

REGULAR ARTICLE

Analysis of secreted proteins from *Aspergillus flavus*

Martha L. Medina¹, Paul A. Haynes², Linda Brečić³ and Wilson A. Francisco¹

¹ Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ, USA

² Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ, USA

³ Department of Chemistry, University of Arizona, Tucson, AZ, USA

MS/MS techniques in proteomics make possible the identification of proteins from organisms with little or no genome sequence information available. Peptide sequences are obtained from tandem mass spectra by matching peptide mass and fragmentation information to protein sequence information from related organisms, including unannotated genome sequence data. This peptide identification data can then be grouped and reconstructed into protein data. In this study, we have used this approach to study protein secretion by *Aspergillus flavus*, a filamentous fungus for which very little genome sequence information is available. *A. flavus* is capable of degrading the flavonoid rutin (quercetin 3-O-glycoside), as the only source of carbon via an extracellular enzyme system. In this continuing study, a proteomic analysis was used to identify secreted proteins from *A. flavus* when grown on rutin. The growth media glucose and potato dextrose were used to identify differentially expressed secreted proteins. The secreted proteins were analyzed by 1- and 2-DE and MS/MS. A total of 51 unique *A. flavus* secreted proteins were identified from the three growth conditions. Ten proteins were unique to rutin-, five to glucose- and one to potato dextrose-grown *A. flavus*. Sixteen secreted proteins were common to all three media. Fourteen identifications were of hypothetical proteins or proteins of unknown functions. To our knowledge, this is the first extensive proteomic study conducted to identify the secreted proteins from a filamentous fungus.

Received: June 30, 2004
Revised: November 11, 2004
Accepted: November 12, 2004

**Keywords:**

Aspergillus flavus / Dispensable metabolic pathways / Rutin degradation / Secreted proteins / Tandem mass spectrometry

1 Introduction

Proteomics in filamentous fungi is still at a relatively early stage of development, particularly of secreted proteins, due in part to the lack of available genome sequence data. Protein secretion plays an important role in filamentous fungi, particularly in nutrition. A broad spectrum of structural proteins and enzymes, the majority of which are hydrolytic, is secreted by filamentous fungi [1]. This ability has been widely exploited by the biotechnology industry for the production of enzymes for commercial and industrial use. However, compared to yeast and animal systems, studies on protein secre-

tion by filamentous fungi are limited [2]. Typical studies have focused on the identification, purification, and characterization of single secreted proteins, but studies on the global analysis of fungal extracellular proteomes have not been conducted. Most of the proteins that have been characterized, although not necessarily sequenced, are those of industrial importance such as amylases, cellulases, and proteases [3]. However, other proteins secreted from dispensable metabolic pathways, which are either not required for growth or are only required for growth under a limited range of conditions, have not yet been identified or fully characterized. These enzymes are typically expressed under sub-optimal growth conditions and most likely function to enhance the survival of the fungus in response to nutrient deprivation. Nutrient utilization pathways increase the metabolic versatility of filamentous fungi, enabling them to utilize a variety of complex compounds as alternative sources of nutrients [4].

Correspondence: Dr. Wilson A. Francisco, Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287–1604, USA

E-mail: wfrancisco@asu.edu

Fax: +1-480-965-2747

Flavonols and their glycosides (flavonoids) occur widely throughout the plant kingdom as typical secondary metabolites, and filamentous fungi will occasionally encounter these complex compounds upon infection of a plant. Some species and strains of the filamentous fungus *Aspergillus* are capable of degrading the flavonoid rutin (quercetin 3-O-glycoside) as the only source of carbon *via* an extracellular enzyme system [5, 6]. Some of the extracellular enzymes involved in this metabolic pathway have been purified and characterized, including rutinase, which contains the glycosidase activities (β -glucosidase and α -rhamnosidase) that hydrolyze rutin into quercetin and glucose and rhamnose [6]; quercetin 2,3-dioxygenase (quercetinase), which catalyzes the oxygenolytic cleavage of the O-heterocyclic ring of quercetin to yield carbon monoxide and the depside, 2-protocatechuoylphloroglucinol carboxylic acid [6, 7]; and a phenol carboxylic acid acyl esterase, which hydrolyzes the depside into the last two products of the rutin degradation pathway (protocatechuic acid and phloroglucinol carboxylic acid) [8].

Proteomic analysis has proven to be the most powerful method for identification of proteins in complex mixtures and is suitable for the study of alteration of protein expression in an organism under varying environmental conditions [9]. Advances in mass spectrometric instrumentation have coincided with the availability of increasing amounts of genomic sequence data. Protein identification is now possible by means of peptide mass and fragmentation data generated by mass spectrometric analysis, matched against a database of all possible proteins encoded by a genome [10]. When the genome sequence of the organism being studied is known, data obtained from MALDI-TOF MS can theoretically provide identity of any unknown protein using peptide mass fingerprints. However, when little or no genomic sequence data is available for the organism, a different approach must be taken. Our approach is based on MS/MS of individual peptides rather than peptide mass fingerprints, and has previously been shown to be successful in the direct identification of proteins from large genomes with little or no annotation [11, 12]. Proteins can be identified by MS/MS of single peptides that coincide exactly with peptides found in protein sequences known from other fungi, or other sequenced organisms. The sequences of two species of *Aspergilli* are currently available, *Aspergillus fumigatus* and *Aspergillus nidulans*, and genome sequencing projects for *Aspergillus oryzae* and *Aspergillus flavus* are currently in progress.

In this study, the *A. flavus* strain PRL 1805, which was selected by Simpson *et al.* [13] from culture collections and enrichment techniques for its ability to degrade rutin, was used in the identification of differentially secreted proteins when the fungus was grown on the flavonoid rutin, or the simple medium glucose or the rich medium, potato dextrose broth. In a previous study, we identified a small number of proteins secreted by *A. flavus* when grown on rutin or potato dextrose using MALDI-TOF MS. Nine unique proteins were identified from rutin-grown *A. flavus*, and five unique pro-

teins were identified from potato dextrose-grown *A. flavus*. Although that study aided in the identification of a number of *A. flavus* secreted proteins, a large number of visualized proteins remained unidentified [14].

The MALDI-TOF MS technique is most effective when the genome sequence of the organism is known. In the current study, nanoLC-MS/MS was used to positively identify a large number of secreted proteins from *A. flavus*. A total of 88 samples were analyzed successfully, producing 51 unique protein identifications. Approximately one-third of these proteins were identified in secretions from all three culture conditions, and one-third of the proteins were unique to rutin-grown fungus. Although there have been some proteomic studies done on filamentous fungi [15–17], this is, to our knowledge, the first complete proteomic study of secreted proteins conducted on a filamentous fungus.

2 Materials and methods

2.1 Materials

Rutin was obtained from ACROS Organics (New Jersey, USA). Potato dextrose broth and mineral salts were obtained from Sigma-Aldrich (St. Louis, MO, USA). All other reagents were purchased from Bio-Rad (Hercules, CA, USA), unless otherwise noted. Molecular weight markers were from New England BioLabs (Beverly, MA, USA).

2.2 Expression of secreted proteins

A. flavus strain PRL 1805 was grown in rutin-containing and nonrutin-containing (glucose and potato dextrose) media and culture solutions were collected at 3, 7, 10, and 14 days.

2.3 Fungal strains and culture condition

A. flavus PRL 1805 was obtained from the American Type Culture Collection (#24457) and grown on potato dextrose slants. For the proteomic analysis, *A. flavus* was grown in a mineral salts medium containing rutin and potato dextrose broth as described previously [14], and in a mineral salts medium containing glucose. The rutin-containing and glucose-containing media were prepared as follows: 75 mL of an autoclaved mineral salts solution (0.469 g K_2HPO_4 , 0.143 g KH_2PO_4 , 0.6 g $(NH_4)_2SO_4$, 75 mg $MgSO_4$, 75 mg $NaCl$, 1.5 mg $CuCl_2 \cdot H_2O$, 7.5 μg $ZnSO_4 \cdot 7H_2O$, 7.5 μg $FeSO_4 \cdot 7H_2O$, 7.5 μg $MnSO_4 \cdot H_2O$, pH 6.8) was added to 60 mL of autoclaved 12.5 mM rutin or glucose (for a final concentration of 5 mM), 10 mL of autoclaved deionized water, and 5 mL of an aqueous suspension of spores from *A. flavus* slants. The potato dextrose culture medium was made with 145 mL of potato dextrose broth and 5 mL of an aqueous suspension of spores. All cultures were grown in 250 mL baffled flasks on a rotary shaker at 30°C and 200 rpm for a total of 14 days. The rutin-induced and noninduced culture supernatants after 3,

7, 10, and 14 days were collected separately by filtration through a No. 2 Whatman filter paper and lyophilized overnight. After lyophilization, the cultures were resuspended in 5 mL of deionized water and stored at -20°C until further analysis.

2.4 SDS-PAGE and protein 2-DE (IEF/SDS-PAGE)

An aliquot of the concentrated protein sample from each growth condition was precipitated with an equal volume of TCA (200 g/L) at -20°C for 1 h and subsequently centrifuged for 5 min ($15\,000 \times g$) at 4°C . The pellet was resuspended and washed with ice-cold 70% ethanol and recentrifuged two more times. The pellet was air-dried for 30 min and resuspended in 1-D (50 mM Tris-HCl pH 8.8, 6 M urea, 0.1% bromophenol blue, 2% SDS, 30% glycerol) or 2-DE (8 M urea, 2% CHAPS, 50 mM DTT, 0.2% w/v Bio-Lyte 3/10 carrier ampholytes (Bio-Rad), and a trace of bromophenol blue) sample buffer.

For SDS-PAGE, 12% gels ($20 \times 20\text{ cm}^2$) were run according to established methods using a Bio-Rad Protean II xL Cell apparatus at constant current (24 mA/gel). For 2-DE, IPG strips, pH 4–7 (Bio-Rad), were rehydrated with the precipitated samples from day 14, containing 2-DE buffer overnight at room temperature. IEF was performed using a Protean IEF Cell apparatus (Bio-Rad) at 20°C (250 V for 15 min and a linear gradient reaching 10 000 V for a total of 60 kVh, maintaining 50 μA per IPG strip). After IEF, the strips were stored at -80°C . Prior to the second dimensional run, the strips were equilibrated for 15 min in 6 M urea, 2% SDS, 0.375 M Tris-HCl pH 8.8, 20% glycerol containing 130 mM DTT, and then for 15 min in the same buffer containing 135 mM iodoacetamide. Equilibrated strips were transferred onto 12% gels ($20 \times 20\text{ cm}^2$) without stacking gels and the second-dimensional separation was carried out as in SDS-PAGE. All gels were fixed and stained with CBB R-250. All gel experiments used in comparing different growth conditions were performed in triplicate to ensure consistency, with only a single gel shown in the figures for clarity of presentation.

2.5 Sample preparation for nanoLC-MS/MS analysis

Protein bands from 1-D gels or spots from 2-DE were manually excised and transferred to 96-well plates. The plates were transferred to a Perkin Elmer Multiprobe-II liquid handling robot for destaining [18] and in-gel digestion with trypsin [19]. Following digestion, tryptic peptides were extracted from the gel pieces with 5% formic acid/5% ACN.

2.6 nanoLC-MS/MS

A microbore HPLC system (Surveyor; ThermoFinnigan, San Jose, CA, USA) was modified to operate at capillary flow rates using a simple T-piece flow splitter. Columns (6 cm $\times$ 100 μm id) were prepared by packing 100 Å, 5 μm Zorbax C18 resin at 500 psi pressure into columns with integrated electrospray tips made from fused silica, pulled to

a 5 μm tip using a laser puller (Sutter Instrument, Novato, CA, USA). Peptides were eluted in a gradient using buffer A (5% v/v ACN, 0.1% formic acid) and buffer B (90% v/v ACN, 0.1% formic acid), at a flow rate of 400 nL/min. Following an initial wash with buffer A for 10 min, peptides were eluted with a linear gradient from 0 to 100% buffer B over a 30 min interval. Samples were introduced onto the analytical column using a Surveyor autosampler (Surveyor, ThermoFinnigan). The HPLC column eluent was eluted directly into the ESI source of a ThermoFinnigan LCQ-Deca XP Plus IT mass spectrometer. Spectra were scanned over the range 400–1500 mass units. Automated peak recognition, dynamic exclusion, and daughter ion scanning of the top three most intense ions were performed using the Xcalibur software as previously described [10].

2.7 Database searching and data interpretation

MS/MS data were analyzed using SEQUEST, a computer program that allows the correlation of experimental data with theoretical spectra generated from known protein sequences [20], and the data were filtered and organized by DTASelect [21]. In this study, the criteria for a preliminary positive peptide identification for a doubly-charged peptide were a correlation factor (Xcorr) greater than 2.0, a delta cross-correlation factor (ΔXcorr) greater than 0.08, a minimum of one tryptic peptide terminus, and a high preliminary scoring [12]. For triply-charged peptides, the correlation factor threshold was set at 3.0. All matched peptides were confirmed by visual examination of the spectra. All spectra were searched against a composite database containing the translated genome sequences of *Saccharomyces cerevisiae* and *Escherichia coli*, known fungal sequences extracted from the NCBI nonredundant protein and Swiss-Prot databases, and translated, unannotated genomic sequence data from *A. fumigatus* (preliminary sequence data obtained from The Wellcome Trust Sanger Institute, www.sanger.ac.uk), *A. nidulans* (Aspergillus Sequencing Project, Center for Genome Research, <http://www.broad.mit.edu>), and *Coccidioides immitis* (obtained from <http://www.tigr.org>). In cases where peptides were identified from unannotated sequence data, protein function was assigned where possible by BLAST homology searching [22]. The database entry that contained the identified peptide was searched using BLASTP against the latest release of the NCBI nonredundant protein database, and an expected value cutoff score of 10^{-22} was applied.

3 Results and discussion

3.1 Visualization and identification of secreted proteins

The secreted proteins from *A. flavus* grown in rutin, glucose, and potato dextrose media, harvested at 3, 7, 10, and 14 days, and separated by SDS-PAGE, are shown in Fig. 1. In pre-

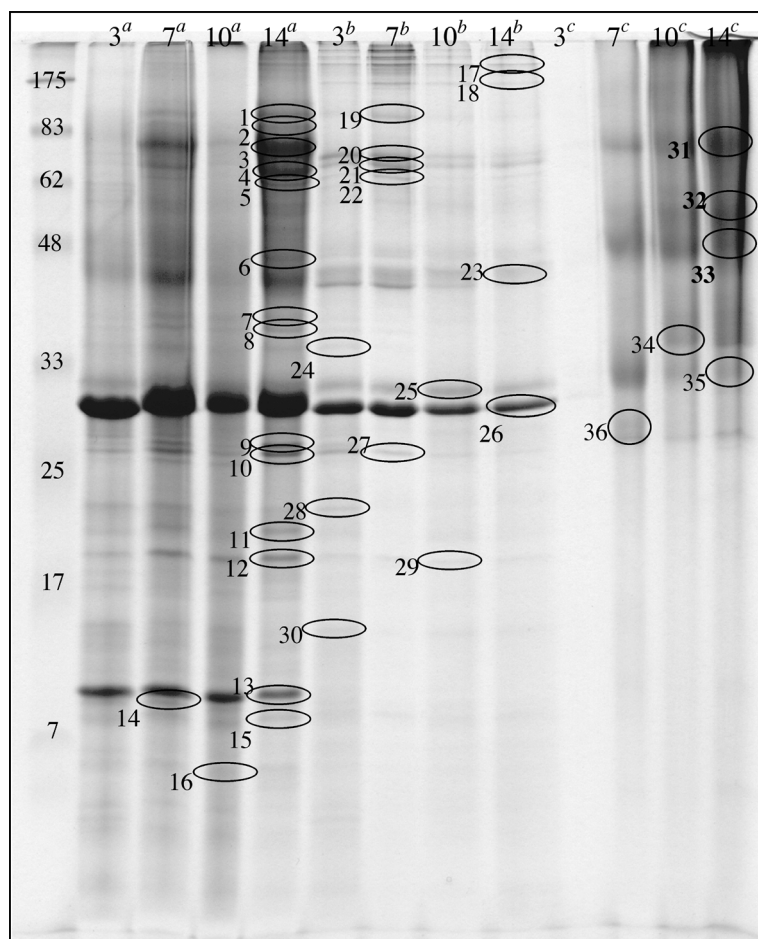


Figure 1. SDS-PAGE analysis of secreted proteins from *A. flavus* under different growth conditions over a period of 14 days, with standards indicating approximate protein molecular weights. Samples from the 3rd, 7th, 10th, and 14th day are shown. Rutin (a); glucose (b); potato dextrose (c). Spot identification is annotated. Proteins were separated using a 12% gel and visualized with CBB staining. Bands were cut from the gel as indicated and subjected to in-gel digestion with trypsin prior to nanoLC-MS/MS analysis.

vious studies, it was determined that all visible secreted proteins showed *pI* values between pH 4 and 7 on a pH 3–10 IPG strip [14]. 2-DE gels, using pH 4–7 IPG strips in the first dimension, of secreted proteins from the 14-day cultures of *A. flavus* grown in rutin, glucose, and potato dextrose media are shown in Fig. 2. More than 50 distinct spots were detected in the 2-DE gels of proteins from rutin media (Fig. 2A), while only 19 well resolved spots were detected in the gels of proteins from glucose media (Fig. 2B). The spots from the 2-DE gel of proteins from potato dextrose media were not well resolved; therefore, 14 sections of the gel as indicated were cut out (Fig. 2C). Gel images showing all the unique spots separated, with annotation of which spots were analyzed and identified by nanoLC-MS/MS analysis are depicted in Figs. 1, 2. Proteins identified from these gels are listed in the supplementary table, according to their gel of origin and spot number. From these combined experiments, 60 of the well resolved 2-DE gel spots and 36 of the 1-D gel bands were excised from the gels and digested with trypsin, and the resulting peptides were subjected to analysis by nanoLC-MS/MS. Peptides were identified from proteins present in all but 12 of the gel samples. Upon visual examination of the MS data from these 12 samples, four samples were found to

contain good quality MS/MS spectra. The remaining eight unidentified samples produced poor spectra, possibly due to low concentration of the protein in the gel. The four samples with good data, but no peptide matches in the database used, may be novel proteins. The spectra from these four samples will be researched against a complete *A. flavus* genome sequence translation when such information becomes publicly available.

3.2 Classification of differentially secreted proteins

Proteins were categorized into six functional classifications as presented in Fig. 3. The complete list of secreted proteins identified is presented in the supplementary table. Carbohydrate metabolism forms the largest class of proteins with a known function, comprising 27% of the total identified proteins. A significant portion of the proteins identified in this study are of hypothetical or unknown function, meaning that the peptide sequence matched either to an ORF that has not previously been shown to be expressed or to a protein whose function is unknown. This category also contains 27% of the total identified proteins. Proteins involved in proteolysis and

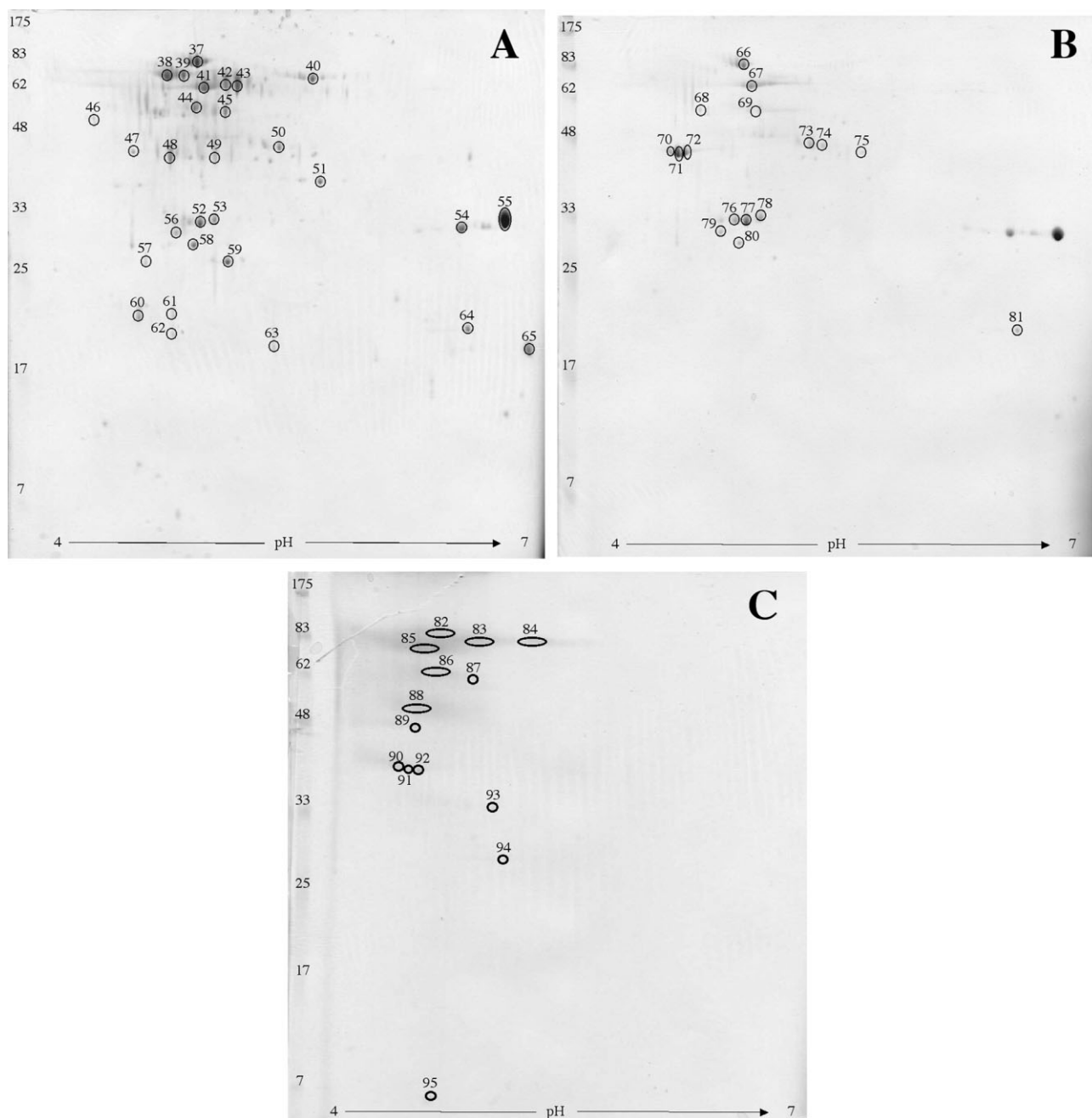


Figure 2. 2-DE of secreted proteins from *A. flavus* grown on rutin (A), glucose (B) and potato dextrose (C), harvested after 14 days, with standards indicating approximate protein molecular weights. Samples were separated using first dimension 17 cm 4–7 pI range IPG strips, followed by second dimension 12% linear SDS-PAGE, and visualization of proteins with CBB staining. Spot identification is annotated. Spots were cut from the gels as indicated and subjected to in-gel digestion with trypsin prior to nanoLC-MS/MS analysis.

peptidolysis are also highly abundant, comprising 22% of all proteins identified. There are two proteins each in the categories of redox enzymes (4%) and electron/proton transport proteins (4%), and the remainder of the identified proteins are functionally diverse and are therefore grouped together under the heading “miscellaneous proteins.”

3.3 Comparative analysis of differentially secreted proteins

Identification of the proteins secreted from the dispensable metabolic pathway in the degradation of rutin was one of the major goals of this study. Glucose and potato dextrose media

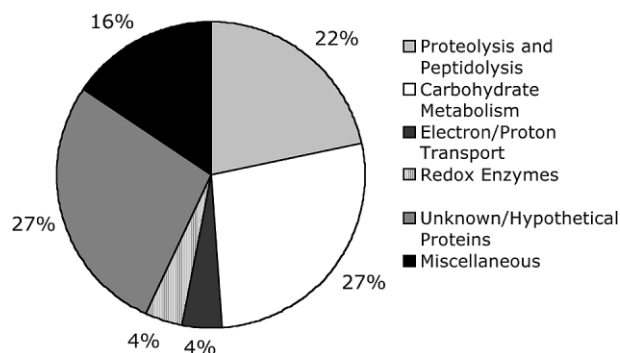


Figure 3. Functional classification of identified secreted proteins from *A. flavus*. Figure shown for each functional category represents the percentage of the total number of unique proteins identified from all three growth conditions as shown in the supplementary table. Proteins identified from one or more unique peptide hits are included.

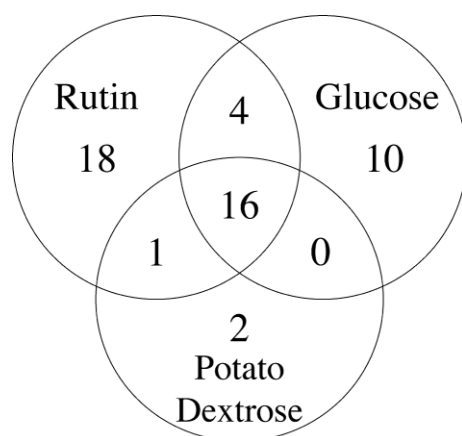


Figure 4. Culture media specificity of identified proteins. All of the unique secreted proteins listed in the supplementary table and Table 2 are classified by culture media differential secretion, with areas of overlap between culture media as indicated. Unique unknown/hypothetical proteins are included. Proteins identified from one or more unique peptide hits and unknown/hypothetical proteins are included.

were used to determine which proteins were differentially secreted by the fungus, in response to growth on different carbon sources. Distribution of differentially secreted proteins is shown in Table 1 and Fig. 4. Sixteen distinct proteins were secreted in all three media. These include several proteins involved in proteolysis and peptidolysis, carbohydrate metabolism, and proteins involved in electron/proton transport. Glutaminase was also found to be common to all three metabolic pathways. Only one protein was unique to the potato dextrose medium, ribonuclease T2 precursor. Five proteins were unique to the glucose medium: two of which are involved in proteolysis and peptidolysis, and three of which are involved in carbohydrate metabolism. One secreted protein was common to both rutin and potato dextrose media, and four to rutin and glucose media.

Table 1. Culture media specificity of identified unique secreted proteins from *A. flavus*

Protein ^{a)}	Rutin	Glucose	Potato dextrose
Proteolysis and peptidolysis			
Leucine aminopeptidase	x	x	
Carboxypeptidase S1	x	x	x
Aminopeptidase	x	x	x
Alanyl dipeptidyl peptidase	x	x	x
Alkaline protease	x	x	x
Prolyl peptidyl peptidase	x	x	x
Oryzin precursor	x		x
Carboxypeptidase D	x	x	x
Metalloproteinase 23 kDa precursor	x		
Aminopeptidase Y		x	
Neutral protease I		x	
Carbohydrate metabolism			
Beta-galactosidase	x		
Alpha-amylase	x	x	x
Beta-glucosidase	x	x	x
Glucoamylase	x	x	x
Beta(1-3)endoglucanase	x	x	x
Beta-N-acetylglucosaminidase	x	x	x
Alpha-mannosidase	x	x	x
Exo-1,3-beta glucanase	x	x	x
Beta-glucosidase 1	x	x	
Endo-1,4-beta-xylanase	x		
Arabinan endo-1,5-alpha-L-arabinosidase A precursor	x	x	
Alpha-glucosidase		x	
Alpha-glucosidase B		x	
Chitinase	x		
Putative glucanase precursor		x	
Electron/proton transport			
Isoamyl alcohol oxidase	x	x	x
Catalase	x	x	x
Redox Enzymes			
Manganese superoxide dismutase	x		
Cu, Zn superoxide dismutase	x		
Miscellaneous			
Glutaminase	x	x	x
ATP synthase beta subunit	x		
Quercetin 2,3-dioxygenase	x		
Phosphoglyceromutase	x	x	
Probable cyanate lyase	x		
Malate dehydrogenase	x		
Ribonuclease T2 precursor			x

a) Proteins with known function identified from a single peptide are included.

One interesting finding was that ten proteins were secreted only in the degradation of rutin, with the most obvious of these being quercetin 2,3-dioxygenase, not expected to be secreted in glucose or potato dextrose. Another protein secreted only in rutin was metalloproteinase 23 kDa precursor found in spot number 13. It should be noted that

the identification of this protein did not fulfill our DTASelect filtering criteria, because the peptides identified did not have at least one tryptic peptide terminus. However, because there were ten distinct high scoring peptides that matched the protein, we chose to include it in our final data set. We believe that the nontryptic ends to the identified metalloproteinase peptides may be a result of autodigestion or other degradation during the trypsin in-gel digestion procedure.

Another surprising finding was the protein chitinase, found in spot 51. Chitin is the characteristic component of fungal cell walls and all chitinous fungi produce chitinolytic enzymes at all stages of their growth and development to contribute to spore germination and side-branch formation. *N*-acetylglucosaminidases, as well as exo- β -glucanase have also been considered “chitinases” [23]. Therefore, it is interesting that both of these enzymes were secreted by *A. flavus* in all three media, but chitinase was only found in rutin-containing media.

Five proteins identified are proteins that do not follow the classical secretory pathway, as shown in the supplementary table: ribonuclease T2 precursor, uniquely secreted in potato dextrose media (nine peptides identified from an *A. oryzae* protein sequence); ATP synthase beta subunit, uniquely secreted in rutin media (four peptides identified from a *Scoiopus hallii* protein sequence); phosphoglyceromutase, secreted in rutin and glucose media (two peptides identified from an *A. oryzae* protein sequence); a similar protein to a probable cyanate lyase, uniquely secreted in rutin media (two peptides identified from an *A. nidulans* unannotated protein sequence and identified by BLAST homology searching); and malate dehydrogenase, also uniquely secreted in rutin media (two peptides identified from an *E. coli* protein sequence). These proteins do not contain an N-terminal signal peptide, which is typically believed to be required for targeting a protein to the extracellular space [24]. According to several recent studies, there are some proteins that can be secreted without an N-terminal signal peptide, *via* a nonconventional/nonclassical secretory pathway [25]. The online sequence-based program, SecretomeP, allows for prediction of secretory proteins targeted to the nonclassical secretory pathway (<http://www.cbs.dtu.dk/services/SecretomeP>). Using this method, ATP synthase beta subunit, probable cyanate lyase, ribonuclease T2 precursor, and malate dehydrogenase obtained scores (0.557, 0.587, 0.851, and 0.585, respectively) indicating that these proteins may be secreted *via* the nonclassical secretory pathway. These predictions are an extrapolation as they are based on protein sequences from related organisms, not complete *A. flavus* sequences. The results, however, indicate that there is a possibility that these proteins could be secreted into the extracellular space of *A. flavus*, and our data strongly suggest that this may indeed be the case. Phosphoglyceromutase obtained a SecretomeP score of 0.455, indicating that the probability of it being secreted is lower than for the other proteins. There were 14 proteins identified as being either unknown or hypothetical (9 proteins were identified from more than one peptide hit).

Of these, eight were identified from the rutin medium, five from the glucose medium, and one from the potato dextrose medium. Interestingly, none of the secreted hypothetical proteins we identified were found in more than one growth media. This brings the total number of unique proteins to 2 in the potato dextrose medium, 10 in the glucose medium, and 18 in the rutin media (Fig. 4).

3.4 Assignment of protein identification through the use of different protein databases

In this proteomic study, we have presented data identifying and classifying a large number of secreted proteins from *A. flavus*. However, there is little genomic sequence data currently available for *A. flavus*. Out of over 1700 *Aspergillus* sequence entries found in the Swiss-Prot and NCBI databases, about 100 entries are protein sequences from *A. flavus*, of which only seven are of known secreted proteins.

Few EST sequences for *A. flavus* are available; however, they are not suitable for assigning functional information to expressed proteins as they lack such annotation. For this reason, we employed the strategy of using SEQUEST to search a combined database incorporating all known fungal proteins from Swiss-Prot and NCBI nonredundant protein databases, combined with complete annotated genome databases from *S. cerevisiae* and *E. coli*, and partially annotated genome databases from *A. fumigatus*, *A. nidulans*, and *C. immitis*. A combined summary of nanoLC-MS/MS results from all experiments performed in this study is shown in Table 2. Protein identities were assigned based on exact peptide matching. In cases where the peptide matched was present in a known protein from another species, the peptide was considered to belong to the *A. flavus* version of that protein. In cases where the peptide matched to raw *A. flavus* genome sequence entries, the sequence of the expressed ORF was matched at the protein level to known proteins using BLAST homology searching [22]. It should be emphasized that while we are very confident of our peptide identification data, we are not identifying the complete protein sequence from *A. flavus* because the genome sequence is incomplete. Especially in cases where only one or two peptides are identified, it is always possible that the *A. flavus* protein may actually contain these two peptides in one domain, but may also contain other domains that confer a different function. We have also included protein identifications based on single peptide matches in our final data set (not shown in the supplementary table), even though these are lacking the confirmation provided by identification of additional peptides. The additional data provided by this approach, even if considered as preliminary in nature, reveals several interesting biological insights.

For example, although quercetin 2,3-dioxygenase from *A. flavus* has been purified and characterized, its amino acid sequence is still unknown. In this study, the protein bands and spots corresponding to quercetin 2,3-dioxygenase matched several peptides from an entry in the *A. nidulans*

Table 2. Combined summary of nanoLC-MS/MS results

	Rutin	Glucose	Potato dextrose	Total
Total number of samples analyzed	45	30	20	95
Total number of samples identified	37	27	19	83
Total number of unique proteins ^{a)}	39	30	19	88
Total number of non-redundant proteins ^{b)}	18	10	2	30

a) Includes overlap between growth conditions

b) Total unique proteins identified (no overlap)

database. Based on homology BLAST searching, this protein was identified as being 56% identical to quercetin 2,3-dioxygenase from *Aspergillus japonicus*. Although the sequence of the *A. japonicus* quercetinase was present in the composite database used with SEQUEST to search for identifications, no gel bands/ spots matched any of these peptides. This confirms our suggestion [14] that there are a number of significant differences at the amino acid sequence level between the quercetinases of *A. japonicus* and *A. flavus*, even though both sequences are clearly recognizable as closely related enzymes. The sequence from *A. nidulans* was also similar to the protein YxaG, a quercetin 2,3-dioxygenase from *Bacillus subtilis*, but with a much lower degree of homology.

Almost all of the proteins in this study were identified from peptides present in other *Aspergillus* species or other fungal organisms, such as *Penicillium citrinum*, *Candida molischiana*, *S. hallii*, and *Pseudomonas syringae*, as well as from *Emericella nidulans*, the teleomorph of *A. nidulans*, and also from *E. coli*. Many of these protein identifications were reinforced by identification of peptides from unannotated genomic sequences of *A. fumigatus* and *A. nidulans*, which were shown to have similar functions by BLAST homology searching. None of the proteins identified were found by matching to peptide sequences of either *S. cerevisiae* or *C. immitis*.

It should be noted that although unique proteins were found repeatedly throughout the gel bands and spots, these bands or spots did not always correspond to isoforms of the proteins. Protein isoforms can occur due to different glycan composition such as different lengths of outer mannan chains [26] or variable numbers of sialic acid [27] as well as other PTMs such as sialylation, methylation, phosphorylation, sulfation, palmitoylation, and farnesylation. Additionally, the presence of dimeric and monomeric forms of a protein can be found on a gel, leading to multiple spots for a single protein [28]. In this study, the repetitive identification of proteins is not always due to isoforms, but rather, it is most likely due to breakdown products of the intact proteins,

probably catalyzed by secreted proteases also present in the samples. Addition of a fungal protease inhibitor cocktail immediately after filtration of the mycelia from the cultures after the third day of growth inhibited the visualization of many proteins as seen on an SDS-PAGE gel (data not shown). There are two cases, however, where protein isoforms do appear to be present, as there are groups of protein spots of very similar mass that are well resolved in the IEF dimension. In the case of the 2-DE gel of proteins secreted in glucose, spots 70–72 correspond to isoforms of the enzyme alpha-amylase, and in the 2-DE gel of proteins secreted in rutin, spots 38 and 39 correspond to alanyl dipeptidyl peptidase isoforms. Also, some proteins, such as quercetin 2,3-dioxygenase, are known to exist as dimers, providing a reason for finding multiple spots of this protein.

4 Concluding remarks

The aim of this study was to learn more about the dispensable metabolic pathway expressed when *A. flavus* is grown on rutin as the only source of carbon. Using a combination of SDS-PAGE and 2-DE and nanoLC-MS/MS, a large number of proteins secreted when the fungus is grown in rutin, glucose, and potato dextrose were identified. Eighteen proteins secreted were unique to the rutin metabolic pathway (including hypothetical/unknown proteins). We have further demonstrated that using nanoLC-MS/MS is an efficient mean of high-throughput protein identification in samples where little or no genomic sequence data are available, and thus the majority of identifications are based on protein sequences from databases containing homologous protein sequences. Finally, the availability of complete genome sequences from other *Aspergilli* has been an important factor in this study. As more genome sequences become available, identification and functional classification of proteins from filamentous fungi will become easier to perform, and this approach will become an important research tool in functional fungal biology.

This work was supported in part by a grant from the National Science Foundation (MCB-0317126 to W.A.F.) and the MGE@MSA headquartered at Arizona State University, an Alliance for Graduate Education and the Professoriate, funded by the NSF. M.L.M. was supported in part by a fellowship through the Research Training Group in Optical Biomolecular Devices provided under a grant from the NSF (DBI-9602258–003). The *A. fumigatus* Sequencing Consortium was funded by the National Institute of Allergy and Infectious Disease, The Wellcome Trust, and Fondo de Investigaciones Sanitarias. We gratefully acknowledge Jeffrey Thresher for assistance with IEF. P.A.H. would like to thank Peter Lehmann for his continued support.

5 References

- [1] Peberdy, J. F., *Trends Biotechnol.* 1994, 12, 50–57.
- [2] Wallis, G. L. F., Swift, R. J., Hemming, F. W., Trinci, A. P. J., Peberdy, J. F., *Biochim. Biophys. Acta* 1999, 1472, 576–586.
- [3] Hemming, F. W., *Biochemical Soc. Trans.* 1995, 23, 180–185.
- [4] Keller, N. P., Hohn, T. M., *Fungal Genet. Biol.* 1997, 21, 17–29.
- [5] Hund, H.-K., Breuer, J., Lingens, F., Hohn, T. M., Hüttermann, J. et al., *Eur. J. Biochem.* 1999, 263, 871–878.
- [6] Krishnamurthy, H. G., Simpson, F. J., *J. Biol. Chem.* 1970, 245, 1467–1471.
- [7] Fusetti, F., Schröter, K. H., Steiner, R. A., van Noort, P. I., Pijning, T. et al., *Structure* 2002, 10, 259–268.
- [8] Child, J. J., Oka, T., Simpson, F. J., Krishnamurthy, H. G., *Can. J. Microbiol.* 1971, 17, 1455–1463.
- [9] Chu, P.-W., Yap, M.-N., Wu, C.-Y., Huang, C.-M., Pan, F.-M. et al., *Electrophoresis* 2000, 21, 1740–1745.
- [10] Haynes, P. A., Fripp, N., Aebersold, R., *Electrophoresis* 1998, 19, 939–945.
- [11] Küster, B., Mortensen, P., Andersen, J. S., Mann, M., *Proteomics* 2001, 1, 641–650.
- [12] Andon, N. L., Hollingworth, S., Koller, A., Greenland, A. J., Yates, J. R. I. et al., *Proteomics* 2002, 2, 1156–1168.
- [13] Westlake, D. W. S., Talbot, G., Blakley, E. R., Simpson, F. J., *Can. J. Microbiol.* 1959, 5, 621–629.
- [14] Medina, M. L., Kiernan, U. A., Francisco, W. A., *Fungal Genet. Biol.* 2004, 41, 327–335.
- [15] Grinyer, J., McKay, M., Herbert, B., Nevalainen, H., *Curr. Genet.* 2004, 45, 170–175.
- [16] Bruneau, J.-M., Magnin, T., Tagat, E., Legrand, R., Bernard, M. et al., *Electrophoresis* 2001, 22, 2812–2823.
- [17] Chen, Z.-Y., Brown, R. L., Damann, K. E., Cleveland, T. E., *Phytopathology* 2002, 92, 1084–1094.
- [18] Gharahdaghi, F., Weinberg, C. R., Meagher, D. A., Imai, B. S., Mische, S. M., *Electrophoresis* 1999, 20, 601–605.
- [19] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, 68, 850–858.
- [20] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, 67, 1426–1436.
- [21] Tabb, D. L., McDonald, W. H., Yates, J. R. I., *J. Proteome Res.* 2002, 1, 21–26.
- [22] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., *J. Mol. Biol.* 1990, 215, 403–410.
- [23] Gooday, G. W., Zhu, W. Y., O'Donnell, R. W., *FEMS Microbiol. Lett.* 1992, 100, 387–392.
- [24] Bendtsen, J. D., Jensen, L. J., Blom, N., von Heijne, G., Brunak, S., *PEDS Advance Access* 2004, 17, 349–356.
- [25] Rubartelli, A., Sitia, R., *Secretion of Mammalian Proteins that Lack a Signal Sequence*, RG Landes, Austin, TX 1997, pp. 87–114.
- [26] Packer, N. H., Keatinge, L., in: Pennington, S. R., Dunn, M. J. (Eds.), *Proteomics from Protein Sequence to Function*, BIOS Scientific Publishers, Oxford 2001, pp. 257–275.
- [27] Löster, K., Kannicht, C., in: Kannicht, C. (Ed.), *Posttranslational Modifications of Proteins*, Humana Press, Totowa, NJ 2002, pp. 301–316.
- [28] Porubleva, L., Velden, K. V., Kothari, S., Oliver, D. J., Chitnis, P. R., *Electrophoresis* 2001, 22, 1724–1738.