Creation of a genome-scale metabolic model for the fungal pathogen *Cryptococcus neoformans*

Samantha Gonyea
BBSI 2006-2007
Final Report

Abstract

In recent years, *Cryptococcus neoformans* has become a more pervasive fungal threat to those individuals who are immunocompromised. *C. neoformans* affects the central nervous system and causes cryptococcosis which is very similar symptomatically to meningitis. The pathogenicity of *C. neoformans* is mostly due to a complex polysaccharide capsule that evades the body's immunological response. The approach being taken in this research is to systematically study the metabolism of *C. neoformans* with particular interest in metabolic pathways associated with polysaccharide synthesis reactions. Specifically, a computational genome-based metabolic model has been constructed that will allow researchers to comprehensively study *C. neoformans* metabolism and to perform *in silico* gene deletion simulations. These effects can be tested experimentally to validate the model's accuracy. If the model is incorrect in its predictions, this may lead to further clues about unusual metabolic pathways of *C. neoformans*. We are interested specifically in the capsular synthesis and assembly pathways. The way in which the capsule is synthesized and attached to the cell is important to fungal biology and in the development of drugs to target this fungal pathogen. The creation of an *in silico* metabolic model, in addition to other capsule-based research, may eventually aid in a greater understanding of how *C. neoformans* works and a way in which to cure those individuals infected with cryptococcosis.

Introduction

As technology in biological research continues to advance, large amounts of new data have become available in such fields as genomics and proteomics. In order to comprehend this flood of information, the data must be interconnected so the organism can be understood as a whole, instead of many different parts (Edwards *et al*., 2002). To accomplish this seemingly difficult task, an *in silico* model of metabolism is beneficial. This computer simulation is based on physicochemical laws and principles to recreate the complex interconnected pathways of a cell's metabolism (Kauffman *et al*., 2003).

The process of metabolic reconstruction is known as flux-balance analysis (FBA) and is thoroughly described by Kauffman *et al*. (2003). The model requires all of the metabolic reactions and metabolites that the organism uses for survival and growth. This includes all enzymatic reactions, as well as the transport mechanisms performed by the cell. At this point, the mass balance for the metabolites is calculated based on the flux going in and out of each reaction. Biochemical reactions are represented as a stoichiometric matrix which can be analyzed using linear algebra and optimization to identify the fluxes in the organism's steady-

state. Using these fluxes and the constraints placed on the organism (flux limitations, energy balance, and mass balance) a solution space can be calculated. This solution space defines the boundaries of an organism's biological properties (Kauffman *et al.,* 2003). *In silico* metabolic models have already been constructed and used for organisms such as *Escherichia coli* (Reed and Palsson, 2003) and *Saccharomyces cerevisiae* (Duarte *et al*., 2004). The genome-scale *E. coli* model (iJE660) was used to predict growth behavior at the end of the adaptive evolution (Edwards *et al.* 2001, Ibarra *et al*. 2002, Fong *et al.* 2003, Fong & Palsson 2004). The same *E. coli* model correctly predicted two double gene deletions and one quadruple gene deletion strain as designs for increased output of lactic acid as a by-product of cellular growth (Fong *et al.* 2005). The *Saccharomyces cerevisiae* metabolic model demonstrates eukaryotic compartmentalization to include extracellular space, cytosol, mitochondria, peroxisome, nucleus, Golgi apparatus, endoplasmic reticulum, and vacuoles (Duarte *et al*. 2004).
The growth patterns of *S. cerevisiae* predicted by its metabolic model were in agreement with experimental results 87.8% of the time (Famili *et al*. 2003).

The focus of this work is to create a model for the fungal pathogen *Cryptococcus neoformans,* using *S. cerevisiae* as starting model, with sequences from the *C. neoformans* database and information from the literature to further create this new model.

Cryptococcus neoformans is a yeast-like fungus that can cause potentially fatal, if not treated, cryptococcosis in immunocompromised individuals, especially those with AIDS (Bose *et al*., 2003). This yeast has several important virulence factors that aid in its pathogenicity such as the ability to grow at 37ºC, the ability to produce melanin on specific substrates, and a complex polysaccharide capsule (Casadevall *et al.,* 2000). With the proposed new *in silico* metabolic model, further clues about the virulence of *C. neoformans* may be discovered that may expand our information about fungal biology and the pathogenesis of this organism.

Materials and Methods

Initial Analysis

Cryptococcus neoformans has a partially annotated genome at The Institute for Genomic Research database (TIGR) website. The TIGR database showed the TC (tentative consensus) annotation numbers in relation to the metabolic pathways of *C. neoformans*. These relationships were organized further into an Excel worksheet by enzyme classification (EC) number, subsystems, and the enzymes related to the pathway. Therefore, the known genes of *C. neoformans* could be directly related to the enzyme and pathway they affect. Once the already known *C. neoformans* data was organized, these pathways were compared to a pre-existing metabolic model of the non-pathogenic yeast *Saccharomyces cerevisiae.* Since *S. cerevisiae* already had a working model, known as iND750, it made a logical starting point rather than creating a brand new model. The metabolic pathways were compared by using the computerized version of the metabolic pathways in *S. cerevisiae* and highlighting the corresponding known pathways in *C. neoformans*.

Sequence Homology Gap Analysis

After analyzing the results from the TIGR annotations, there were many noticeable gaps in the metabolic pathways. Gap analysis was performed to try to complete the metabolism of *C. neoformans*. All of the genes and gene products that were not located in *C. neoformans* with the TIGR database but were shown to be found in *S. cerevisiae* based on its existing model, had to be located in *C. neoformans*. Starting with the *S. cerevisiae* genome database (http://db.yeastgenome.org/cgi-bin/seqTools,) the gene product names from the existing iND750 *S. cerevisiae* model that were not located with TIGR were entered into the database. Under the "DNA of Region" section, the FASTA format of DNA sequence was selected. A sequence in FASTA format consists of one line starting with a ">" sign, followed by a sequence identification code and one or more lines containing the sequence itself (http://www.cmbi.kun.nl/bioinf/tools/crab_fasta.html). The FASTA DNA sequence codes for a specific gene product in *S. cerevisiae.* This DNA sequence was copied and pasted into the TIGR BLAST (Basic Local Alignment Search Tool) search database (http://tigrblast.tigr.org/er-blast/index.cgi?project=cna1) that has the entire *C. neoformans* genomic sequence. A tblastx search was performed on the *S. cerevisiae* DNA sequence and the entire *C. neoformans* genome was searched. Tblastx converts a nucleotide query sequence into protein sequences in all 6 reading frames and then compares this to an NCBI nucleotide database which has been translated on all six reading frames (http://www.incogen.com/public_documents/vibe/details/NcbiTblastx.html). All results found in *C. neoformans* with a smallest sum probability number of 1×10^{-3} or less were accepted as reactions in the *C. neoformans* metabolism. An example of the resulting Excel datafile is shown in Figure 1.

ABBREVIATION	REACTION NAME	REACTION	E.C. #	SUBSYSTEM	ORF	PROTEIN	TC of *C.neoformans* or BLAST search results	Best Smallest Sum Prob. (taken from BLAST)
TKT2	transketolase	[c] : e4p + xu5p-D <==> f6p + g3p	EC-2.2.1.1	Pentose Phosphate Cycle	YPR074C or YBR117C	(Tkl1) or (Tkl2)	TC8518, TC9045	N/A
TKT1	transketolase	[c] : r5p + xu5p-D <==> g3p + s7p	EC-2.2.1.1	Pentose Phosphate Cycle	YBR117C or YPR074C	(Tkl2) or (Tkl1)	TC8518, TC9045	N/A
TALA	transaldolase	[c] : g3p + s7p <==> e4p + f6p	EC-2.2.1.2	Pentose Phosphate Cycle	YLR354C or YGR043C	(Tal1) or (Tal2)	TC8110, TC8112, TC8116	N/A
RPI	ribose-5-phosphate isomerase	[c] : r5p <==> ru5p-D	EC-5.3.1.6	Pentose Phosphate Cycle	YOR095C	Rki1	Not found in Crypto.	N/A
RPE	ribulose 5-phosphate 3-epimerase	[c] : ru5p-D <==> xu5p-D	EC-5.1.3.1	Pentose Phosphate Cycle	YJL121C	Rpe1	Found	1.60E-52
RBK	ribokinase	[c] : atp + rib-D --> adp + h + r5p	EC-2.7.1.15	Pentose Phosphate Cycle	YCR036W	Rbk1	Found	8.40E-07

Figure 1: An example of the iND750 excel worksheet using the pentose phosphate cycle metabolic pathway. Data includes abbreviations of the reaction, the reaction name, the actual

reaction, enzyme classification (EC) number, subsystem, open reading frame, protein produced, *C. neoformans* TIGR annotation numbers, and smallest sum probability numbers for those sequences searched in BLAST.

<u>Experimental Gap Analysis</u>

Chemical/Media Solution Preparation

After sequence homology gap analysis, there were still *C. neoformans* pathways that did not completely match up with the *S. cerevisiae* model. Therefore, the extracellular transporters of these pathways were to be tested using the substrates/chemicals the transporters were responsible for. The following formula, $C_1 = C_2$, was used in order to make 5g/L chemical solutions in 4mL of the YNB media without amino acids or a carbon source with C_1 representing the number of grams per 4mL of media and C_2 representing the desired concentration for the chemical solutions.

This calculation resulted in 0.02g of chemical in 4mL of the media. Five chemical solutions (melibose, sucrose, inosine, thymidine, and dextrose) were prepared at this particular concentration. The original concentrations of the chemical solutions were accounted for by multiplying 0.02g by the concentration of the particular solution. Based on these calculations, 0.1mL of the melibose, sucrose, and dextrose solutions were added to 3.9mL of the YNB media, 0.5mL of the inosine solution was added to 3.5mL of YNB, and 1mL of the thymidine solution was added to 3mL of YNB media.

Cryptococcus neoformans Preparation

Two days prior to the microplate experiment, a liquid culture of the *C. neoformans* ags1 mutant was grown in liquid culture of YPD media. Two microcentrifuge tubes were filled with the culture and centrifuged on a short cycle in order to obtain a pellet of cells. The cells were then resuspended in YNB media without amino acids and no carbon source. A 1:10 dilution of the cells were prepared to be read at a 600nm wavelength in the spectrophotometer. The tube with the highest optical density reading was used for the experiment. To obtain a target starting density of 0.05 in 200µL of the media/chemical solutions, the following calculation was used with a correction made for the 1:10 dilution of the cells:

(200µL)(0.05) = (Unknown volume of *C. neoformans* cells)(O.D. of *C. neoformans* culture)(10)

96 Well Microplate Preparation

The 36 external wells of the plate were first filled with 200µL of distilled water and 50µl of mineral oil to prevent evaporation of the water. The rest of the wells were filled with 200µL of a specific chemical/YNB media solution (listed below) and the predetermined volume of the *C. neoformans* cells.

B2-G2 and B3-G3: 200µL dextrose/YNB media
B4-G4 and B5-G5: 200µL inosine/YNB media
B6-G6 and B7-G7: 200µL melibose/YNB media
B8-G8 and B9-G9: 200µL thymidine/YNB media
B10-G10 and B11-G11: 200µL sucrose/YNB media

The 96 well microplate was loaded into a VERSAmax tunable microplate reader at 30°C for 90 hours. The sample's optical density was read every 15 minutes at a wavelength of 600nm and was mixed for 875 seconds between each read.

Results

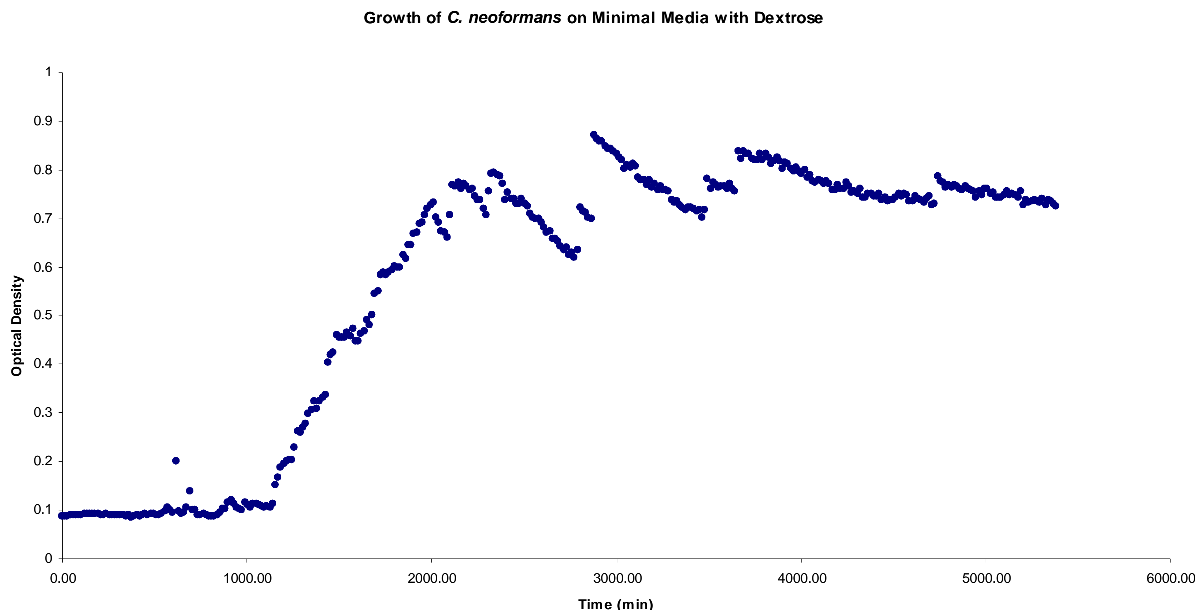

Figure 2: The growth curve of *C. neoformans* on minimal media without amino acids, with the addition of a dextrose substrate over a 90 hour time period.

The experimental gap analysis methods resulted in growth curve graphs for the five chemicals tested. The data for inosine, melibose, sucrose, and thymidine is not shown. Figure 2 shows the growth curve of *C. neoformans* on the YNB minimal media without amino acids with a dextrose supplement. The curve shows that *C. neoformans* was able to grow over time with only dextrose as a carbon source.

Table 1: The number of metabolic pathways found for each subsystem of *C. neoformans* metabolism organized by the way in which the pathways were found (initial analysis, sequence homology gap analysis, or experimental data gap analysis).

Subsystem	Initial Analysis	Sequence Homologies	Experimental Data	Total per Subsystem
Alanine and Aspartate Metabolism	3	6	5	**14**
Alternate Carbon Metabolism	8	9	-	**17**
Aminosugars Metabolism	1	-	-	**1**
Anaplerotic Reactions	4	2	-	**6**
Arabinose Metabolism	-	3	-	**3**
Arginine and Proline Metabolism	11	13	11	**35**
Asparagine Metabolism	2	1	1	**4**
Citric Acid Cycle	8	4	-	**12**

Cysteine Metabolism	3	1	2	**6**
Folate Metabolism	-	5	-	**5**
Fructose and Mannose Metabolism	2	3	-	**5**
Galactose Metabolism	2	5	-	**7**
Glutamate Metabolism	3	6	1	**10**
Glutamine Metabolism	2	3	2	**7**
Glycine and Serine Metabolism	7	9	4	**20**
Glycolysis/Gluconeogenesis	14	4	-	**18**
Glycoprotein Metabolism	3	-	-	**3**
Histidine Metabolism	2	13	1	**16**
Methane Metabolism	1	-	-	**1**
Methionine Metabolism	5	6	5	**16**
NAD Biosynthesis	2	7	-	**9**
Nitrogen Metabolism	-	3	-	**3**
Nucleotide Salvage Pathways	17	13	-	**30**
Other Amino Acid Metabolism	2	9	1	**12**
Oxidative Phosphorylation	13	2	-	**15**
Pantothenate and CoA	1	3	-	**4**
Pentose Phosphate Cycle	8	3	-	**11**
Phospholipid Biosynthesis	2	20	-	**22**
Porphyrin and Chlorophyll	4	1	-	**5**
Purine and Pyrimidine Biosynthesis	10	27	-	**37**
Pyruvate Metabolism	8	-	-	**8**
Quinone Biosynthesis	-	4	-	**4**
Riboflavin Metabolism	-	3	-	**3**
Starch and Sucrose Metabolism	-	1	-	**1**
Thiamine Metabolism	-	6	-	**6**
Threonine and Lysine Metabolism	3	10	5	**18**
Extracellular Transport	-	47	1	**48**
Mitochondrial Transport	-	14	29	**43**
Nuclear Transport	-	-	6	**6**
Peroxisomal Transport	-	-	4	**4**
Tyrosine, Tryptophan, and Phenylalanine	10	26	13	**49**
Valine, Leucine, and Isoleucine	7	15	2	**24**
Xylose Metabolism	-	2	-	**2**
TOTAL	**168**	**309**	**93**	**570**

In order to keep track of the way in which the pathways were found, table 1 organizes the number of pathways by the subsystem of metabolism they are involved in and if they were found by initial analysis, sequence homology gap analysis, or by experimental gap analysis. There were 168 pathways found by initial analysis, 309 pathways by sequence homology gap analysis, and 93 pathways by experimental gap analysis for a total of 570 known *C. neoformans* reactions.

Discussion

The initial analysis of *Cryptococcus neoformans* based on its incomplete annotated genome was still able to confirm 168 metabolic pathways in various subsystems for the model. Sequence homology gap analysis proved to be the most useful with identifying 309 pathways, also in a large range of subsystems. The transporter experiment results for dextrose were the most promising of the five chemicals tested. The growth of *C. neoformans* on minimal media without amino acids with only dextrose as the carbon source demonstrated *de novo* ability to synthesize all essential amino acids. Since dextrose is a precursor to all these amino acids and *C. neoformans* must be able to produce them to survive, all the previously un-annotated amino acid pathways and the transporters associated, totaling 93 pathways, were updated. Based on these three methods, 570 total metabolic pathways were confirmed for *Cryptococcus neoformans* which is an excellent starting point for the *in silico* model.

With these 570 pathways confirmed, they can now be added into a metabolic model program in order to run gene deletion simulations. MetModel, developed by Wil Burns of Virginia Commonwealth University, will be the software used to create the model. All reactions will be added to MetModel to determine the completeness of the model. If there are still pathways that are required to allow a functional *C. neoformans* metabolic model, they will be added based on necessity.

Once the metabolic model for *Cryptococcus neoformans* is completed and a particular phenotype of interest is determined, gene deletion simulations will be run. These simulations will give rise to possible strains that may lead to possible gene targets of interest. These specific strains will be then be created experimentally. Fusion PCR will be used to generate DNA segments that will introduce a selectable marker between flanking regions of gene interest (Davidson, *et al.,* 2002). The resulting DNA will then be introduced into *C. neoformans* by biolistic gene gun method to disrupt the gene of interest by homologous recombination and genomic integration (Toffaletti *et al.,* 1993). Resulting strain genotypes will be verified by PCR and DNA blotting methods and the resulting strain phenotypes can then be analyzed as has been done previously in *C. neoformans* (Reese, *et al*., 2007). These methods should allow for the testing and experimental validation of the metabolic model's predictions.

References

Bose, I., A.J. Reese, J.J. Ory, G. Janbon, and T.L. Doering. (2003) A yeast under cover: the capsule of *Cryptococcus neoformans*. *Eukaryotic Cell* **2**: 655-663.

Casadevall, A., A.L. Rosas, and J.D. Nosanchuk. (2000) Melanin and virulence in *Cryptococcus neoformans*. *Current Opinion in Microbiology* **3**: 354-358.

Davidson, R.C., J.R. Blankenship, P.R. Kraus, M. de Jesus Berrios, C.M. Hull, C. D'Souza, P. Wang, and J. Heitman. (2002) A PCR-based strategy to generate integrative targeting alleles with large regions of homology. *Microbiology* **148**: 2607-2615.

Duarte, N.C., M.J. Herrgard, and B.O. Palsson. (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research* **14**: 1298-1309.

Edwards, J.S. and B.O. Palsson. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences* **97**: 5528-5533.

Edwards, J.S., R.U. Ibarra, and B.O. Palsson. (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology* **19**: 125-130.

Edwards, J. S., M. Covert, and B. Palsson. (2002) Metabolic modeling of microbes: the flux-balance approach. *Environmental Microbiology* **4**(3): 133-140.

Famili, I., J. Forster, J. Nielsen, and B.O. Palsson. (2003) *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences* **100**: 13134-13139.

Fong, S.S. and B.O. Palsson. (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Genetics* **36**: 1056-1058.

Fong, S.S., A.P. Burgard, C.D., Herring, F.M. Knight, F.R. Blattner, C.D. Maranas, and B.O. Palsson. (2005) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnology and Bioengineering* **91**: 643-648.

Fong, S.S., J.Y. Marciniak, and B.O. Palsson. (2003) Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale *in silico* metabolic model. *Journal of Bacteriology* **185**: 6400-6408.

Ibarra, R.U., J.S. Edwards, and B.O. Palsson. (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**: 186-189.

Kauffman, K.J., P. Prakash, and J.S. Edwards. (2003) Advances in flux balance analysis. *Current Opinion in Biotechnology* **14**: 491-496.

Reed, J.L. and B.O. Palsson.2003. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *Journal of Bacteriology* **185**: 2692-2699.

Reese, A.J., A. Yoneda, J.A. Berger, A. Beauvais, H. Liu, C.L. Griffith, I. Bose, M-J. Kim, C. Skau, S. Yang, J.A. Sefko, M. Osumi, J-P. Latge, E. Mylonakis, and T.L. Doering. (2007) Loss of cell wall alpha(1-3) glucan affects *Cryptococcus neoformans* from ultrastructure to virulence. *Molecular Microbiology* **65**(5): 1385-1398.

Toffaletti, D.L., T.H. Rude, S.A. Johnston, D.T. Durack, and J.R. Perfect. (1993) Gene transfer in *Cryptococcus neoformans* by use of biolistic delivery of DNA. *Journal of Bacteriology* **175**:1405-1411.