A functional genomics study of extracellular protease production by Aspergillus niger

Braaksma, M.

SUMMARY

The filamentous fungus *Aspergillus niger* is widely used in industrial biotechnology for the production of many substances, including citric acid and a broad range of enzymes. Because of its natural ability to secrete large amounts of proteins, it also has been explored as a host organism for the commercial production of enzymes from fungal as well as non-fungal origin. However, although heterologous fungal proteins are efficiently expressed, so far the levels of most non-fungal proteins produced are too low to be commercially interesting. Extracellular degradation of non-fungal proteins by the fungus' native protein-degrading enzymes, so-called proteases, has long been recognized as one of the bottlenecks that reduce yield.

Currently, approaches for the development and use of *A. niger* as a production host are rapidly changing by recent advances in fungal genomics and related functional genomics tools such as microarrays, proteomics and metabolomics. The aim of the work presented in this thesis was to use functional genomics tools to study the production of extracellular proteases in *A. niger*.

Chapter 1 reviews the role of proteases in strain and process development of *A. niger* and other aspergilli and provides an outlook on how functional genomics techniques may play a role in further understanding the proteolytic system of aspergilli. To this day, classical mutagenesis and molecular genetic methods both are successfully applied to generate strains with reduced protease activity. With the resulting mutants and disruptants a significant improvement of heterologous protein production levels can be reached. Disruption strategies have not only focused on individual protease-encoding genes, but also on specific regulators of protease genes as well as wide-domain regulators. However, the latter approach seems unsuitable to generate protease-deficient fungal host strains for protein production due to pleiotropic growth defects of wide-domain regulatory mutants. In addition to strain development, the selection and development of fermentation conditions that repress protease production can improve heterologous protein production, but this has never been investigated systematically. The few available studies on this subject have mainly focused on the environmental parameters pH, carbon and nitrogen source.

A systematic study of the influence of several environmental factors on the production of extracellular proteases of *A. niger* in controlled batch cultivations is described in **Chapter 2**. In the first part of the study, a change-one-factor-at-a-time approach was used to establish the factors and their levels that affect protease activity. Subsequently, four selected factors, i.e. carbon source, nitrogen source, nitrogen

concentration and pH, were investigated further using a full two-level factorial experimental design. Amongst others, the results showed a clear interaction effect between nitrogen source and nitrogen concentration. Interaction effects between environmental factors in relation to protease secretion of *A. niger* have not been reported before. Due to the occurrence of interaction effects, the selection of environmental factors to reduce protease activity is not straightforward, as unexpected antagonistic or synergistic effects may occur. Furthermore, in addition to maximum protease activity the effects of the process parameters on five other protease-related phenotypes, such as maximum specific protease activity and maximum protease productivity, were investigated. The results indicated significant differences in the effect of the environmental parameters on the various protease-related phenotypes. For instance, pH significantly affected final levels of protease activity, but not protease productivity. The experiments of the full factorial design showed large and evenly distributed variation of protease activity, and were therefore a suitable starting point for a full-scale systems biology approach (Chapters 3-5).

The availability of the sequenced genome of *A. niger* has allowed the prediction of the fraction of potentially secreted proteins by scanning for the presence of a N-terminal signal peptide (SP). Due to gene-model errors, the *in silico* predicted secretome is not an accurate description of the real secretome. Moreover, not all proteins with a SP are actually secreted; some are resident endoplasmic reticulum proteins. In **Chapter 3** an improved list of potential SP directed proteins encoded by the *A. niger* genome is presented. When compiling this list, in addition to SP predictions of *A. niger* CBS 513.88, SP predictions of the best homologs of *A. niger* ATCC 1015 and three neighbouring *Aspergillus* species were taken into account as well. We propose that the SP prediction of an *A. niger* non-SP protein was likely to be incorrect, when a SP prediction was present for the majority of the homologous proteins and vice versa. Four of those likely false negative SP predictions and one likely false positive SP prediction were re-evaluated by aligning N-terminal ends. In all cases of the false negative SP predictions, selection of an alternative start codon in the most likely reading frame would add a predicted SP feature to the alternative N-terminal end of the predicted CBS protein.

As a complement to the *in silico* data, shotgun proteomics approach was used to determine the secretome associated with *A. niger* growth and upon carbon source depletion. In our study, more than 200 proteins with a predicted SP were identified. Additionally, at least two secreted proteins, including an aspartic protease (An01g00370) with a strong similarity to aspergillopepsin apnS of *A. phoenicis*, were identified that apparently use a non-classical route for secretion. Of the other 19

proteases identified in this study and that did have a predicted SP, one aspartic protease (ATCC 53364) lacked a CBS gene-model.

The secretome state was observed to change with the growth condition. Growth on sorbitol lead to the secretion of a range of carbohydrate active enzymes, but a pectinolytic subset was specifically induced during growth on galacturonic acid. Carbon source exhaustion induced the expression of proteases, as was already observed in Chapter 2. However, it was not the number of different proteases, but the relative contribution of specific proteases that increased upon carbon source starvation. These included the most prominent protease in *A. niger*, the aspartic protease aspergillopepsin A, for which the number of spectral counts was almost upto seven times higher under carbon source exhaustion compared to the growth conditions. Also data sets from other studies confirmed the relevance of the SP-classifier approach.

One of the application fields of functional genomics tools is the optimization of microbial production processes. In **Chapter 4** we investigated the influence of the choice of the quantitative phenotype to be optimized on the outcome of a optimization strategy using metabolomics. To this end, we evaluated the production by *A. niger* of the industrially relevant products glucoamylase and protease. For both products different quantitative phenotypes associated with production were defined, taking the different time points of sampling into account as well. The information content of the metabolomics data set in relation to all these different quantitative phenotypes defined was evaluated using the multivariate data analysis tool partial least squares (PLS). The results showed that the effect of different ways to define the quantitative phenotype on the information content and resulting targets for production process optimization is much smaller than the effect of the time point of sampling.

A detailed analysis of specific metabolites identified by PLS as important to the two products under study revealed that for glucoamylase activity various sugar-derivatives were correlating. However, the identified disaccharides can be either inducers of glucoamylase secretion or formed from glucose by transglucosylation activity from glucoamylase. Especially in the first case, the disaccharides are relevant in relation to strain improvement strategies. Glucose-6-phosphate isomerase was another potential target identified by PLS for optimization of glucoamlyase in *A. niger*. For the reduction of protease activity no obvious targets were found. It was anticipated that the relation between intracellular metabolite concentrations and extracellular protease activity would not be straightforward, because multiple enzyme activities are involved in proteolytic degradation, as was also found in Chapter 3.

Chapter 5 provides a large-scale, global view of the transcriptional response of *A. niger* by the clustering of co-expressed genes. Gene co-expression networks were constructed on the basis of DNA microarray data sets from two experimental approaches. In one approach *A. niger* cultures were relatively mild perturbated by the addition of low amounts of inducer. In the other approach *A. niger* cultures were severely pertubated by the imposed environmental conditions, which included carbon source starvation.

Initially, gene co-expression networks for both the individual and combined data sets were constructed using only a set of conserved genes. Comparative analysis revealed the existence of modules, some of which were present in all three networks, while others were condition-specific. Next, all protein-coding *A. niger* genes, including hypothetical and poorly conserved genes, were integrated to the co-expression analysis by the application of module-derived consensus expression profiles. Evidence for the biological relevance of the discovered modules was provided by the overrepresentation of specific Gene Ontology terms within the modules, the overrepresentation of genes which related to specific biochemical pathways, and the presence of conserved motifs in the upstream region of many genes in several of the modules. Some of the conserved sequence motifs detected represented known binding sites for transcription factors. These included the amino acid metabolism-related transcription factor CpcA and the fatty acid metabolism-related transcription factors, FarA and FarB. In addition, not previously described putative transcription factor binding sites were discovered for the module containing genes encoding cytosolic ribosomal proteins and the module with genes related to 'gene expression'.

Finally, in **Chapter 6** a top-down systems biology approach, based on the information gathered with functional genomics technologies and in combination with multivariate data analysis tools, is discussed as a method to achieve unbiased selection and ranking of targets for both strain improvement and bioprocess optimization.