Title: NOVEL MEANS OF TRANSFORMATION OF FUNGI AND THEIR USE FOR HETEROLOGOUS PROTEIN PRODUCTION

Abstract: A recombinant Aspergillus sojae comprising an introduced acetamidase S (amdS) gene as selectable marker is disclosed. An Aspergillus sojae exhibiting growth with medium comprising uracil and fluoro-orotic acid, said Aspergillus sojae further not exhibiting growth on medium comprising uridine and fluoro-orotic acid i.e. said Aspergillus sojae exhibiting uracil auxotrophy, said Aspergillus sojae being unable to utilize uridine, said Aspergillus sojae being pyrG negative, said Aspergillus sojae exhibiting resistance to fluoro-orotic acid, said uracil auxotrophy and said fluoro-orotic acid resistance being receivable upon complementation with an active introduced pyrG gene, is described. The Aspergillus sojae further comprises a nucleic acid sequence encoding a phytase or a protein having phytase activity or any other heterologous protein or polypeptide and can be used for the biotechnological production of said phytase or said other heterologous proteins or polypeptides. Additional mutants exhibiting amended morphology are also disclosed. Methods of producing such expression hosts are described.
Novel means of transformation of fungi and their use for heterologous protein production.

SUMMARY OF THE INVENTION

The invention relates to novel means of transformation of fungi and to their use for production of heterologous proteins. The means involve genetically engineered fungi belonging to the taxonomic group *Aspergillus sojae*. Suggestions have been provided in the past to use *Aspergillus sojae* as a host strain for transformation. However to date no data are provided on successful transformation and/or expression of heterologous proteins.

In addition it has been found, that so far certain proteins, such as phytase which were difficult to express in large amounts, due to several reasons including proteolytic degradation in expression hosts other than *Aspergillus sojae*, can surprisingly be expressed in *Aspergillus sojae*. Production levels for heterologous proteins in *Aspergillus sojae* have been found to exceed those levels achieved for the same proteins in *Aspergillus niger* and *Aspergillus awamori*. In addition to the above, the subject of the invention further covers a process for obtaining improved *Aspergillus sojae* strains for expression purposes, characterized by, on the one hand, a decreased proteolytic activity, and, on the other hand improved fermentation characteristics related to the morphology of the fungi.

BACKGROUND OF THE INVENTION

Suggestions have been provided in the past to use *Aspergillus sojae* as a host for transformation. However, to date no data are provided on successful transformation and/or production of heterologous proteins and, more specifically, nothing is revealed concerning expression of phytase. Previously, expression levels were too low in expression hosts other than *Aspergillus sojae*, mainly due to proteolytic degradation. We have now found expression levels for the protein in *Aspergillus sojae* that exceed those levels achieved for the same protein in other strains, e.g. *Aspergillus niger*, *Aspergillus awamori* and *Trichoderma*. It is surprising to find such an improvement in closely related strains. Thus, prior art disclosures concerning phytase production exhibit shortcomings. Prior art disclosures on the use of *Aspergillus sojae* for expressing heterologous proteins or polypeptides were inadequate.

The fact that until now hardly any successful attempts for *A. sojae* transformation have been reported is remarkable in view of the fact that numerous successful transformations of closely related strains of the taxonomic group *Aspergillus oryzae* have
been described in the past. On the basis of this close relationship the skilled person would anticipate (and in fact did anticipate) that analogous methods to those used for *Aspergillus oryzae* are also applicable for *Aspergillus sojae* strains.

WO97/04108 for example describes the isolation of a protease encoding nucleic acid sequence, specifically a leucine aminopeptidase encoding sequence, and the transformation of a variety of host organisms, i.a. *Aspergillus sojae*, with a leucine aminopeptidase encoding sequence. However no illustration of this particular transformation actually having been carried out is provided. It is merely suggested as one possibility among many other strains such as *Trichoderma reesei*, *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus phoenicus*, and *Aspergillus oryzae* as a potential host strain to be used for transformation. In the cited document, 3 transformation protocols readily used in the art are suggested for the strains. Specifically any of the selection markers acetamidase S (=amdS) (e.g. as maintained on vector p3SR2), argB or hygromycin B (e.g. using the vector pAN7-1) are suggested as being suitable markers to be used according to the transformation protocols described therein.

The use of vector p3SR2 with the amdS marker has frequently been described in the literature as being useful for transforming various strains, for example *Aspergillus oryzae* (in EP 0.238.023), *Trichoderma reesei* (in EP 0.244.234) and *Aspergillus niger* (EMBO Journal 4, pages 475-479). Consequently, the analogous use for transforming *Aspergillus* in general is put forward in WO97/04108 on the basis of these previous publications.

Quite specifically on page 17 of WO97/04108 it is described that *Aspergilli* and *Trichoderma* that prior to transformation grew slowly on minimal medium comprising solely the substrate acetamide as source of nitrogen could be selected after transformation with the vector p3SR2 due to a clear growth advantage. Subsequently, the thus obtained transformants would need to be further subjected to selection for leucine aminopeptidase (=LAP) productivity in order to find a desired transformant. As stated above this is merely put forward as speculative means of transformation applicable over the two aforementioned genera in toto based on a few successful transformations of strains other than *Aspergillus sojae*.

The suggested transformation protocol is, however, unsuccessful with *Aspergillus sojae*. The selection criteria described in the prior art are insufficient to ensure practical selection of desirable transformants when using the vector p3SR2. We have conducted the
experiments and found the described method inoperable due to excessive background growth eliminating practical selectability.

Another routinely used selection method for fungal transformants is that of transformation of orotidine-5-monophosphate decarboxylase (=PyrG) mutants. Mattern *et al.* in Mol. Gen. Genet. 210, pages 460-461 disclose transformation of *Aspergillus oryzae* using the *Aspergillus niger pyrG* gene. Standard practice is to isolate pyrG mutants based on direct resistance to fluoro-orotic acid as a positive selection marker. This has resulted in isolation of numerous pyrG mutants for a variety of fungi to date.

From experience with a number of different filamentous fungi, the auxotrophic pyrG-based system has many favourable characteristics. Experiments were carried out to obtain *A. sojae* pyrG mutant strains, using a standard procedure based on direct selection for resistance to fluoro-orotic acid (FOA) on plates containing uridine to support growth of the mutant strain (Van Hartingsveldt *et al.* in Mol. Gen. Genet. (1987) 206, pages 71-75). However, use of the analogous method on *Aspergillus sojae* strains did not lead to pyrG mutants. The usual method did lead to fluoro-orotic acid resistant strains but all the strains were able to grow without uridine. Thus, none of these strains were pyrG mutants. Normally, the isolation of the pyrG mutants can be done directly from the fluoro-orotic acid resistant strains on a uridine selection medium. For *Aspergillus sojae* this method however turned out to be inoperable.

Clearly, *Aspergillus sojae* exhibits different traits than the closely related *Aspergillus oryzae* when it comes down to transformation. The standard protocols using amdS or pyrG as selectable markers do not suffice. Unfortunately, the method of argB as selectable marker is not an attractive option either, since this requires isolation of a corresponding argB mutant for every host strain one wishes to use. This is an arduous task based on trial and error. The required argB mutant can be obtained through random mutagenesis followed by screening of tens of thousands of colonies. The situation for pyrG is better in that the mutant is itself selectable. In the case of amdS no mutant is required as the presence of amdS works as dominant selectable marker.

Additional problems dissuading the skilled person from use of *Aspergillus sojae* as expression host for recombinant proteins or polypeptides exist. In JP-A-02-234666 for example an ArgB based selection of *Aspergillus sojae* is described using an analogous protocol to that described for other fungi. Such a process has been described for *Aspergillus oryzae* in Biotechnology (1988) 6, pages 1419-1422. The cited article also refers to successful analogous transformation of *Aspergillus nidulans* and *Aspergillus*
niger. However, when the *Aspergillus sojae* strain ATCC42251 disclosed in the Japanese patent application was analysed, an undesirable protease profile was found. The protease profile of this strain is incompatible for application as a production host. So even though a transformation protocol has been suggested in the prior art for this particular *Aspergillus sojae* strain it could not possibly lead to a high level of expression of heterologous protein even if the protocol for transformation was successful.

It is in fact due precisely to the explicit characteristics of *Aspergillus sojae* strains to produce excessive amounts of alkaline proteases and amylases that they currently find application in practice. They are used specifically in processes requiring degradation of complex polymeric substrates. It was thus at best to be expected that any transformants of *Aspergillus sojae* that are finally successful will not lead to good expression levels unless the product is an *Aspergillus sojae* protein that is impervious to its own proteases.

In summary the problems facing the skilled person in finding a means to use *Aspergillus sojae* strains for expressing heterologous recombinant proteins on an industrial scale are manifold. Firstly, a number of processes for introducing the desired nucleic acid material to be expressed are not applicable in the manner used for other fungi. This includes pyrG- and amdS-based processes that are useful for the closely related *Aspergillus oryzae*. Secondly, it remained to be seen whether high level production of heterologous proteins would be feasible despite the known excessive proteolytic activity of the host strain *Aspergillus sojae*.

Unexpectedly, it has been found that the problems addressed above can be solved, thus resulting in novel expression hosts for producing proteins and novel methods of production of heterologous proteins. We describe transformation of *A. sojae* strains with the amdS and pyrG selection markers. In addition efficient gene expression is described, including expression of a phytase gene.

DESCRIPTION OF THE INVENTION

As stated the subject invention is directed at *Aspergillus sojae* strains and the application thereof for production of recombinant proteins and polypeptides. Firstly, a description of *Aspergillus sojae* strains is provided.

*Aspergillus sojae* determination.

The fungal taxonomy is a complex issue. The *Aspergillus* genus comprises *Aspergillus sojae* in the Flavi/Tamarii section (see Table 1). *A. sojae* is clearly shown to be
distinct from *A. oryzae* which is located in the same section (see Table 2). Currently, strains belonging to *Aspergillus sojae* can be distinguished from taxonomically closely related *Aspergillus oryzae* and also closely related *Aspergillus parasiticus* strains in a number of manners recognised in the art. Reference is made to the random PCR fragments, **ver-1, aflR** and rDNA sequences as described, respectively, in Ushijama *et al.* (1981), Chang *et al.* (1995), Yuan *et al.* (1995), Kusomoto *et al.* (1998) and Watson *et al.* (1999). In addition it has been found that *Aspergillus oryzae* further differs from *Aspergillus sojae* upon comparison of the *alpA* sequence of these strains. Inter alia (there are other sequence differences between *A. oryzae* and *A. sojae* *alpA* which could be used as a determination tool), it has been found that *Aspergillus sojae* comprises an *XmnI* restriction site at a specific location in the *alpA* gene. The corresponding location in the *alpA* gene of several *Aspergillus oryzae* strains does not possess such a restriction site. Thus, this provides an additional discrimination point between the two types of fungal strains. Consequently, numerous methods are available to the skilled person to assess whether a strain is an *Aspergillus sojae*. Currently more than 10 strains are deposited with the ATCC that are defined as *Aspergillus sojae*. The 10 oldest deposits have been analysed. Two out of 10 did not pass the last mentioned determination test. One of them is the ATCC20235 which according to Ushijama *et al.* (1981) also did not fulfill the requirements for classification as an *Aspergillus sojae* on the basis of morphological parameters. The other is ATCC46250. The definition of *Aspergillus sojae* as used throughout the patent application is meant to imply a strain that preferably fulfils all the requirements described in the cited references in combination with the presence of the *XmnI* restriction site in the *alpA* gene. Specific homologous primers for both the *Aspergillus oryzae* and *Aspergillus sojae* sequences are also provided. They can be used to test for the presence of the *XmnI* restriction site by way of example of a screening test useful for distinguishing *Aspergillus oryzae* from *Aspergillus sojae* (Primer sequences are SEQ ID No.1 MBL1784: 5'-CGGAATTTCGAGCGCAACTACAAGATCAA-3' and SEQ ID No.2 MBL1785: 5'- CGGAATTCAGCCCAGTTGAAGCCCGTC-3'). They are derived from the coding region of the *alpA* gene. It will be obvious to the skilled person on the basis of the known sequence data that alternative probes or primers are conceivable. PCR amplification using these primers on *Aspergillus* DNA, followed by restriction enzyme digestion of the resulting DNA fragments with *XmnI* provides a way to discriminate *A. sojae* strains from *A. oryzae* strains. Having established the definition of *Aspergillus sojae* strains we can proceed further with the detailed description of the invention.
The invention in one aspect covers a recombinant *Aspergillus sojae* comprising an introduced acetamidase S (*amdS*) gene as a selectable marker. Such an *A. sojae* is selectable on a medium comprising a substrate for the introduced amdS protein as sole source of nitrogen, said medium further comprising a carbon substrate and said medium being free of endogenous amdS inducing substrate. A suitable medium comprises acrylamide as substrate for the introduced amdS as sole source of nitrogen. A suitable medium at least further comprises minimum substrates required for growth of *Aspergillus sojae*. A suitable category of *A. sojae* according to the invention is formed by *A. sojae* that are not selectable on acetamide comprising medium. An *A. sojae* according to the invention is suitably an *A. sojae* selectable on a medium free of glucose, i.e. a medium wherein the carbon source is not glucose. Such a medium can be a medium having sorbitol as carbon source. Best results in the case of sorbitol are achieved when sorbitol is the sole carbon source.

An *Aspergillus sojae* according to the invention may comprise a further introduced nucleic acid sequence, said further introduced sequence preferably encoding a protein or polypeptide. The further introduced sequence may be adapted for optimised codon usage to the host strain codon usage or may have the original codons from the host from which it has been derived. The introduced sequence is in principle any sequence the skilled person wishes to express. The introduced sequence can suitably be heterologous, i.e. foreign to the *Aspergillus sojae* into which it is introduced. It can also be native but introduced in the form of one or more additional copies.

One of the subjects of the invention is aimed at expressing phytase or proteins having phytase activity. Numerous sequences are known to the skilled person concerning sequence data of phytases. We refer to and incorporate by reference the contents of EP 684.313, EP 897.010, WO 99/49022, EP 911.416 and EP 897.985. These documents describe various natural and modified phytase sequences. They also describe a consensus sequence. A suitable embodiment is formed by phytase sequences from *Peniophora* being either the natural sequences or modified versions thereof. The new system is more flexible than prior systems and thus heterologous sequences, including heterologous sequences encoding phytase or proteins having phytase activity that were difficult to express in the prior art fungal systems can be expressed in the novel system according to the invention.

An *Aspergillus sojae* according to the invention as defined in any of the embodiments defined above comprising an introduced *amdS* gene as selectable marker may suitably have no active endogenous *amdS* gene. The *Aspergillus sojae* according to such an embodiment
may by way of example have an endogenous \textit{amdS} gene comprising an endogenous \textit{amdS} inactivating mutation. Any type of inactivating mutation known or conceivable to the skilled person may have occurred. A suitable example of such inactivating mutation may be a deletion or disruption. The mutation may inactivate the gene or the gene product. The skilled person will realise that numerous options are available to achieve this and that they can readily be achieved.

In an alternative embodiment the invention is also directed at a recombinant \textit{Aspergillus sojae} free of an active endogenous \textit{amdS} gene and further comprising an introduced \textit{amdS} gene as selectable marker. The recombinant \textit{Aspergillus sojae} according to the invention is selectable on a medium comprising a substrate for the \textit{amdS} as sole source of nitrogen, said medium further comprising a carbon substrate. A suitable medium at least further comprises minimum substrates required for growth of \textit{A. sojae}. In a suitable embodiment the endogenous \textit{amdS} gene can for example have been inactivated. This inactivation can be any type of inactivation known or conceivable to a person skilled in the art that still leaves the \textit{A. sojae} viable. By way of example the endogenous \textit{amdS} gene can comprise an inactivating mutation in the form of a substitution, deletion or insertion of the gene or part thereof, or by virtue of a mutation affecting expression of the gene such as to render it inactive. The complete endogenous \textit{amdS} gene can also be absent.

An \textit{Aspergillus sojae} in any of the described embodiments according to the invention may be an \textit{A. sojae} into an \textit{amdS} gene has been introduced. This can be achieved e.g. by transformation or transfection. The resulting \textit{Aspergillus sojae} according to the invention must then subsequently have been separated from non transformed or transfected \textit{A. sojae}. Any of the embodiments described above as such or in combination are covered by the invention.

The invention not only covers \textit{Aspergillus sojae} as such, but also covers a method of introducing a nucleic acid sequence into \textit{A. sojae}. The method comprises subjecting \textit{Aspergillus sojae} to introduction of a nucleic acid sequence in a manner known per se for introduction of a nucleic acid sequence into a fungus. Such a manner can e.g. be transformation or transfection of the \textit{A. sojae}. The method comprises the introduction of the \textit{amdS} gene as the nucleic acid sequence followed by selection of the resulting transformed or transfected \textit{A. sojae} on a medium free of endogenous \textit{amdS} inducing substrate, said medium further comprising a substrate for the introduced \textit{amdS} as sole source of nitrogen and said medium further comprising a carbon substrate, said medium enabling the desired \textit{A. sojae} comprising introduced \textit{amdS} gene to grow whilst eliminating growth of \textit{A. sojae}
devoid of a functional \textit{amdS} gene. A suitable embodiment of such a method involves applying a medium comprising a substrate for \textit{amdS} other than acetamide. Suitably, such a medium comprises acrylamide as substrate for the introduced \textit{amdS} as sole source of nitrogen. Suitably, a medium for the method according to the invention comprises a carbon source other than glucose. Suitably, a medium for use in a method according to the invention comprises sorbitol as carbon source, preferably as sole carbon source. A suitable medium at least further comprises minimum substrates required for growth of \textit{A. sojae}.

A method according to the invention as defined above in any of the embodiments comprises introduction of an additional nucleic acid sequence besides the \textit{amdS} gene. The additional nucleic acid sequence for example encodes a protein or polypeptide, such as a phytase or proteins having phytase activity. The sequence does not necessarily have to be a non \textit{Aspergillus sojae} sequence, but can also include \textit{A. sojae} derived sequences. It is however intended to indicate that the sequence that is introduced is absent in the non-transformed strain or else is present in a lower copy number than in the \textit{A. sojae} according to the invention.

Naturally, the subject invention also covers any \textit{Aspergillus sojae} obtained by the method described above. Basically, the method is directed at introducing a sequence capable of realising the presence of sufficient active \textit{amdS} to function as selectable marker as opposed to the \textit{A. sojae} into which the sequence is introduced which cannot for some reason or another produce sufficient active \textit{amdS} to enable growth on a substrate for \textit{amdS} as sole source of nitrogen.

A method of selecting transformed or transfected \textit{A. sojae} also falls within the scope of the invention. The method comprises subjecting \textit{A. sojae} (with no active endogenous \textit{amdS} gene as defined according to any of the embodiments described) to a method of transformation or transfection of the \textit{A. sojae} in a manner known per se for transformation or transfection of fungi with a nucleic acid sequence. The method comprises the introduction of an \textit{amdS} gene as the nucleic acid sequence, followed by selection of the resulting transformed or transfected \textit{A. sojae}, said selection occurring on a medium comprising a substrate for the introduced \textit{amdS} as sole source of nitrogen, said medium further comprising a carbon substrate, said medium enabling the desired \textit{A. sojae} to grow whilst eliminating growth of non transformed or transfected \textit{A. sojae} due to inability of such to grow without the introduced \textit{amdS} gene on the selection medium. A suitable medium at least further comprises minimum substrates required for growth of \textit{A. sojae}.

The invention is also directed at a method for producing recombinant \textit{Aspergillus}
sojae. This method comprises introducing a desired nucleic acid sequence e.g. by transformation or transfection in a manner known per se into an A. sojae, said desired nucleic acid sequence being flanked by sections of the endogenous amdS gene of a length and homology sufficient to ensure recombination. The introduction is followed by selection of the recombinant A. sojae having the desired nucleic acid sequence. The selection occurs for a selectable marker comprised in or transformed in cotransformation with the desired nucleic acid sequence, said selectable marker being absent in the A. sojae prior to introduction of the desired nucleic acid sequence. The flanking sequences may also be sequences corresponding to the endogenous amdS gene sufficient to ensure recombination. The skilled person can readily assess which sequences will suffice on the basis of hybridisation knowledge and the sequence data of the endogenous amdS gene. The recombination event eliminates the endogenous amdS activity in both cases. The selectable marker can quite suitably be pyrG, with, however, uracil instead of uridin in the selection medium.

A further embodiment of the invention comprises Aspergillus sojae exhibiting growth on medium comprising uracil and fluoro-orotic acid, said A. sojae further not exhibiting growth on medium comprising uridine and fluoro-orotic acid. This means that the A. sojae exhibits uracil auxotrophy, is unable to utilize uridine, is pyrG negative and exhibits resistance to fluoro-orotic acid. The uracil auxotrophy and the fluoro-orotic acid resistance are relievable upon complementation with an active introduced pyrG gene. Such an A. sojae according to the invention can be free of active endogenous pyrG genes. The pyrG negative A. sojae according to the invention may comprise an endogenous pyrG gene with a mutation inactivating it. The mutation can be any mutation known or conceivable to a person skilled in the art, said mutation inactivating a pyrG gene or the expression product thereof. Such a mutation can by way of example be in the form of an insert of a nucleic acid sequence in the gene, a substitution of a part of the encoding sequence of the gene, a deletion of a part of the encoding sequence of the gene or a deletion of the whole encoding sequence of the gene. The mutation can also occur in the regulating part of the gene. In the case of Aspergillus sojae according to the invention with a mutated pyrG gene, said Aspergillus sojae can have a nucleic acid sequence for the mutated pyrG gene different to that of the wild type A. sojae pyrG gene. A further embodiment comprises pyrG negative A. sojae according to the invention as described in any of the above embodiments which further comprise any of the characteristics described for any of the amdS variant A. sojae according to the invention as such or in combination.
A method of selecting transformed or transfected *Aspergillus sojae* also falls within the scope of the invention. The method comprises subjecting *A. sojae* of the pyrG negative type according to any of the embodiments of the invention as described above to a method of transformation or transfection with a nucleic acid sequence, said method comprising introducing an active pyrG gene into the pyrG negative *A. sojae* in a manner known per se for transformation or transfection. The introduction step is then followed by selection of the resulting transformed or transfected *A. sojae* on a medium free of uracil and fluoro- orotic acid, said medium at least further comprising minimum substrates required for growth of *A. sojae*, said medium enabling the desired *A. sojae* to grow whilst eliminating growth of non-transformed or -transfected *A. sojae* due to inability of such to grow without uracil due to the inactivated pyrG gene. In a suitable embodiment of such a method the active pyrG gene that is introduced is flanked by identical nucleic acid sequence fragments, and the pyrG positive *A. sojae* resulting from the introduction of the pyrG gene and the flanking sequences is selected on a medium free of uracil and fluoro-orotic acid.

Subsequently the pyrG positive *A. sojae* is cultivated on medium comprising uracil and fluoro-orotic acid, thereby eliminating the pyrG gene that had been introduced and thus resulting in a pyrG negative *A. sojae* that is selectable by growth on uracil comprising medium and fluoro-orotic acid resistance. In a suitable embodiment of the aforementioned method the flanking sequences and the pyrG gene are further flanked by sequences that direct integration of the pyrG gene and the flanking sequences into a specific location, due to the fact that the integration directing sequences are homologous to a specific sequence of the *A. sojae* to be transformed. This enables knock out (if desired) of the gene associated with the specific sequence. The process of knock-out mutant creation as such is well known to the person skilled in the art. Any of the embodiments of the selection method just described may further comprise the step wherein the *Aspergillus sojae* is transformed or transfected with a further heterologous nucleic acid sequence. The further heterologous nucleic acid sequence preferably encodes a protein or polypeptide and the same remarks are valid here as made elsewhere in this description for the nature of such further nucleic acid sequences for the other embodiments of *Aspergillus sojae* and fungi in general according to the invention. The further sequence can be introduced with the active pyrG gene either on the same vector or by cotransformation with the active pyrG gene that is introduced. The method of selecting transformed or transfected *A. sojae* as described may also be carried out in combination with the method for introducing a nucleic acid comprising introduction of a heterologous amds gene in any of the embodiments according
to the invention disclosed therefore above. Naturally, the invention covers any recombinant *A. sojae* obtained by the method of selecting transformed or transfected *A. sojae* according to the invention.

The invention is also directed at a method for producing recombinant *Aspergillus sojae*, said method comprising transformation or transfection in a manner known per se of a pyrG positive *Aspergillus sojae* with a nucleic acid sequence comprising the sequence to be introduced flanked by sections of the pyrG gene or corresponding sequences of a length and homology sufficient to ensure recombination eliminating the pyrG gene and introducing the desired sequence, followed by selection of the recombinant *Aspergillus sojae* with the desired sequence by selecting for the *A sojae* with a pyrG negative phenotype. Determination of the corresponding sequences lies within the reach of the skilled person by virtue of their knowledge of hybridisation processes with nucleic acid sequences and their knowledge of required sequence data of the pyrG genes.

In particular the invention also covers such *Aspergillus sojae* exhibiting the characteristics of the amdS variant *A sojae* according to the invention as defined above. Thus any *Aspergillus sojae* strain obtained by either the amdS and/or pyrG introduction method according to the invention is a novel strain falling within the scope of the invention as is any subsequent use of such a novel strain. Such a novel strain can comprise nucleic acid sequences that do not occur in the original corresponding *Aspergillus sojae* strain or even do not occur in *Aspergillus sojae*, *Aspergilli* or fungi. The sequences can be of mammalian origin or derived from any animal, plant or microbe. Nucleic acid sequences can also be expressed that are naturally present in the *Aspergillus sojae* strain but that are present in a lower copy number in the corresponding non-transformed *A sojae*. Thus the production of homologous proteins is also covered by the invention when pyrG and/or amdS *Aspergillus sojae* strains according to the invention are involved. A preferred embodiment is that wherein the particular protein or polypeptide to be produced is absent in the corresponding non-treated *A. sojae* and/or is present in a lower copy number in the corresponding non-treated *A. sojae*, i.e. the *A. sojae* prior to introduction of the nucleic acid sequence. Expression of heterologous proteins by any of the novel strains of *Aspergillus sojae* in a manner known per se for producing protein or polypeptide in a fungus thus covers both sequence native to the strain and foreign to the strain. Basically, only the native non-transformed or -transfected *A sojae* is excluded from protection. A process of production comprises cultivating the fungus under suitable conditions for expression of the desired sequence to occur. The process of production optionally includes
the step of isolation of the resulting polypeptide or protein in a manner known per se for protein or polypeptide production by fungi. Preferably the protein or polypeptide will be secreted into the culture medium.

A preferred protein or polypeptide is a protein or polypeptide susceptible to degradation upon expression by *Aspergillus niger* or *Aspergillus awamori*. A number of such proteins and polypeptides have already been disclosed in the prior art and a large number remain yet to be determined. Such determination is however a matter of routine for the skilled person. Another preferred embodiment of the protein or polypeptide to be expressed is one whereby the protein or polypeptide differs from an *Aspergillus sojae* protease and amylase. A preferred embodiment involves a non-*Aspergillus sojae* protein or polypeptide.

A particularly interesting embodiment comprises a combination of the two processes for introducing nucleic acid sequences according to the invention as described above. The advantage thereof lies in the fact that the frequency of transformation obtained with the pyrG marker is clearly much higher than that of the amds marker. However, secondary screening of the pyrG+ strains for the best growth on acrylamide selective plates allows the identification of those recombinant *Aspergillus sojae* showing the highest copy number and thus most likely the highest level of gene expression.

As indicated in the examples homologous and heterologous expression regulating sequences can be used by *Aspergillus sojae* i.e. natively occurring sequences of the strain itself or sequences foreign to the strain can be used. Thus the transformants according to the invention can comprise any such regulatory sequences. The selection of the suitable regulatory region is a matter of choice that lies well within the range of the standard capabilities of the skilled person and will depend on the particular application. The regulating sequences can be constitutive or inducible. The regulating sequences can be fungal or non-fungal. A broad range are exemplified in the examples. A large number of expression regulating sequences are regularly used in the art for other systems, in particular fungal systems such as *Aspergilli*, and can routinely be applied without undue burden in the *Aspergilli* according to the invention.

For introducing the desired nucleic acid sequences into *Aspergillus sojae* any vector may be used that is suitable for introducing nucleic acid sequences into fungal host cells. Numerous examples are available in the art. In particular vectors that have been found suitable for transformation, transfection or expression in *Aspergilli* such as *Aspergillus niger, Aspergillus awamori* and *Aspergillus oryzae* can suitably be applied.
In addition to the above the subject invention describes efficient protein production for recombinant *Aspergillus sojae*. Such efficient production is disclosed in those strains having a protease profile superior to ATCC42251 or at least as good as any of ATCC9362, ATCC11906 and ATCC20387. The subject description thus reveals that some known strains of *A. sojae* are well suited already as such for production of proteins, polypeptides and metabolites. These *Aspergillus sojae* strains exhibit a lower proteolytic activity than the reference strain *A. sojae* ATCC42251. In particular the two known strains ATCC11906 and ATCC20387 are well suited. So preferred *A. sojae* strains for production of proteins, polypeptides and metabolites will be those expressing equal to or less proteolytic activity than the two preferred strains. Strain ATCC11906 is the best embodiment of the deposited ATCC *A. sojae* strains according to the prior art. Suitable proteins or polypeptides will be produced. Now that the subject invention has enabled introduction of nucleic acid sequences, such can serve to provide any protein or polypeptide of choice using an *A. sojae* as expression host.

The subject invention offers an improvement over existing expression systems. A number of existing protein production systems exhibit expression problems due to proteolysis. In particular the new system is better than the currently frequently applied expression systems *Aspergillus niger* and *Aspergillus awamori*. The subject invention now renders it possible to provide a recombinant *Aspergillus sojae* comprising a introduced nucleic acid sequence encoding a protein or a polypeptide for expression, said protein or polypeptide being susceptible to degradation upon expression by *A. niger* or *A. awamori*. The invention also provides a recombinant *A. sojae* comprising an introduced nucleic acid sequence encoding a protein or polypeptide for expression, said protein or polypeptide being other than *A. sojae* protease and amylase. A preferred embodiment is that wherein the introduced nucleic acid sequence encodes a non-*A. sojae* protein or polypeptide. Such recombinant *A. sojae* strains also fall within the scope of the invention.

In addition, illustration of *Aspergillus sojae* strains that have been modified in order to enhance their suitability as expression hosts is currently provided. These modifications can be reduced proteolytic activity as induced by any means. Specifically, the use of UV random mutagenesis is illustrated. Also specific mutation of one or more protease genes is illustrated. The means by which mutations can be introduced are common knowledge to the skilled person, and numerous alternative embodiments are thus readily available to arrive at the desired mutants. A suitable embodiment is formed by mutants in which alkaline proteolytic activity has been reduced. In particular elimination of activity of
specifically the major 35 kDa alkaline protease is illustrated as ensuring increased expression of proteins and polypeptides. Specifically the invention thus also covers novel strains exhibiting reduced proteolytic activity, specifically reduced alkaline proteolytic activity. Such strains are obtainable using any specific mutation route known or conceivable to the skilled person. A preferred embodiment of such expression hosts exhibiting reduced proteolytic activity as described above further comprises a selectable marker. Quite suitably the selectable marker will be amdS, pyrG or a combination thereof.

The invention in particular covers a method of producing protease deficient mutants of *Aspergillus sojae* by knocking out the 35 kDa alkaline protease gene. There are numerous ways in which this can potentially be achieved on the basis of the sequence data provided for this gene. In particular a method using recombination with a pyrG selection marker linked to two flanking regions eliciting cross over of the 35 kDa alkaline protease gene, whereby the resulting strain has the pyrG selection marker and misses the 35 kDa alkaline protease gene is an elegant one. Subsequently the pyrG selection marker can be eliminated, thus providing a 35 kDa alkaline protease negative *Aspergillus sojae* mutant that can be used for expression purposes of any desired sequence to be introduced therein. Naturally, the sequence to be introduced can have been incorporated in the previous steps already either on the same vector as the pyrG marker or in a cotransformation event. Also the method can be carried out analogously where a different protease gene than the 35 kDa alkaline protease gene is to be knocked out. The analogous measures to be taken are obvious to the skilled person on the basis of the illustration provided herein in combination with knowledge of other protease sequences. Also analogously the amdS selectable marker can be used in accordance with the invention as described elsewhere in this description.

Mutant fungi exhibiting improved fermentation characteristics are also provided as an additional aspect of the subject invention. Specifically, the invention is directed at a fungus comprising a mutation inhibiting the activity of proprotein convertase or an equivalent protein. Numerous proprotein convertases are known in the art. In particular we refer to Figure 1 providing sequence data of a number of such proteins. A fungus according to the invention is suitably selected from *Agaricus, Aspergillus, Trichoderma, Rhizopus, Mucor, Phanerochaete, Trametes, Penicillium, Cephalosporium, Neurospora, Tolypocladium* and *Thielavia*. Particularly suitable fungi are *Aspergillus niger, Aspergillus foetidus, Aspergillus sojae, Aspergillus awamori, Aspergillus oryzae, Trichoderma reesei, Penicillium chrysosporum, Cephalosporium acremonium, Neurospora crassa, Tolypocladium geodes and Thielavia terrestris*. A preferred embodiment covers the mutant
when it is an *Aspergillus sojae*, most particular preference is extended to *Aspergillus sojae*

as defined above according to the invention, i.e. comprising heterologous nucleic acid

sequences, e.g. in combination with the selectable markers amdS and/or pyrG.

A suitable equivalent of a proprotein convertase is a protein or polypeptide

exhibiting an amino acid sequence with more than 40%, preferably more than 45% 
similarity or identity with the inferred amino acid sequences of the DNA sequences given 
in SEQ ID No. 3 (= gene fragment encoding *A. niger* proprotein convertase amino acid 
sequence), SEQ ID No. 4 (= partial gene fragment encoding *Aspergillus sojae* proprotein 
convertase amino acid sequence) or with any of the sequences given in Seq ID Nos. 5 to 9.

The functionally equivalent protein may suitably have a nucleic acid sequence capable of 
hybridising under stringent conditions to a nucleic acid sequence according to SEQ ID 
Nos. 3 to 9. Stringent hybridisation conditions can readily be determined by the skilled 
person. A suitable example of stringent hybridisation conditions are hybridisation at 50°C 
and preferably at 56°C and final washes at 3xSCC. *PE4*, *PCL1* and *PCL2* are specifically 
mentioned as examples of suitable oligonucleotide mixtures corresponding to the coding 
strand (i.e. SEQ ID Nos. 10, 11 and 12). For the noncoding strand *PE6*, *PCL2*-rev, *PCL3* 
and *PCL4* are mentioned (i.e. SEQ ID Nos. 13, 14, 15 and 16, respectively). Use of these 
primers in amplification procedures common in the art will provide equivalent sequences 
and such use and the resulting newly found sequences and application thereof in the 
manner analogous to that described in the subject description fall within the scope of the 
invention. The sequences for which the oligonucleotides were made are well conserved as 
could be determined from comparison of the various amino acid sequences for the proteins 
provided (see Figure 1). Any other nucleic acid sequences exhibiting the same or higher 
degree of identity, similarity or homology with the sequences provided in the subject 
patent application for the proteins or relevant active parts thereof are covered by the 
invention as is the use thereof as primers or probes to find other proprotein convertase or 
equivalent protein encoding sequences and/or for subsequently introducing mutations in 
such protein encoding sequences. By way of example Maniatis *et al.* (1982) Molecular 
Cloning, A Laboratory manual, Cold Spring Harbor Laboratory, New York or any other 
handbook on cloning and/or screening nucleic acid sequences has been referred to. The 
equivalent protein or polypeptide will exhibit the activity of a proprotein convertase as the 
one having an amino acid sequence according to SEQ ID Nos. 3 to 9. The mutant fungus 
can comprise a substitution, insertion or deletion in the encoding sequence of the 
proprotein convertase or equivalent protein. The mutant fungus can suitably comprise a
mutation in the regulation of the expression of the gene encoding proprotein convertase or equivalent protein. A mutant fungus according to the invention in a suitable embodiment exhibits reduced viscosity vis a vis the corresponding non mutated fungus under equivalent cultivation conditions. A mutant fungus according to any of the above embodiments exhibiting increased expression of a desired introduced nucleic acid sequence encoding a protein or polypeptide is included within the scope of the invention, said fungus exhibiting increased production of a protein or polypeptide under equivalent conditions vis a vis the corresponding wild type fungus. The activity site for the A. sojae proprotein convertase has been ascertained to be comprised within the amino acid sequences inferred by SEQ ID Nos. 3 and 4.

A process for producing a phytase or protein having phytase activity or any other protein or polypeptide, preferably a recombinant phytase or any other heterologous protein or any other polypeptide, said process comprising cultivating a mutant fungus according to any of the embodiments described above falls within the scope of the invention. A process for obtaining the resulting protein or polypeptide either from the cell as such or after secretion thereof from the cell is also included.

The use of any of the described novel strains for transformation of any nucleic acid sequence encoding a phytase or protein having phytase activity or any other protein or polypeptide thereto and any subsequent expression of any nucleic acid sequence introduced therein and also optionally any following processing and/or secretion and/or isolation is covered by the invention.

Any phytase or phytase-like or any other heterologous protein or polypeptide encoding sequence can suitably be used. This can be of fungal or non-fungal origin. A preferred embodiment is formed by acid labile protein or polypeptide encoding sequences. Suitably the protein encoding sequence encodes non protease-like proteins. The examples show a phytase sequence and a number of heterologous sequences suitable for use in transformation and also for expression in Aspergillus sojae hosts. Further examples of suitable proteins to be expressed are obvious to a person skilled in the art.

The invention is further illustrated by the examples below. The examples are not to be considered restrictive to the interpretation of the scope of the invention. Alternative embodiments are readily envisageable to the skilled person on the basis of the description and knowledge of the relevant field of technology. The content of references cited in the description are incorporated by reference. The claims serve to illustrate the intended scope of the invention.
EXPERIMENTAL DETAILS CONCERNING THE INVENTION
CONSTRUCTION OF AN *Aspergillus sojae* GENE LIBRARY.

Genomic DNA of *A. sojae* was isolated from protoplasts obtained from ATCC11906 using a previously described protocol (Punt, van den Hondel, 1992). After isolation, DNA was extracted from the protoplasts using the protocol described by Kolar et al., 1988. Subsequently the DNA was partially digested with MboI to result in DNA fragments of an average size of 30-50 kb.

Vector pAOpyrGcosarp1, which was used for the construction of the gene library, was constructed by ligation of a 3 kb BamHI-HindIII fragment from pANsCos1 (Osiewacs, 1994) and a 3.2 kb Acc651-HindIII fragment from pAO4.2 (De Ruiter-Jacobs, 1989) in Acc651-BamHI digested pHELP1 (Gems et al., 1991). This cosmid vector carries the *A. oryzae* pyrG selection marker and is self-replicating in filamentous fungi. MboI digested genomic DNA was ligated to BamHI-digested pAOpyrGcosarp1, and the ligation mixture was packaged into phage particles using the Stratagene Supercos1 vector kit. In total 30,000 individual clones were obtained representing an approximate 30-fold representation of the *A. sojae* genome. Stocks (in 15% glycerol) of pools of the resulting clones were stored at -80°C for later use.

AMDS TRANSFORMATION METHOD AND TRANSFORMANTS.

Two currently used protoplasting protocols and transformation protocols [the modified OM-method (Yelton et al., P.N.A.S. 81 (1984) 1470-1474) and the NaCl-method (Punt and Van den Hondel, Meth. Enzym. 216 (1993) 447-457)] were tested on the *Aspergillus sojae* strain ATCC9362. Both methods resulted in protoplasts, but the yield of viable protoplasts with the OM-method was clearly better. The overall yields were lower than normally obtained for *A. niger*. A pilot protoplasting/transformation experiment was carried out with all *A. sojae* strains using the OM method.

For transformation, vector p3SR2 (carrying the amdS marker) was used in combination with pAOpyrGcosARP1. This latter vector is a derivative of the autonomously replicating *Aspergillus* vector Arp1, which in all *Aspergillus* species tested so far, resulted in highly increased numbers of (instable) transformants when used as a cotransforming vector. For nearly all strains sufficient protoplasts (about 10E6-10E7 per transformation) were obtained. Analysis of appropriate AmdS selection conditions for the various *A. sojae* strains revealed vigorous growth of most strains on the commonly used
selective acetamide medium. Clearly, the acetamide selection conditions proposed for A. sojae amdS transformants as reported in WO97/04108, were not appropriate for the selection of A. sojae transformants. Our experiments revealed, surprisingly, that AmdS+ transformants could only be selected with acrylamide selection. Even on selective acrylamide plates, a considerable background from non-transformed protoplasts was observed. Selection of primary transformants requires around three weeks and many of the initially selected putative transformants turned out to be false positives, only showing background growth after transfer to fresh selective acrylamide plates. To optimize selection of transformants attempts had to be made to reduce this background growth. Improved results were obtained by omitting glucose from the selective plates. In Table 3 the composition of the improved selection medium and the usual media is given. Figures 2a, b and c show the background growth observed for selected strains on the selection medium described in WO97/04108 and the improved acrylamide selection medium described in Table 3.

Further transformation experiments with the three selected A. sojae strains revealed that protoplasting efficiencies for ATCC11906 and ATCC20387 were better using the NaCl-method. Successful protoplasting was obtained using various commercially available protoplasting enzyme preparations such as NOVOZYM, Caylase, Glucanex, etc. Based on the NaCl transformation protocol the three selected A. sojae strains were transformed with amdS selection vector p3SR2 or derivatives thereof. Using the modified acrylamide selection plates a number of vigorously growing transformants were obtained, while no growth was observed in a control transformation without DNA. Another approach to circumvent background growth of non-transformed mycelium is the elimination of the activity of the wild type A. sojae amdS gene. This can be achieved for example by disruption of the A. sojae amdS gene. As a first step specific DNA fragments carrying ATCC11906 amdS sequences were PCR-amplified using primers derived from published A. oryzae amdS sequences (Gomi et al.; 1991, Gene 108, 91-98). Previous experiments had shown that cloning by stringent hybridisation would be unsuccessful due to a low level of sequence conservation between A. nidulans and A. sojae amdS sequences. The expected fragment of about 1.6 kb, which should carry most of the coding region of the amdS gene, was obtained. Sequence analysis from both ends of the cloned PCR fragment (Figures 3a and 3b) confirmed the cloning of a part of the A. sojae amdS gene. The stringent hybridisation occurred at 56°C with final washes at 3xSSC. The cloned sequence was very similar to the published A. oryzae amdS sequence. Several hybridising clones (7 out of
10,000) were isolated from the ATCC11906 cosmid library in pAOpyrGcosarp1 using the cloned ATCC11906 amdS fragment as a probe. After subcloning a fragment carrying the complete amdS gene, a part of the amdS gene was replaced by a re-usable pyrG selection marker to generate an amdS replacement vector. Transformation of this vector to Aspergillus sojae ATCC11906PyrG resulted in pyrG+ transformants. After subsequent analysis of these transformants on acetamide and acrylamide selection plates several of these transformants showed reduced background growth. Southern analysis of a few of these strains revealed that the expected gene replacement had occurred. One of these strains was used for subsequent transformation with the A. nidulans amdS gene using acrylamide selection plates and resulted in a number of amdS+ transformants.

PYRG TRANSFORMATION METHOD AND TRANSFORMANTS.

(1) Initial experiments

For A. sojae, the standard experiments used in the prior art for other fungi to generate pyrG mutants as described in the introduction resulted in numerous fluo-orotic acid (FOA)-resistant strains. However, all of these strains were able to grow on medium without uridine and were therefore not considered pyrG mutants. With our final aim to isolate the appropriate mutant strains a number of alternative approaches were followed.

(2) Near-homologous gene disruption

Based on the expectation that the pyrG genes from A. sojae and A. oryzae are very similar in sequence (which was confirmed by Southern hybridisation carried out under stringent conditions), experiments were carried out to disrupt the A. sojae pyrG gene with a mutant version of the A. oryzae pyrG gene using an approach previously described by Gouka et al. (1996). The stringent hybridisation occurred at 65°C with final washes at 0.3 x SSC. An A. oryzae pyrG disruption vector was constructed in which an 0.5 kb Clai fragment carrying part of the pyrG coding region was deleted (Figure 4). The XbaI pyrG fragment from this new vector was used for transformation and direct selection for FOA resistant transformants. None of the FOA resistant colonies obtained was uridine requiring.

(3) UV mutagenesis and filtration enrichment

Another approach to improve the yield of specific mutant strains is the use of a filtration-enrichment step (Bos et al. 1986, Thesis, Agricultural University Wageningen). The UV mutagenized spores are used for inoculation of a minimal medium (MM) liquid
culture. From the resulting repeated overnight culture those spores unable to germinate in minimal medium (a.o. pyrG mutant spores) are separated from the grown mycelium by filtration through myracloth. The spores obtained after several enrichment steps were tested for their PyrG phenotype, by inoculating the spores on plates containing FOA. Again none of the resulting FOA resistant colonies was uridine requiring. Also none of the colonies obtained after this enrichment on MM plates containing uridine was shown to be uridine requiring.

(4) Modified selection conditions

Our previous attempts to isolate pyrG mutants from A. sojae had failed suggesting the inability of the required pyrG mutants to utilize exogenous uridine, which is used in the FOA selection medium for the analysis of uridine auxotrophy. A modified selective FOA medium, now containing uracil next to uridine, was used in a new isolation attempt. From this approach several FOA resistant mutants were obtained which were uracil requiring. Retesting of these strains showed that these were unable to grow on uridine supplemented minimal medium. Subsequent transformation experiments with some of the uracil-requiring strains showed that these mutants could indeed be complemented with a fungal pyrG gene (e.g. vector pAB4.1; A. niger pyrG). The inability of pyrG mutants to grow in minimal medium supplemented with only uridine was an unprecedented observation for related Aspergillus species (A. nidulans, A. niger, A. oryzae) and various other fungal species.

(5) Re-usable selection marker

Versatile genetic modification of A. sojae requires the possibility to modify, disrupt and express a number of different genes in a single fungal strain, which would require the availability of a (series) of different selection markers. However, the availability of a marker such as pyrG, which allows selection of both the mutant (FOA selection) and the transformant (Uracil-less medium), provides the possibility of repeated use of the same marker in subsequent experiments. For this approach a pyrG marker gene was designed, in which the complementing sequence was flanked by a direct repeat sequence originating from the 3' flanking end of the pyrG gene. The resulting plasmid is pAB4-1rep. The construction of this vector is detailed in Figure 5. The full sequence of the vector is given in SEQ ID No. 17. Transformation of A. sojae pyrG mutants with this vector results in a similar number of PyrG+ transformants as with the vector pAB4-1. However, subsequent
plating of spores of selected pAB4-1 and pAB4-1rep transormants to tFOA selection plates resulted in many more FOA resistant/uracil requiring colonies for the pAB4-1rep transformant. Southern analysis of these FOA resistant/uracil requiring clones showed that in most of the pAB4-1rep strains the A. niger pyrG marker gene had been deleted leaving only the small 0.7 kb repeat region at the locus of integration, while in the pAB4-1 strains the A. niger gene was still present and had presumably acquired a mutation resulting in the pyrG-negative phenotype.

EXPRESSION HOSTS: STRAIN SELECTION.

Protease production

Very important characteristics of a fungal expression system are the level and type of fungal proteases produced under various culture conditions. Sometimes strains which can be readily transformed are not suitable as expression hosts due to production of proteases or acidification of the culture media which is detrimental to the expressed product. Analysis of the growth behaviour of the various A. sojae strains revealed that, in contrast to what was observed for A. niger, acidification of the culture medium did not occur either on agar based plates (MacConkey) or in shake flask cultures. In fact in shake flask cultures, irrespective of the three medium types analyzed (Table 4), in most cases even an alkaline pH was obtained in the cultures. Based on these results and literature data it is thus expected that primarily alkaline proteases will be present in the A. sojae culture fluid. To analyse protease activity of the culture fluids of the various strains, a milk clearing assay was performed. In addition medium samples were incubated with different proteins (e.g. bovine serum albumin (BSA)), and degradation of these proteins was followed in time in order to assess the suitability of the tested strains as expression hosts for a range of products. BSA was chosen as in our previous experiments with A. niger. This protein was shown to be very susceptible to proteases. A. terreus phytase was chosen as example of another proteolytically instable protein. Degradation of milk proteins as shown by the formation of a milk clearing zone at the periphery of growing colonies is a generally accepted criterion for protease activity. Detection of BSA was carried out by Coomassie staining of SDS-PAGE gels. For phytase, Western analysis using specific antibodies, was carried out. As shown in Table 4, clear differences of degradation in A. niger culture fluid are evident when this is compared with that in A. sojae culture fluid. In A. niger culture fluid (pH 3-4) rapid degradation of BSA occurs. In A. sojae culture fluids
from richer media, degradation of BSA occurs, albeit less than in *A. niger* culture fluid. In most *A. sojae* culture fluids (pH 7-8) rapid degradation of *A. terreus* phytase occurs, with the exception of ATCC9362, ATCC11906 and ATCC20387 culture fluids. In general, the strains with the lowest phytase degradation also show low BSA degradation under the conditions tested. In particular the two *A. oryzae* strains ATCC20235 and ATCC46250 show much higher proteolytic activity than most *A. sojae* strains.

To exclude that differences in the pH of the culture fluid cause the observed effects, similar degradation experiments were also carried out with culture fluids of which the pH was adjusted to pH 4.5 (50 mM Na/HAc), pH 5.8 (50 mM Na/HAc) and pH 8.3 (50 mM Tris/HCl). Table 5 gives the degradation data obtained with these samples. As can be seen in the table *A. oryzae* ATCC20235, which had the highest proteolytic activity at pH 7-8 also shows high proteolysis at other pH values. Degradation of *A. terreus* phytase occurs primarily at pH 8. Similarly to what was found before, ATCC11906 and ATCC20387 showed low phytase degrading activity. ATCC9362 showed phytase degradation in rich media. BSA degradation by *A. sojae* showed no significant differences with the data presented in Table 4.

In conclusion, these protease assays resulted in the identification of three low protease *A. sojae* strains, namely ATCC9362, ATCC11906 and ATCC20387. Thus, *A. sojae* can clearly be used as expression host for a range of proteins and provides a series of advantages over prior art transformation and expression systems.

**STRAIN IMPROVEMENT**

Once the potential of transformability and expression had been ascertained for *Aspergillus sojae*, means by which additional strains could be created with enhanced characteristics for expression were considered. Two different approaches which can be used as such or in combination were developed to provide novel improved strains for expression of proteins.

On the one hand the possibility of developing protease deficient mutants was investigated and the impact of such on levels of expression was assessed. On the other hand strains with amended morphology were developed with a view to improve fermentation characteristics. To achieve this a hitherto non-disclosed or suggested route was followed which is applicable not only to *Aspergillus sojae* but to *Aspergilli* and in fact to fungi in general.
Development of protease deficient mutants

To obtain protease deficient *A. sojae* strains two approaches were followed. In a first approach spores from ATCC11906 and ATCC11906-derived strains were mutagenized with UV. In a second approach gene disruption of the major alkaline protease was carried out.

UV mutants

Freshly harvested spores from *A. sojae* ATCC11906 or one of its pyrG derivatives was UV-mutagenized in a Biorad UV-chamber with a dose resulting in 20-50 % survival. Serial dilutions were plated onto skim-milk plates (Mattern et al., 1992). From 5000 UV-surviving strains four mutant strains with a considerably reduced milk-clearing halo were obtained

AlpA gene disruption

In this approach the endogenous alpA (alkaline protease) gene cloned from ATCC11906 was disrupted using a disruption vector carrying the re-usable pyrG selection marker as described in this description.

An ATCC11906 cosmid library (in a PyrG cosmid) was constructed. From 10,000 independent cosmid clones initially 4 were found to hybridize under homologous conditions with an *A. sojae* alpA fragment obtained by PCR with primers MBL1784 and MBL1785. Rescreening of the 4 clones revealed only strong hybridisation with one clone. A 4 kb EcoRI and a 2.5 kb HindIII fragment from this clone, together expected to carry the complete gene, were subcloned and characterised by restriction enzyme digestion and sequence analysis. Based on these subclones a new gene-replacement vector was constructed as described in Figure 6. For transformation of an ATCC11906pyrG derivative the vector was digested with EcoRI, and the 8.7 kb alpA deletion fragment was used for transformation (see Figure 6). Transformation of the replacement cassette to ATCC11906pyrG5 resulted in a number of transformants with a reduced milk-halo. Southern analysis of these strains revealed the successful deletion of the alpA gene. To allow subsequent use of the pyrG marker for transformation of one of these strains, spores from this strain were plated on FOA containing medium selective for pyrG mutants. From strains with the correctly integrated disruption cassette with the re-usable pyrG marker a large number of FOA resistant colonies were obtained. In contrast to the results obtained for spontaneous FOA resistant mutants of wild type strains, the FOA strains obtained from
these disruption strains were virtually all uracil requiring and turned out to be PyrG negative again. Southern analysis was used to confirm the desired removal of the pyrG marker gene at the alpA locus, leaving only the 700 bp "footprint".

Analysis of protease activity in UV and disruption mutants

To analyse the levels of protease production in the different low protease derivatives of ATCC11906 controlled batch fermentation experiments were carried out. From the resulting culture supernatants protease activities were determined at various pH values. Deletion of the alpA gene resulted in a strong reduction of proteolytic activity at alkaline pH. Analysis of the protease activity in one of the UV mutants showed almost complete absence of proteolytic activity at both pH 6 and pH 8. Consequently the level of proteolysis towards secreted proteins produced in these strains was considerably less than observed for the parent strain.

Development of low viscosity mutants

Initial controlled batch or fed batch fermentation trials with A. sojae resulted in considerable biomass yield, however both the culture viscosity and sporulation phenomena in the fermenter vessel represented less favourable characteristics.

Therefore attempts were made to improve these characteristics in the desired host strain. Various patent applications teach that low viscosity mutants can be isolated by various ways of screening. WO96/02653 and WO97/26330 describe non defined mutants exhibiting low viscosity. However here we describe a new unexpected case of a completely characterised and fully defined low viscosity mutant from A. sojae. It was found that a proprotein processing mutant from this organism had an unexpected aberrant growth phenotype (hyper-branching) while no detrimental effect on the productivity of proteins was observed. Controlled fermentation experiments with this strain revealed that increased biomass concentrations were obtained at considerably lower viscosity values. The observed characteristics were not only present in A. sojae but other fungi as well, e.g. in A. niger.

(1) Construction of an A. niger proprotein processing mutant

To clone the proprotein convertase encoding gene from A. niger, heterologous hybridisation using specific probes from the Saccharomyces cerevisiae KEX2, Schizosaccharomyces pombe KEX1 and the Xenopus laevis PC2 genes was carried out.
However, no specific hybridisation signals were obtained even at very low stringency hybridisation conditions (47°C, washes at 6xSSC), precluding the use of this approach to clone the corresponding *A. niger* gene.

As an alternative approach to clone the proprotein convertase encoding gene from *A. niger*, PCR was used. Based on the comparison of various proprotein convertase genes from various yeast species and higher eukaryotes (Figure 1) different PCR primers were designed (SEQ ID Nos. 10, 13 and 18-23) which are degenerated, respectively, 2048, 49152, 4, 2, 2, 512, 2048, and 4608 times. From the amplification using primers *PE4* and *PE6*, two individual clones were obtained of which the encoded protein sequence did show significant homology to the *S. cerevisiae KEX2* sequence (SEQ ID No. 24). These clones were used for further experiments.

Based on the observed homology to other proprotein convertase genes of the cloned PCR fragment, the corresponding *A. niger* gene was designated *pclA* (from proprotein convertase-like). Southern analysis of genomic digests of *A. niger* revealed that the *pclA* gene was a single copy gene with no closely related genes in the *A. niger* genome, as even at stringent hybridisation conditions (50°C; washes at 6xSSC), no additional hybridisation signals were evident. A first screening of an EMBL3 genomic library of *A. niger* N401 (van Hartingsveldt et al., 1987) did not result in any positively hybridising plaques although about 10-20 genome equivalents were screened. In a second screening a full length genomic copy of the *pclA* gene was isolated from an *A. niger* N400 genomic library in EMBL4 (Goosen et al., 1987). Of the 8 hybridising plaques which were obtained after screening 5-10 genome equivalents, 6 were left after a first rescreening. All these 6 clones most likely carried a full copy of the *pclA* gene, as in all clones (as was observed for the genomic DNA) with the PCR fragment two hybridising *EcoRV* fragments of 3 and 4 kb were present (note that the PCR fragment (SEQ ID No. 24) contains an *EcoRV* restriction site). Based on a comparison of the size of other proprotein convertases, together these fragments will contain the complete *pclA* gene with 5'- and 3'-flanking sequences. The two *EcoRV* fragments and an overlapping 5 kb *EcoRI* fragment were subcloned for further characterisation. A detailed restriction map of the DNA fragment carrying the *pclA* gene is given in Figure 7.

Based on the restriction map given in Figure 7 the complete DNA sequence of the *pclA* gene was determined from the *EcoRI* and *EcoRV* subclones (SEQ ID No. 3). Analysis of the obtained sequence revealed an open reading frame with considerable similarity to that of the *S. cerevisiae KEX2* gene and other proprotein convertases. Based
on further comparison two putative intron sequences (SEQ ID No. 3, from position 1838 to 1889 and from 2132 to 2181) were identified in the coding region. Subsequent PCR analysis with primers flanking the putative introns, on a pEMBLyex based A. niger cDNA library revealed that only the most 5' of these two sequences represented an actual intron. The general structure of the encoded PclA protein was clearly similar to that of other proprotein convertases (SEQ ID No. 25 and Figure 8). The overall similarity of the PclA protein with the other proprotein convertases was about 50% (Figure 1).

To demonstrate that the cloned pclA gene is a functional gene encoding a functional protein, the construction of strains devoid of the pclA gene was attempted. Therefore, pPCL1A, a pclA deletion vector, in which a large part of the pclA coding region was replaced with the A. oryzae pyrG selection marker gene, was generated. Subsequently, from this vector the 5 kb EcoRI insert fragment was used for transformation of various A. niger strains. From these transformations (based on pyrG selection) numerous transformants were obtained. Interestingly, a fraction of the transformants (varying from 1-50%) displayed a very distinct aberrant phenotype (Figure 9). Southern analysis of several wild-type and aberrant transformants revealed that these aberrant transformants, which displayed a severely restricted growth phenotype, had lost the pclA gene. All strains displaying wild-type growth were shown to carry a copy of the replacement fragment integrated adjacent to the wild-type pclA gene or at a non-homologous position.

Analysis of protein production in selected pclA mutant strains carrying various glucoamylase fusion genes revealed the production of unprocessed fusion protein. The production of high levels of unprocessed glucoamylase-interleukin-6 fusion protein in a pclA mutant was achieved. Protein analysis revealed that in pclA mutant strains also no fully processed endogenous glucoamylase is formed but only pro-glucoamylase.

To further improve the yields of fusion proteins controlled batch and fed-batch fermentations were also carried out. Surprisingly the fermentation characteristics of pclA mutant strains were clearly superior to those of the parent strain resulting in a much reduced viscosity/biomass ratio, without loss of productivity.

(2) Construction of an A. sojae proprotein processing mutant

To construct the corresponding mutant in A. sojae, functional complementation of the low-viscosity mutant of A. niger genomic cosmid clones from the ATCC11906 cosmid library were isolated, which comprised the A. sojae proprotein processing protease pclA gene. Partial sequence analysis of the isolated sequences SEQ ID No. 4 confirmed the
cloning of the A. sojae pclA gene. Figure 10 shows the comparison of the DNA sequences of part of the A. niger and A. sojae pclA genes. Based on the A. sojae sequence and a partial restriction map with the coding region of the A. sojae pclA gene, a replacement vector was generated using the EcoRV-site in the A. sojae pclA gene to clone the re-usable pyrG marker as a SmaI fragment inside (Figure 11). The resulting vector was partially digested with ClaI to obtain the delta-pcl fragment of 10.5 kb (see Figure 11). This fragment was isolated to be used for transformation of A. sojae pyrG strains. The gene replacement vector was used to generate pclA mutants in ATCC11906 and ATCC11906 derivatives. The resulting strains were used for the expression of homologous and heterologous proteins. Controlled fermentation experiments with some of the resulting transformants revealed improved fermentation characteristics, in particular a lower viscosity/biomass ratio of the culture.

(3) Cloning of fungal genes homologous to Aspergillus pclA

Based on the comparison of amino acid sequences inferred from the A. niger and A. sojae pclA genes with those of other proprotein processing enzymes (Figure 1) several oligonucleotide mixtures corresponding to the coding or non-coding strand of well conserved sequences were designed (SEQ ID Nos. 10 to 16).

These oligonucleotide mixtures were used in PCR with chromosomal DNA from Trichoderma reesei QM9414, Fusarium venenatum ATCC20334, Penicillium chrysogenum P2, Trametes versicolor, Rhizopus oryzae ATCC200076, and Agaricus bisporus HORST. Depending on the template DNA used, PCR amplifications (30 cycles; 1 min 94°C; 1 min 40°C; 2 min 68°C) with one or more of the combinations of coding and non-coding strand oligonucleotides resulted in specific PCR products. Table 6 gives the results of the various amplification reactions. Sequence analysis was carried out with a number of the obtained PCR fragments using either of the two oligonucleotide mixtures used for amplification. These analyses resulted in the identification of the various pclA homologues from these different fungi. Figure 12 gives the inferred aminoacid sequences corresponding with the various DNA fragments (SEQ ID Nos. 5 to 9).

(4) Examples of biomass and viscosity determinations

The following operating parameter data ranges have been determined for fungal fermentations using a number of different fungal strains.
Viscosity:

Viscosity is determined on a Haake Viscotester VT500 using sensor system MV DIN (vessel number 7), operated at 20°C. Obtain a fresh sample of fermentation broth and place 40 ml of the broth in the measuring cell. After a small period of equilibration (4 min) at a set spindle speed the viscosity is measured. This measurement is repeated for ten different spindle speeds. Multiplication of the spindle speed with the measuring cell factor results in the shear rate. Viscosity \( \eta \) (in centipoise = cP) is plotted against shear rate \( \gamma \) (1/s) and gives the viscosity profile of the fermentation broth.

Viscosity ranges have been determined for fermentations using the specified fungal organism using the above procedure (Table 7).

Biomass:

Biomass is determined by the following procedure:

Preweigh 5.5 cm filter paper (Whatman 54) in an aluminium weighing dish. Filter 25.0 ml whole broth through the 5.5 cm paper on a Buchner funnel, wash the filter cake with 25.0 ml deionised water, place the washed cake and filter in a weighing pan and dry overnight at 60°C. Finish drying at 100°C for 1 hour, then place in desiccator to cool. Measure the weight of dried material. Total biomass (g/l) is equal to the difference between the initial and final weights multiplied by 40.

Protein:

Protein levels were determined using the BioRad Assay Procedure. The data presented above represent values determined 48 hours into the fermentation process until fermentation end; all values of Aspergilli and Trichoderma are for commercially relevant fungal organisms and reflect actual commercial data.

A fungal strain such as A. sojae lfiA and A. sojae pclA has the advantage that the low viscosity permits the use of lower power input and/or shear the in the fermentation to meet oxygen demands for those cases where shear stress on the product may be detrimental to productivity due to physical damage of the product molecule. The lower biomass production at high protein production indicates a more efficient organism in the conversion of fermentation media to product. Thus A. sojae mutants provides better biomass and viscosity data whilst also delivering at least as much protein, and in fact a lot more protein than the two commercially used systems which obviously are better than for typically deposited Aspergillus or Trichoderma reesei strains in general public collections.
The high protein production with low biomass concentration produced by *A. sojae* _lfvA_ would allow development of fermentation conditions with higher multiples of increase in biomass, if increasing biomass results in increased productivity, for the desired product before reaching limiting fermentation conditions. The present high levels of biomass and viscosity produced by the *T. longibrachiatum* and *A. niger* organisms restrict the increase of biomass as the present levels of biomass and viscosity are near limiting practical fermentation conditions.

**EFFICIENT GENE EXPRESSION**

(1) *Heterologous regulatory sequences*

The three selected *A. sojae* strains were cotransformed with three GUS reporter vectors carrying different fungal expression signals (*A. nidulans* PgpdA; pGUS54, *A. niger* PgliA; pGUS64, *A. niger* Pbipa; pBIPGUS) and the amdS selection vector p3SR2 or derivatives thereof. Expression of the GUS gene was analysed in representative transformants (Table 8). From the results it is clear that under the conditions tested the *gpdA* promoter was by far the best promoter resulting in about 5000 U GUS/mg protein. This corresponds to about 5% of the total amount of cellular protein. The *bipA* promoter results in about 30% of the *gpdA* activity, which corresponds to expression data obtained in *A. niger*. Surprisingly, the *glaA* promoter which is very active in *A. niger* (at least as active as *gpdA*) results in less than 1% of the *gpdA* activity in *A. sojae*.

(2) *A. sojae regulatory sequences*

We also isolated an *Aspergillus sojae* homologous promoter and assessed the applicability of such in an expression system. In some instances of expression it will be preferable to use a homologous promoter rather than a heterologous promoter. It was also interesting to assess whether the homologous promoter would be more efficient than a heterologous one.

PCR cloning of three efficiently expressed *A. sojae* genes, namely *alpA* (alkaline protease; inducible), *amyA* (amyrase; inducible) and *gpdA* (glyceraldehyde-3-phosphate dehydrogenase; constitutive) was attempted using primers based on sequences available from *A. oryzae* (SEQ ID Nos. 26 to 31). Figures 13 a, b and c give the sequences and the position in published *A. oryzae* sequences of the various PCR primers used for this approach. Genomic template DNA from *A. sojae* ATCC11906 was used for PCR amplification. Initial PCR amplifications (30 cycles; 1 min 94°C; 1 min 40°C; 2 min 68°C) resulted in a specific PCR product of the expected size (400 bp) for the *gpdA*. For the other
two PCR reactions no product was obtained. Therefore, for alpA PCR conditions were made less stringent (10 cycles; 1 min 94°C; 1 min 25°C; 2 min 68°C + 20 cycles; 1 min 94°C; 1 min 40°C; 2 min 68°C), which resulted in a specific alpA PCR product of about 1000 bp.

The complete sequence of the cloned genes was determined. As shown in Figure 14 the A. sojae ATCC11906 gpdA promoter region has a very high homology with other gpdA promoter sequences and the alpA promoter was virtually identical to the A. oryzae alpA promoter (SEQ ID Nos. 32 and 33). Expression vectors carrying expression cassettes comprising these A. sojae promoters show significant levels of gene expression.

HETEROLOGOUS PROTEIN PRODUCTION

A number of heterologous proteins were tested which were known to be susceptible to acidic proteolysis and thus could not be expressed efficiently in other well known expression systems. Also proteins that are already efficiently expressed in alternative systems were tested in order to assess by way of comparison the levels of expression achieved with Aspergillus sojae vis a vis other known expression systems such as Aspergillus niger.

Phytase production

DNA fragments carrying various fungal phytases (Wyss et al. (1999) Appl. Environ. Microbiol. 65, 359-366) were ligated as 5' NcoI or BsrHI sites introduced at the ATG codon - 3' blunt-ended fragments downstream of the A. nidulans gpdA promoter in pAN52-1NotI. The resulting vectors were used in cotransformation experiments of A. sojae using the amdS and or pyrG selection marker. Phytase producing transformants were screened using the described phytase plate-assay.

Further improved phytase expression vectors were generated using a multicopy cosmid approach. In this approach several copies of a phytase expression cassette, recloned in a multiple cloning site vector (pMTL series, Chambers et al., (1988) Gene 68, 139-149) to allow its isolation as a EcoRI fragment. Several copies of these EcoRI fragments were cloned into cosmid vector pAN4cos1 through packaging (Verdoes et al. (1993) Transgenic Research 2, 84-92), resulting in cosmid clones carrying a number of expression cassettes. The resulting clones were introduced into A. sojae using the amdS selection marker. AmdS+ clones were screened for phytase production using the phytase plate-assay.
Further phytase expression vectors were generated using the GLA fusion approach (e.g. Broekhuijsen et al. (1993) J. Biotech. 31, 135-145). To this end phytase gene fragments, encoding the mature \textit{A. fumigatus} phytase protein were fused, using convenient restriction sites and fusion PCR, to the 3'-end of the glucoamylase carrier gene in vector pAN56-1 (Genbank accession number Z32700). Between the glucoamylase and phytase part of the gene-fusion a sequence encoding a proprotein processing site (Asn-Val-Ile-Ser-Lys-Arg) was introduced. The resulting vectors were used in cotransformation experiments of \textit{A. sojae} using the amdS and/or pyrG selection marker. Phytase producing transformants were screened using the described phytase plate-assay.

Shake flask fermentation was carried out resulting in significant levels of active phytase. Yield were significantly higher than those reported in literature for \textit{A. niger} (van Hartingsveldt et al. (1993) Gene 127, 87-94; Van Gorcom et al. (1991) EP420358). On average, the levels obtained with the multicopy cosmid vectors were higher than those obtained with the single copy vectors. Phytase levels obtained with the glucoamylase-phytase fusion vectors resulted in high levels of both glucoamylase and phytase. Controlled batch and fedbatch fermentations from a selected number of phytase producing \textit{A. sojae} transformants revealed a further increased level of phytase.

Glucoamylase production

An example of an efficiently produced fungal protein is provided by the expression of the \textit{A. niger glaA} gene. Vector pGLA6S (Figure 15) is derived from pGLA6 (Punt et al. (1991) J. Biotech. 17, 19-334) by introducing a 5 kb EcoRI fragment carrying the \textit{A. nidulans} amdS gene as selection marker into the unique EcoRI site of pGLA6. Vector pGLA6S (Figure 15) carrying the amdS selection marker and the glucoamlyase gene under control of the \textit{A. nidulans gpdA} promoter was introduced into \textit{A. sojae} ATCC11906pyrG using cotransformation with vector pAB4.1. Starch plate-assays demonstrated the production of increased levels of amylolytic activity in these transformants. From the resulting transformants those showing proficient growth on acrylamide medium were analysed for glucoamylase production. On a Coomassie Briljant Blue-stained SDS PAGE gel from the culture supernatant of some of these transformants a single dominant protein band corresponding to glucoamylase was observed. Western analysis using a monoclonal antibody against glucoamylase (Verdoes et al. (1993) Transgenic Research 2, 84-92) was used to confirm the identity of this protein band.
Interleukin-6 production

Production of interleukin-6, which is an example of a highly sensitive protein for proteolytic degradation, was shown to be virtually impossible in A. niger without the use of the gla-fusion strategy and protease deficient strains. Even with all these improvements the yields of IL-6 were only a few mg per litre culture fluid. Introduction of the IL-6 vector pAN56-4 (Broekhuijsen et al. (1993) J. Biotech. 31, 134-145) into A. sojae by cotransformation with the pyrG or amdS marker resulted in transformants expressing the IL-6-fusion gene present in this vector. From the resulting transformants a few were selected for further analysis. Shake flask fermentation experiments were carried out with these transformants. SDS-PAGE and Western analysis of culture supernatants of several of these strains surprisingly showed levels of correctly processed IL-6 which were about 5-10 fold higher than the levels obtained in the best reported cases in A. niger. The use of the various types of protease deficient and fermentation-optimized mutants from A. sojae further increased the level of IL-6 production to be obtained from controlled fermentations (Broekhuijsen et al. (1993) J. Biotech. 31, 134-145).

Green fluorescent protein (GFP)

Another type of acid labile protein we have attempted to produce in A. sojae is GFP from the jelly fish Aequoria victoria. This protein is not only proteolytically sensitive but furthermore it loses its activity at acid pH. Vectors carrying GFP or GLA-GFP fusion genes (driven by the A. nidulans gpdA promoter) were introduced into A. sojae by cotransformation using either the pyrG or amdS selection marker. Expression resulted in brightly fluorescent A. sojae transformants for both vector types. Based on the observed fluorescence and the subsequent analysis of culture supernatants from selected, shakeflask-cultured transformants using SDS-PAGE and Western analysis it was ascertained that the yields of intact cytoplasmic GFP and secreted GLA-GFP are much higher than those obtained in A. niger protease deficient hosts (Siedenberg et al. Biotechn. Prog. (1999) 15, 43-50; Gordon et al., Microbiology (2000) 146, 415-426). In contrast to the situation in A. niger culture supernatants also the secreted GFP showed significant fluorescence.

DESCRIPTION OF THE FIGURES

Figure 1: This figure provides a comparison of amino acid sequences of KEX2-like processing proteases from X. laevis (XENPC2 and XNFURIN), S. cerevisiae (SCKEX2),
K. lactis (KLKEX1), C. albicans (CAKEX2), S. pombe (SPKRP) and Y. lipolytica (YLKEX2). The primers, which encode for the amino acid sequences with the highest overall identity (indicated with lightblue boxes), are indicated: MBL793, MBL1208, MBL 794, MBL1158, PE6, PCL1, PCL2(rev), PE6, PCL3, MBL789, PCL4 and MBL1219. Regions of overall identity (4 out of 7 entries) are indicated with purple boxes. Gaps are indicated with -; no sequence data are indicated with ~; asteriks indicate the stop codon of the protein.

Figure 2: This consists of 2a, b and c

Figure 2a provides the background growth of the A. sojae strain described in patent WO97/04108 after 5 days of incubation at 33°C. The top picture reveals growth on non selection medium. The bottom left picture shows selection medium according to WO97/04108 and the bottom right picture shows the results using improved medium (acrylamide) according to the invention.

Figure 2b provides the background growth of the A. sojae strain ATCC11906 after 5 days of incubation at 33°C. The top picture reveals growth on non selection medium. The bottom left picture shows selection medium according to WO97/04108 and the bottom right picture shows the results using improved medium (acrylamide) according to the invention.

Figure 2c provides the background growth of the A. sojae strain ATCC20387 after 5 days of incubation at 33°C. The top picture reveals growth on non selection medium. The bottom left picture shows selection medium according to WO97/04108 and the bottom right picture shows the results using improved medium (acrylamide) according to the invention.

Figure 3 (a and b): This figure provides a comparison of A. sojae ATCC11906 and A. oryzae amdS sequences from both ends. A and B indicate the two ends. The cloned 1.6 kb A. sojae sequence was used. Underlined bold bases differ between species/strains. Intron I sequences are indicated in small letters.

Figure 4 (a and b): This figure illustrates the construction of a pyrG disruption vector via pAO4-13 and pAO4-13deltaCla.

Figure 5: This figure illustrates the construction of pAB4-1rep going from pAB4-1 via
isolation of \textit{XhoI} fragment and \textit{HindIII} fragment followed by cloning into pMTL24.

Figure 6: The construction of the \textit{alpA} gene replacement vector is disclosed in this figure. A 4.4 kb \textit{EcoRl-StuI} fragment from pAS1-1 with the ATCC11906 genomic fragment, the 2.6 kb \textit{SmaI-NcoI} fragment from pAB4-1rep and the 4.4 kb \textit{NcoI-EcoRl} fragment from pAS1-2A are ligated in a 3 way ligation thus providing pAS1-deltaalp.

Figure 7: This figure provides the restriction map of the DNA fragment carrying the \textit{A. niger pclA} gene.

Figure 8: This figure provides the structure (functional organisation) of the \textit{A. niger pclA} encoded protein. It shows pre, pro, activity and P domains from left to right. The light coloured triangles indicate KR sites. The dark coloured triangles indicate glycosylation sites. The vertically striped light box is an S/P/T rich region. The dark weavepatterned box at the right end is a D/E rich region.

Figure 9: This figure illustrates growth phenotype of an \textit{A. niger pclA} mutant strain.

Figure 10: This figure provides a DNA sequence comparison between the \textit{A. sojae} and \textit{A. niger pclA} genes. A vertical bar indicates identity; : indicates 5; · indicates 1. 72.139% similarity and 72.073% identity were found.

Figure 11: The construction of the \textit{pclA} gene replacement vector is disclosed in this figure. A 7.6 kb \textit{Clal} fragment, which is a ATCC11906 genomic fragment, was cloned into pMTL23p. In this construct the 2.6 kb \textit{SmaI} fragment from pAB4-1rep was cloned into the \textit{EcoRl}-site, thus providing pAS2-delta pcl.

Figure 12: This figure shows the amino acid sequence comparison of the various PclA homologous from \textit{S. cerevisiae} (Sckex2), \textit{K. lactis} (Klkex1), \textit{A. sojae} (AspclA), \textit{A. niger} (A. niger), \textit{P. chrysogenum} (Penpcl1), \textit{A. bisporus} (Agarmbl129), \textit{T. reesei} (Trichpcl1), \textit{R. oryzae} (Rhizpcl1), \textit{F. venenatum} (Fuspcl1), \textit{S. pombe} (Spkrg), \textit{C. albicans} (Cakex2) and \textit{Y. lipolytica} (Ylkex2). Regions of overall identity (8 out of 12 entries) are indicated with yellow boxes. Gaps are indicated with .. ; no sequence data are indicated with ~ .
Figure 13: Sequence data are provided in figure 13a for the *A. oryzae alpA* promoter sequences (Q11755). The primer position for PCR cloning is indicated. In figure 13b the sequence data are provided for the *A. oryzae amyA* promoter sequences also including primer positions (A02532). Figure 13c provides the ATCC42149 *A. oryzae* derived *gpdA* promoter sequences (EP0.436.858 a1) also including primer positions.

Figure 14: This figure provides a comparison between various *gpdA* promoter sequences of *Aspergillus*: From top to bottom, *A. sojae* ATCC11906, *A. oryzae*, *A. niger* and *A. nidulans*. Asterisks indicate the putative intron present in the 5' untranslated region of the promoters. Arrowheads indicate the CT rich regions. Bold underlined letters indicate the differences between the *A. oryzae* and *A. sojae* sequences.

Figure 15: This figure shows a map of the vector pGLA6S of 12700bp.

SEQUENCE LISTING

SEQ ID No. 1

MBL1784: 5'-CGGAATTCGAGCGCAACTACAAGATCAA-3'

SEQ ID No. 2

MBL1785: 5'-CGGAATTCAGCCCAGTTGAAGCGGTC-3'

SEQ ID No. 3

The sequence of the *Aspergillus niger* gene encoding proprotein convertase

The start codon and the stop codon are indicated with bold underlined letters

The intron is indicated with underlined small letters

```
1  CCATGGCAAG CCTCTACTTT GCCCTGATTA CATCGTCTG AGAGAGAGAG
51  TCTACCCAACTTCTCCCCCA AAGGATGGTCTTACAGGTT GTGTGGGCTGC
```
GGCTCTGGGC CTCTCGCGTG CTGCGCTCGGC TTCTCTCCAT CCCCATCGTT
CCTACGGAGAC CCATGATTAC TTGCCTCTAC ACCTTGTAGA ATCCACCTCG
CCGCGCCAGC TGCCCAACGC ACTAGGTGTCT CGCACAAGAAG GCCCCGTCGG
AGAAATTCCCC TCACATCATA CTTTCTCGAT ACCCGGTGAA AACAGTGACG
ATGTCATGTC CGTGCTGGAT CAATTGCCGGG ATCGTGCGAG GTACGCGCGC
CGCTCCGGAG ATGAGCCTGGC TGTCCTTCCC TCTTGGTTCG GCCGAGACGA
AGGTCTAGGT GCCATCTCTTT GGTCCGAGAA GCTGGCTCCC CAGAGAAAGC
TCCATAAAAAG AGTGCCGCGG ACAGGATATG CTGCAAGATG GCCGTCGCAAC
ACTCAGAAAATG ACCCCCAAGC GCTTGCGCGG CAGAAACGCA TTGGCTCGGA
ATTTGGAATCTC GCGGACCCCA TCTCCGGCGA ACAATGGCAT TTGTATAATA
CTGTCAGCAGT GGGCAGATAG CTAAACGATG CGGATCTGGT CGTGGAGGCC
GTACAGGCGG AGGTTGTCAC GACCGCCATT GTGATAGACG GTTGGACAT
GTACAGCAAC GATTTAGGCA CGAACTATTT TGCGCCGCGT TCTTATGACT
ATAACGACAA AGTACAGAGAG CGAGGCGGCCG GCTGGAGCGA TGACCGCCAC
GGTAGCTAGAT GCAGCGGTTGA ATCAGGTGCG GCAGAAGAACG AGGTGTGCGG
GGTGGTTGTT GCATAGTATA GTCCGATGCC TGTTATTCCG ATTCTCTCCG
CACCACAGTA CGACACTGAT GAGGCTCGGG CTAATTTACTA CGCCTATCAG
GAGAACGATA TCTACTCGTG TTCTTGGGTT CCCATATGAT ATGGGCACAC
AATGGGAAAC CCGGGCACTC TGATCAAGCG GCACATGTTGC AATGGTATCC
AAAATGGTCG AGGTGAAAGA GCCTCGCTTT TTGTAATTGC GCCTGGTAGAC
GTG GCCATTC ACGAGGATAA CGTAGTTCTTT GACGGTTACA CCAACAGTAT
CTACAGCATC ACGGTCGGTT CCATTGATCG GGAGGTAAC CATCCTCGGT
ATTCGGAATC CTGCTCGGC CCACTGGTG G T TGCCTACAG CAGCGGGCGC
AGTGATGCAA TTCTACACAC GGACGTCGGC ACAGACAAGT GCTCGACTAC
CCATGGTGGA ACTTCGCGGC CGGCCCAGT CGCTGCGGGA ACCGTTGGCGC
TGCCCCTCAG TGTCGGCCCG GAACACGACCT GCGGTGACGT TCAGTATTTG
ATGATTGAGG CGGCAGTGCC TGTTCTGAA GAATGATGGAA GCTGCGACGA
CACAAGAAGC GGAAAGAAGT TCAGCCATGA CTGGGGATAT GGTAAAGTGC
ACACATATAAC GTCTGTGAAA CGGGCAGAGA CCTGGGATCT GGTGAAGCCT
CAAGCCTGGC TCAATTCCCC CGGCCAGCCG GGGTGGCATG AGATCCCACA
GGCGAGCAG GGGCTGCTTA GTTCTGACGA GTGACGGGAG GATATGTTGG
AGGGAGCCAA CCTGGAGCGG CTGGAGCATG TCACGCTCAC CATGAATGTG
AACCCACACC GCGAGGCGGA TCTCAGCGGT GAGTTACGGA GCCCTGATGG
TCGGTGCTGT CACCTCAGTA CGCCCCGGCG GCAGATAAT CAAGAGTTGG
GCTATGTGGA TTGGACCTTC ATGAGCCGTTG CTCAGCTGta agtaaaaaact
tttctcgtttgctcggttct tctgctaata catactagG GGCAGTCCCG
GGATTGGCAA ATGGACTGTG ATGTCAGGG ACAAAGAATGT CAACGACCAT
ACTGGGCAAT TCATCGATTG GCGACTCAAC TTGGTGCGGC AGGCGATTGA
CGGAGCCCG CAGCCCTCTCC ACCCATGCC TACTGAAACAC GATGACGCACC
ACAGCTATGA GGAAGGAACG GTGGCTACCA CGAGCATCAG CGCGTTCCC
ACGAAAACC AGCTGGCTGA CAAGCCCGCT GGTGGCGTTG ATGCACCGGT
GAACGTTAAG CCTACAACAT CGCGATGCC GACCCGTAGT CCTACAGACC
CCATCGATGA TGAAGAACC CAGAGGACCC CTAGTACAGA GGCAAGCTCA
ACACCAAGTC CTTCCTGCAC CACCGCGTCA GATAGTATCC TGGCTTCTTT
CTTCCCCACG TTGGGTGCAG CGAAGCGGAC CGAAATTTTG ATCTACGCTG
CGATGCGGTC ATCATATTGT TTTCGACTTG GCCTGGGCCT CTACTTCCAT
GTGCAAGCGCC GCAAACGTAT TGCGAGCGAC AGCCGGGATG ACTACGATTT
CGAGATGATG AAGGACGGGA ATGACGTACA GGCAATGAAC GGACGGTCTG
ACCGTTCCAG TGCGCGGGGT GGCAGAGCTG ACAATGCTTT TGCGGGCGAG
AGCGATGAGG AACCATATT ACGTGATGAG GATGATGAAC CGTATCGGGA
TCGGGGGATC AGCGCGAAC AAGAACGGGA GGGCGCAGAT GGAGACATT
CTCGGAGATG AAGTGCAGT AGATGAGGTT TGACTTATT TGCGACAGTG
TTTCTAACTTT GGTTGATGAC TGCGGTGGAA CAATATTTCT GCTGTGTATG
CTGCACTAGG AAGCGTGATAT ATACCATGTA TGTTGACATC ATCGTGATCG
GGTTATCAT TCTTCATCTG CCATGTTTTG TGATCTCCGG AATAGTACCA
AAGGAACACT AAATTAAGGG TCTTGGCGAT GACGCTTCCC GTCGCTGCTT
TTGACTTCCG CGCAATCTCG TCTCTCTGTC TGTTGACCGC GCGCAAACCA
ACCTCCATCT CCTCACTCCT CCCACCTTAA TCTTCCTGTTG CTGCTTCCTG
AACCCCCCAGTTAACAA AACCAGGCTT TTCTAGCTTC ACATATTG
ACCTCGCACT GATCCCCCATC TCCGCCCACT CCAACGCTAC CGACCCAGGC
TTCTCTGGCG GCTCCAGGCG GCAGGCAAATC AAACCAACCC TCGATGGAT
CAGCAGCGAC ACTTCGACAG GTGTCCTGAC AGGCATGACC GGAGACCGGA
TCTCTGGCGA CCCAGCAGCT CCAGAACAGA CACAGAGGAA CAGGGCCCAT
AAGTCAATGGC AAACGGAGGA TGAGCAACCG CTCAAGAAAG
CCCTCCAGCTGG AACACCTGCG CGGATGCCCCT GCACCTGGCG GGCATCGCGA
SEQ ID No. 4

The partial sequence in the coding region of the *Aspergillus sojae* gene encoding proprotein convertase:

1  CGCGGATCCA TGGAAACACGA TGTGCAGGGTGT AAATTGGGAGC AGCTAGGAAT
15  GATGTCCTGTG GAGTAGGTGT TGCATACGAC AGCCAAGTGT CGGGAATTCG
20  GATTGTGTCC GCACCCATTG ACGAGCGAGA TGAGGTGTCT GCCATCAACT
25  ATGGCTTCCA GCAGAAATGAT ATATATTCAT GCTCCTGGGG CCGTCGGGAT
30  GAGGCGGCAA CGATGGAGGC GCCAGGGATT CTTATCAAAAC AGCTATGATT
35  CAACGCTATC AAAAAATGCCC GAGGAGTTAA AGTTCTATTC TCTGCTCTTG
40  CAGCTGGAAA TGTTGCGAGG TACGATGACA ACTGCAATTG CGACGGTTAT
45  ACAAAACAGCA TTTACAGCAT ACCGGTGGGC GCTATTGATC GAGAGGCAA
50  ACATCCCGAC TACTCGGAAT CATGCTCTGC CCAGTTGGTT GTCGGTATTA
55  GCAGTGCGTC GAGTGACCGG ATTACATCCA CGGACGTTGG AACTGATAAA
60  TGTTTTTCAC TNTCAGGCG GCCAATCTTG CAACTGGACC GCTAGCTGCG
65  GGTACATTG CCGTGCCTCT TACTGCCCAGA CCGGAACCTAA CTTGCGGAGA
70  TGCCCATGCT CGATGATGAG AGACCGCAGT TCCCGTCCAC GAAGACGACG
75  GGAGCTGGCA GACCTACAAA ATGGGGAAGA AGTTTAGCCA TGACTGGGTT
80  TTTGGGAAAG TAGATGCTA TTTCACTTGC CGCTCGGCA AGACGGTGGGA
SEQ ID No. 5

The partial sequence in the coding region of the *Trichoderma reesei* QM9414 gene encoding proprotease converetase

1  GCTGGTGAAA CCACAGGCGT GGTTCCTCCT ACCTGGTGCCT CGGGTGAGC
2  ATTAAATCCC ACAAGTGAC CAGGCCCTTG CCAGCTCATA CGAAATTACC
3  AAGGATATGA TGTACAGGCC CAATGTCGAG AAATTGGAAC ATGTCACTGT
4  GACCATGAAT GTAAATCACA CTCGCGCGAG CGATATCAGC GTGGAGTTGC
5  GCAGCCCGGA AGGTATCGTC AGTACATCTGA GTACAGCGCG GCGGTACAGAT
6  ATGCAAAGG GCTGGCATGA AGATTTGACG TTTATGACGT TGGCTCATTTG
7  GTATGTATTT GCTCCCCCCAA TTTAGTTTTTC GTGCTCAGTC CTGACATTTA
8  CATTTAGGAGG TAGGTCCGGG GTGGAAAGTG GACCGTGTAT TGTGAAGGAT
9  ACCAATGTCA ATGATCATGT TGGAGAATTC ATCGACTGGC GGCTCAACCT
10 CTGGGAGCTTT TGATCGAGCG GCTCCACGCCA GCCCCCTCAT CCTATGCCCG
11 ATGACATGCA CGAGACCCAC TCGATTGAAG ATGCCATTGT TGTTACCACCT
12 AGTGGTGACC CTATCCCCAA TAAGACTGAA GCCCCACCTG TCCCAACTGA
13 TCCGTGGGAT CGTCTCTGTA ACGCAAAGCC ATCTGCGCGA CCAACGATGC
14 GAGCAGGAGGC TGCTGCTCAA GAGACATCTG AAGGTCCCAAC CCGAGCGAAA
15 CCTAGTTCTA CTGAATCACC TCTTTACCCAC CTCTCTTGCG GATAGCTTTT
16 TGCCATCCCT CTTCCTTTCAGG GCTGCGCTG GTGGAGATC CAAGCTTGGG
17 TACGT

1  GCTGTCCGCA CTGATGGGTCGCGCC CCGCTCGGGCG GTGCTCAGC ACTCCAAGAT
SEQ ID No. 6

The partial sequence in the coding region of the *Fusarium venenatum* ATCC20334 gene encoding proprotein convertase

SEQ ID No. 7

The partial sequence in the coding region of the *Penicillium chrysogenum* P2 gene encoding for proprotein convertase
SEQ ID No. 8

The partial sequence in the coding region of the *Rhizopus oryzae* ATCC200076 gene encoding proprotein convertase

```
1 ACTNGGGGCA TTGGTGAAAT NTGCTTTGTG GNNTTGGTGT GCTTACGACG
20
51 CAAAAATATC TGGTATAAGT ATATTATCAG GTGAATACAC AGAGGCGACG
101 GAGGCTGCTG CTTTGAAATTA CAATATCAA GAAAATCAA TCTACTCCTG
25
151 CTCNTGGGGC CCA
```

SEQ ID No. 9

The partial sequence in the coding region of the *Agaricus bisporus* HORST gene encoding proprotein convertase

```
1 ATGTTGGTCTT GGTCTCGCCT ACAGATCCCA GGTCGCTGGT GTTCGCAATAT
35
51 TGTCTGGTCC CATAACGGAAC GCAGTGAAGCG ACTGCCTGCT CAACTATGCT
101 TTCCAAATAATG TATCTATCTTT CAGCCTGACT TGGGGGCCCAC CTGACAAATGG
151 TATGTCCATG GAAGCCCATG GATACCTCAT CAAAAAGCT GTCGCAACG
```

SEQ ID No. 10

coding strand  

BamHI-site is underlined

PE4  5'- CG CGGATC CA(T/C) GGX ACX (C/A)GX TG(T/C) GCX GG -3'
degenerated 2048 times

SEQ ID No. 11

coding strand

PCL1  5'-CA(T/C) GGX ACX (C/A)GX TG(T/C) GCX GGX GA-3'
degenerated 8192 times

SEQ ID No. 12

coding strand

PCL2  5'-AT(C/T/A) TA(T/C) TCX TG(T/C) TCX TGG GGX CC-3'
degenerated 768 times

SEQ ID No. 13

non coding strand  

BamHI-site is underlined

PE6  5'- CGC GGA TCC XCC (A/G) TT XCC X(C/G)(A/G) XGC
      (G/A/C)(C/A) A XAC -3'
degenerated 49152 times

SEQ ID No. 14

non coding strand

PCL2rev  5'-GG XCC CCA XGA (A/G)CA XGA (A/G) TA (A/T/G) AT-3'
degenerated 768 times
SEQ ID No. 15
non coding strand
PCL3  5'- (A/G)TT XGT (A/G)TA XCC (A/G)TC (A/G)(A/T)A (A/G)TT-3'
        degenerated 1024 times

SEQ ID No. 16
non coding strand
PCL4  5'-GC XGC XGA XGT XCC XCC (A/G)TG-3'
        degenerated 2048 times

SEQ ID No. 17

The sequence of pAB4-1rep

...59-499 bp........................................: 0.4 kb HindIII fragment of pAB4-1

1-58 bp.......500-513 bp........2873-2930 bp : polylinker sequence of pMTL24
(indicated with underlined small letters)

........................................514-2872 bp................................:
            2.3 kb XhoI fragment of pAB4-1

1  ggccca tgtgaa ttcgagctcg gtacccgggg atctcttaga gtcgacctgc

51  aggcacatgc AA GCTTGGTCA GAGTACCACA CGCCCCGATC GGCTATCGGC

101  CGGGGTGTCTG ACTTCATTAT CGCGGGTCGC GGTATCTACG CGCGCGCGGA

151  TCCGGTGACG GCTGCGCAAC AGTATCAGAA GGAGGGGTGG GAAGCCTACC

201  TGGGGCGGTGT CGCGGGAAC TAATACTATA AAAGGAGGAT CGAAGTTCTG

251  ATGGTTATGA ATGATATGAA AATGCAAACCT GCCCGCAAACG ATACGGGAAGC

301  GAAAACCGAC CAATGTCGAG CACGGGTAGT CAGACTGCGG CATCGGATGT

351  CCAAAACGTTA TTGTATCTCG AGGCTTCTAT GGTGTGCGAC AAGGATCAAT

401  GCGGTACGAC GATTTGATGC AGATAAGCAG GCTGCGGAAGT AGTAACCTTT

451  GCGTAGAGAA AATGGCGAAG GGTGGGCGTA TAAGGCCTG GATAAGCTTq
catgcctgca ggcCTCGAGC TAACATACAT TCCGAAACCGT GCAGCCCCAG
GCCGACGAGT TCAACTGCGC TCAGCGCGCT CATGCCAACCT TCCTTGAAGAA
CTCCAGCCAA ACTATGCTCT TCCTCCTGCT AGCTGAGACTG AAGTACCC
AGTTGGCGAC TGGCCTCGGA AGCATCTGGG TCCCTCGGTGC CTCACATGTT
CTTTACGGAT ATGTTGACTC CGGAACACCG CGGGGTCCCG GTCGTGTTGA
CGGCAGCTTC TACTTGCTTG CACAGGGAGC TCTCTGGGCG NTGACGTCTT
TTGGAGTGTG GCAGGAGTTG ATTTCCCTACT TCTAAGTTTG GACTTGAATC
CGTGATTGTA TTGAGGTGAT TGGCGATGTT TGGCTATACC AGCTATATG
AATAATCTCT ACTGTATACT ACTAATTCAAC GCATTTTACT ATGCGTGCTG
CTAGGGTCGG CAATGACAAT GGCAATCTGGA CTGACGGGTG GCTATTTCTCC
ATGTGCAAAGCA GAAATCCGCG TTTGCAGGTG TGCGCAGACC CCAAGGTTCG
GACGCGGAGC AATCCACCCC CGATGTTGGCT GGTGCCTGGA GGGGCCTCGG
ATGATTTTTAC TGAGCTTGGCT TTTCTTGTCG ACAATGGACA TTGTCCCTTG
TCTTCCCTCA GATTTAAGGG TCAGTCACTG CTACATTTCT CAGTAGATAC
CGCGCAGCCTC TCGGATTAAA GAAATCGAGG TCCACGAGTC GAAACTGCGA
ACTACTCTCA TTATAACATC CTCTTTCCTAT TCCGCCATTTC ACCCCTCAT
CAACCACATG TCCTCCAAGT GCAGTTGAC TACACTGCC CGTGCCGAGCA
AGCATCCCAA TGCTCTGGCG AAGAGGCTGT TCGAGATTGC CGAGGCCAGG
AAGACCAAATG TGACTGCTTC GGCTGACGGT ACCACCACTA AGGAGCCTACT
AGATCCTGCT GACCGTAGGC CGACCGCCTA CTCTGCTGTA TTATGCCTGCA
TGCAAACCTTA TTACGGGTGA TACCAGCTG CAGGTCTCGG TCCCTACATT
1601  GCCGTGATCA AAACCCACAT CGATATCCTC TCTGATTTCA GCAACGAGAC
1651  CATTGAGGGA CTTAAGGCTC TCGCCGAGAA GCACAACCTT CTCATCTTCG
1701  AGGACCGCAA GTCTATGGAC ATCGGCAACA CGGTCCAGAA GCAATACCAC
1751  GGCCTGTACCC TCCGTATCTC GGAATGGGCC CACATCATCA ACTGCAAGCAT
1801  TCTCCTGTGT GAGGTATTCG TCGAGGCTCT CGCTCAGAGC GCGTCTGCAC
1851  CGGACTTGGC TACAGGCCCC GAACCGGTCG TGTTGTACTT GGCAGAGATG
1901  ACCTCTAAGG GCTCTTTGGC TACCGGCGAG TACACTACTT CCTGGTCTGA
1951  TTATGCCCCG AAATAACAAGA ACTTCGTTAT GGGATTCGTC TGACCCGGCG
2001  CGTTGAGTGA GGTGCAGCTG GAAGTCGCTG CTCTTCCGGA TGAGGAGGAC
2051  TTGTGTGTCT TCAGACTGCG TGTAACATT TCTTCCAAGG GAGATAAGCT
2101  TGTCAGCAGC TACAGACGCG CCAGATCGGC TATCGGCCGG GGTGCTGACT
2151  TCATTATCAG GGTGCCGGGT ATCTACCGCG CGCCCGATCC GGTGCAAGCT
2201  CGGCAACAGT ATCGAAGGGA GGGTGGGGAA GCCTACCTGG CCCGTTGGCG
2251  CGGAAGCTAA TACTATAAAA GGAGGATCGA AGTTCTGATG TTATGAATG
2301  ATATAGAAAT GCAACTTGGC GCAACGGATA CGGAAGCGGA AACGGGACAA
2351  TGTCGAGCAC GGTTAGTGCA ACTGCGCCAT CGGATGTTCA AACGTTATTG
2401  ATCTCTGCAGG CTACATGCGTG GTGGCACAAA GATCAATGCG GTACGACGAT
2451  TTGATGCAGA TAAGCAGGCT GCCAAGTAGT AACTCTTCGG TAGAGAAAAT
2501  GGCAGCAGGT GGCTGTAAGA GGGCGGTGAT AAGCTTAATT GTCACTGCGAG
2551  ATACGCAGCT CGTCTTGGCA TCCAAGTCAG CGTCAGCAGA AAATACGGGAC
2601  TTCCGAAAGT ATATGGCAGG ATTAAGAAGAC TGGACTCTGC AGCAATGTGTT
SEQ ID No. 18

15 MBL 789  EcoRI is underlined

5'- GGAA TTC (A/G)GA ATA (T/A)GG AGG ATG TAG -3'
degenerated 4 times

SEQ ID No. 19

20 MBL 793  BamHI is underlined

5'- CGGATCCG CAG TGG CAC TTG (G/A)TC AAT CCA A -3'
degenerated 2 times

SEQ ID No. 20

25 MBL 794  EcoRI is underlined

5'- GGA ATT CTT AAA A(T/G)C CCA AGA ACC TTC A -3'
degenerated 2 times

SEQ ID No. 21

30 MBL 1158  EcoRI is underlined

5'- G GAA TTC (T/C)TC (T/G)CC (T/G)GC (A/G)CA (C/G)C(T/G)  
(C/G)GT (T/G)CC (A/G)TG -3'
degenerated 512 times

SEQ ID No. 22

MBL 1208  Clal is underlined
5'- CGG ATC GA(T/C) GGX ACX (C/A)GX TG(T/C) GCX GG -3'  
degenerated 2048 times

SEQ ID No. 23
MBL 1219 BamHI is underlined
5'- CGG ATC (C/T)TG XA(G/T/C) (A/G)TC XC(T/G) CCA XGT  
(C/A/G)AG -3'  
degenerated 4608 times

SEQ ID No. 24
Restriction sites are bold
Primers are underlined

BamHI \hspace{1cm} PE4 primer
GGATCCATGG CACGAGATGT GCAGGTGAAA TCGGTGCGGC GAAAGAAAAAC  
AACGTGTGCG  60
GGGGTGGTGT TGCGATGAT AGTACGTACGCT GATTTCTCTCC  
ACACCCCATCG  120

EcoRV
ATGACACTGA TGAGGCTGCG GCTATTAACT ACAGCTATCA GGAGAACGAT  
ATCTAATCCTGT  180

GTTCTGGGG TCCCTATGAT GATGGCGCCA CAATGGAAGC CCGGGCACT  
CTGATCAAGC  240

GGGCATGGT CAATGCTATC CAAAATGGTC GAGGTTGAAA AAGCTCGGTT  
TTTGTCTGGCG  300 \hspace{1cm} PE6 primer

CCGGGAAA TGGTGGAATC

320 \hspace{1cm} BamHI

SEQ ID No. 25
Aspergillus niger PclA protein sequence
1  Met Arg Leu Thr Gly Gly Val Ala Ala Ala Ala Leu Gly Leu Cys Ala

16  Ala Ala Ser Ala Ser Leu His Pro His Arg Ser Tyr Glu Thr His

31  Asp Tyr Phe Ala Leu His Leu Asp Glu Ser Thr Ser Pro Ala Asp

46  Val Ala Gln Arg Leu Gly Ala Arg His Glu Gly Pro Val Gly Glu

61  Leu Pro Ser His His Thr Phe Ser Ile Pro Arg Glu Asn Ser Asp

76  Asp Val His Ala Leu Leu Asp Gln Leu Arg Asp Arg Arg Arg Leu

91  Arg Arg Arg Ser Gly Asp Asp Ala Ala Val Leu Pro Ser Leu Val

106  Gly Arg Asp Glu Gly Leu Gly Gly Ile Leu Trp Ser Glu Lys Leu

121  Ala Pro Gln Arg Lys Leu His Lys Arg Val Pro Pro Thr Gly Tyr

136  Ala Ala Arg Ser Pro Val Asn Thr Gln Asn Asp Pro Gln Ala Leu

151  Ala Ala Gln Lys Arg Ile Ala Ser Glu Leu Gly Ile Ala Asp Pro

166  Ile Phe Gly Glu Gln Trp His Leu Tyr Asn Thr Val Gln Leu Gly

181  His Asp Leu Asn Val Thr Gly Ile Trp Leu Glu Gly Val Thr Gly

196  Gln Gly Val Thr Thr Ala Ile Val Asp Asp Gly Leu Asp Met
Tyr

211 Ser Asn Asp Leu Arg Pro Asn Tyr Phe Ala Ala Gly Ser Tyr Asp

226 Tyr Asn Asp Lys Val Pro Glu Pro Arg Pro Arg Leu Ser Asp Asp

241 Arg His Gly Thr Arg Cys Ala Gly Glu Ile Gly Ala Ala Lys Asn

256 Asp Val Cys Gly Val Gly Val Ala Tyr Asp Ser Arg Ile Ala Gly

271 Ile Arg Ile Leu Ser Ala Pro Ile Asp Asp Thr Asp Glu Ala Ala

286 Ala Ile Asn Tyr Ala Tyr Gln Glu Asn Asp Ile Tyr Ser Cys Ser

301 Trp Gly Pro Tyr Asp Asp Gly Ala Thr Met Glu Ala Pro Gly Thr

316 Leu Ile Lys Arg Ala Met Val Asn Gly Ile Gln Asn Gly Arg Gly

331 Gly Lys Gly Ser Val Phe Val Phe Ala Ala Gly Asn Gly Ala Ile

346 His Asp Asn Cys Asn Phe Asp Gly Tyr Thr Asn Ser Ile Tyr

361 Ser Ile Thr Val Gly Ala Ile Asp Arg Glu Gly Asn His Pro Pro

376 Tyr Ser Glu Ser Cys Ser Ala Gln Leu Val Val Ala Tyr Ser Ser

391 Gly Ala Ser Asp Ala Ile His Thr Thr Asp Val Gly Thr Asp Lys
406  Cys Ser Thr Thr His Gly Gly Thr Ser Ala Ala Gly Pro Leu Ala 

5  421  Ala Gly Thr Val Ala Leu Ala Leu Ser Val Arg Pro Glu Leu Thr 

436  Trp Arg Asp Val Gln Tyr Leu Met Ile Glu Ala Ala Val Pro Val 

10  451  His Glu Asp Asp Gly Ser Trp Gln Asp Thr Lys Asn Gly Lys Lys 

466  Phe Ser His Asp Trp Gly Tyr Gly Lys Val Asp Thr Tyr Thr Leu 

15  481  Val Lys Arg Ala Glu Thr Trp Asp Leu Val Lys Pro Gln Ala Trp 

20  496  Leu His Ser Pro Trp Gln Arg Val Glu His Glu Ile Pro Gln Gly 

511  Glu Gln Gly Leu Ala Ser Ser Tyr Glu Val Thr Glu Asp Met Leu 

25  526  Lys Gly Ala Asn Leu Glu Arg Leu Glu His Val Thr Val Thr Met 

541  Asn Val Asn His Thr Arg Arg Gly Asp Leu Ser Val Glu Leu Arg 

30  556  Ser Pro Asp Gly Arg Val Ser His Leu Ser Thr Pro Arg Arg Pro 

35  571  Asp Asn Gln Glu Val Gly Tyr Val Asp Trp Thr Phe Met Ser Val 

586  Ala His Trp Gly Glu Ser Gly Ile Gly Lys Trp Thr Val Ile Val 

40
Lys Asp Thr Asn Val Asn Glu His Thr Gly Gln Phe Ile Asp Trp
Arg Leu Asn Leu Trp Gly Glu Ala Ile Asp Gly Ala Glu Gln Pro
Leu His Pro Met Pro Thr Glu His Asp Asp Asp His Ser Tyr Glu
Glu Gly Asn Val Ala Thr Thr Ser Ile Ser Ala Val Pro Thr Lys
Thr Glu Leu Pro Asp Lys Pro Thr Gly Gly Val Asp Arg Pro Val
Asn Val Lys Pro Thr Thr Ser Ala Met Pro Thr Gly Ser Leu Thr
Glu Pro Ile Asp Asp Glu Glu Leu Gln Lys Thr Pro Ser Thr Glu
Ala Ser Ser Thr Pro Ser Pro Ser Pro Thr Thr Ala Ser Asp Ser
Ile Leu Pro Ser Phe Phe Pro Thr Phe Gly Ala Ser Lys Arg Thr
Glu Val Trp Ile Tyr Ala Ala Ile Gly Ser Ile Ile Val Phe Cys
Ile Gly Leu Gly Val Tyr Phe His Val Gln Arg Arg Lys Arg Ile
Arg Asp Asp Ser Arg Asp Asp Tyr Asp Phe Glu Met Ile Glu Asp
Glu Asp Glu Leu Gln Ala Met Asn Gly Arg Ser Asn Arg Ser Arg
Arg Arg Gly Gly Glu Leu Tyr Asn Ala Phe Ala Gly Glu Ser
Asp

Glu Glu Pro Leu Phe Ser Asp Glu Asp Asp Glu Pro Tyr Arg Asp

Arg Gly Ile Ser Gly Glu Gln Glu Arg Glu Gly Ala Asp Gly Glu

His Ser Arg Arg

SEQ ID Nos. 26 to 31

PCR-primers for *A. sojae* promoter cloning

Restriction sites are underlined

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5' - 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ ID</td>
<td></td>
</tr>
<tr>
<td>No. 26</td>
<td></td>
</tr>
<tr>
<td>Alp-1</td>
<td>GGAATTCCGCGGCCGCGTTATTCTGCAGAAAAGC</td>
</tr>
<tr>
<td></td>
<td><em><strong>EcoRI</strong></em></td>
</tr>
<tr>
<td></td>
<td><em><strong>NotI</strong></em></td>
</tr>
<tr>
<td>SEQ ID</td>
<td></td>
</tr>
<tr>
<td>No. 27</td>
<td></td>
</tr>
<tr>
<td>Alp-2</td>
<td>GGAATTCCCATGGTGAGAGATGTTGAAAG</td>
</tr>
<tr>
<td></td>
<td><em><strong>EcoRI</strong></em></td>
</tr>
<tr>
<td></td>
<td><em><strong>NcoI</strong></em></td>
</tr>
<tr>
<td>SEQ ID</td>
<td></td>
</tr>
<tr>
<td>No. 28</td>
<td></td>
</tr>
<tr>
<td>Amy-1</td>
<td>GGAATTCCGCGCGAGATCTGCCCTTATAAA</td>
</tr>
<tr>
<td></td>
<td><em><strong>EcoRI</strong></em></td>
</tr>
<tr>
<td></td>
<td><em><strong>NotI</strong></em></td>
</tr>
<tr>
<td>SEQ ID</td>
<td></td>
</tr>
<tr>
<td>No. 29</td>
<td></td>
</tr>
<tr>
<td>Amy-2</td>
<td>GGAATTCCATGGATGCTTCTGTTTGTGG</td>
</tr>
<tr>
<td></td>
<td><em><strong>EcoRI</strong></em></td>
</tr>
<tr>
<td></td>
<td><em><strong>NcoI</strong></em></td>
</tr>
<tr>
<td>SEQ ID</td>
<td></td>
</tr>
<tr>
<td>No. 30</td>
<td></td>
</tr>
<tr>
<td>AOGPDA-1</td>
<td>GGAATTCCGCGCGCTATGAAACCAGAAAG</td>
</tr>
<tr>
<td></td>
<td><em><strong>EcoRI</strong></em></td>
</tr>
<tr>
<td></td>
<td><em><strong>NotI</strong></em></td>
</tr>
<tr>
<td>SEQ ID</td>
<td></td>
</tr>
<tr>
<td>No. 31</td>
<td></td>
</tr>
<tr>
<td>AOGPDA-2</td>
<td>GGAATTCTAGGTTTATGATGTG</td>
</tr>
<tr>
<td></td>
<td><em><strong>EcoRI</strong></em></td>
</tr>
<tr>
<td></td>
<td><em><strong>NcoI</strong></em></td>
</tr>
</tbody>
</table>

SEQ ID No. 32

The sequence of the *Aspergillus sojae gpdA* promoter region
SEQ ID No. 33

The sequence of the *Aspergillus sojae* alpA promoter region

1  GCCTGCGGGG TTATTCGCG GAAGCGGACC CCCCCCTTCC GCCCCAACAG
51  GGGAAATGTT GCAAGCTTCA GATACCTATCA GAAGACCTCC AGGAGCACAT
151  GCCTGTTCGG ATAAACCTGCG TGTAGCAGG CAAAATGCTT AGCTTACATT
251  TTGCGACTGA GGGGCCAGGA AGTGCTTTATG GCAAAGATTC CACTTCTTTG
351  TGTGATAGCC CCTCGCGGCG CCGCTTTGAC AGCGTTCTCT GCTATCCAAT
401  ATGGAAAGCG TGATATTATA GGTGCACATG GTATATATCC TTTTTTTTTTT
451  TCTCTCTCTT TGTCTTCTCG GCAAACCCCAT AGCTGCGCCA ATTTGGCTAC
501  ACCTTGCGGC TCATCTCTTC AAGTTTAGAT TCCGAACAGA CCTCACCCAC
TGTTAGGATG ATCTTCCTCA CTAAGGCGAT CGCCCGCGCC GCTAGGTGCT
TCCTGTCCAG GATAAGTCTTT ACTCCTCTCA TTATGGCGAG CTACTTTCGGTG
AATTAAATTGA CTGAGGGATA TACCACCTTC CTTTGAAGGG TACCAAGGCA
CTACCTTGGAG CGTTAGTTAC TTTTTCGAGG AAAGCGCTCT ATGCTGGTCT
CCGCCAAAAACC CTGGAACAATC TGGCATAGCC TTTGTGTTTCT CATGGTCTAT
CGGAGTACCC GTTCATGACT GAAGCAGGGTG AGCGTCCGTG GTGTCATCA
TCATTCTCAT CTTTCATCAT GCGCGCTGAT TGATAGAGTA ATTTCCGGTG
GAGCAACAAGG CCGTCTCTGT AGATGCAATG TCACCCCTGTA AGTTTACAATC
ACACTCTGTA GTACAGAGCA TCCCTGTCAT TGCTGCTGTC GCAAGTGATC
TAAATCCGTA GAATCTGCTC GAGAACGGGG AAATATAGAA CTCCTGAAAGG
TTATAATATC CACATGCATC CCTCGCTCCAT CCTCAATCC ATCATCAAGC
CGCGGTCTTC TATGCTCCGA TTTGAGTCTG TCTCGCGCAT TTTTCAATTC
TTCTCACCAGT GG
Table 1. **Taxonomic scheme of the genus* Aspergillus* (Samson, 1992)**

<table>
<thead>
<tr>
<th>GENUS</th>
<th>SUBGENUS</th>
<th>SECTION</th>
<th>SELECTED SPECIES&lt;sup&gt;a&lt;/sup&gt;</th>
<th>&quot;SUBSPECIES&quot;&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspergillus</td>
<td>Circumdati</td>
<td>Wentii</td>
<td>A. wentii (glucosidase)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Flavi/Tamarii <em>A. oryzae</em> (amylase, protease)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. tamarii&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>A. sojae</em> (fermented food, protease)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. parasiticus&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. flavus&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Nigri</td>
<td></td>
<td></td>
<td><em>A. niger</em> ---------------------→ A. pulverulentes</td>
<td>(fermented food, A. phoenicis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>various proteins, A. awamori</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>organic acids) A. foetidus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. kawachii A. usamii</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. ficuum</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>A. japonicus</em>------------------&gt; A. aculeatus</td>
<td>(endoglucanase,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>glucosidase, galactanase)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. ellipticus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>A. tubingensis</em> --------→ &quot;A. niger&quot;</td>
<td></td>
</tr>
<tr>
<td>Circumdati</td>
<td></td>
<td></td>
<td><em>A. ochraceus</em>&lt;sup&gt;tox&lt;/sup&gt; (xulanase)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. alliaceus&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Candidi</td>
<td></td>
<td></td>
<td>A. candidus (lipase, glucosidase)</td>
<td></td>
</tr>
<tr>
<td>Cremei</td>
<td></td>
<td></td>
<td><em>A. itaiconicus</em> (organic acid)</td>
<td></td>
</tr>
<tr>
<td>Sparsi</td>
<td></td>
<td></td>
<td>A. sparsus</td>
<td></td>
</tr>
<tr>
<td>Aspergillus</td>
<td>Aspergillus</td>
<td>Aspergillus</td>
<td>A. glaucus (fermented food)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Restricti <em>A. restrictus</em>&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Fumigati</td>
<td></td>
<td></td>
<td><em>A. fumigatus</em>&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Cervini</td>
<td></td>
<td></td>
<td>A. giganteus&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Ornati</td>
<td></td>
<td></td>
<td>Clavati</td>
<td></td>
</tr>
<tr>
<td>Clavati</td>
<td></td>
<td></td>
<td>A. giganteus&lt;sup&gt;tox&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>
Nidulantes
Versicolores A. sydowii (lipase)
Usti
Terrei A. terreus\textsuperscript{tox} (glucansae)
Flavipes

\textsuperscript{a} For the species selected for this list either the production of proteins/organic acids/fermented foods (indicated between brackets) and/or a DNA-mediated transformation procedure (indicated by underlining) is described, except for \textit{A. tamarii}, \textit{A. sparsus} and \textit{A. ellipticus}. Species recorded to produce toxins are indicated by \textsuperscript{tox}

\textsuperscript{b} Based on various methods the listed names may be considered synonymous to the given SPECIES name.
Table 2. The classification of the different ATCC strains

<table>
<thead>
<tr>
<th>Strains</th>
<th>Morphology</th>
<th>Aflatoxin Production</th>
<th>RAPD&lt;sup&gt;2&lt;/sup&gt;</th>
<th>PCR&lt;sub&gt;αftr&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt;</th>
<th>PCR&lt;sub&gt;αlpA&lt;/sub&gt;</th>
<th>Classification&lt;sup&gt;4&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCC 9362&lt;sup&gt;5&lt;/sup&gt;</td>
<td>ND</td>
<td>no&lt;sup&gt;7&lt;/sup&gt;</td>
<td>A. sojae type I</td>
<td>A. sojae</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 11906&lt;sup&gt;6&lt;/sup&gt;</td>
<td>A. sojae</td>
<td>no&lt;sup&gt;1,7&lt;/sup&gt;</td>
<td>A. sojae type I</td>
<td>ND</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 20235</td>
<td>A. oryzae</td>
<td>no&lt;sup&gt;1&lt;/sup&gt;</td>
<td>A. sojae type II</td>
<td>ND</td>
<td>A. oryzae</td>
<td>A. oryzae</td>
</tr>
<tr>
<td>ATCC 20245</td>
<td>A. sojae</td>
<td>no&lt;sup&gt;1,7&lt;/sup&gt;</td>
<td>A. sojae type I</td>
<td>A. sojae</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 20387</td>
<td>A. sojae</td>
<td>no&lt;sup&gt;1&lt;/sup&gt;</td>
<td>ND</td>
<td>ND</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 20388</td>
<td>A. sojae</td>
<td>no&lt;sup&gt;1&lt;/sup&gt;</td>
<td>ND</td>
<td>ND</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 42249</td>
<td>A. sojae</td>
<td>no&lt;sup&gt;1&lt;/sup&gt;</td>
<td>A. sojae type II</td>
<td>ND</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 42250</td>
<td>A. sojae</td>
<td>no&lt;sup&gt;1&lt;/sup&gt;</td>
<td>ND</td>
<td>ND</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 42251</td>
<td>A. sojae</td>
<td>no&lt;sup&gt;1&lt;/sup&gt;</td>
<td>ND</td>
<td>A. sojae</td>
<td>A. sojae</td>
<td>A. sojae</td>
</tr>
<tr>
<td>ATCC 46250</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>A. oryzae</td>
<td>A. oryzae</td>
</tr>
<tr>
<td>IFO 4177</td>
<td>A. oryzae</td>
<td>No&lt;sup&gt;8&lt;/sup&gt;</td>
<td>ND</td>
<td>ND</td>
<td>A. oryzae</td>
<td>A. oryzae</td>
</tr>
<tr>
<td>(CBS 205.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Legend: ND = not determined


4) Conclusion on classification drawn by TNO based on data presented in this table

5) This strain was deposited at ATCC as *A. oryzae*, but later reclassified as *A. sojae* based on Yuan et al, 1995\(^3\) and Chang et al, 1995\(^3\)

6) This strain was deposited at ATCC as *A. parasiticus*, but later reclassified as *A. sojae* based on Ushijima et al, 1981\(^1\) and Yuan et al, 1995\(^7\)

7) REF: ATCC catalogue

Table 3. Composition of selection media

<table>
<thead>
<tr>
<th>Composition</th>
<th>Non-selection medium</th>
<th>Selection medium (WO97/041 08)</th>
<th>Acrylamide selection medium</th>
<th>Improved acrylamide selection medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>KH₃PO₄</td>
<td>1.5 g/l</td>
<td>1.5 g/l</td>
<td>1.5 g/l</td>
<td>1.5 g/l</td>
</tr>
<tr>
<td>KCl</td>
<td>0.5 g/l</td>
<td>0.5 g/l</td>
<td>0.5 g/l</td>
<td>0.5 g/l</td>
</tr>
<tr>
<td>MgSO₄.7H₂O</td>
<td>0.5 g/l</td>
<td>0.5 g/l</td>
<td>0.5 g/l</td>
<td>0.5 g/l</td>
</tr>
<tr>
<td>NaNO₃</td>
<td>6 g/l</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>glucose</td>
<td>10 g/l</td>
<td>10 g/l</td>
<td>10 g/l</td>
<td>----</td>
</tr>
<tr>
<td>sorbitol</td>
<td>1.2 M</td>
<td>----</td>
<td>1.2 M</td>
<td>1.2 M</td>
</tr>
<tr>
<td>saccharose</td>
<td>----</td>
<td>1 M</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>mineral solution¹</td>
<td>0.1% v/v</td>
<td>0.1% v/v</td>
<td>0.1% v/v</td>
<td>0.1% v/v</td>
</tr>
<tr>
<td>acetamide</td>
<td>----</td>
<td>10 mM</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>acrylamide</td>
<td>----</td>
<td>----</td>
<td>10 mM</td>
<td>10 mM</td>
</tr>
<tr>
<td>CsCl</td>
<td>----</td>
<td>15 mM</td>
<td>15 mM</td>
<td>15 mM</td>
</tr>
<tr>
<td>agar</td>
<td>15 g/l</td>
<td>15 g/l</td>
<td>15 g/l</td>
<td>15 g/l</td>
</tr>
</tbody>
</table>

¹ mineral solution:

- CuSO₄.5H₂O: 0.16 g/l
- FeSO₄.7H₂O: 0.5 g/l
- ZnSO₄.7H₂O: 2.2 g/l
- MnCl₂.4H₂O: 0.5 g/l
- CoCl₂.6H₂O: 0.17 g/l
- Na₂MoO₄.2H₂O: 0.15 g/l
- H₂BO₃: 1.1 g/l
- EDTA: 5 g/l
Table 4. Protease activity in different media

<table>
<thead>
<tr>
<th>STRAINS</th>
<th>Degradation of proteins after incubation</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimal Medium</td>
<td>Complete Medium</td>
<td>Minimal Medium + Trusoy</td>
<td>Skim milk</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>BSA</td>
<td>Phytase (A.terreus)</td>
<td>pH</td>
<td>BSA</td>
</tr>
<tr>
<td>ATCC 9362</td>
<td>6.75</td>
<td>-</td>
<td>-</td>
<td>7.5</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 11906</td>
<td>7.00</td>
<td>-</td>
<td>-</td>
<td>8.28</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 20235 (= A. oryzae)</td>
<td>8.40</td>
<td>-</td>
<td>+</td>
<td>8.38</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 20245</td>
<td>8.05</td>
<td>-</td>
<td>+</td>
<td>8.18</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 20387</td>
<td>7.45</td>
<td>-</td>
<td>-</td>
<td>7.5</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 20388</td>
<td>8.20</td>
<td>-</td>
<td>+</td>
<td>7.5</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 42249</td>
<td>8.30</td>
<td>-</td>
<td>+</td>
<td>7.5</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 42250</td>
<td>8.30</td>
<td>-</td>
<td>+</td>
<td>7.5</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 42251</td>
<td>8.40</td>
<td>-</td>
<td>+</td>
<td>7.5</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 46250 (= A. oryzae)</td>
<td>7.75</td>
<td>-</td>
<td>+</td>
<td>8.0</td>
<td>+</td>
</tr>
<tr>
<td>A. niger</td>
<td>3.85</td>
<td>+</td>
<td>-</td>
<td>4.25</td>
<td>+</td>
</tr>
</tbody>
</table>
Legend:  
+ (partial) degradation of proteins after 4 hours incubation, large milk clearing zone  
- no degradation of proteins after 4 hours incubation, small/no milk clearing zone

Incubation at 30°C:  
27 µl medium sample  
2.5 µl BSA (25 mg/ml)  
0.5 µl Phytase (*A. terreus*, 3-4 g/l)  

BSA and phytase were added after the mediums sample was taken from the culture. This sample was incubated at 30°C and after certain timepoints the sample was analysed for the degradation of BSA and phytase.
Table 5.  Protease activity at different pH values

<table>
<thead>
<tr>
<th>STRAINS</th>
<th>Degradation of proteins in Minimal Medium + Trusoy after incubation</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pH = 4.5</td>
<td>pH = 6</td>
<td>pH = 8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BSA</td>
<td>Phytase (A. terreus)</td>
<td>BSA</td>
<td>Phytase (A. terreus)</td>
</tr>
<tr>
<td>ATCC 9362</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 11906</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ATCC 20235 (= A. oryzae)</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 20387</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Legend:  
+  (partial) degradation of proteins after 4 hours incubation  
-  no degradation of proteins after 4 hours incubation

Incubation at 30°C:  
25 µl medium sample  
50 mM NaAc pH = 4.2  
2 µl buffer (50 mM)  
50 mM NaAc pH = 5.8  
2.5 µl BSA (25 mg/ml)  
50 mM Tris/Cl pH = 8.3  
0.5 µl Phytase (A. terreus, 3-4 g/l)

BSA, phytase and buffer were added after the medium sample was taken from the culture. This sample was incubated at 30°C and after certain timepoints the sample was analysed for the degradation of BSA and phytase.
Table 6. PCR results for cloning fungal pclA genes

<table>
<thead>
<tr>
<th>Primercombination</th>
<th>Expected size PCR product</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 pcl1 + pcl2rev</td>
<td>180 bp</td>
</tr>
<tr>
<td>2 pcl1 + pcl3</td>
<td>350 bp</td>
</tr>
<tr>
<td>3 pcl1 + pcl4</td>
<td>500 bp</td>
</tr>
<tr>
<td>4 pcl1 + MBL1372</td>
<td>300 bp</td>
</tr>
<tr>
<td>5 pcl2 + pcl3</td>
<td>200 bp</td>
</tr>
<tr>
<td>6 pcl2 + pcl4</td>
<td>350 bp</td>
</tr>
<tr>
<td>7 pcl2 + MBL1372</td>
<td>150 bp</td>
</tr>
<tr>
<td>8 MBL1298 + pcl2rev</td>
<td>180 bp</td>
</tr>
<tr>
<td>9 MBL1298 + pcl3</td>
<td>350 bp</td>
</tr>
<tr>
<td>10 MBL1298 + pcl4</td>
<td>500 bp</td>
</tr>
</tbody>
</table>

Strain | Primercombination
--- | ---
| Trichoderma reesei QM9414 | 1 2 3 4 5 6 7 8 9 10 |
| Penicillium chrysogenum P2 | + + - - - - + - - |
| Fusarium venenatum ATCC20334 | + + + - - + + + |
| Trametes versicolor TV1 | - - - - - - - - |
| Rhizopus oryzae ATCC200076 | + - - - - + - - |
| Agaricus bisporus HORST | - + - - - - - + - |
| Aspergillus sojae ATCC11906 | + + + - - + - - |
| positive control | + + + + + + + + |

Legend: + specific PCR product
- aspecific or no PCR product

this PCR product was used for sequencing
Table 7. The viscosity ranges of the various *A. sojae* strains

<table>
<thead>
<tr>
<th>Strains</th>
<th>Viscosity (cP)</th>
<th>Biomass (g/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shear rate 6.5 l/s</td>
<td>Shear rate 83.2 l/s</td>
</tr>
<tr>
<td><em>A. sojae</em> wild type</td>
<td>&gt;&gt;2000</td>
<td>1505</td>
</tr>
<tr>
<td><em>A. sojae pclA</em></td>
<td>2000</td>
<td>751</td>
</tr>
<tr>
<td><em>A. sojae IfvA</em></td>
<td>1565</td>
<td>94</td>
</tr>
</tbody>
</table>

Table 8. Promoter strength in *A. sojae* transformants

<table>
<thead>
<tr>
<th>Transformants</th>
<th>Promoter</th>
<th>GUS activity (U/mg) in Minimal Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5% xylose</td>
</tr>
<tr>
<td>ATCC11906 wild type</td>
<td>____</td>
<td>0</td>
</tr>
<tr>
<td>ATCC11906[pGUS54]</td>
<td><em>gpdA</em></td>
<td>9141</td>
</tr>
<tr>
<td>ATCC11906[pGUS64]</td>
<td><em>glaA</em></td>
<td>33</td>
</tr>
<tr>
<td>ATCC11906[pBIPGUS]</td>
<td><em>bipA</em></td>
<td>2914</td>
</tr>
</tbody>
</table>
CLAIMS

1. A recombinant *Aspergillus sojae* comprising an introduced acetamidase S (*amdS*) gene as selectable marker.

2. An *Aspergillus sojae* according to claim 1, said *Aspergillus sojae* being selectable on a medium comprising a substrate for the introduced *amdS* as sole source of nitrogen, said medium further comprising a carbon substrate and said medium being free of endogenous *amdS* inducing substrate.

3. An *Aspergillus sojae* according to claim 1 or 2, wherein the source of nitrogen is acrylamide.

4. An *Aspergillus sojae* according to any of the preceding claims wherein the *Aspergillus sojae* has no active endogenous *amdS* gene, for example because the endogenous *amdS* gene comprises an endogenous *amdS* inactivating mutation, e.g. a deletion or a disruption.

5. A method of introducing a nucleic acid sequence into *Aspergillus sojae*, said method comprising subjecting *Aspergillus sojae* to a method of introduction of a nucleic acid sequence e.g. transformation or transfection of the *Aspergillus sojae* in a manner known per se for introduction of a nucleic acid sequence into fungi, said method comprising the introduction of the *amdS* gene as the nucleic acid sequence (henceforth the introduced *amdS* gene) followed by selection of the resulting transformed or transfected *Aspergillus sojae* on a medium free of endogenous *amdS* inducing substrate, said medium further comprising a substrate for the introduced *amdS* as sole source of nitrogen and said medium further comprising a carbon substrate, said medium enabling the desired *Aspergillus sojae* comprising the nucleic acid sequence to grow whilst eliminating growth of *Aspergillus sojae* free of the so-called introduced nucleic acid sequence due to inability of such *Aspergillus sojae* to grow without the introduced *amdS* gene on the selection medium, said medium suitably comprising a substrate for *amdS* other than acetamide, for example acrylamide as substrate for the introduced *amdS* as sole source of nitrogen.

6. An *Aspergillus sojae* obtained by the method of claim 5.

7. A method of selecting transformed or transfected *Aspergillus sojae* said method
comprising subjecting *Aspergillus sojae* according to any of claims 1-4 and 6 to a method of transformation or transfection of the *Aspergillus sojae* in a manner known per se for transformation or transfection of fungi with a nucleic acid sequence, said method comprising the introduction of the *amdS* gene as the nucleic acid sequence followed by selection of the resulting transformed or transfected *Aspergillus sojae* on a medium comprising a substrate for the introduced *amdS* as sole source of nitrogen and said medium further comprising a carbon substrate, said medium enabling the desired *Aspergillus sojae* to grow whilst eliminating growth of non-transformed or -transfected *Aspergillus sojae* due to inability of such to grow without the introduced *amdS* gene on the selection medium.

8. A method for producing recombinant *Aspergillus sojae*, said method comprising introduction of a nucleic acid sequence into an *Aspergillus sojae* e.g. by transformation or transfection in a manner known per se according to any of claims 1-4 and 6, said nucleic acid sequence comprising a desired sequence to be introduced flanked by sections of an endogenous *amdS* gene or corresponding sequences of a length and homology sufficient to ensure recombination thus simultaneously eliminating the endogenous *amdS* gene and introducing the desired sequence, followed by selection of the recombinant *Aspergillus sojae* with the desired sequence by selecting for a selectable marker comprised in or transformed in cotransformation with the desired sequence, said selectable marker being absent in the *Aspergillus sojae* prior to introduction of the nucleic acid sequence, suitably the selectable marker being *pyrG*.

9. An *Aspergillus sojae* exhibiting growth with medium comprising uracil and fluoroorotic acid, said *Aspergillus* further not exhibiting growth on medium comprising uridine and fluoroorotic acid, i.e. said *Aspergillus sojae* exhibiting uracil auxotrophy, said *Aspergillus sojae* being unable to utilize uridine, said *Aspergillus sojae* being *pyrG* negative, said *Aspergillus sojae* exhibiting resistance to fluoroorotic acid, said uracil auxotrophy and said fluoroorotic acid resistance being relievable upon complementation with an active introduced *pyrG* gene, suitably said *Aspergillus sojae* being free of active endogenous *pyrG* genes; e.g. the *Aspergillus sojae* endogenous *pyrG* gene comprises a mutation in the form of an insertion, substitution or deletion in the gene or in a gene regulating sequence, e.g. a deletion of the whole coding
sequence of the gene.

10. An *Aspergillus sojae* according to claim 9 in combination with the characteristics of an *Aspergillus sojae* according to any of claims 1-4 and 6.

11. A method of selecting transformed or transfected *Aspergillus sojae*, said method comprising subjecting *Aspergillus sojae* according to claim 9 or 10 to a method of transformation or transfection with a nucleic acid sequence, said method comprising introducing an active *pyrG* gene into the *Aspergillus sojae* in a manner known per se for transformation or transfection of fungi followed by selection of the resulting transformed or transfected *Aspergillus sojae* on a medium free of uracil and fluoro-orotic acid, said medium at least further comprising minimum substrates required for growth of *Aspergillus sojae*, said medium enabling the desired *Aspergillus sojae* to grow whilst eliminating growth of non-transformed or -transfected *Aspergillus sojae* due to inability of such to grow without uracil due to the inactivated *pyrG* gene.

12. A method according to claim 11, wherein the active *pyrG* gene that is introduced is flanked by identical nucleic acid sequence fragments, and the *pyrG* positive *Aspergillus sojae* resulting from the introduction of the *pyrG* gene and the flanking sequences is selected on a medium free of uracil and fluoro-orotic acid and subsequently the *pyrG* positive *Aspergillus sojae* is cultivated on medium comprising uracil and fluoro-orotic acid thereby eliminating the *pyrG* gene that had been introduced thus resulting in a *pyrG* negative *Aspergillus sojae* that is selectable by growth on uracil comprising medium and by fluoro-orotic acid resistance, suitably the flanking sequences and the *pyrG* gene being further flanked by sequences that direct integration of the *pyrG* gene and the flanking sequences into a specific location due to the fact that the integration directing sequences are homologous to a specific sequence of the *Aspergillus sojae* to be transformed, thereby enabling knock out, if desired, of the gene associated with the specific sequence.

13. A method according to claim 11 or 12, wherein the *Aspergillus sojae* according to claim 9 or 10 has a further nucleic acid sequence introduced therein, preferably said further nucleic acid sequence encoding a protein or polypeptide, said further nucleic acid sequence being introduced with the active *pyrG* gene either on the same vector or by cotransformation with the active *pyrG* gene that is introduced.
14. A method of selecting transformed or transfected *Aspergillus sojae* by carrying out the method according to any of claims 11-13 in combination with the method of claim 5.

15. A method for producing recombinant *Aspergillus sojae*, said method comprising introducing a nucleic acid sequence into a pyrG positive *Aspergillus sojae*, e.g. by transformation or transfection in a manner known per se, said nucleic acid sequence comprising the desired sequence flanked by sections of the *pyrG* gene or corresponding sequences of a length and homology sufficient to ensure recombination eliminating the *pyrG* gene and introducing the desired sequence, followed by selection of the recombinant *Aspergillus sojae* with the desired sequence by selecting for *Aspergillus sojae* with a pyrG negative phenotype.

16. A recombinant *Aspergillus sojae* obtained by a method according to any of claims 11-15, optionally further comprising the characteristics of an *Aspergillus sojae* according to any of claims 1-4, 6, 9 and 10.

17. A recombinant *Aspergillus sojae* comprising an introduced nucleic acid sequence encoding a protein or polypeptide for expression, said protein or polypeptide being susceptible to degradation upon expression by *Aspergillus niger* or *Aspergillus awamori*.

18. A recombinant *Aspergillus sojae* comprising an introduced nucleic acid sequence encoding a protein or polypeptide for expression, said protein or polypeptide being other than *Aspergillus sojae* protease and amylase, said protein or polypeptide preferably being a non-*Aspergillus sojae* protein or polypeptide.

19. A mutant or recombinant *Aspergillus sojae* comprising a mutation inactivating a protease gene, suitably an alkaline protease gene.

20. A mutant or recombinant *Aspergillus sojae* comprising a mutation inactivating the major protease gene, suitably a mutation inactivating the major alkaline protease gene, e.g. the gene encoding major alkaline protease gene of 35 kDa.

21. A method for producing recombinant *Aspergillus sojae*, said recombinant *A. sojae* exhibiting reduced proteolytic activity, said method comprising introduction into an *A. sojae*, e.g. by transformation or transfection in a manner known per se, of a nucleic acid sequence comprising a selectable marker encoding sequence to be introduced flanked by sections of the protease gene to be eliminated and further said flanking
sequence and the selectable marker encoding sequence being comprised within sequences of a length and homology sufficient to ensure recombination at the protease gene thus simultaneously eliminating the protease gene and introducing the desired selectable marker encoding sequence, the introduction being followed by selection of the recombinant *A. sojae* by selecting for the selectable marker, whereby the *A. sojae* prior to the introduction of the nucleic acid sequence, e.g. by transformation or transfection, is free of the selectable marker to be introduced, e.g. the *A. sojae* prior to the introduction of the nucleic acid sequence being mutated such that the *A. sojae* cannot produce active selectable marker, suitably the selectable marker being the pyrG gene, suitably the method being carried out together with the method according to any of claims 11-15.

22. A recombinant *Aspergillus sojae* obtained according to the method of claim 21.

23. A mutant or recombinant *Aspergillus sojae* according to any of claims 17-20 or 22 comprising a selectable marker, preferably amdS as defined in any of claims 1-4 or 6 and/or pyrG as defined in claims 9, 10 or 16.

24. A recombinant *Aspergillus sojae* according to any one of claims 1-4, 6, 9, 10, 16-20, 22 and 23, comprising an introduced nucleic acid sequence encoding phytase or a protein having phytase activity.

25. A process of expression of an introduced nucleic acid sequence encoding a protein or polypeptide comprised in a recombinant or mutant *Aspergillus sojae* as defined in any of the claims 1-4, 6, 9, 10, 16-20, 22-24 or obtained via a method according to any of claims 5, 7, 8, 11-15 and 21, said process comprising cultivating the recombinant or mutant *A. sojae*, suitably the introduced nucleic acid sequence encoding a protein or polypeptide being absent in the corresponding non-transformed or wild-type *A. sojae* and/or being present in a lower copy number.

26. A recombinant fungus comprising a mutation in a gene encoding a proprotein convertase or a functionally equivalent protein.

27. A fungus according to claim 26 exhibiting increased production of a protein, polypeptide or metabolite under equivalent conditions when compared to the corresponding wild-type fungus.

28. A fungus according to claims 26 or 27 said mutation being obtained by specific gene
modification using transformation or transfection in a manner known per se.

29. A fungus according to claims 26-28, said proprotein convertase or functionally equivalent protein being encoded by a nucleotide sequence of which a fragment can be amplified by in vitro DNA amplification using any of two mixtures of nucleotides given in SEQ ID Nos. 10 to 16.

30. A fungus as described in claim 27, said proprotein convertase or functionally equivalent protein being encoded by a nucleotide sequence allowing functional complementation of the growth phenotype of an *Aspergillus niger* mutant comprising a mutation which inhibits the activity of a proprotein convertase or a functionally equivalent protein.

31. A fungus according to any of claims 26-30, said fungus being an *Aspergillus sojae*.

32. A fungus according to any of claims 26-30, said fungus further containing an introduced *amds* gene or *pyrG* gene.

33. A process for expressing a protein or polypeptide, preferably a recombinant protein or polypeptide, encoded by a nucleotide sequence, said process comprising cultivating a fungus according to any of the claims 26-32.

34. A process for producing a protein or polypeptide, preferably a recombinant protein or polypeptide, said process comprising a process of expression according to claim 33, optionally including processing and/or secretion and/or isolation of the expressed protein or polypeptide.

35. A process for producing a phytase or a protein having phytase activity, preferably a recombinant phytase or recombinant protein having phytase activity, said process comprising a process of expression according to claim 33, optionally including processing and/or secretion and/or isolation of the expressed phytase or protein having phytase activity.
Fig 1 (1)

XENPC2

XNFURIN

SCKEX2

KLKEX1

CAKEX2

SPKRP

YLKEX2 MLRFILGLL LASQAVAQLP HKERDYDSRV VVALSLRDGL DPREFEASVS

51 100

XENPC2

XNFURIN

SCKEX2

KLKEX1

CAKEX2

"MLPIKLLIF

SPKRP

YLKEX2 GLDHGQWTFE HPVGTPNTY VFSAPKEYAP IENIRDQDLR EVAGGVLAKR

101 150

XENPC2

XNFURIN
Fig 1 (2)

SCKEX2 ~~~~~~~ M KVRKYITLCF WWAFSTSLAV
KLKEX1 ~~~~~~~~~ MILSSLQLM LALIAVSGYG
CAKEX2 ILGYLLSPTL QQYQQIPPRD YENKNYFLVE LNTTNSQKPL IDFISHYRGH
SPKRP ~~~~~~~~~~~~ ~~~MHP
YLKEX2 ELRKREKLQK KYGMSEEDVE KRLVALERLD YDWSERGLGS LEVLSERRIH

151 200

XENPC2 ~~~~~~~~~
XNFURIN ~~~~~~~~~

SCKEX2 SSQQIPLKDH TSRQYFAVES NETLSRLEEM HPNWKYEHDV RGPLNHVFS
KLKEX1 KAMQVPKKDH ENRQYFAIES YDDVGNLLAE HSDWSFEHDV RGLANHYVFS
CAKEX2 YNFEHQLSSL DNHYVFSDK SHPHNSFLGN HNSNEYNLMK RQLGHEQDYD
SPKRP ALLCGPILAI FLQFLVSSCS PLENDLFLV QVEPEVDPVV AAEAIGAKYV
YLKEX2 KRAPVNWTEE EMELYKEIKR RAEEAQKAQD DKGDKKEDQK DDKKEGQEAQ
Fig 1

201 250

XENPC2 --------------  KVNKEQEQE GFHRKKKR... 
XNFRUIN --------------  QVHWLEQQV AKKRKKKR... 
SCKEX2 KELLKLGKRS SLEELQCDNN DHLLSVHDLF PRNDLFKRLP VP.....APP 
KLKEX1 KPLQSLGKRD AIDTGYSEN...IIDFHDL. PPVQLHKRLP IG....... 
CAKEX2 ELISHVESIH LLPMKKLSKR IPVPIEMEDV VFDNRDDTGS DN.....HEA 
SPKRK RPLLNLKYHH LIKLHGSGD SVQSSIRKRG IDAGILELER QTPRWRYKRD 
YLKEX2 KEGDKEDNKG DDKGEDDEED DDEDEDEDD ASPAMPVQWK PVDESMYGGM 

251 300

XENPC2 ......GYR.. DINDIEINMN DPLFTKQWYL INTGQDADGT PGLDINVAEA 
XNFRUIN ............. ...DIYTDP DPKFMQQWYL LDTRNHR.... ...DLHVEKA 
SCKEX2 MDSSLLPV... KEAEDKLISN DPLFERQWHL VNPSPFPGS... ...DINVDL 
KLKEX1 .DSSMEQI.. QNARILFNIS DPLFDQQWHL INPNYPGN.. ...DVNVTGL 
CAKEX2 TDEAHQKL.. IIEIAKLDIH DPEFTTQWHL INLKYPGH.. ...DVNVTGL 
SPKRK ASESDELL.. NEFSNHFGIS DPLFYQWHL FNSNPNPGH.. ...DLNREV 
YLKEX2 PDDSLYDVYR KYYPDEVGK DPNLWKQWYL HNVHKAGH.. ...DLNVTGL
XENPC2 WELGYTGRGV TIAIM.DDGI DYLHPDLASN YNAEASYDFS SNDPYPYPRY
XNFURIN WEQGFRTGKGI VVSILSDDGGI EKNHPDLQAN YDPAAASYDVN DQDPPQPPKY
SCKEX2 WYNONTGAGV VAAI.V.DDGL DYNENELKDN FCAEGSWDFN DNTNLPKPRPL
KLKEX1 WKENITGYGV VAALV.DDGL DYENEDLKDN FCEVGSWDFN DNNPLPKPRPL
CAKEX2 WLLEDILQQGI VTALV.DDGV DAESDDIKQN FNSEGWSDFN NKGKSPLPLRL
SPKRP WDAGYFGENV TVAFV.DDGI DFKHPDLQAA YTLSLGSWDFN DNIADPLPKL
YLKEX2 WLRNVWGTVGVTAVV.DDGL DMNAEDIKAN YFAEGSWDFN FNKSDPKPSS

XENPC2 TDDWFNSHGT RCAGEVSASA NNNICGVGVA YNSKVAGIRM LDQPFMTDII
XNFURIN TQLNDNRHGT RCAGEVAVA A NNGICGVGIA YNANIGVIRM LDGE.VTDAV
SCKEX2 SDDY...HGT RCAGEIAAKK GNNFCGVGVA YNAKISGIRI LSGD.ITTED
KLKEX1 KDDY...HGT RCAGEIAAFR .NDICGVGVA YNSKVSGIRI LSGQ.ITAED
CAKEX2 FDDY...HGT RCAGEIAAVK .NDVCIGVGA WKSQVSGIRI LSGP.ITSSD
SPKRP SDDQ...HGT RCAGEVAAA. WNDVCGVGIA PRAKVGRLRI LSAP.ITDAV
YLKEX2 HDDY...HGT RCAGEIAAVR .NNVCGVGVA YDSKVAGIRI LSKE.IAEDI
XENPC2 EASSISHMPQ VIDIYSASWG PTDDGKTVDG PRELTLQAMA DGVNKGRRGGK
XNFURIN EARSGLNPN HIHIYSASWG PEDDGKTVDG PAKLAEAEFY RGVTQGRGGL
SCKEX2 EAASLIYGLD VNDIYSASWG PADDGRHLQG PSDLVKKALV KGVTEGRDSK
KLKEX1 EAASLIYGLD VNDIYSASWG PSDDGKTMQA PDTLVKKIAI KGVTEGRDAK
CAKEX2 EAEAMVYGLD TNDIYSASWG PTDNGKVLSE PDVIVKKAMI KGIQEGRDKK
SPKRP ESEALNYGFQ TNHIYSASWG PADDGRAMDA PNTATRRALM NGVLRNRRNL
YLKEX2 EALAINYEMD KNDIYSASWG PPNGQTMAR PGKVVKDAMV NAITNGRQGGK

XENPC2 GSIYVWASGD GG SYDDCNC DGYASSMWTI SINDAINDGR TALYDESCSS
XNFURIN GSIYVWASGN GGREHDSCNC DGYTNSIYTL SISSTTQMGN VPWYSEACSS
SCKEX2 GAIYVFASGN GGTRGDNCNY DGYTNSIYSI TIGAIDHKDL HPPYSEGCSA
KLKEX1 GAYVFASGN GMFGDSCNF DGYTNSIFSI TVGAIDWKGL HPPYSESCSA
CAKEX2 GAIYVFASGN GGRFGDSCNF DGYTNSIYSI TVGAIDYKGL HPPYSEACSA
SPKRP GSIFVFAFGN GGHYHDNCFN DGYTNSIFSA TIGAVDAEHK IPFYSEVCAA
YLKEX2 GNVVFASGN GGSRGDNCNF DGYTNSIYSI TVGALDFNDG HPPYSEACSA
XENPC2 TLASTFSNGR KRNPAGVAT TDLY...... .GNCTLRHS GTSSAAPEAAG
XNFURIN TLATTYSSGN QN..EKQIVT TDLR...... .QKCTDSHT GT~~~~~~~~
SCKEX2 VMAVTYSSGS GEY....IHS SDIN...... .GRCSNSHG GTSSAAPLAAG
KLKEX1 VMVVTYSSGS GNY....IKT TDLD...... .EKCSNTHG GTSSAAPLAAG
CAKEX2 VMVVTYSSGS GEH....IHT TDI...... .KKKCSATHG GTSSAAPLASG
SPKRP QLVSAYSSGS HLS....ILT TN.P...... .EGTCTRSHG GTSSAAPLASA
YLKEX2 NMVVTYSSGS EHY....IVG TDINAIDDKS AAPRCQNQHG GTSSAAPLAAG

XENPC2 VFALALEANP GLTWRDLQHL SVLTSK~~~~ ~~~~~~~~~~~~~
XNFURIN ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~
SCKEX2 VYTLLLEANP NLTWVDVQYL SILSAVGLE. KNADGDWRDS AMGKKYSHRY
KLKEX1 IYTLVLEANP NLTWVDVQYL SILSSEEIN. PH.DGKWQDT AMGKRYSHTY
CAKEX2 IYSLILSAANP NLTWVDVQYI SVLSATPIN. EE.DGNYQTT ALNRRYSHKY
SPKRP VYALALSIRP DLWRRDIQHI TVYSASPFDS PSQNAEWQKT PAGFQFSHHF
YLKEX2 VFALALSVRP DLWRRDMQYL ALYSAVEIN. .SNDDGWQDT AGQRPFHQF
XENPC2  ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~
XNFURIN ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~
SCKEX2 GFGLDAHL IEMSXSWENY NAQTFQYLPT LUVSQSTNSN .......... 
KLKEX1 GFGLDAYNI VHMAKSWINV NPQGWLQRT PVEKQSNISNS .......... 
CAKEX2 GTGKTDAYKM VHSACTWYVNV KPOAWYYSIDT IEXNQSTITTT PEQKAPSKRD 
SPKRPGFGLDAKFR VEOAVDKQVNV NPOQTWIAPE INYMNKGDSV NNETITE ...
YLKEX2 GTGKLDSKIV VELAEGWNLV NQTSFHSF CVTSQKV .......... 

XENPC2  ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~
XNFURIN ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~ ~~~~~~~~~
SCKEX2 ..EETLESVI TISEKSLQDA NFKEIRHVTQ TVRDMTEIRG TTVTDLISPA 
KLKEX1 ..DEVIESV TVSAEEFQON NLKRELHVTQ TVRIDAPYRG HVLVSDLISPD 
CAKEX2 SPQKIIHSSV NVSEKDLKIM NVERVEHITQ KVNIIDSYRG RVGMDIISPT 
SPKRPG......MVSEF TVTKDMEKNS NFKRELEHVTG RVCIPFNRG ALEELLEPS 
YLKEX2 KYNPELKSVT TVTRDDLDKQV NFKRAEHITQ VLNLEASYRG HVRVLLKGPR
Fig 1

701 750
XENPC2
XNFURIN
SCKEX2 GIISNLGVVR PRDVSSEGFK DWTFMSVAHW GENGVGDWKI KV...KTTEN
KLKEX1 GVTSTLATAR RLKDNYRFQ NWTFMSVAHW GSSGVGSKWL KV...KSTHD
CAKEX2 GVSDLATFR VNDASTRGFQ NWTFMSVAHW GETGIGECKV EVFVDDSKGD
SPKRP GIRSILASER PYDENSKGFL DWTFMTQWHW AEPPEGVWKL ..LVNDRSGG
YLKEX2 GVVSLEALLR RRDRSKDYGD NWAFMSVAHW ADEEGDWEL TV...ENTGE

751 800
XENPC2
XNFURIN
SCKEX2 GHRIDFHSWR LKLFGESIDS SKTETVFGN DKEEVEPAAT ESTVSQYSAS
KLKEX1 NEIVTLKSWR LKMGETIDA KKKVISYGN DKEDEVKST E....S
CAKEX2 QVEINFKDQW FRIFGESIDG DKADEVYDTK DYAAIR...R ELLEKEQNS
SPKRP KHEGTFENQW LALWGESENQ SNTAPLQYDT LEIPKEMVLG IYSEPNSDLT
YLKEX2 QDQVELVNWQ LNVDGQKDK REENKEGESK PEDENKEGEK EGEKKPESDEN
Fig 1(9)

801 850
XENPC2  ********  ********  ********  ********  ********
XNFURIN  ********  ********  ********  ********  ********
SCKEX2 STSISISATS TSSISIGVET SA..IPQTTR ASTDPDSDPN TPKKLSSPRQ
KLKEX1 KTTTPTAQTS SFTTTSGEET SG.......... ........... .ANKLPRPEQ
CAKE2 KSTTTSSSTT TATTTSGGEQ DQ..KTTTSA ENKEETTKVD N.SASITTSQ
SPKRP NSSTLLSPTS TSFTSVTSA TA..TPTSTS HIPITVLPP TQPVLEPSYR
YLKEX2 KEENGKEDDK GDQXEDKPED KPEDKPEDTP EDKPEDKPED APEDKPSDEK

851 900
XENPC2  ********  ********  ********  ********  ********
XNFURIN  ********  ********  ********  ********  ********
SCKEX2 AMHYFLTIIFL IGATFLVLYF MFFMKSRRRI RRSRAETYEF DIIDTDSEYD
KLKEX1 AAQLYLAIFV IGAIIVYYY LFFLKSRRII RRSRAEAYEF DIIDTDSEYD
CAKE2 TASLTSSNEQ HQPTESNSDS DSDTDDENQ EGEEDNNDN DNGNKKANSND
SPKRP EIVAFITFFL LFAFIFVAVI WTWISAFWKA KAPPLSQQE IA* -------
YLKEX2 KPEEKPEEKPV DNNSDSSSDS SDSHTSWWPD LSSKKSAWLY GAVLLVGGFI
Fig 1(10)

901 950

XENPC2 ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

XNFURIN ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

SCKEX2 STLDNGTSGI TEPEEVEDFD FDLSEDHDHA SLSSSENGDA EHTIDSVLTN

KLKEX1 ASINKLQS~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

CAKEX2 NTGFYLMSIA VVGFIAVLLV MKFHKTPGS GRRRRRDGYE FDIIPGEDYS

SPKRP ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

YLKEX2 AVIGIYACVT RRNRVRRNRS KDAPSASSFE FDLIPHDSD DDFVYPEDETH

951 1000

XENPC2 ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

XNFURIN ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

SCKEX2 ENPFSDPIKQ KFPNDANAES ASNKLQELQP DVPPSSGRS* ~~~~~~~~~~~~~~~~

KLKEX1 ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

CAKEX2 DSDDEDSSD TRRADDSFD LGHRNDQRVV SASQQQRQYD RQQDEARDRL

SPKRP ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~

YLKEX2 RRSGNDRDLY DPFAEVEDDD DMFRISDEGE DAHVPEDEL RVSMEADKRD
Fig 1

1001 1053

XENPC2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~

XNFURIN ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~

SCKEX2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~

KLKE1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~

CAKE2 FDDLNAESLP DYENDMKIG DEEEEEEEE EGQSAKAP SNGSGS TKS

SPKRP ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~

YLKE2 NDRQNLLG* ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~
Fig 2a

Non-selection medium:

Selection medium: (WO97/04108)

Improved acrylamide selection medium:
Fig 2b

Non-selection medium:

Selection medium: (WO97/04108)

Improved acrylamide selection medium:
**Fig 2c**

Non-selection medium:

Selection medium: (WO97/04108)

Improved acrylamide selection medium:
Fig 3a

A. oryzae 1  GTCAAGTCCCA ATAGAAGGCT CGGTGATCGA TCTACCTGAG AAGTCTGGGA
ATCC11906 GTCAAGTCCCA ACGAARGGCT CGGTGATCGA TCTACCTGAG AAGTCTGGGA
A. oryzae 51  TTCTGTCCGCT TTCTGAAATA AAGATTACAA ACTCGTCTGC CACAGAAGCTT
ATCC11906 TTCTGTCCGCT TTCTGAAATA AAGATTACAA ACTCGTCTGC CACAGAAGCTT
A. oryzae 101 GTCGCTCAAT TAGCCGATGG CACGTTGAG TGCCGTGGATG TGRCATCGGC
ATCC11906 GTCGCTCAAT TAGCCGATGG CACGTTGAG TGCCGTGGATG TGRCATCGGC

INTRON I

A. oryzae 151 ATTCTGTAAA AGAGCTGCAC TGGCTCATCA ACTTgtgggt ataacct~
ATCC11906 ATTCTGTAAA AGAGCTGCAC TGGCTCATCA ACTTgtgggt ataacctgc
Fig 3b

A. oryzae 1
CCCCAAGGCAGATTC AATGAGGTAT GGGACGGGCA GCTGCAAAAA

ATCC11906 TGGCAGTTATC AATGTAATA CCTGACAAG TGGCCGGATG GGGAGGAAGC

A. oryzae 51 TGGCAGTTATC AATGTAATA CCTGACAAG TGGCCGGATG GGGAGGAAGC

ATCC11906 TGGCAGTTATC AATGTAATA CCTGACAAG TGGCCGGATG GGGAGGAAGC

A. oryzae 101 GCACCGGGGCAAG GAGCTTGGACG CATATCATCC CCCCCTGGGCG GGGAGAAGCTG

ATCC11906 ACGGGGCRAG GAGCTTGGACG CATATCATCC CCCCCTGGGCG GGGAGAAGCTG

A. oryzae 151 CAGTCCGC

ATCC11906 CAGTCCGC
4.4 kb EcoRI-StuI fragment from pAS1-1

2.6 kb Smal-NcoI fragment from pAB4-1rep

4.4 kb NcoI-EcoRI fragment from pAS1-2A

3-way ligation

pAS1-Δalp

11400bp
Fig 10 (2) 24/45

947 TCAGGAGAAGATATCTACTGCTTCTCTGGGCTCCCTATGATGATGGCG 996

158 CCAGCGCAATGATATATATATCTGCTCGGCGCTCGGATGATGGCG 207

997 CCACAATGGAAGCCCGGGCCACTCTGATCAAGCGGCGCATGGTCAATGGT 1046

208 CCACGATGGAGGCGCCAGGGATTATTATCAAACGACTATGTCACCGGT 257

1047 ATCCAAAATGGTGAGGGAAGAAGGCTCGGTTTTGTAATGGCGGCTGG 1096

258 ATCCAAAATGGCGGAGGAGTAAAGGTCTATCTCTGCTTTTCGAGCTGG 307

1097 TAACCGGTGAGCATTCTGACTGAACTTCTTTGAGCGTTACACCCAACA 1146

308 AAATGGTGCGAGGATGCACACTGCAATTTGCAGGGTTATACAAACA 357
Fig 10

1345 TGGCGCTGGCCCTCAGTGTGCGGCCGGAACCTCACCTGGCCTGACGTTGACGCTGAGTGCCTGGCAG 1394

558 TTGCCCTCGCTCTTTAGTGCCGGACCGGAACCTAAGTTGGCAGATGCCCAG 607

1395 TATTTGATGATGGGAGGCGCGCAGTGCTGTTATGAAAGATGATGGAAGCTG 1444

608 TACCTGATGATAGAGACCCAGTTCGCCTGCACAGAAAAGACACGGAGGGCTG 657

1445 GCAGGACACATAGAAGACAAGGAAGATTCAGCAGCCATGACTGGGATATGTA 1494

658 GCAGACTACCAAAATGGGGRAGAGATTATGACCTGACTGGGTTTTGGGA 707

1495 AGGTCGACACATATACGCTGGTGAAGACGGGCAGAGACCTGGGACTCTGGTG 1544

708 AAGTACATGCATATTCTAGGGTCAGTGGCAGACGTGGGAGCTGGTG 757
Fig 10

1545 AAGCCTCAAGCCTGGCTCACCACCTTTCCCCCTCAGGGTTGAGCATGAGAT 1594

758 AAACCAAGGCGTTGCTCACCACCTGGCTCAGGCTCGTTGAAGCATGAAAT 807

1595 CCCACAGGGCGACAGGGCTTGCTAGTTCTGTCAGGATGACGGAAGGATA 1644

808 CCCACAAAGTTGACCAGGGCCTTGCCAGCTCATACTGAAATTACCAAGGATA 857

1645 TGTTGAAGGGAGCCCAACCTGGAAACGGCTGGGAGCATGTCAACGTCACCATG 1694

858 TGATGTACAGCGCATTGCGAGAATTGGARCATGTCACTGTGACCATG 907

1695 ARTGTTAACCCAACCCCGGCGAGGGATCTCAGCGTTGAACGAGCAGCCC 1744

908 AATGTCATACCATCGCGAGGCGATATACGCGGGAGTTGCGCCAGCCC 957
Fig 10

1745 TGAATGGTGCTCTACAGTCATCAGCTCACCTCACGCCCACCGGGCCAGGATATCAAG 1794

958 CGAAGGTATCGTCAGTCATCTGAGTCAGCAGCGGCGGGTGCGAGATAATGCAA 1007

1795 AGGTGGGCTATGTGGATTGGACCTTCATGACGCTGTTGCTCAGTCTGtaagta 1844

1008 AGGCTGGCTATGARGATTGGACGTTTTATGACTGTGGCTCATTGGTATG. 1056

1845 aaaaactttttttttctcgggtgtcgggtttctttctgtct............aatacatat 1885

1057 ....ATTTGCTCCCGTAATTAGTTTTTCGTGCTCAGTCCTGACATTTACAT 1103

1886 ctagGGGCGATCCTCAGGGATTGGCAATGGGACTGTGATTCAGGCAACC 1935

1104 TTAGGGTGAGTCGGTGGGAGAAAGTGCAGGGTCAGTGGCACGATTACCC 1153
Fig 10

1936 AATGTCAACGAGCATACTGGGCAATTTCATCGATTGGCGACTCAACTTGTG 1985

1154 AATGTCAATGATCATGTTGGAGAAATTCATCGACTGGCGGCTCACCCTCTG 1203

1986 GGGCGAGGGCGATTGACGGAGCCGAGCAGCCTCTCCACCCCATGCTACTG 2035

1204 GGGACTTTTCGATCGACGAGGCTCCAGCCAGCCCTTCATCTATGGCCGATG 1253

2036 AACACGATGACGACCACAGCTATGAGGGAAACGCTGGCTACCAAGGACG 2085

1254 AGCATGACGATGACCCACTCGATTGAAGATGCCATTGTTGATTACACTAGT 1303

2086 ATCAGCGCCCTCCACGAAAAACCGAGCTGGCTGACAAGCCACTGGTG 2134

1304 GTTGACCCTATCCCAACTAAGACTGAAGCCCAACCTGTCCCAACTGATCC 1353
Fig 10 (b)

2135 CGTTGATCGCCCGTGAACGGTAAAGGCTACACARATCCCGGATCGCGACCC 2184

1354 CGTGATAGCTGCTGTAACGCAARAGC......CRTCTGCGCAGCCAAC... 1395

2185 GTAGTCTTACAGGCCATCGATGATGGAARRCTCCAGAA...ACCCCTA 2232

1396 GATGCCTTCAGAGGCTCTGCTCAAGA...GACATCTGAAGTCTCCACCCC 1442

2233 GTACAGGGCAARGCTCACAACCACAGTCCTTC.TCCGACCACCGGTCCAGA 2281

1443 CGACGAAACCTAGTTCTACTGARCTACCTTCTCTACCCCTACCTCTGCGGA 1492

2282 TAGTAGCCTGCTCCTTCTTCTTCTCCACGTTGGTGGTCCAGAAGCGGACCG 2331

1493 TAGCTTTTTGCGATACCTTCTTCTCCACGTTGGCTGGTTCGGACTCC 1541
Fig 10

2332 AAGTTTGGAT 2341

1542 AAGCTTTGGGT 1551
**Fig 11**

2.6 kb Smal fragment from pAB4-1rep was cloned into the EcoRV-site of pAS2-3
Fig 11

pAS2-3Δpcl

13374bp

pclA

ATG

STOP

ClaI (10480)

NcoI (13370)

SmaI (13360)

EcoRI (13350)

XbaI (13345)

HindIII (13340)

BamHI (13305)

EcoRV (13300)

BglII (10490)

EcoRI (2300)

XbaI (3000)

EcoRI (4000)

MscI (4644)

EcoRV/SmaI (4841)

BamHI (4844)

XbaI (4850)

HindIII (4874)

BamHI (4963)

ClaI (6435)

EcoRV (6439)

HindIII (6911)

HindIII (7346)

BamHI (7000)

BglII (6316)

HindIII (5309)

HindIII (7718)

NcoI (7692)

BglII (7686)

EcoRI (7967)

EcoRV/SmaI (7725)
Fig 12

301 350

Sckex2 TGA GVAAIV D DGLDYENED L KDNFCAEGS WDFNDN TNL P K PRLSD DYHG
Kl kex1 tgygvvalv ddgldyened ldknfcvegs wdfndnnplp kprlkddyhg
Asp c1a ---------- ---------- ---------- ---------- ----------
Anp c1a TGGGVTTAIV DDGLDMYNSD LRPN YFAAGS YDYNDKV PEP RPR MDDR HG
Penpc1l ----------------- ----------------- ----------------- ---------------
Agarmbl129 ----------------- ----------------- ----------------- ---------------
Trichpc1l ----------------- ----------------- ----------------- ---------------
Rhizpc1l ----------------- ----------------- ----------------- ---------------
Fuspc1l ----------------- ----------------- ----------------- ---------------
Spkrp FGENTVTAVF DDGIDFKHPD LQAAY TSLGS WDFNDNI ADP LPKLSD DQHG
Cakex2 lg ggivtalv d dgvaed ddd ikq fns eg s wdfmngkgs p lpr lfddyhg
Yl kex2 TGWG VTTAVV DDGLDMNAED IKANYFAEGS WDFNFKSDP KPSHDYHG

351 400

PCL1/MBL1298

Sckex2 TRCAGEIAAK KGNFCGVGV GYNAKISGIR LSGDITTED EAASLI YGLD
Fig 12

Klkex1 trcageiaaaf r.ndicgvgv aynskvsgir ilsgqitaed eaasliygld
Aspcla ~~~~~~~ ~~~DDVCGGVG AYDSQVAGIR ILSAPIDDAD EAAAANYGFQ
Anpcla TRCAGEIGAA K.NDVCGVGV AYDSRIAGIR ILSAPIDDTD EAAAINYAYQ
Penpcll ~~~~~~~~~ ~~~~~~VGV AYDKSVSIRG ILSKAIDDVD EAAAINFAFQ
Agarmbl129 ~~~~~~~~~ ~~~~~~CGLGL AYESKVAGVR ILSGPITDVE EATALNYGFQ
Trichpcll ~~~~~~~~AV R.TDACGLGV AYDSKIAGIR ILSSAISDAD EAEAMITYKFQ
Rhizpccl1 ~~~~~~LGAL V.KXCLWXGV AYDAKISGIR ILSGEITEAD EAAALNYKYQ
Fuspccl1 ~~~~~~~~~ ~~~~~~VXPLV LLRLQVAGIR ILSKLISDAD EAEALMYKYH
Spkrp TRCAGEVAAA W.NDVCGVG1 APRAKVAGLR ILSAPITDAV ESEALNYGFQ
Cakex2 trcageiaav k.ndv cgigv awksesqvsgir ilsgpitssd eaeamvygld
Y1kex2 TRCAGEIAAV R.NNVCGVGV AYSKVAGIR ILSKIAEEDI EALAINYEMD

401 450

PCL2/PCL2rev

Sckex2 VNDIYSCSWG PADDGRHLQG PSDLVKKALV KGVTEGRDSK GAIYVFASGN
Klkex1 vndiyscswg psddgktmqa pdtlvkka ii kgvtegrrak galyvfasgn
Aspcla RKYIYSCSWG PPDDGATMEA PGILIKRAMV NGIQNRRGGK GSIFVFAAGN
Anpcla ENDIYSCSWG PYDDGATMEA PGTLIKRAMV NGIQNRRGGK GSVFVFAAGN
Penpcll DNDIYSCSWG PPDDGATMDA PGLLIKRAMV NGVXEGRGGK GSIFVXAAGN
Agarmbl1129 NVSIFSCSWG PPDCMSMGK PGLIKKAVV GQNGQRGGK GSIFVFASGN
Trichpcl1 DNQIYSCSWG PPDDGRSMEA PVLIRRAMLK GQVEGRGGL XSIXXFASGN
Rhizpcl1 ENQIYCSWGG P----------- --------- --------- ---------
Fuspc11 DNHIYSCSWG PSDDQRTMEA PDVIRRAML KAIQGRGGL GSYVFASGN
Spkdp TNHIYSCSWG PADDGRAMDA PNTATRALM NGVNGRGL GSIFVFASGN
Cakex2 tndiyscswg ptdngkvlse pdvivkkami kgiqegrdkk gaiyvfasgn
Ylke2 KNDIYSCSWG PPDNGQTMAKPGKVVDAMV NAITNGQGK GNVFVFASGN
451 500

PCL3

Sckex2 GGTGRGNCNY DGYTNSIYSI TIGAIDKDL HPPYSEGCSA VMAVTVSSGS
Klkex1 gmgfngscnf dgytnsifs tvgaidwkggl hppysescsa vmvvytssgs
AspclA GARYDDNPNF DGYXNSIYRV TVGAIDREAN IPPYSESLA QLVAAIGSGS
AnpclA GAIHDDNCFN DGYTNSIYSI TVGAIDREGN HPPYSESCSA QLVVAYSSGA
Penpcl1 GALFGDNCNF DGYNK--- --------- --------- ---------
Agarmbl1129 GAAADDQCNY DGYTNN---- --------- --------- ---------
Trichpcl1 GAASGDNCNX DGYXN--- --------- --------- ---------
Rhizpcl1 --------- --------- --------- --------- ---------
Fuspc11 GAGQGDNCNX DGSTK--- --------- --------- ---------
Spkdp GGHYHDNCFN DGYTNSIFSA TIGAVDNEH PKFYSEVCAA QLVAYSSGS
Cakex2 ggrrfgdscnf dgytnsiysi tvgaidykgl hpgyseacsa vmvvytssgs
Ylke2 GGSRGDNCFN DGYTNSIYSI TVGALDFNDG HPPYSEACSA NMVVTYSSGS
Fig 13a

1 ccatGGTTAT TCTGCCGAAG CGaaccacc ctcccacccca aacagggcct 
51 atgtgccccag gtcctgtatac catcagaaga cctccaggag cacatgcttg 
101 ttcgcataac cgtggtgtag caccaggata tgcttagctt agcttctttcg 
151 actggggggc cagaaagtgc ttatcgcaaa gatccccactt ctttgtgtga 
201 tagccccctcc cgccggccctt gatcagccg ttctcgctcg cccataccga 
251 aaccggagata ttatatggtgc acatatggtta tatcttttttt ctttttcttt 
301 tctctttgtt tctcatgagcc cccatacggt gccaatttg gctacactt 
351 ggggctctatt cttcgaagtt tagatccgga caagacccca gcaccccaac 
401 aaaaccccttg attctctgata aagagcggttg aaaaaagcgg atatcgagtg 
451 aggatgcaca gcaaggggaa tgggtcacat tgtctctgtg cgccgtgtta 
501 ggtatgtcct cactcctaaaa ggcattgggcc ggggctattag gcccttcctg 
551 tccaagatat cggttactcc tctcattagt gcgcggctact tttgtaatta 
601 attgacttgag ggtatccaca ccttccccctt gaaggtaccg agccactacc 
651 tttgacgtta gttacttttt cggaggaagc attctatgcg agtctctgccc 
701 aatcactgca gcgtgcacaa cttgccatag ctttggttgc ttcaggggct 
751 atcgggaacac ccggtcatga ctaaagggg tcacggtccg tgtggtgcaa 
801 catcatttctc atttttcact atgccccgcttg attgatagag taatctccgg 
851 tggagccaa ccggtccttc tgaagatgcaa tgtcacccttg taagtttcaa 
901 ctacaatctg tagtacagag cactcttgct attgcatgct gtgcaagtga
951 tccaaatccg tagaacttgct cgagaacag ggaatatag aacctctgaa
1001 gttttataaat accacatgca tcctctgcttc atcctcactt ccatcatcaa

1051 gccagcgggt tctatccctcc gacctgagtt ttcttgagc g atCTTACAA

ALP-2

1101 TCTTCTCATCT ATGcagtccca tcaagcgtac ctgtgctctct ctcggagcta

Fig 13b

AMY-1

1 AGATCTGCCC TTATAAATCT CCtagtctga tcgtgacgc attccgaata
51 cgagggctga ttaatgatta catacgcctc cggtgtagtag accgagcgc
101 cgagccagtt cagcgcctaa aacgccttat acaattaagc agttaaagaa
151 gttagaatct acgcctaaa acgactttaa aaatcgatct gcaggtcccg
201 attcgcctat caaaaccagtt ttaaatcaac tgattaagg tgccgaacga
251 gctataaatg atataacacat attaaagcat taattagagc aatactcaggc
301 cgccgacgaa aggcaactta aaaaagcgaaa gcgcgtctact aaacagatta
351 cttttgaaaaa aggacatcga gtatattaag cccgaatcct tattaagcgc
401 cgaaatcagg cagataaagc catacaggca gtagacctc tacctattaa
451 atccggtctct aggccgctctc catcataaatg ttctggctgt ggtgtacagg
501 ggcataaaat tacgcactac ccgaatcgat agaactactc atttttatatat
551 agaagtcaga attcatagtg ttttgatcat ttttaatttt tatatggcgg
601 gtggtgggca actcgcttgca gcgggcacct cgctttaccga ttacgtaggg
651 gctgatatatt acgtgaaaat cgctcaagggctgcaagaca aagttgttaa
701 acccnggaag tcaacagcat ccaagccccctgcaatctagcgcacgggac
751 acgcgtccaca tcacgagcga aggaccacct ctaggcatcgc gacgcaccat
801 ccaattagaa gcagcaaagc gaaacagccc aagaaaaaggtccgccccctc
851 ggccctttct tcacgagcctt gacgttacctgca ttaaaagggctttaaatttca tcaaccaca
901 agagtagcta gggccgagata ttaaaagggctttaaatttca tcaaccaca
951 aatcagtcgc ctttaactcctgtccgtta ttaaattccataaatgcaatttaaacttttca
1001 tgcngaatcgcttggtttcccgcttc ataactctttatatgtgctcgagcggatcgtatcataacactataaatactagcaggggatgcca
1051 cctttgtcagat gcggatgtacataacactataaatactagcaggggatgcca
1101 tgcctttggagg atagcaaccc agacactccactacaagctccttcctttctcctg
/AMY-2
1151 aacaatataaac CCCACAGAAG GCATTATAGtggctgctgtgtggtctcata
1201 tttttntacatgccttcaggtgcgggcactcctgttttgctgcgacgcctgc
1251 gacgtgggca tcgcaatccatcccacctttatccttcttcagctcgccttggca
1301 gacgggtacggtgcgac
Fig 13c

1 gatatctcgg cccggaaacg gaaaggctag accgagtcgcc ccttcatttttt
51 ccattgttct catccattaa gctttggtgtg gattgctgtg tcctgtagttg
101 tagtctgtat ggccagattg taattacat cagcccccctc tatggggatag
151 cttcaggtat gggaccccaag ggattcatttt cccccctcaat tgctggaact
201 acggaacaaa ggacaaaaag atagagtaat agccgggatc gttttcctcgtg
251 tagnctaggt agtacctcccc cttcgattcc gaaaaactgg caaagatttc
301 acgagatggtt aggattgagt acccggcagta ctggatttga ggcaacgtta
351 ttgcacagac cggtagctgc cgaggagagg cagagtccccca aatatctgtga

AOGPDA-1

401 gtctctctgtct ttgccccggatg TATGAAACCG GAAAGgtag ctgggagctg
451 gggacgggca caagcggggg aacagctga caagacccca ttccactctg
501 gattcttggag agagctgttag cttttgcctct gttgtcctcag ccggtgaactg
551 gatttagtgcg ctggtctgttg cgtcagtttta cattgctttt ttttatctc
601 cccctcccccc gcggcctcag cttttctcctct tttttctactc tttctgtaata
651 ctcaaccactg caatcacctt atccctttgt ctttttaacta aaagtgagtc
701 gtctccccggc catcatcccc ttggatcttt caagtttcaag tgcctaccgt
/

751 ttcccccttcc acacagattg actgacagct acccggccac accaacagaC

AOGPDA-2

801 ACATCTAAAC AATGGCTA
Fig 14

901 gpd-box 950

Atcc11906gpdapr GCCGCCGCTA TGAACCGGA AAGGACTGCT .GAAGCTGG GGAACCGGCC

Gpdaorypr GCCCGGTGTA TGAACCGGA AAGGGCTGCT .GGGAGCTGG GGAGCGGCC

Gpdanigpr GCCCGGTGTA TGAACCGGA AAGGACTGCT GGGGACTGG GGAGCGGCC

Gpdanidpr GCCCGGTGTA TGAACCGGA AAGG.CCGCT CAGGAGCTGG CCAAGCGGCC

951 1000

Atcc11906gpdapr AGGCCGGAA .AACAGCTGA CARGGACCCCA TTTCACTCTG GATCTTGGG

Gpdaorypr AGGCCGGAA .AACAGCTGA CARGGACCCCA TTTCACTCTG GATCTTGGG

Gpdanigpr AGGCCGGAA TCCCGGTGTA CATTTGACCC ATCTCTATGC CGTGGCAGAG

Gpdanidpr AGGCCGGAA CACARGCTGG CAGTGTACCC ATCCGGTGTCT CTGCACTCGA

1001 1050

Atcc11906gpdapr AGAGCTGTAG CTTTGCCCC GTCTGTCCAC CCGGTGACTG GATTAG....

Gpdaorypr AGAGCTGTAG CTTTGCCCC GTCTGTCCAC CCGGTGACTG GATTAG....

Gpdanigpr CTTGAGGTAG CTTTGCCCC GTCTGTCCAC CCGGTGACTG CATTCCAGCTG

Gpdanidpr CCGTCTGAGG TCCCTCGTGC CCGTGTGGGC AGCTTTGCCA GTCTGTCCG
Fig 14(2) 42/45

Atcc11906gpdapr ................. ................. ................. .................

Gpdaorypr ................. ................. ................. .................

Gpdanigpr GCGCGGCCAT CTGTCGCTCC TCCAGGAGCG GAGGACCCAG TAGTAAGTAG

Gpdanidpr CCCGTTGTGT CGGCG................. ................. ...............G

1101 1150

Atcc11906gpdapr ...TGACCTG GTGTTGCGT CAGTCAA......CAT TGCTCCTTTTT

Gpdaorypr ...TGACCTG GTGTTGCGT CAGTCAA......CAT TGCTCCTTTTT

Gpdanigpr GCCTGACCGT GTGTTGCGT CAGTCCAGAG GTTCCCTCCC CTACCCCTTT

Gpdanidpr GGTTGACAAG GTGTTGCGT CAGTCCA......ACATT TGTTGCCATA

1151......... ................. ................. ................. 1200

Atcc11906gpdapr TTATCTCCCC CTCCCCCGCC GTCGGACTTT TCTCCCCCTTT T..........

Gpdaorypr TTATCTCCCC CTCCCCCGCC GTCGGACTTT TCTCCCCCTTT T..........

Gpdanigpr TCTRACTCCCC CTCCCCCGCC GCTCAGCTTT TCTTCCCTTT TTACTTTTCT

Gpdanidpr TTTTCCGCT CTCCCCACCA GCTGCTCTTT TCTTTTCTCT T..........
Fig 14(3)

1201 ......... ********* .... 1250

Atcc11906gpdapr ......CTAC TCTCTTCGTA TACTCACAC TGCAATCATC TTATCCCTTT

Gpdaorypr ......CTAC TCTCTTCGTA TACTCACAC TGCAATCACCC TTATCCCTTT

Gpdanigpr TCTCTCTTCC TCTCATCCCA TCTCTCTTCC ATCACTTTCC TCTTCCCTTC

Gpdanidpr ......... TCTTTTCCCCA TCTTCAGTAT ATTCATCTTC CCAT.CCAAG

1251 ** ********** ********** *****1300

Atcc11906gpdapr GTC...TTCT TACTTAAAGT GAGTCGTC.. TCCCGCCCAT CTTTCCCTTT

Gpdaorypr GTC...TTTT TACTTAAAGT GAGTCGTC.. TCCCGCCCAT CATTTCCCTTT

Gpdanigpr ATCCAATTCA TCTTCCAAGT GAGTCTCTCTT CCCCATCTGT CCTTCCATCT

Gpdanidpr AACCTTTATT TCCCTAAGT AAGTACTTTG CTACRTCCAT ACTCCATCCT

1301***** ********** ********** ********** *****1350

Atcc11906gpdapr GAACCTTGTG AATCAGAGCC ACTTCCAAGT GCTTACCGTT T.CCTTTCCA

Gpdaorypr GGATCTTT... ........C ACTTCCAAGT GCCTACCGTT TCCCTTTCCA

Gpdanigpr TTCCCATCAT CATCTCCCTT CCCAGCTCCT CCCCTCCCTCT CGTCTCCTCA
Fig 14(4)

Gpdanidpr TCCCATCCCT TATTCCTTGG AACCTTTGAG TTGAGCTTTT CCCACCTTCCAT

1351****** ************ ************** ******* 1400

Atcc1906gpapr CATAGATTGA CTGACAGCTA CCCCGCCACA CCRACGACACA CATCTAAACC ATG

Gpdaorypr CACAGATTGA CTGACAGCTA CCCCGCCACA CCRACGACACA CATCTAAACA ATG

Gpdanigpr CGAAGCTTGA CTAACCATTCA CCCC GCCACA .TAGACA CATCTAAACA ATG

Gpdanidpr CGCAGCTTGA CTAACAGCTA CCCGGCTTGA .GCAGACA .TCACA ATG