CHAPTER 3
Materials and Methods

3.1 *Candida tropicalis* culture

*C. tropicalis* (MTCC 184) culture was procured from the Microbial Type Culture Collection and Gene Bank, IMTECH, Chandigarh, India. The cells were cultured in Yeast Nitrogen Base (YNB) media (HiMedia Laboratories, Mumbai, India). This media was specifically selected over others such as YPD (Yeast extract, Peptone, Dextrose) and SDB (Sabouraud’s Dextrose Broth) as the goal of the study included experiments on conditioned media. Presence of protein in the media is a hindrance in proteomic experiments. Further, presence of peptides from yeast leads to ambiguity in database search based protein identifications. Table 3.1 gives the detailed composition of YNB media.

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>Concentration (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ammonium sulphate</td>
<td>5000</td>
</tr>
<tr>
<td>Boric acid</td>
<td>0.5</td>
</tr>
<tr>
<td>Calcium chloride</td>
<td>100</td>
</tr>
<tr>
<td>Calcium pantothenate</td>
<td>0.4</td>
</tr>
<tr>
<td>Copper sulphate</td>
<td>0.04</td>
</tr>
<tr>
<td>D-Biotin</td>
<td>0.002</td>
</tr>
<tr>
<td>Ferric chloride</td>
<td>0.2</td>
</tr>
<tr>
<td>Folic acid</td>
<td>0.002</td>
</tr>
<tr>
<td>Inositol</td>
<td>2</td>
</tr>
<tr>
<td>Magnesium sulphate</td>
<td>500</td>
</tr>
<tr>
<td>Manganese sulphate</td>
<td>0.4</td>
</tr>
<tr>
<td>Niacin</td>
<td>0.4</td>
</tr>
<tr>
<td>PABA</td>
<td>0.2</td>
</tr>
<tr>
<td>Potassium iodide</td>
<td>0.1</td>
</tr>
<tr>
<td>Potassium phosphate, monobasic</td>
<td>1000</td>
</tr>
<tr>
<td>Pyridoxin, HCl</td>
<td>0.4</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>0.2</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>100</td>
</tr>
<tr>
<td>Sodium molybdate</td>
<td>0.2</td>
</tr>
<tr>
<td>Thiamine HCl</td>
<td>0.4</td>
</tr>
<tr>
<td>Zinc sulphate</td>
<td>0.4</td>
</tr>
</tbody>
</table>

**Table 3.1:** Composition of Yeast Nitrogen Base media
### 3.2 Preparation of culture filtrate

Cultures were maintained at 37°C with shaking (300 RPM) for around 8-10 hours, until OD$_{600}$ of the culture reached ~1. Approximately 6 billion cells were pelleted down by centrifuging for 15 minutes at 1500 g at 4°C. The pellet was stored at −80°C until further use. The supernatant fraction was processed further to study the secreted fraction of proteome, as follows: The supernatant was filtered through a 0.22 µm membrane filter (Millipore Corporation, Billerica, MA) and then concentrated using a 3 kDa molecular cut-off filter. (Millipore Corporation). Protein estimation was carried out using bicinchoninic acid (BCA) assay (Thermo Fisher Scientific, Rockford, IL). The assay involved construction of a standard curve with known concentrations of Bovine Serum Albumin (BSA) and thus estimating protein concentration in the sample.

### 3.3 Protein extraction from cell lysate

Cell pellets were suspended in cracking buffer (5% SDS, 8M urea lysis buffer, 0.1mM EDTA, 40mM Tris-HCl; pH 6.8) [59]. Cell lysis was carried out using a cell disruptor (Disruptor Genei SI-D267, Scientific Industries Inc. NY). For this, cells suspended in cracking buffer were subjected to disruption with glass beads for 60 min. Samples were then centrifuged at 20,000 g for 15 min and the clear supernatant was collected. Protein estimation was carried out using BCA assay.

### 3.4 In-gel digestion

For both total proteome and secreted fraction, 300 µg of protein was resolved on a 10% 15x15 cm SDS-PAGE and stained with colloidal Coomassie blue. The gel was fixed using 40% methanol and 15% acetic acid in water for 10 minutes and stained using colloidal Coomassie blue stain. Excess stain was removed with the help of destainer (40% methanol in water). The protein lane was cut into 23 and 8 bands respectively for cell pellet and conditioned media as marked in Figure 3.1. The lane of conditioned media was divided into only 8 bands as the proteome was not complex. Bands were not uniformly cut but, each visibly discernible band was cut into a separate fraction. In-gel tryptic digestion and peptide extraction was carried out as described below –
i. The destained gel was rinsed with water and placed onto a clean glass plate. The lane of the protein sample of interest was separated out by cutting along the border. A clean scalpel blade was used to cut the visibly discernible bands.

ii. Bands were excised into 1mm X 1mm pieces and transferred into micro-centrifuge tubes containing destaining solution (40mM ammonium bicarbonate – pH 8, 40% acetonitrile). Tubes were placed on a shaker and gently agitated to aid in destaining. The solution was discarded with a 200 μl pipette tip and the procedure was repeated until the gel pieces were completely destained.

iii. Once the gel pieces were completely destained, tubes were spun and supernatant was discarded. 500 μl of 100% acetonitrile was added and incubated for 10-15 min until gel pieces were completely dehydrated and had become opaque.

**Figure 3.1:** SDS PAGE profile used for in-gel digestion
iv. The tubes were spun and acetonitrile was removed. Any remaining acetonitrile was removed with the help of gel loading tips.

v. Freshly prepared reducing reagent (5mM dithiothreitol in 40mM ammonium bicarbonate) was added to completely cover the pieces. Tubes were incubated at 60°C for 30 min.

vi. Tubes were brought to the room temperature and reducing reagent was removed and freshly prepared alkylation reagent (20mM iodoacetamide (IAA) in 40mM ammonium bicarbonate) was added. Tubes were incubated in dark for 10 min.

vii. Alkylating reagent was removed and gel pieces were dehydrated by adding 100% acetonitrile. Acetonitrile was removed after gel pieces were completely dehydrated and the tubes were kept on ice. Sequencing grade trypsin (Promega) dissolved in chilled 40 mM ammonium bicarbonate buffer was added to the tubes containing gel pieces. Amount of trypsin solution was decided based on number and size of gel pieces in the tubes.

viii. Tubes were left on ice for 45 min until gel pieces sufficiently imbibed the trypsin solution. More trypsin was added if necessary. Once the gel pieces were completely rehydrated, excess trypsin was removed and replaced with 40 mM ammonium bicarbonate to cover gel pieces. The tubes were incubated at 37°C overnight.

ix. Peptide Extraction – Tubes were cooled to room temperature and 100 µl of 5% formic acid was added. Tubes were spun and supernatants were collected in a fresh set of tubes labelled according to fraction number. 100 µl of extraction buffer (5% formic acid; 40% acetonitrile) was added to the gel pieces and the tubes were kept on shaker for 10 min. Supernatants were pooled into the tubes containing supernatant from the earlier step. This step was repeated or twice. Final extraction was carried out by adding 100% ACN. After gel pieces had completely dehydrated, supernatants were pooled into respective tubes and the dehydrated gel pieces were discarded.

x. Peptide extracts were dried down using a vacuum dryer and desalting was carried out.

3.5 In-solution digestion

2 mg of protein lysate from the cell pellet and ~ 900 µg of protein from the conditioned media were separately subjected to buffer exchange with 50 mM triethylammonium bicarbonate (TEAB buffer) using 30,000 Da molecular mass cut-off spin filters (Millipore Corporation, Billerica, MA). This buffer exchange step was necessary to facilitate trypsin digestion as the denaturing agents (SDS and urea) present in cracking buffer hinder digestion. In-solution digestion was carried out as follows –
i. Reducing agent – dithiothreitol solution (50 mM) was added to the protein solution such that final concentration of dithiothreitol was 5 mM. After gentle mixing the tube was kept at 60°C for 60 minutes.

ii. After incubation the tubes were brought to room temperature and alkylating agent – iodoacetamide solution (100 mM) was added to the protein solution so that the final concentration was 20 mM. The tube was kept in dark for 10 minutes.

iii. An aliquot equivalent of 20 μg of protein was taken in a separate tube (pre-digest) to be resolved on SDS-PAGE to check for protein integrity.

iv. L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK) treated trypsin (Worthington Biochemical Corporation, Lakewood, NJ, USA) was added at 1:10 (w/w) final concentration and digested for 16 hours at 37°C.

v. The reaction was quenched by acidification with formic acid (0.5% final concentration). An aliquot equivalent to 20 μg of protein was taken out (post-digest).

vi. The pre and post-digest aliquots were resolved on SDS-PAGE to check for digestion efficiency. Complete digestion was indicated by absence of bands in the post-digest lane but not in the pre-digest lane.

vii. The tryptic peptides were lyophilized and basic pH reverse phase liquid chromatography was carried out to achieve peptide level fractionation.

3.6 bRPLC fractionation

Basic pH reversed-phase liquid chromatography (bRPLC) was carried out at pH 9.5 on a XBridge C18 column (250 x 4.6 mm, 5 μm, 200 Å, Waters Corporation, Milford, MA, USA). Peptides were reconstituted in bRPLC solvent A (10mM Trimethyl ammonium bicarbonate, pH 9.5) and loaded on C18 column using an Agilent 1200 series HPLC system. Peptides were fractionated with the following gradient: 95% solvent A for 5 min, continued by a short (30 seconds) gradient of 5-8% of solvent B (10mM trimethyl ammonium bicarbonate, 90% acetonitrile, pH 9.5), followed by a gradient of 8-40% of B in 35 min and a 40-100% gradient for 1 min. Fractionation was completed by a gradient of 100% B for 4 min and then equilibrated in 95% of solvent A for 5 min. The flow rate was maintained 1.5 mL/min for the entire run. Fractions were collected into a 96-well plate from 4.5-50 min. The bRPLC profiles of cell lysate and conditioned media are shown in Figure 3.2. As shown in Figure 3.3, the fractions were concatenated into 24 in cell lysate and 12 in conditioned media, respectively.
Pooled samples were lyophilized, desalted and stored at – 80°C till they were subjected to LC-MS/MS analysis.

**Figure 3.2**: Basic pH Reverse Phase Liquid Chromatography (bRPLC) profiles

Top panel – cell lysate; Bottom panel – conditioned media

### 3.7 Desalting of peptides

Individual in-gel digested protein bands and bRPLC fractions were desalted using C$_{18}$ stage tips. C$_{18}$ STop-And-Go-Extraction (STAGE) protocol was followed as described by [60]. The C$_{18}$ STAGE tips were made in-house using Empore™ C$_{18}$ disc. The samples were
reconstituted in 30 µl of solvent A (0.1% Formic Acid) and desalted using steps mentioned below:

i. Column activation – 100% Acetonitrile

ii. Column equilibration – solvent A

iii. Load sample that is reconstituted in solvent A

iv. Column wash – Solvent A (twice)

v. Elution – Solvent B (40% Acetonitrile in 0.1% Formic Acid)

The eluted peptides were vacuum dried using Speed-Vac (Eppendorf) and stored at -20°C till mass spec analysis.

**Figure 3.3:** Pattern followed to concatenate fractions obtained by bRPLC

### 3.8 LC-MS/MS analysis

The peptide fractions were analysed on LTQ-Orbitrap Velos (Thermo Scientific, Bremen, Germany) interfaced with an Easy-nLC (Thermo Scientific, Bremen, Germany). In total, 67 fractions from all the methods were analyzed by mass spectrometry. The LC system consisted of an enrichment column (75 µm × 2 cm, C18 material 5 µm, 100 Å) and an analytical column (75 µm × 10 cm, C18 material 5 µm, 100 Å). Both the columns were packed in-house. A flow rate of 350 nl/min and a linear gradient of 7% to 30% solvent B (0.1% formic acid, 90% ACN) was used to separate the peptides on the analytical column. The spray
voltage was 2.0 kV while the heated capillary temperature was 225°C. Mass spectrometry data was acquired in a data dependent mode with a resolution of 60,000 at 400 m/z for MS scans and 7500 for MS/MS scans. The 20 most intense precursor ions from each survey scan were selected for fragmentation. Peptide fragmentation was performed by higher-energy collision dissociation (HCD) with normalized collision energy of 39%. The automatic gain control and maximum ion injection time for FT MS was 0.5 million ions and 100 ms respectively; while it was 0.1 million ions and 200 ms for FT MS/MS.

3.9 Database searches for peptide and protein identification

The MS/MS data acquired was searched against two databases – i) *C. tropicalis* protein database (6,258 entries; 3,034,039 amino acids), and ii) six-frame translated genome database (992,921 entries; 25,161,168 amino acids). The whole genome and protein databases were downloaded from the resources of Broad Institute (http://www.broadinstitute.org/annotation/genome/candida_group/MultiDownloads.html). The six-frame translated genome database was created using in-house Python scripts. Six-frame translation of the genomic sequence was carried out from stop codon to stop codon and the sequences that were not unique and/or shorter than 7 amino acids in length were removed. Sequences of commonly encountered protein contaminants such as BSA, trypsin and keratins were added to both the databases.

The searches were performed using SEQUEST and Mascot through Proteome Discoverer (Version 1.4) software suite (Thermo Scientific, Bremen, Germany). The workflow adopted for the searches is represented in Figure 3.4. As depicted, the raw files were processed using spectrum selector; these filtered spectra were then searched using two search engines followed by target-decoy-based FDR calculation. While the first parse searches used *C. tropicalis* protein database, the second parse used the six-frame translated genome database.
Figure 3.4: Workflow used in Proteome Discoverer 1.4 software suite

Two different search algorithms were used to maximize the peptide and protein identifications. The search parameters included trypsin as the proteolytic enzyme with 2 missed cleavages and semi-tryptic identifications allowed. Oxidation of methionine, deamidation of asparagine and glutamine, carbamylation of the peptide N-terminus and lysine, and acetylation of the protein N-terminus (only for Mascot) were set as dynamic modifications while carbamidomethylation at cysteine was set as a static modification. Precursor and fragment mass tolerance were set to 10 ppm and 0.05 Da, respectively. Only those peptide-spectrum matches (PSMs) that qualified a 1% false discovery rate (FDR) were considered as authentic identifications. The FDR was calculated using target-decoy database searches as follows –

A reverse sequence database (decoy) was searched separately in addition to forward (target) database and false discovery rate (FDR) was calculated at every PSM score value using the formula –

\[
\% \text{ FDR} = \frac{\text{Number of hits in the reverse DB at or above the score}}{\text{Total number of hits in the target and reverse DB at or above the score}} \times 100
\]
A score threshold for 1% FDR was applied to search results from individual data files. Only first rank peptide sequences matching to each spectrum were considered. Since four different types of searches were performed with the same data set (genome and protein database searches using Mascot and SEQUEST), peptide sequences assigned to a single spectrum in each search were compared. Spectra which were assigned different sequences in different searches were omitted from further analysis. Protein identification list was generated by grouping proteins based on shared peptides and using maximum parsimony principles (Figure 3.5), as described by [61].

**Figure 3.5:** Protein grouping was achieved using the principles of maximum parsimony

### 3.10 Intensity-based absolute quantification (iBAQ) calculation

iBAQ was calculated as described earlier [62]. Briefly, iBAQ value for a given protein was computed by summing up the intensities of all identified unique peptides. To account for the varied lengths of proteins and peptides that were not identified, this value was divided by the total number of theoretically observable peptides. To calculate the number of theoretically observable peptides, *C. tropicalis* protein database was digested *in-silico* using a PERL script generated in-house. All tryptic peptides with a length of 6 – 30 amino acids were counted.

\[
iBAQ = \frac{\text{Sum of intensities of all unique peptides identified}}{\text{Total number of theoretically observable unique peptides}}
\]

The resulting iBAQ values were log transformed and plotted.
3.11 Proteogenomic analysis

The peptides identified uniquely in the six-frame translated genome database were used for proteogenomic analysis. These peptides, referred to as Genome Search Specific Peptides (GSSPs), were considered for analysis only if they qualified the 1% FDR threshold and matched uniquely with six frame translated genome database. The quality of MS/MS spectra of these GSSPs was manually verified to ensure correct peptide assignments. The major criteria considered for manual evaluation included –

i. All intense peaks should have been assigned. Intense unassigned peaks were checked if they were arising from internal fragment ions.

ii. Majority of the y series of ions should have been identified.

iii. The spectrum should have high signal to noise (S/N) ratio.

iv. If an ammonium ion indicated the presence of an amino acid which was not present in the assigned peptide sequence, PSM was rejected.

v. Presence of y1 ion confirming peptide ending either with K (m/z 147.11) or with R (m/z 175.12) was checked.

vi. If the presence of any un-assigned fragment ion especially from higher m/z range indicated the presence of an amino acid which was not a part of assigned sequence, PSM was rejected.

vii. If many assigned peaks were from the noise level then the PSM was rejected.

The GSSPs in the intergenic region were analysed to identify novel exons and novel protein coding regions while the GSSPs that were in close proximity with the existing genome annotation were used to refine the existing gene models.

Proteogenomic method used in this study was essentially a semi-manual annotation strategy. A visualization platform like a genome browser forms an integral part of such analysis. The genome of C. tropicalis, annotated genes, gene models derived from publicly available RNA-Seq data (www.ncbi.nlm.nih.gov/sra/SRX470927), gene predictions by Augustus, ortholog genes of C. albicans, peptides identified from known proteins, and GSSPs were overlaid in the Integrated Genome Viewer (IGV) Version 2.3 [63] for analysis as depicted in Figure. 3.6.
Figure 3.6: Integrative Genomics Viewer was used for proteogenomic analysis