGAGTAACCGCCTTTCC
ATGTCTTGTGCTCTGCTGCTCTCACCCTCCCCCCCAAAATTGAACTCTGGGTCTGCAATATCC
ACTTTGTTTGCTTCCGCATATCTCCAGTTTCTATCCCTGGATCTGGATACGGGCTTGATCACC
TTGACCATGGCTGCTATTTAGTACCTCTTGTCCCTTTACTCTGTACTCTCTCTACCTGCCTCGT
CCTCTCTATCCTTTGAGCTTTTCTCAACGTCCTGGCTGTGACCTCTTCTTTACATAGAGTGGC
CCAACCCTTGCCGCTTGGCCCCAAGAGATTGCCACCCAGCCGGCGTTGACATTAACCCCCCC
AATGCTCGAAGCACCAATCATGGTATCTTCGGGTTCTTCAGACCTGCCACTGCAGCTGAATCC
CTGGGAAGCCCAAGCTCTAAACGTAAGTCACCCGGTCCTCAGCCGCGTTGCCTTGGTTTTGA
ATGCAGACGGAACCCGAGATTAACTCATCATGCCATCAGTACGGGTTTCCACAAGAACCTTC
TACCAATGGCCATGCGGGAAGCAACAGCAACAACAACAACAATACCCCGGCTTGGATACAGC
TTCCCGACCCCATCATATATCCTGGCATATATTCAGCTAGCGGCATCGATGTGATGGGTATAC
TGGTAAGCATAACATCTCTCTCGCTCTTGCTTAACCGTTGCTCGATCTTAAGCCATGCTAGAC
ATTGCAACGCTAACAGTTCTATGCCTAGTTACGACTCGAGATAACGGTGAGACTAGCGGCCG
GTCCCTTATCCAGCTGTCCACGTTGgcctgccctcagttagcgctcaactcatgccctcactggcgaggcgagggcaaggatgga
ggggcagcatcgcctgagttggagcaaagcgccgccatgggagcagcgaaccaacggagggatgccgtgctttgtcgtggctgctgtggccaatccgg
gcccttggttggctcacagagcgttgctgtgagaccatgagctattattgctaggtacagtatagagagaggagagagagagagagagtgtgagaggga
aaaggtgaggttgaagtgaggtagtaagtgAAAAAAAAAAAAAAAATCAACACTGACGGCTGCCGTCTGCCACC
CTCCTCCACCCAGCCACCTGCACACTCAGCGCGCAGCATCACCTAATCTTGGCTCGCCTTCC
GCAGCTCAGGTTGTTTTTTTTTTTTCTCTCTCCTCGTCGAAGCCGCCCTTGTTCCCTTATTTAT
TTCCCTCTCCTACCTTGTCTGCCTTTGGTCCATCTGCCCCTTTGTCTGCATCTCTTTTGCACGC
ATCGCCTTATCGTCGTCTCTTTTTTCACTCACGGGAGCTTGACGAAGACCTGACTCGTGA<u>GCC</u>
<u>TCACCTGCTGATTTCTC</u>TCCCCCCCTCCCGACCGGCTTGACTTTTGTTTCTCCTCCAGTACCT
TATCGCGAAGCCGGAAGAACCCTCTTAACCTCTAGAATGAAAAAGCCTGAACTCACCGCGAC
GTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGG
AGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTA
AATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCG
CTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTC
CCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGC
AGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTT
CGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTCATATGCGCGA
TTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTC
GCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGT
GCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTG
ACTGGAGCGAGGCGATGTTCGGGGATTCCC<u>AATACGAGGTCGCCAACATC</u>TTCTTCTGGAGG
CCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGC
AGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCT
TGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGA
TCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCG
ATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCA
AAGGAATAATGCATGGCTTTCGTGACCGGGCTTCAAACAATGATGTGCGATGGTGTGGTTCC
CGGTTGGCGGAGTCTTTGTCTACTTTGGTTGTCTGTCGCAGGTCGGTAGACCGCAAATGAGC
AACTGATGGATTGTTGCCAGCGatactataattcacatggatggtctttgcgatcagtagctagtgagagagagagaacatctatcca
caatgtcgagtgtctattagacatactccgagaataaagtcaactgtgtctgtgatctaaagatcgattcggcagtcgagtagcgtataacaactccgagt
accagcaaaagcacgtcgtgacaggagcaggctttgccaactgcgcaaccttgcttgaatgaggatacacggggtgcaacatggctgtactgatccatc
gcaaccaaaatttctgtttatagatcaagctggtagattccaattactccaccTCTTGCGCTTCTCCATGACATGTAAGTGCAC
GTAGGAAACCATACCCAAATTGCCTACAGCTGCGGAGCATGAGCCTATGGCGATCAGTCTGG
TCATGTTAACCAGCCTGTGCTCTGACGTTAATGCAGAATAGAAAGCCGCGGTTGCAATGCAA
ATGATGATGCCTTTGCAGAAATGGCTTGCTCGCTGACTGATACCAGTAACACCTTTGCTTGGC
CGTCTAGCGCTGTTGATTGTATTCATCACAACCTCGTCTCCCTCCTTTGGGTTGAGCTCTTTG
GATGGCTTTCCAAACGTTAATAGCGCGTTTTTCTCCACAAAGTATTCGTATGGACGCGCTTTT
GGCTGTATTGCGTGAGCTACCAGCAGCCCAATTGGCGAAGTCTTGAGCCGCACTCGCATAGA
ATAATTGATTGCGCATTTGATGCGATTTTTGAGCGGCTGTTTCAGGCGACATTTCGCCCGCCC
TTATTTGCTCCATTATATCATCGACGGCATGTCCAATAGCCCGGTGATAGTCTTGTCGAATAT
GGCTGTCGTGAATAACCCATCGGCAGCAGATGATAATGATTCCGCAGCACAAGCTTGATGGG
ATAGATGCTAGGCGTACCTCGAAGGCCTCCGAGGGAACCCTTCTTCTCATTCTGCATTTTAAC
ATCTGCATCCGCTTCCCTGGCAACCACTGTCAGAGAAATACCACGCCAGCTCGACGCTGTTT

GCTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGACAGCCAGCACTCTGCCACTCTACCCTGGACA
TAACATTTGTATTCGGTGCACTCGAGATTTCATCAAGGATCGTACATACAGTCACACATGAGG
TCTGTAGTTCGTTTTCCCCATCGAGATGCCGTGTCGGCACCGGCATCCCGACAACTGACGGT
CAACCCTGGAGCTCGATACCACAGACTTTGATTTTCTTGCAGCAGCCAGGAGAAGAGGGCGG
CGCACACGATTATGATTTGCATGATATTCTAGCATGGTTTTTGTTTCCCCTTCCAGCGATTTA
GCCTAGCTGCGCGGTTTTACATTGGAGTTCTTTCAAACAATTCTACGGGTGGTACACGAAAA
GGGTTTTGGGAGATATCCAATAGATGCGTACACGATAGTTGTGTGGAAAGACCAATATTTGC
AATCAGTTTGCCAAGTAGATGAGCCATACATATGCCACGCTCTGCGAAACACAATATACGAA
AAAAAGCTTGCTCTGTACAGGGGAACCTTCATCATCAGACGCTTGTGCCAACAGTACGTGCC
ATCACTTCCTCAGCCGAGTAGATCGACGCCGTCCTTCTTCCACCACCTCCTCTTTCACAAGAG
CAGCTTTTGGTTTCTCCGTCGCTCCTTTGCCCCGTCTCTTTTTAGCAGGCGGACTCACGTCAT
CCGGCTCATCAGCCTCTTCCCTGATTGCCACCTTTCGCGCTCTCTTGGGCTTCTTCTCCTCAA
CCACCTCTTCTTCCTCTCCCTCATTGTCCTCTTGCTTCAACTCGTTCTTCACGGCCTTGCTCTT
CTTCCCTGATCCGCCCTTTTCAACAAGGGGCTCCTCCTTTGCGCCAGAAACAACCGCGCCAG
TCTTCTTCTGCACGCCCGGCGCATAACAGCTCGTTCTCCCGCCCACCGTAATGAATGCAATCT
TCTCTCCGTTTGGCAGCTTCGTCGCCGCGCCCTTGTCGCCCTTTCCCCAGCGGTGCTTGAAG
AGCCAGTGATCCGGGAACTGGTCCGAGTCGCCCAGCTTGTCGACggccagctggcagacgtaccgaatggcc
tcgtgcagccgcttgatttccgcagcggagaatgtatccgagtactgctcgggatggagcttggcctggtagagcgtctcgtcggcgacccagttgccgatgccg
ctgatcatggcctggtcgagcaggaacgccttgacggggacgtggcggttcttcatcttgtcccgcaggtaggtttccgtgaagatgtccttgtcgacgacggggtc
ggggccgttctccacgagggggagtgctggcggatggcgtcgccgggacagtcgacgaggcggacgcggccgaagcggcgggcgtcggtgaaggcgact
tggacggggtcgccgtctttggcgtcggtggtgaggtggaacttccagaacttgggggccactgggcgatgtcgctttctttcattttcttgtagtagtttgtgtaggc
ggttcgttcgccgttgatgtggacccagcctggacgttgaggggtgaggatggttggcctagaggcttccaggttcatgttgtctgagtgtggacttacccgtcatgc
caaagtgcattaccagatgtggcggtttgtcaaatgttatcctttgaaag_sc*26:94000

AN ACCURATE NORMALIZATION STRATEGY FOR RT-QPCR IN *HYPOCREA JECORINA* (*TRICHODERMA REESEI*)

MATTHIAS G. STEIGER, ROBERT L. MACH AND ASTRID R. MACH-AIGNER

Journal of Biotechnology 145 (2010) 30–37

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*)

Matthias G. Steiger, Robert L. Mach, Astrid R. Mach-Aigner*

Gene Technology, Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, TU Wien, Getreidemarkt 9/166/5/2, A-1060 Wien, Austria

ARTICLE INFO

Article history:
Received 20 July 2009
Received in revised form 12 October 2009
Accepted 16 October 2009

Keywords:
Trichoderma reesei
Reference gene
RT-qPCR
Gene expression
Transcript analysis
Normalization

ABSTRACT

Hypocrea jecorina is an important, filamentous fungus due to its effective production of hydrolytic enzymes. Gene expression studies provide deeper insight into environment sensing and cellular response mechanisms. Reverse transcription-quantitative PCR is a gene-specific and powerful tool to measure even minor changes in mRNA composition. An accurate normalization strategy is absolutely necessary for appropriate interpretation of reverse transcription-quantitative PCR results. One frequently applied strategy is the usage of a reference gene. Adequate reference genes for *Hypocrea* have not been published so far. By using the NormFinder and geNorm softwares, we evaluated the most stable genes amongst six potential reference genes in 34 samples from diverse cultivation conditions. Under those experimental conditions, *sar1* encoding for a small GTPase was found to be the most stable gene, whereas *act* encoding for actin was not amongst the best validated ones. The influence of the reference system on the expression data is demonstrated by analysis of two target genes, encoding for the Xylanase regulator 1 and for Xylanase II. We further validated obtained xylanase 2 transcription rates with the corresponding enzyme activity.

1. Introduction

Hypocrea jecorina (anamorph: *Trichoderma reesei* (Kuhls et al., 1996)) is a filamentous ascomycete widely used because of its high secretory capacity for hydrolyzing enzymes. Hydrolases secreted by this fungus are used in a broad range of industrial applications covering, for instance pulp and paper industry (Buchert et al., 1998; Welt and Dinus, 1995; Noé et al., 1986), food and feed industry (Galante et al., 1993; Walsh et al., 1993; Lanzarini and Pifferi, 1989), and textile industries (Koo et al., 1994; Kumar et al., 1994; Pedersen et al., 1992) as well as biofuels and bioenergy (Himmel et al., 2007; Hahn-Hägerdal et al., 2006; Ragauskas et al., 2006).

Detailed information on the regulatory mechanisms of hydrolase expression in *Hypocrea*, will enable the systematically engineering of existing production strains in order to increase their expression efficiency. Transcription analysis is the basis for the understanding of regulatory mechanisms governing gene expression. One powerful tool to measure the mRNA content is reverse transcription-quantitative PCR (RT-qPCR). Major advantages of this method are its high sensitivity, large dynamic range, and accurate quantification (Huggett et al., 2005). However, an accurate and robust normalization system is needed when performing relative quantification of qPCR data. Normalizing to a stably expressed gene of the target organism, often called reference or housekeeping gene, is a powerful method for internal error prevention. In general, when applying mRNA quantification techniques, an error is caused by the multistage process required to extract, process (i.e. *in vitro* reverse transcription) and detect mRNA.

A major challenge is to find a suitable reference gene. Even genes like β-actin, believed to be stable, turned out to be inapplicable under certain conditions (e.g. comparing different mouse tissues) (Barbu and Dautry, 1989). In *Hypocrea*, as in other filamentous fungi, only a single gene encoding for actin is described (Matheucci et al., 1995). So far, only less is known about its suitability as a reference gene in fungi. A study performed in *Saccharomyces* reveals that actin scores at the third best position when tested with NormFinder and geNorm (Stahlberg et al., 2008). In *Aspergillus niger* an investigation done by means of geNorm demonstrated that under the tested conditions the actin encoding gene belongs to the most stable ones (Bohle et al., 2007). An important requirement for such a reference gene is its robust expression under all conditions applied in a certain experiment. Therefore, a careful evaluation is necessary to obtain one or even more suitable genes for normalization. Vandesompele and co-workers suggest the use of multiple reference genes rather than relying on a single one (Vandesompele et al., 2002). However, it depends on the experimental setup how many reference genes have to be included in order to obtain consistent and reliable results. Especially for the

* Corresponding author. Tel.: +43 664 60588 7253; fax: +43 1 58801 17299.
E-mail address: aaigner@mail.tuwien.ac.at (A.R. Mach-Aigner).

0168-1656/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2009.10.012

setup of new experimental conditions, it has to be evaluated if a certain set of reference genes is suitable or not. Different approaches have been published, which allow the evaluation of multiple reference genes (Andersen et al., 2004; Pfaffl et al., 2004; Vandesompele et al., 2002). The software geNorm ranks the genes according to the similarity of their expression profiles by a pairwise comparison (Vandesompele et al., 2002). The authors calculated the average pairwise variation of a particular gene with all other control genes and denominated it M, the internal control gene-stability measure. Within this system, genes with the lowest *M* values have the most stable expression. Another calculation method is used by the program NormFinder (Andersen et al., 2004), which uses a model-based approach for the estimation of expression variation. The advantage of this algorithm compared to geNorm is that it can deal with systematic differences in the data set like different tissues or strains.

The relative expression software tool (REST; version: 2008) developed by Pfaffl and co-workers allows the estimation of gene expression using qPCR amplification data (i.e. the threshold cycle values). This software enables the measurement of uncertainty in expression ratios by introducing randomization and bootstrapping techniques. Confidence intervals for expression levels allow measurement of not only the statistical significance of deviations but also of their likely magnitude even in the presence of outliers (Pfaffl et al., 2002).

In this study we analyzed six potential reference genes with geNorm and NormFinder in order to find suitable normalization conditions for *H. jecorina*. With REST software we calculated the expression profiles of two target genes (*xyn2*, *xyr1*) applying different combinations of reference systems. *xyr1* encodes for the Xylanase regulator 1 (Xyr1), which is the main activator of hydrolase expression in *H. jecorina* (Stricker et al., 2006, 2007). Xyr1 itself is regulated at the transcriptional level by a repression/derepression mechanism. The constitutive expression of *xyr1* leads to a significant elevation/deregulation of hydrolytic enzymes encoding genes transcription in comparison to what is observed in the parental strain (Mach-Aigner et al., 2008). *xyn2* encodes for Xylanase II (XYNII) and this enzyme is responsible for >50% of xylanolytic activity in the culture filtrate of *H. jecorina* grown on xylan (Törrönen et al., 1992, 1993, 1994). In this study, we illustrate the normalization strategy choice impact upon the relative gene expression output. Examples of results obtained by using stable and unstable reference genes for transcript analysis are highlighted. Furthermore, we evaluated the normalization system by a correlation analysis of transcription rates with enzyme activity.

2. Materials and methods

2.1. Strains

The following *H. jecorina* (*T. reesei*) strains were used throughout this study: QM9414 (ATCC 26921, a cellulase hyperproducing mutant derived from wild-type strain QM6a (Mäntylä et al., 1992)), nx7 strain (a QM9414 recombinant strain constitutively expressing *xyr1* under the *nag1* promoter control) (Mach-Aigner et al., 2008), Δ*ace2* strain (a QM9414 *ace2* deletion strain) and Re*ace2* strain (an *ace2* retransformation strain of the Δ*ace2* strain) (Stricker et al., 2008). All strains were maintained on malt agar.

2.2. Growth conditions

For replacement experiments mycelia of the strains QM9414 and nx7 were pre-cultured in 1-liter Erlenmeyer flasks on a rotary shaker at 30 °C for 18 h in 250 mL of Mandels–Andreotti (MA) medium (Mandels, 1985) containing 1% (w/v) glycerol as the sole carbon source. Conidia (final concentration, 10^8 per liter) were used as the inoculum. Pre-grown mycelia were washed and thereafter equal amounts were resuspended in 50 mL MA medium containing 1% (w/v) glucose or D-xylose, or 1.5 mM xylobiose, or a medium without a carbon source. Incubation was continued at 30 °C and shaking (250 rpm). 15-mL samples were taken after 3, 5, and 8 h (Mach-Aigner et al., 2008).

Cultivation of Δ*ace2* and Re*ace2* strain in a bench top bioreactor (Applikon Biotechnology, Schiedam, Netherlands) was carried out using 1 L medium adjusted to pH 4.5 comprising 20 g beech wood xylan (Lenzing AG, Lenzing, Austria), 2.8 g $(NH_4)_2SO_4$, 1 g $MgSO_4 \cdot 7H_2O$, 4 g KH_2PO_4, 0.5 g NaCl, 0.5 g Tween 80, 0.1 g peptone, 5 mg $FeSO_4 \cdot 7H_2O$, 1.7 mg $MnSO_4 \cdot H_2O$, 1.4 mg $ZnSO_4 \cdot 7H_2O$, and 2 mg $CaCl_2 \cdot 2H_2O$ in distilled water. Some drops glanapon (Becker, Vienna, Austria) were added to the medium to avoid excessive foam formation. For 1 L, 10^8 conidia were used as inoculum. Cultivation was performed at 30 °C, pH 4.5, 0.3 l^{-1} min^{-1} aeration rate and 500 rpm agitation rate. Timing of sample drawing was planned according to the expected time points of *xyn2* transcript formation. 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). Computer-aided process control and monitoring was performed using the LIME Process Control software (ATS, Vienna, Austria) (Stricker et al., 2008).

All mycelia samples were harvested using a textile filter (Miracloth, Calbiochem, USA), washed and flash-frozen in liquid nitrogen.

2.3. Xylanase enzyme assay

Endo-xylanase activity was measured applying Xylazyme AX Tablets (Megazyme, Wicklow, Ireland) according to the manufacturer's instructions at a pH of 4.7 which is within the range of the pH optimum of Xylanase II (pH optimum 4.5–5.5) but not of Xylanase I (pH optimum 3.5–4.0) (Tenkanen et al., 1992). One Unit of activity is defined as the amount of enzyme required to release one micromole of xylose reducing-sugar-equivalents per minute under the defined assay conditions (Stricker et al., 2008).

2.4. RNA-extraction and reverse transcription

Harvested mycelia were homogenized in 1 mL peqGOLD TriFast DNA/RNA/protein purification system (PEQLAB Biotechnologie, Erlangen, Germany) using a FastPrep FP120 BIO101 ThermoSavant cell disrupter (Qbiogene, Carlsbad, USA). DNA and RNA were simultaneously isolated in a two-step-process according to the manufacturer's instructions. Total RNA quantity was determined by means of UV-VIS-spectrophotometry. Quality of extracted RNA was determined as the 260 nm/280 nm ratio. Samples with a ratio between 1.8 and 2.2 were further processed immediately (Pfaffl et al., 2004). Synthesis of cDNA from mRNA was carried out applying RevertAidTM H Minus First Strand cDNA synthesis Kit (Fermentas, Burlington, Canada) according to the manufacturer's instructions. 0.5 μg of each mRNA sample was used for reverse transcription. cDNA samples were diluted 1:100 prior to qPCR.

2.5. Selection of reference genes and primer design

Candidate reference genes were selected according to previously published reference gene analysis performed in *Aspergillus niger* (Bohle et al., 2007) and *Hypocrea atroviride* (Pucher, 2006). Corresponding genes in *H. jecorina* were found by means of a tblastn analysis (standard settings) using the published genome sequence

Table 1
Genes of reference (RG) and target genes (TG) in *H. jecorina*.

Gene	Location (*H. jecorina* genome)	Predicted/known function	Usage
zwf1	Trire2/scaffold_4:393661-395884	Glucose-6-phosphate 1-dehydrogenase	RG
tef1a	Trire2/scaffold_6:764788-767286	Translation elongation factor 1a	RG
sar1	Trire2/scaffold_9:389249-390386	SAR/ARF type small GTPase	RG
glk1	Trire2/scaffold_17:541672-543636	Glucokinase	RG
cox4	Trire2/scaffold_17:758104-759394	Cytochrome c oxidase subunit IV	RG
act	Trire2/scaffold_1:3249300-3251565	Actin	RG
xyr1	Trire2/scaffold_11:201774-204723	Xylanase Regulator 1	TG
xyn2	Trire2/scaffold_27:125447-126445	Xylanase II, Glycoside Hydrolase Family 11	TG

Table 2
Primer sequences and amplification properties of genes used throughout this study.

Gene	Forward primer (5′ → 3′)	Reverse primer (5′ → 3′)	Amplicon size (bp)	Primer optimized elongation temperature (°C)	Usage
zwf1	CCACACCCAGCCCCAAAC	TTGACCCGTCGCAGGAACT	198	62	RG
tef1a	TCGACAAGCGTACCATTGAGAAG	GCCGGGAGCGTCAATGA	182	64	RG
sar1	TGGATCGTCAACTGGTTCTACGA	GCATGTGTAGCAACGTGGTCTTT	115	64	RG
glk1	AAGCCAACGCGGGATTAAAG	GTGCCGCCCAGGTCTACAG	118	63	RG
cox4	CGCCAGTCTCGCCCTCTT	TCGGCGTTTTGTTGCTTGAG	114	60	RG
act	TGAGAGCGGTGGTATCCACG	GGTACCACCAGACATGACAATGTTG	103	64	RG
xyr1	CCCATTCGGCGGAGGATCAG	CGAATTCTATACAATGGGCACATGGG	97	64	TG
xyn2	GGTCCAACTCGGGCAACTTT	CCGAGAAGTTGATGACCTTGTTC	75	64	TG

version 2.0 at the U.S. Department of Energy Joint Genome Institute homepage (http://www.jgi.doe.gov). Table 1 lists all selected reference genes and the two target genes and their putative or known physiological function.

Either the forward or reverse primer was placed across an exon–exon junction. The primer pairs presented in Table 2 produced a single amplification product. This was checked by dissociation (or melting) curve analysis and gel electrophoresis. Raw expression data of the six selected reference genes were measured in 34 samples. Two different experimental setups were used: a replacement experiment (24 samples) and a bioreactor cultivation approach (10 samples). These different conditions were chosen to ensure a wide range in sample spectrum.

2.6. Quantitative PCR

All PCRs were performed in a Mastercycler® ep realplex 2.2 system (Eppendorf, Hamburg, Germany). The Mastercycler® ep realplex 2.2 software was used to compile PCR protocols and to define plate set-ups. All reactions were performed in triplicates. qPCR amplification mixture (final volume 15 μL) contained 7.5 μL 2 x iQ SYBR Green Mix (Bio-Rad Laboratories, Hercules, USA), 100 nM forward and reverse primer and 2 μL cDNA (diluted 1:100). Primers sequences are given in Table 2. Each run included a no-template control and a no-amplification control (0.015% SDS added to the reaction mixture). The cycling conditions comprised 3 min initial denaturation and polymerase activation at 95 °C, followed by

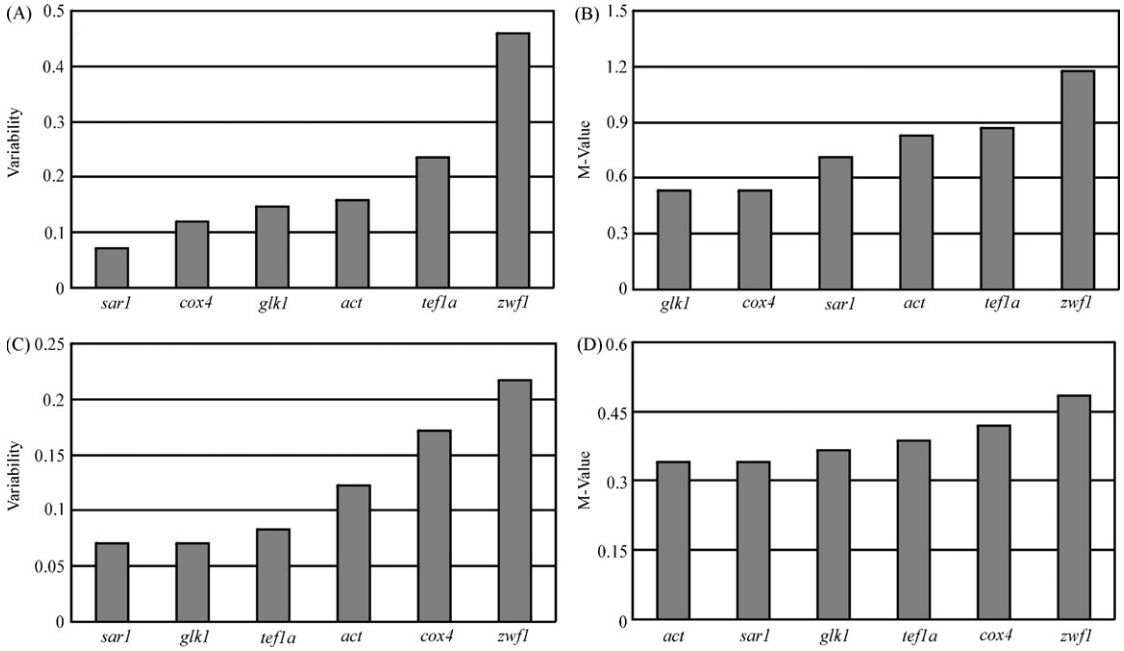


Fig. 1. Identification of reference genes in *H. jecorina*. Analysis of the stability of six different reference genes in 34 different samples calculated with NormFinder (A and C) and geNorm (B and D). Two experimental setups were analyzed: 10 samples of a bioreactor approach (A and B) and 24 samples of a replacement experiment (C and D). The lower the variability (NormFinder) or the *M*-value (geNorm) is for a certain gene, the more stable it is when evaluated by the corresponding algorithm (NormFinder or geNorm).

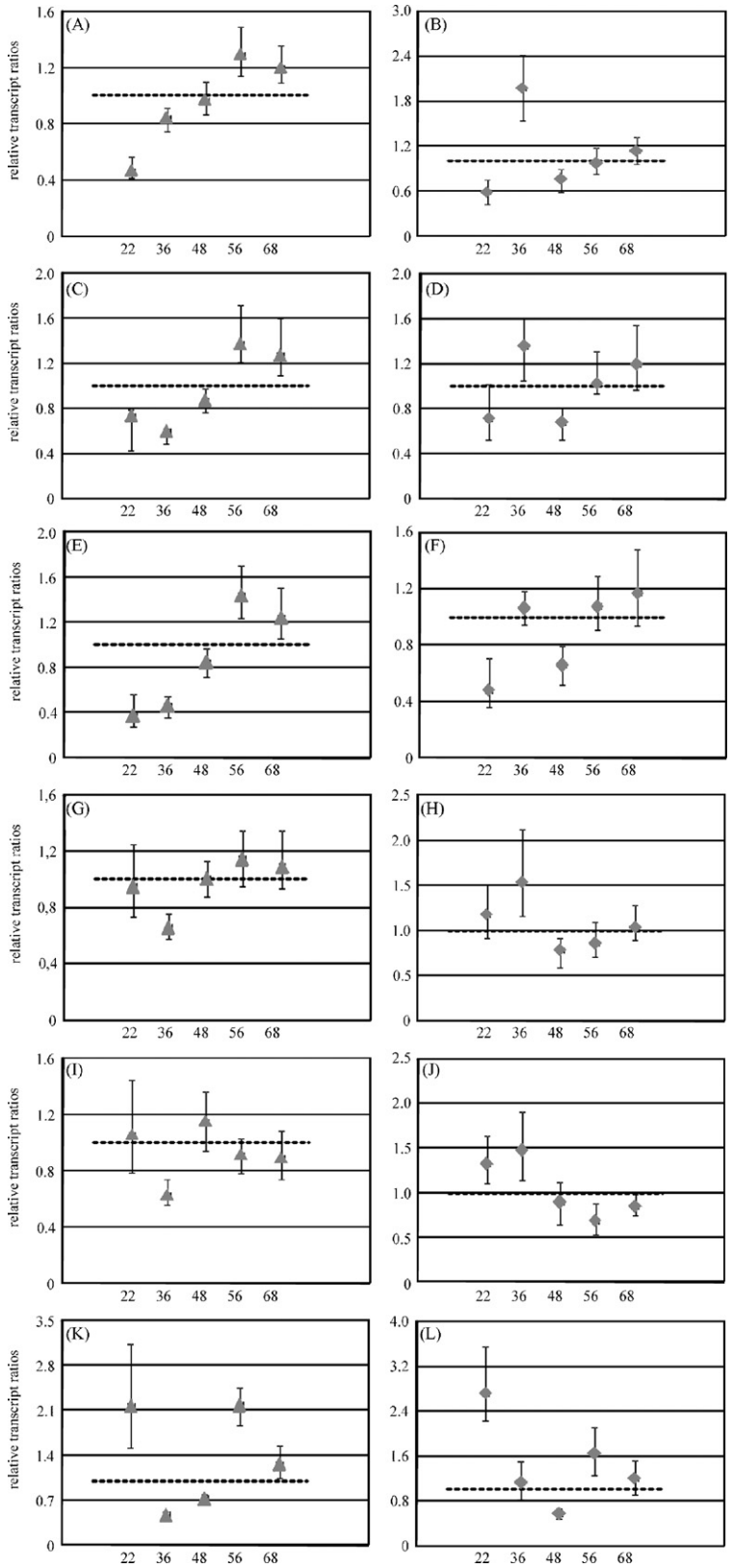


Fig. 2. Comparative transcript ratio analyses of samples from a bioreactor cultivation using different normalization strategies. Transcript ratios of *xyn2* (A, C, E, G, I, K) (triangle) and *xyr1* (B, D, F, H, J, L) (diamonds) from a bioreactor experiment on beech wood xylan, comparing a *H. jecorina* R*ace2* strain (control) with a Δ*ace2* strain (sample), were calculated with REST 2008. The dashed line (ratio of 1) indicates equal transcript amounts in both strains. Values above 1 indicate more transcripts in the R*ace2* strain compared to Δ*ace2* strain and values beneath 1 indicate fewer transcripts. Five cultivation time points (22, 36, 48, 56, 68 h) were analyzed. Different (combinations) of reference genes were applied: *sar1*/*glk1*/*cox4*/*act*/*tef1a* (A and B); *sar1*/*cox4* (C and D); *sar1* (E and F); *cox4*/*glk1* (G and H); *glk1* (I and J); *zwf1* (K and L). Error bars indicate a 95% confidence interval and stress the semantic significance of the data.

50 cycles of 15 s at 95 °C and 120 s at the optimized elongation temperature for each primer pair (Table 2). PCR efficiencies were calculated by linear regression of efficiency (LRE) analysis (Rutledge and Stewart, 2008a; Rutledge and Stewart, 2008b). All qPCR efficiencies were above 90%.

2.7. Data analysis

The two experimental setups, a bioreactor cultivation and a replacement experiment, were separately calculated. Threshold cycle (Ct) values were determined by Mastercycler® ep realplex 2.2 software (settings: threshold 100, automatic baseline). For geNorm and NormFinder analysis, Ct values were transformed to quantities via the comparative Ct method. For NormFinder analysis the data sets were divided into two subgroups. Each strain is defined as a subgroup. Applying the softwaretool REST 2008, Ct values and PCR efficiencies were used to calculate relative expression quantities (number of randomization: 10,000) of *xyr1* and *xyn2* by different normalization systems. Bootstrapping techniques applied by REST 2008 were used to provide 95% confidence intervals for expression ratios. The confidence interval provides a range that can be checked for semantic significance.

3. Results and discussion

3.1. Identification of stable reference genes

We applied in parallel two different gene-ranking algorithms to calculate the most stable reference gene. One is used by geNorm, the other implemented in NormFinder software.

We compared two experimental setups: a replacement and a bioreactor cultivation approach. Under bioreactor growth conditions, NormFinder ranked *sar1*, *cox4* and *glk1* as the most stable genes (Fig. 1A) and geNorm indicated *glk1*/*cox4* and *sar1* (Fig. 1B). Although both calculation methods resulted in the same best three genes, the order is not the same. *sar1* scores better when applying NormFinder algorithm. As the best combination of genes, NormFinder proposes *sar1* and *cox4* (Fig. 1A).

Under replacement cultivation conditions *sar1*, *glk1* and *tef1a* performed best using NormFinder (Fig. 1C), whereas geNorm ascertained *act*/*sar1* and *glk1* (Fig. 1D). Generally, NormFinder and geNorm produce similar results. Only *act* scores notably better when applying the geNorm calculation (Fig. 1D). As the best combination of genes, NormFinder suggests the use of *sar1* and *glk1* (Fig. 1C).

sar1 seems to be highly stable under all conditions tested (Fig. 1A, C, and D). This is in good accordance with a previous reference gene analysis from *Aspergillus niger* (Bohle et al., 2007). Only when the geNorm algorithm is applied to the samples of the bioreactor cultivation experiment, it is not ranked at the first position (Fig. 1B). Exactly in this case we will later on demonstrate that the NormFinder rating is more accurate. Evidence for this is given in the next section. In our study *zwf1* is always classified as the least reliable reference gene (Fig. 1A–D). Remarkably *act*, which is widely used as a reference gene, both in Northern blot and RT-qPCR, is not always among the most stable genes (Fig. 1A–C).

3.2. Evaluation of reference genes

With NormFinder and geNorm we have ranked the candidate reference genes. In order to demonstrate the suitability of those rankings we measured the transcription of two target genes. *xyr1* (Xylanase regulator 1-encoding) and *xyn2* (Xylanase II-encoding) were selected because their transcription profile under the applied cultivation conditions is well known and characterized (Stricker et al., 2008). Thus they serve as a perfect model for the evaluation of the reference genes.

In the first example, the reference system is based on the five most stable genes (Fig. 2A and B). *zwf1* is excluded because it is too unstable (compare Fig. 1 A–D). The idea of using multiple reference genes for normalization is based on the assumption that the variation in a single gene is higher than the variation in the average of multiple genes. The next reference systems used are *sar1* and *cox4*, because NormFinder classifies them as the best combination of genes (Fig. 2C and D). In the third set, the normalization is based on *sar1* only (Fig. 2E and F). This gene was found to be the most stable one when using NormFinder. These first three data sets show the same trend for *xyn2* transcript formation (Fig. 2A, C, and E) and for *xyr1* transcript formation (Fig. 2B, D, and F). At early cultivation time points, less *xyn2* transcript is found in the Re*ace2* strain compared to the Δ*ace2* strain. At late sampling time points a trend reversal is observed: the Re*ace2* strain has more *xyn2* transcript than the Δ*ace2* strain. These results are in good accordance with the previously published transcription levels for *xyn2* and *xyr1* (Stricker et al., 2008).

In the next experiment, *cox4* and *glk1* were studied as reference system (Fig. 2G and H) because they were classified by geNorm as the most stable genes (Fig. 1B). The trend observed in this data set differs substantially from the first three reference systems presented. In case of *xyn2* transcription ratios, only at one measuring point (36 h) they are significantly down-regulated in the Re*ace2*

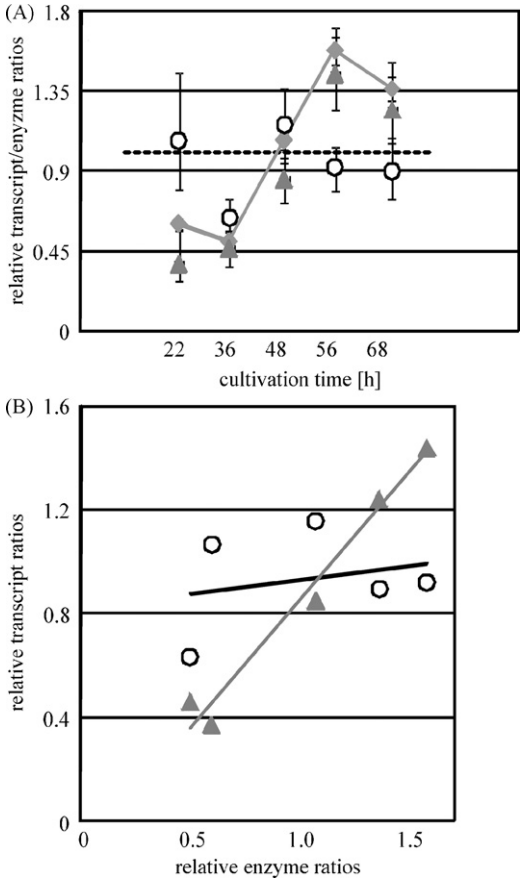


Fig. 3. Correlation of *xyn2* transcript ratio with xylanase enzymatic activity ratio comparing a *H. jecorina* Re*ace2* strain with a Δ*ace2* strain. (A) *xyn2* transcript ratios of the bioreactor experiment on xylan are normalized either with *sar1* (triangle) or with *glk1* (cycles) and are presented together with the volumetric activity ratios of xylanase (diamonds). (B) Linear correlation of relative *xyn2* transcript ratios with xylanase enzymatic activity ratio. Normalization was performed against *sar1* (triangles; $R^2 = 0.97$) or *glk1* (cycles; $R^2 = 0.06$).

strain compared to the Δ*ace2* strain. All other measuring points indicate no significant difference between the compared strains. These results are mainly caused by the influence of *glk1* when used for normalization. This fact is illustrated when using *glk1* as single reference gene (Fig. 2I and J). Careful analysis of NormFinder results reveals that in the Δ*ace2* strain-subgroup the intragroup variation for *glk1* is much higher than in the Re*ace2* strain-subgroup (data not shown). Assuming that both strains have the same expression profile for a certain gene, the gene variation would be the same in both strains. It seems that *glk1* is differentially regulated in the Δ*ace2* strain compared to the Re*ace2* strain. Thus, *glk1* is not suitable as a reference gene. Further evidence is provided below.

The last two graphs depict the results using the most unstable gene *zwf1* as reference (Fig. 2K and L). This clearly demonstrates the importance of choosing a stable reference gene. Otherwise inconclusive data like the results for the first time points might occur, leading to misinterpretation of data (Fig. 2K and L).

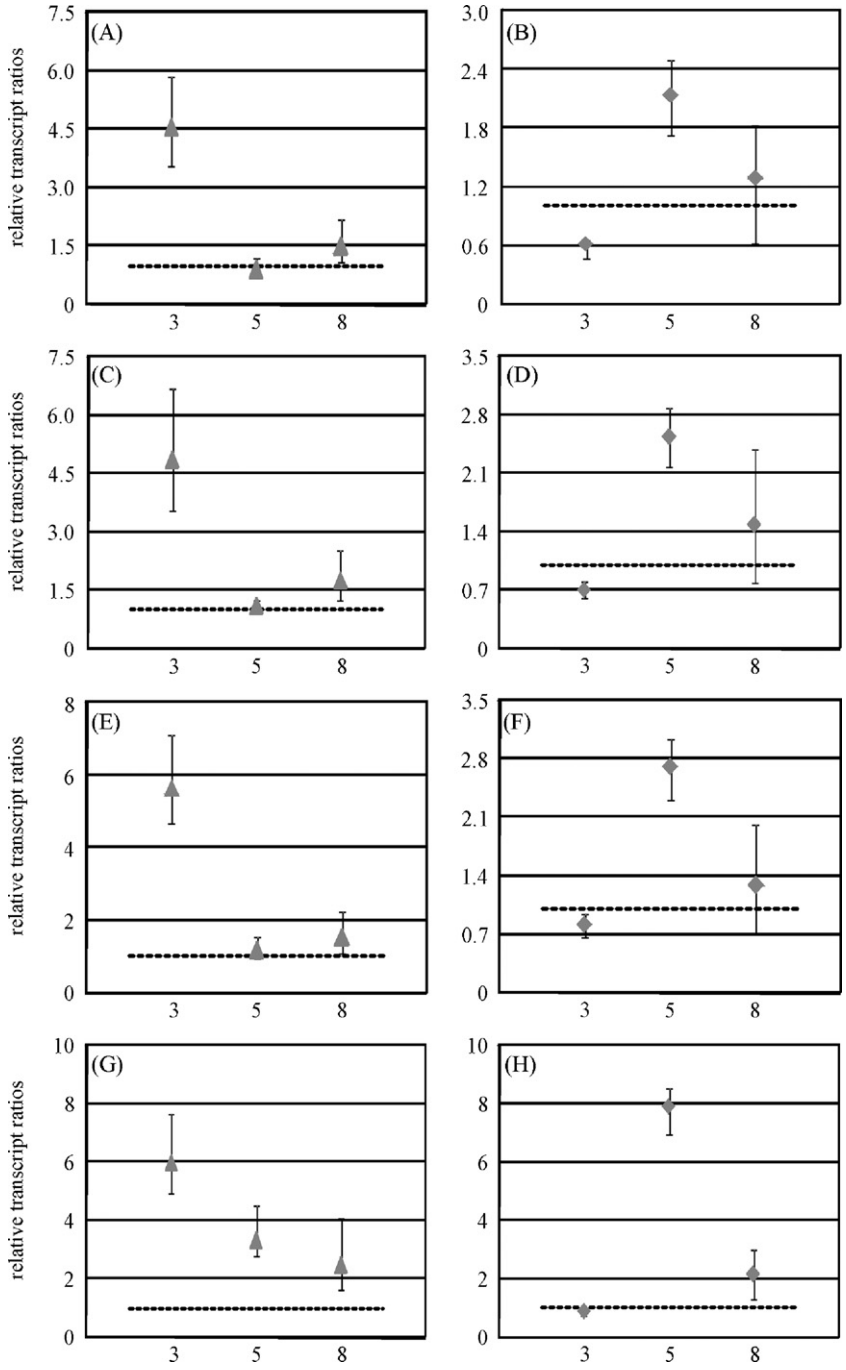


Fig. 4. Comparative transcript ratio analyses of samples from a replacement experiment using different normalization strategies. Transcript ratios of *xyn2* (A, C, E, and G) (triangles) and *xyr1* (B, D, F, and H) (diamonds) from a replacement experiment on xylose, comparing a nx7 strain (sample) with a QM9414 strain (control), calculated with REST 2008. Data of three different cultivation times are given (3, 5, 8 h). The dashed line (ratio of 1) indicates equal transcript amounts in both strains. Values above 1 indicate more transcripts in the nx7 strain compared to QM9414 and values beneath 1 indicate fewer transcripts. Different combinations of reference genes were applied: *sar1/glk1/cox4/act/tef1a* (A and B); *sar1/glk1* (C and D); *sar1* (E and F); *zwf1* (G and H). Error bars indicate a 95% confidence interval and stress the semantic significance of the data.

3.3. High correlation between xyn2 transcript formation and Xylanase II activity is obtained using sar1 as reference system

For further evaluation of the obtained results, we compared the transcription data with the xylanase enzymatic activity. We measured this xylanase activity in a pH range which is optimal for Xylanase II, but not for Xylanase I (Tenkanen et al., 1992). Additionally, Xylanase II is responsible for over 50% of the xylanolytic activity in *Hypocrea* (Törrönen et al., 1994). Thus, we measured predominantly Xylanase II in the applied enzyme assay. Again, the transcript ratios of *xyn2* transcription either with *sar1* or *glk1* as reference system are depicted. Additionally, the measured xylanase activity ratio is given (R*ace2* strain compared to Δ*ace2* strain), which perfectly fits to the *xyn2* transcript ratio normalized with *sar1* (Fig. 3A). In the second graph, a linear correlation analysis of xylanase activity ratio versus *xyn2* transcript formation ratio either normalized on *sar1* or *glk1* is presented (Fig. 3B). A high correlation ($R^2 = 0.97$) is found with *sar1* as the reference gene, whereas *glk1* as reference does not correlate ($R^2 = 0.06$) with the xylanase activity ratio (Fig. 3B). A causal correlation between *xyn2* mRNA formation and xylanase enzymatic activity is very likely. Accordingly, in this case *sar1* turned out to be an appropriate reference gene. Using *glk1* as a reference gene would lead to a misinterpretation of transcription data.

The observed correlation between *xyn2* transcript ratio and xylanase enzymatic activity ratio allows the assumption that *xyn2* transcript stability does not interfere with the transcript analysis. Although *xyr1* transcript stability cannot be evaluated by an enzymatic approach reflecting its transcript level, it should be considered that all mRNA samples were immediately processed into cDNA and all were treated in the same way. Consequently, the transcript profile obtained by RT-qPCR should not be affected by a transcript instability effect.

3.4. Under certain experimental conditions sar1 can be used as single reference gene

In the other experimental setup, the replacement experiment, *sar1* was also found to be the most stable gene, this time with both algorithms (compare Fig. 1C and D). Fig. 4 presents the transcript ratios of *xyn2* and *xyr1* from the replacement experiment using D-xylose. This time two other strains, nx7 strain (constitutively expressing *xyr1*) and QM9414 (parental strain), were investigated. Again, different normalization systems (reference genes) were applied. The first one comprised the five best performing genes (Fig. 4A and B), the second one uses *sar1*/*glk1* found by NormFinder as the best combination of genes (Fig. 4C and D). In the third one, the best scoring gene *sar1* was used for normalization (Fig. 4E and F). The first three data sets perfectly fit together, showing the same trend. This time *glk1* (Fig. 4C and D) could be used as reference gene together with *sar1*, because the intragroup variation calculated by NormFinder for *glk1* is comparable in both strain subgroups (data not shown). This example demonstrates that this time *sar1* could be even used as single reference gene (Fig. 4E and F). The biological information is not altered or improved by introducing more reference genes into the normalization calculation (compare Fig. 4A and B and E and F). *sar1* as a stable housekeeping gene coincides with the results published by Saloheimo et al. (2004). They did not observe clear differences in *sar1* expression levels between different carbon sources. However, in the case of induction of the unfolded protein response (UPR) with dithiothreitol (DTT), they reported an increased mRNA level of *sar1* (Saloheimo et al., 2004). This fact once more highlights the necessity of a careful reference gene validation.

Therefore, it is generally advisable to include at least a second reference gene into the normalization. This strengthens the robustness of the calculation and prevents the occurrence of inconclusive data. Especially, when looking at minor changes in transcription rates, this leads to more precise results. Moreover, it is indispensable that the used reference genes are not co-regulated. A method to estimate the number of reference genes which should be included into the normalization is provided in the geNorm software with the pairwise variation value, and in the NormFinder software with the accumulated standard deviation. For instance, in the replacement experiment, it is sufficient to take another reference gene besides *sar1*, like *glk1*, because the pairwise variation V2/3 calculated by geNorm has a value of 0.11. This value is beyond the cut-off value of 0.15 suggested by Vandesompele et al. (2002). With NormFinder a similar result is found. The minimum accumulated standard deviation (0.095) is calculated using two genes. Here, the inclusion of a third gene into the normalization system is not necessary.

The last two graphs were generated using *zwf1* as the least stable reference gene (Fig. 4G and H). Once more, this example demonstrates that results and drawn conclusions can be substantially altered by using the wrong reference gene. Obviously, *xyn2* as well as *xyr1* transcript formation would be systematically overestimated in the nx7 strain compared to the QM9414 parental strain, if one would solely rely on this normalization strategy (Fig. 4G and H).

4. Conclusion

RT-qPCR is a powerful tool to analyze transcription profiles in *H. jecorina*. Like in other organisms, it is essential to use a stabile and robust reference system. Since comparison of different strains is often a focus of gene expression studies, we suggest the use of NormFinder for the evaluation of reference genes. The algorithm of geNorm has substantial problems when using more than one *H. jecorina* strain. We found that *sar1* can be used as a robust reference gene applying different experimental setups and different *H. jecorina* strains. In addition to *sar1*, at least another reference gene like *glk1*, *cox4*, *tef1a* or *act* should be used to increase the reliability of the results of transcript analysis. In any case, it has to be evaluated whether the transcription profile of a reference gene is altered or affected by the experimental conditions.

Acknowledgments

This study was supported by two grants from the Austrian Fonds zur Förderung Wissenschaftlicher Forschung (P21287, P20192), and by a grant from the Vienna University of Technology, "DemoTech–Innovative Project" to R.L.M., which are gratefully acknowledged.

References

Andersen, C.L., Jensen, J.L., Orntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 64, 5245–5250.

Barbu, V., Dautry, F., 1989. Northern blot normalization with a 28S rRNA oligonucleotide probe. Nucleic Acids Res. 17, 7115.

Bohle, K., Jungebloud, A., Gocke, Y., Dalpiaz, A., Cordes, C., Horn, H., Hempel, D.C., 2007. Selection of reference genes for normalisation of specific gene quantification data of *Aspergillus niger*. J. Biotechnol. 132, 353–358.

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

Galante, Y.M., Monteverdi, r., Inama, S., Caldini, c., De Conti, A., Lavelli, V., Bonomi, F., 1993. New applications of enzymes in wine making and olive oil production. Italian Biochem. Soc. Trans. 4, 34.

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

Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., Foust, T.D., 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315, 804–807.

Huggett, J., Dheda, K., Bustin, S., Zumla, A., 2005. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun 6, 279–284.
Koo, H., Ueda, M., Wakida, T., Yoshimura, Y., Igarashi, T., 1994. Cellulase treatment of cotton fabrics. Textile Res. J. 64, 70–74.
Kuhls, K., Lieckfeldt, E., Samuels, G.J., Kovacs, W., Meyer, W., Petrini, O., Gams, W., Borner, T., 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.
Kumar, A., Lepola, M., Purtell, C., 1994. Enyme finishing of man-made cellulosic fabrics. Textile Chem. Colorist 26, 25–28.
Lanzarini, G., Pifferi, P.G., 1989. Enzymes in the fruit juice industry. In: Cantarelli, C., Lanzarini, G. (Eds.), Biotechnology Applications in Beverage Production. Elsevier Science, London, pp. 189–222.
Mach-Aigner, A.R., Pucher, M.E., Steiger, M.G., Bauer, G.E., Preis, S.J., 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.
Mandels, M., 1985. Applications of cellulases. Biochem. Soc. Trans. 13, 414–416.
Mäntylä, A.L., Rossi, K.H., Vanhanen, S.A., Penttilä, M.E., Suominen, P.L., Nevalainen, K.M., 1992. Electrophoretic karyotyping of wild-type and mutant Trichoderma longibrachiatum (reesei) strains. Curr. Genet. 21, 471–477.
Matheucci Jr., E., Henrique-Silva, F., el-Gogary, S., Rossini, C.H., Leite, A., Vera, J.E., Urioste, J.C., Crivellaro, O., el-Dorry, H., 1995. Structure, organization and promoter expression of the actin-encoding gene in *Trichoderma reesei*. Gene 161, 103–106.
Noé, P., Chevalier, J., Mora, F., Comtat, J., 1986. Action of enzymes in chemical pulp and fibers. Part II: Enzymatic beating. J. Wood Chem. Technol. 6, 167–184.
Pedersen, G.P., Screws, G.A., Cereoni, D.A., 1992. Biopolishing of cellulosic fabrics. Can. Text. J. Dec, 31–35.
Pfaffl, M.W., Horgan, G.W., Dempfle, L., 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 30, e36.
Pfaffl, M.W., Tichopad, A., Prgomet, C., Neuvians, T.P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. Biotechnol. Lett. 26, 509–515.
Pucher, M., 2006. Evaluierung von Referenzgenen von Trichoderma atroviride für Real-Time PCR basierende Transkriptionsanalysen und deren Anwendung für die mRNA Quantifizierung von G-Protein gekoppelten Rezeptoren. University of Technology, Vienna.
Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick Jr., W.J., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R., Tschaplinski, T., 2006. The path forward for biofuels and biomaterials. Science 311, 484–489.
Rutledge, R.G., Stewart, D., 2008a. Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. BMC Mol. Biol. 9, 96.
Rutledge, R.G., Stewart, D., 2008b. A kinetic-based sigmoidal model for the polymerase chain reaction and its application to high-capacity absolute quantitative real-time PCR. BMC Biotechnol. 8, 47.
Saloheimo, M., Wang, H., Valkonen, M., Vasara, T., Huuskonen, A., Riikonen, M., Pakula, T., Ward, M., Penttilä, M., 2004. Characterization of secretory genes ypt1/yptA and nsf1/nsfA from two filamentous fungi: induction of secretory pathway genes of *Trichoderma reesei* under secretion stress conditions. Appl. Environ. Microbiol. 70, 459–467.
Stahlberg, A., Elbing, K., Andrade-Garda, J.M., Sjogreen, B., Forootan, A., Kubista, M., 2008. Multiway real-time PCR gene expression profiling in yeast *Saccharomyces cerevisiae* reveals altered transcriptional response of ADH-genes to glucose stimuli. BMC Genomics 9, 170.
Stricker, A.R., Grosstessner-Hain, K., Würleitner, E., 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.
Stricker, A.R., Steiger, M.G., Mach, R.L., 2007. Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*. FEBS Lett. 581, 3915–3920.
Stricker, A.R., Trefflinger, P., Aro, N., Penttilä, M., Mach, R.L., 2008. Role of Ace2 (Activator of Cellulases 2) within the xyn2 transcriptosome of *Hypocrea jecorina*. Fungal Genet Biol. 45, 436–445.
Tenkanen, M., Puls, J., Poutanen, K., 1992. Two major xylanases from *Trichoderma reesei*. Enzyme Microb. Technol. 14, 566–574.
Törrönen, A., Harkki, A., 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.
Törrönen, A., Mach, R.L., Messner, R., Gonzalez, R., Kalkkinen, N., Harkki, A., Kubicek, C.P., 1992. The two major xylanases from *Trichoderma reesei*: characterization of both enzymes and genes. Biotechnology (NY) 10, 1461–1465.
Törrönen, A., Rouvinen, J., Ahlgren, M., Harkki, A., Visuri, K., 1993. Crystallization and preliminary X-ray analysis of two major xylanases from *Trichoderma reesei*. J. Mol. Biol. 233, 313–316.
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3.
Walsh, G.A., Power, R.F., Headon, D.R., 1993. Enzymes in the animal-feed industry. Trends Biotechnol. 11, 424–430.
Welt, T., Dinus, R.J., 1995. Enzymatic deinking—a review. Progr. Paper Recycling 4, 36–47.

VII CONCLUSIONS

The work presented in this PhD can be divided into three different sections and is therefore presented in three autonomous manuscripts (section IV, V and VI). Nevertheless, these manuscripts form an entity to fulfil the aims defined in section III. The two papers of section V and VI present methodologies, which are essential prerequisites to fulfil the scientific questions addressed in the first paper in section IV. This paper (section IV) is also the work which resembles the central part of the PhD and therefore bears the same name as the overall PhD thesis namely "Engineering an N-acetylneuraminic acid synthesis pathway into *Hypocrea* (*Trichoderma*)".

In this work, we demonstrate that the saprophytic fungus *H. jecorina* can be used, not only for the production of enzymes, but also as a whole-cell biocatalyst to produce fine chemicals like NeuNAc. The native *H. jecorina* strains are not able to form NeuNAc and thus we had to engineer two bacterial enzymes into *H. jecorina*. Until now, the engineering of such an enzyme cascade was not reported for a filamentous fungus and therefore certain requirements had to be fulfilled until such a system could be put into practice.

First of all, it was necessary to have a potent transformation system at hand. Such a system must meet certain requirements: i) a high rate of homologous integration (gene replacement) events; ii) a recyclable marker system for unrestricted rounds of knock-outs; and iii) a bi-directional positive selection system. The realisation of all those requirements is fulfilled in the transformation system presented in section V. The high rate of homologous integration can be achieved by a gene deletion of *tmus53* leading to a non homologous end-joining deficient strain. A Cre/loxP based system is adapted to *Hypocrea* and allows full marker recycling by exploiting the two genes *amdS* and *pyr4* as bi-directional positive markers. This strategy enables the creation of recombinant strains without the disturbing effects of markers on their physiology. Furthermore, it is possible to perform sequential engineering steps without a shortage of markers and it is possible to generate a tailor made *H. jecorina* strain.

After generation of recombinant strains, it is necessary to characterize them at various levels including DNA, RNA and protein analysis. One powerful tool for quantitative RNA analysis is RT-qPCR. With this method transcription levels can be studied and compared in the organism. Using this methodology, it is essential to use a stabile and robust reference system. The development of such a reference system is described in section VI. Using two different algorithms, geNorm and NormFinder, it was possible to determine stable and reliable reference genes in *H. jecorina*. We found that *sar1* can be used as a robust reference gene applying different experimental setups and different *H. jecorina* strains. In addition to *sar1*, at least another reference gene like *glk1, cox4, tef1a* or *act* should be used to increase the reliability of the results of transcript analysis.

With the knowledge of those two studies (section V and VI), it was possible to construct recombinant *H. jecorina* strains and to characterize transcription levels of the recombinant genes (section IV). With a *H. jecorina* strain expressing a GlcNAc-2-epimerase and a NeuNAc synthase, we were able to show that such a strain gained the ability to form NeuNAc. Although the total amount of NeuNAc formed in the recombinant strain is low (1 µg NeuNAc per g mycelia), this PhD thesis presents for the first time a strategy, how an intracellular enzymatic-cascade of heterologously expressed proteins can be realized in a filamentous fungus. Beside the mere production of NeuNAc, this work can pave the way towards the production of recombinant sialylated glycoproteins in *H. jecorina*. This is especially important for therapeutic protein production which requires a humanized glycosylation pattern.

VIII CURRICULUM VITAE

Persönliche Details

Name	Dipl.-Ing. Matthias Georg Steiger
Geburtsdatum	8. Mai 1982
Geburtsort	Wien
Staatsbürgerschaft	Österreich
Präsenzdienst	abgeleistet im österreichischen Heer (Okt. 2000 – Juni 2001)

Ausbildung und Qualifikation

Okt. 2007 – Aug. 2010	**Technische Universität Wien** Doktoratsstudium in der Arbeitsgruppe von Ao. Prof. Dr. Robert Mach zum Thema: Einbringung eines N-Acetylneuraminsäure-Syntheseweges in *Hypocrea* (*Trichoderma*)
Okt. 2007- Juli 2008	**"TU the TOP",** Teilnehmer am High Potential Programm der Technischen Universität Wien
Okt. 2001- May 2007	**Technische Universität Wien** Diplomstudium der Technischen Chemie mit Spezialisierung im Bereich Biochemie und Gentechnik, Abschluss mit Auszeichnung im Mai 2007, Thema der Diplomarbeit „Der Aktivator Xyr1 und seine Bedeutung für den Laktosemetabolismus von *Trichoderma reesei*
Feb. 2005 - Juni 2005	**Universidad Complutense Madrid, Spanien** Auslandssemester (EU - ERASMUS) an der F*acultad de Ciencias Químicas*,, Mitarbeit an einem Projekt über „Surface proteins of HBV" *Departamento de Bioquímica y Biología Molecular*
Sept. 1992 - Juni 2000	**Humanistisches Gymnasium** Matura mit gutem Erfolg, Bundesgymnasium Gymnasiumstrasse 1190 Wien, Österreich

Arbeitserfahrung und Praktika

seit November 2007	**Technische Universität Wien, Österreich** Anstellung als Projektassistent
Aug. 2006- Sept. 2006	**Sandoz GesmbH,** Kundl, Austria Arbeitsbereich: Cephalosporinaufarbeitung Projekt: "Energie und Massenflussbilanzierung zur Implementierung eines Wärmerückgewinnungssystems"
Juli – Aug. 2004	**Ottakringer Brauerei AG,** Wien, Österreich Praktikum in der Qualitätssicherung
Aug. 2002	**Hoffmann La Roche AG,** Grenzach-Whylen, Deutschland Praktikum in der Finance & Controlling Abteilung
Juli-Aug. 2001	**Wiener Börse AG,** Wien, Österreich Praktikum in der Rechtsabteilung
Juli 2000	**Hoffmann La Roche AG,** Grenzach-Whylen, Deutschlnad Praktikum in der Finance & Controlling Abteilung

Sprachen

- Deutsch Muttersprache
- Englisch Verhandlungssicher
- Spanisch Fließend (in Wort und Schrift)

Wissenschaftliche Publikationen in internationalen Journalen mit Peer-Review System

~Sommer 2010	Engineering an N-acetylneuraminic acid synthesis pathway into *Hypocrea* (*Trichoderma).* Matthias G. Steiger, Astrid R. Mach-Aigner, Rita Gorsche, Erwin E. Rosenberg, Marko D. Mihovilovic, Robert L. Mach *(in Vorbereitung)*
Juni 2010	A modified expression of the major hydrolase activator in *Hypocrea jecorina* (*Trichoderma reesei*) changes enzymatic catalysis of biopolymer degradation. Marion E. Pucher, Matthias G. Steiger, Robert L. Mach, and Astrid R. Mach-Aigner *(eingereicht zum Peer-Review bei Catalysis today)*
April 2010	A transformation system for *Hypocrea (Trichoderma)* that favours homologous integration and that uses reusable bi-directionally selectable markers. Authors: Matthias G Steiger; Marika Vitikainen; Kurt Brunner; Gerhard Adam; Markku Saloheimo; Robert L Mach; Astrid R. Mach-Aigner *(eingereicht zum Peer-Review bei Appl Environ Microbiol.)*
Jänner 2010	An accurate normalization strategy for RT-qPCR in Hypocrea jecorina (Trichoderma reesei). Authors: Matthias G. Steiger, Robert L. Mach, Astrid R. Mach-Aigner. J Biotechnol. 2010 Jan 1;145(1):30-7.
September 2008	Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in Hypocrea jecorina**.** Authors: Astrid R. Mach-Aigner, Marion E. Pucher, Matthias G. Steiger, GudrunE. Bauer, Sonja J. Preis, Robert L. Mach; Appl Environ Microbiol. 2008 Nov;74(21):6554-62.
Juni 2007	Xyr1 receives the lactose induction signal and regulates lactose metabolism in Hypocrea jecorina. Authors: Astrid R. Stricker, Matthias G. Steiger, Robert L. Mach; FEBS Lett. 2007 Aug 21;581(21):3915-20.

Wien, Juni 2010