Semana de Engenharia 2010
Guimarães, 11 a 15 de Outubro

SECRETOME ANALYSIS OF *ASHBYA GOSSYPII*

Tatiana Q. Aguiar[a], Merja Penttilä[b] and Lucília Domingues[a]
[a] IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] VTT Technical Research Centre of Finland, P.O. Box 1000, 02044 VTT Espoo, Finland
E-mail: tatiana.aguiar@deb.uminho.pt

KEYWORDS
Ashbya gossypii, Secretome, Computational Analysis

ABSTRACT

To explore the potential *Ashbya gossypii* extracellular secretome two computational protocols were used. The 4726 protein-encoding genes predicted in the genome of *A. gossypii* strain ATCC 10895 were the data source for the analysis. Depending on the computational methods used, 171 to 185 proteins were predicted to be secreted proteins. Based on the results of the present study, *A. gossypii* is more similar to the closely related yeast *Saccharomyces cerevisiae* than to other filamentous fungi in its secretion ability.

INTRODUCTION

A. gossypii is a filamentous hemiascomycete of considerable importance in biotechnological industry due to its natural ability to overproduce riboflavin (vitamin B2). This fungus presents several interesting unique features making it an appealing microorganism for exploring its potentialities as a cell factory organism: it has the smallest known eukaryotic genome (Dietrich et al. 2004), is prone to easy genetic manipulation (Wendland et al. 2000) and is already used in industrial large-scale processes. However, little is known about the secretion of proteins by *A. gossypii*. Extracellular lipase has been detected, but activity was low in most conditions (Stahmann et al 1997). Although high levels of protein secretion are generally associated with filamentous growth, *A. gossypii* was more similar to the closely related yeast *S. cerevisiae* than to other filamentous fungi in its ability to produce and secrete endoglucanase I (EGI) from *Trichoderma reesei*, demonstrating that filamentous growth alone is not sufficient to ensure good protein secretion (Ribeiro et al. 2010).

Protein secretion is a cell translocation process of major biological and technological significance. Cell communication, as well as intercellular signaling and growth during development in multicellular organisms depends on the secretion pathway. The secretion and downstream processing of proteins by recombinant cells is of great commercial interest.

An ideal system for secreting a desired protein could be developed from analysis of its native secretome. The completed *A. gossypii* genome sequence provides the tools to construct such a system. The term secretome was first used to describe the repertoire of proteins that are processed through the endoplasmic reticulum (ER) secretory pathway (Tjalsma et al. 2000), but is currently more widely used to denote proteins secreted by cells into the extracellular region (Greenbaum et al. 2001). Almost all secreted proteins in eukaryotes are exported via the ER/Golgi pathway, guided by a signal peptide at the N-terminus (Blobel and Dobberstein, 1975). Use of predicted signal peptides as a tool to predict protein secretion is confounded by the fact that proteins destined for the extracellular region comprise only a fraction of those with a secretory signal peptide. Other proteins that possess such a signal include residents of the ER, Golgi complex, vesicles and plasma membrane. In addition, the N-terminal signals that target proteins to the mitochondrion or chloroplast have properties similar to the secretory signal peptide. A number of computational tools and analysis pipelines have been developed to distinguish secreted proteins from proteins targeted to other cellular compartments, reviewed in Emanuelsson et al. (2007) and O'Toole et al. (2006). In this study, two computational protocols are compared in predicting secreted proteins from the entire *A. gossypii* predicted proteins dataset.

BIOINFORMATIC ANALYSIS

To identify which of the genes in the genome sequence of *A. gossypii* strain ATCC 10895 are predicted to encode secreted proteins, we used two computational protocols to analyze its 4726 protein-encoding loci. In the first protocol, Phobius (Kall et al. 2004) was used to identify signal peptides and to discriminate putative transmembrane domains. In the second protocol, SignalP version 3 (Bendtsen et al., 2004) was used to identify the presence of signal peptides, followed by TMHMM version 2 (Krogh et al. 2001) to identify putative transmembrane proteins. Proteins with 1 predicted transmembrane spanning region were kept in the dataset if it was located in the predicted N-terminal signal peptide. Sequences with more than 1 transmembrane spanning region were discarded. In both analyses, TargetP version 1.1 (Emanuelsson et al. 2007) was used to eliminate

proteins that are predicted to be targeted to the mitochondrion.

Although the Phobius and the SignalP/TMHMM protocols use similar algorithms, they generate different results. The Phobius protocol predicted 171 proteins whereas the SignalP/TMHMM analysis pipeline predicted 185 proteins as secreted proteins. An intersecting set of 136 proteins were predicted by both protocols as secreted proteins. A summary of the analysis by these two pipelines is shown in Figure 1.

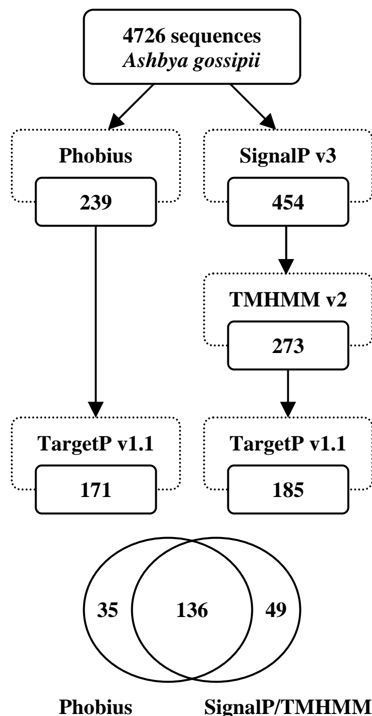

Figure 1: Flowchart of the strategies adopted for mining *A. gossypii* gene sequences for extracellular proteins and the outcome. The numbers in the boxes refer to the numbers of proteins remaining after indicated analysis steps. The relationship of the proteins identified by the two protocols is shown at the bottom of the figure.

CONCLUSIONS AND FURTHER RESEARCH

Many genome-wide approaches have been developed in recent years to explore pertinent information from genome sequence data. Using the selected algorithms Phobius, SignalP, TMHMM and TargetP an extracellular *A. gossypii* secretome of 171 to 185 proteins was predicted. Based on these results, *A. gossypii* seems to be more similar to yeast than to other filamentous fungi in its secretion ability, making it interesting to further experimentally characterize its secretome using analytical approaches.

Bioinformatic prediction, generally, is affected by the choice of tools used, but sequence-based analysis of secretomes ultimately depends on the reliability of the gene model predictions. Proteomic analysis provides us with experimental evidence to verify the gene models, and, in some cases, correct them. On the other hand, computational analysis of the genome provides us information that helps the interpretation of the results obtained with the proteomic analysis.

REFERENCES

Dietrich, F.S.; Voegeli, S.; Brachat, S.; Lerch, A.; Gates, K.; Steiner, S.; Mohr, C.; Pöhlmann, R.; Luedi, P.; Choi, S.; Wing, R.A.; Flavier, A.; Gaffney, T.D.; and Philippsen, P. 2004. “The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome.” *Science* 304, 304–307.

Wendland, J.; Ayad-Durieux, Y.; Knechtle, P.; Rebischung, C.; and Philippsen, P. 2000. “PCR-based gene targeting in the filamentous fungus *Ashbya gossypii*.” *Gene* 242, 381–391.

Stahmann, K.P.; Böddecker, T.; and Sahm, H. 1997. “Regulation and properties of a fungal lipase showing interfacial inactivation by gas bubbles, or droplets of lipid or fatty acid.” *Eur J Biochem* 244, 220–225.

Ribeiro, O.; Wiebe, M.; Ilmén, M.; Domingues, L.; and Penttilä, M. 2010. “Expression of *Trichoderma reesei* cellulases CBHI and EGI in *Ashbya gossypii*.” *Appl Microbiol Biotechnol* 87, 1437–1446

Tjalsma, H.; Bolhuis, A.; Jongbloed, J.D.; Bron, S.; and van Dijl, J.M. 2000. “Signal peptide-dependent protein transport *in Bacillus subtilis*: a genome-based survey of the secretome.” *Microbiol Mol Biol Rev* 64, 515–547.

Greenbaum, D.; Luscombe, N.M.; Jansen, R.; Qian, J.; and Gerstein, M. 2001. “Interrelating different types of genomic data, from proteome to secretome: ‘oming in on function.” *Genome Res* 11, 1463–1468.

Blobel, G. and Dobberstein, B. 1975. “Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma.” *J Cell Biol* 67, 835–851.

Emanuelsson, O.; Brunak, S. ; von Heijne, G.; and Nielsen, H. 2007. “Locating proteins in the cell using TargetP, SignalP and related tools.” *Nat Protoc* 2, 953–971.

O’Toole, N.; Min, X.J.; Butler, G.; Storms, R.; and Tsang, A. 2006. “Sequence-based analysis of fungal secretomes.” In *Applied Mycology and Biotechnology*, D.K. Arora, R.M. Berka and G.B. Singh (Eds.). Elsevier, 277–296.

Käll, L.; Krogh, A.; and Sonnhammer, E.L. 2004. “A combined transmembrane topology and signal peptide prediction method.” *J Mol Biol* 338, 1027–1036.

Bendtsen, J.D.; Nielsen, H.; von Heijne, G.; and Brunak, S. 2004. “Improved prediction of signal peptides: SignalP 3.0.” *J Mol Biol* 340, 783–795.

Krogh, A.; Larsson, B.; von Heijne, G.; and Sonnhammer, E.L. 2001. “Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.” *J Mol Biol* 305, 567–580.