Structural and Biochemical Dissection of the Trehalose Biosynthetic Complex in Pathogenic Fungi

by

Yi Miao

Department of Biochemistry
Duke University

Date:_______________________

Approved:

___________________________
Richard G. Brennan, Supervisor

___________________________
John R. Perfect

___________________________
Jane S. Richardson

___________________________
Maria A. Schumacher

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2016
ABSTRACT

Structural and Biochemical Dissection of the Trehalose Biosynthetic Complex in Pathogenic Fungi

by

Yi Miao

Department of Biochemistry
Duke University

Date:_______________________
Approved:

Richard G. Brennan, Supervisor

John R. Perfect

Jane S. Richardson

Maria A. Schumacher

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2016
Abstract

Trehalose is a non-reducing disaccharide essential for pathogenic fungal survival and virulence. The biosynthesis of trehalose requires the trehalose-6-phosphate synthase, Tps1, and trehalose-6-phosphate phosphatase, Tps2. More importantly, the trehalose biosynthetic pathway is absent in mammals, conferring this pathway as an ideal target for antifungal drug design. However, lack of germane biochemical and structural information hinders antifungal drug design against these targets.

In this dissertation, macromolecular X-ray crystallography and biochemical assays were employed to understand the structures and functions of proteins involved in the trehalose biosynthetic pathway. I report here the first eukaryotic Tps1 structures from Candida albicans (C. albicans) and Aspergillus fumigatus (A. fumigatus) with substrates or substrate analogs. These structures reveal the key residues involved in substrate binding and catalysis. Subsequent enzymatic assays and cellular assays highlight the significance of these key Tps1 residues in enzyme function and fungal stress response. The Tps1 structure captured in its transition-state with a non-hydrolysable inhibitor demonstrates that Tps1 adopts an “internal return like” mechanism for catalysis. Furthermore, disruption of the trehalose biosynthetic complex formation through abolishing Tps1 dimerization reveals that complex formation has regulatory function in
addition to trehalose production, providing additional targets for antifungal drug intervention.

I also present here the structure of the Tps2 N-terminal domain (Tps2NTD) from *C. albicans*, which may be involved in the proper formation of the trehalose biosynthetic complex. Deletion of the Tps2NTD results in a temperature sensitive phenotype. Further, I describe in this dissertation the structures of the Tps2 phosphatase domain (Tps2PD) from *C. albicans*, *A. fumigatus* and *Cryptococcus neoformans* (*C. neoformans*) in multiple conformational states. The structures of the *C. albicans* Tps2PD-BeF₃-trehalose complex and *C. neoformans* Tps2PD(D24N)-T₆P complex reveal extensive interactions between both glucose moieties of the trehalose involving all eight hydroxyl groups and multiple residues of both the cap and core domains of Tps2PD. These structures also reveal that steric hindrance is a key underlying factor for the exquisite substrate specificity of Tps2PD. In addition, the structures of Tps2PD in the open conformation provide direct visualization of the conformational changes of this domain that are effected by substrate binding and product release.

Last, I present the structure of the *C. albicans* trehalose synthase regulatory protein (Tps3) pseudo-phosphatase domain (Tps3PPD) structure. Tps3PPD adopts a haloacid dehydrogenase superfamily (HADSF) phosphatase fold with a core Rossmann-fold domain and a α/β fold cap domain. Despite lack of phosphatase activity, the cleft between the Tps3PPD core domain and cap domain presents a binding pocket for a yet
uncharacterized ligand. Identification of this ligand could reveal the cellular function of Tps3 and any interconnection of the trehalose biosynthetic pathway with other cellular metabolic pathways.

Combined, these structures together with significant biochemical analyses advance our understanding of the proteins responsible for trehalose biosynthesis. These structures are ready to be exploited to rationally design or optimize inhibitors of the trehalose biosynthetic pathway enzymes. Hence, the work described in this thesis has laid the groundwork for the design of Tps1 and Tps2 specific inhibitors, which ultimately could lead to novel therapeutics to treat fungal infections.
Dedication

My sincere thanks to my parents and Jie for their continuous supports!
## Contents

Abstract ................................................................................................................................. iv

List of Tables ......................................................................................................................... xiii

List of Figures ......................................................................................................................... xiv

List of Abbreviations ............................................................................................................. xvi

Acknowledgements ............................................................................................................... xix

1. Introduction: Significance of trehalose biosynthesis in pathogenic fungi .................. 1

   1.1 Invasive mycoses and challenges .............................................................................. 1

   1.1.1 Candidiasis, aspergillosis and cryptococcosis ..................................................... 2

   1.1.2 Treatments for invasive mycoses ......................................................................... 3

   1.1.3 Antifungal drug resistance .................................................................................... 5

   1.2 Discovery and protective function of trehalose ....................................................... 7

   1.3 Emerging regulatory function of trehalose biosynthetic pathway ......................... 9

   1.4 Significance of trehalose biosynthesis in fungal virulence ....................................... 13

   1.5 Outline of research .................................................................................................. 16

   1.5.1 Glycosyltransferases and Tps1 ........................................................................... 16

   1.5.2 HADSF phosphatases and Tps2 ....................................................................... 19

   1.5.3 Tps3 structure and function .............................................................................. 20

   1.5.4 Trehalose biosynthetic complex and regulatory function .................................. 21

2. Structures of C. albicans Tps1 provide insight into antifungal drug design .......... 27

   2.1 Introduction .............................................................................................................. 27
2.2 Results and discussion ................................................................. 29
  2.2.1 Structure of \textit{C. albicans} Tps1 ............................................. 29
  2.2.2 Binding pocket of \textit{C. albicans} Tps1 .................................... 31
  2.2.3 \textit{C. albicans} Tps1 mutant enzymatic activity ......................... 34
  2.2.4 Elucidation of the catalytic mechanism of Tps1 ......................... 34
  2.2.5 Significance of the formation of trehalose biosynthetic complex ....... 36

2.3 Experimental procedures ............................................................ 38
  2.3.1 Protein expression and crystallization ................................... 38
  2.3.2 X-ray data collection and structure determination .................... 39
  2.3.3 Enzyme activity assays ....................................................... 39

2.4 Acknowledgements ...................................................................... 40

3. Structures of trehalose-6-phosphate synthase from \textit{Aspergillus fumigatus} reveal the molecular mechanism of glucose transfer .................................................... 49
  3.1 Introduction ................................................................................ 49
  3.2 Results and Discussion .............................................................. 50
    3.2.1 Structure of \textit{A. fumigatus} Tps1A ....................................... 50
    3.2.2 Structure of \textit{A. fumigatus} Tps1B ....................................... 51
    3.2.3 Substrate binding to Tps1A and Tps1B .................................... 52
  3.3 Future directions in understanding Tps1A and Tps1B ...................... 54
    3.3.1 Activity difference between Tps1A and Tps1B ......................... 55
    3.3.2 Contribution of C-terminal tail of Tps1A in activity and localization 56
    3.3.3 Dissection of oligomeric state of Tps1A and Tps1B .................... 56
3.4 Experimental procedure ................................................................. 57
3.4.1 Protein expression and crystallization .......................................... 57
3.4.2 Data collection and structure determination ................................... 58

4. Structures of Tps2 from pathogenic fungi reveal the mechanisms of substrate specificity and catalysis .......................................................... 65
4.1 Introduction ................................................................................. 65
4.2 Results ....................................................................................... 69
4.2.1 Structure of C. albicans Tps2NTD ............................................ 69
4.2.2 Structure of the C. albicans Tps2PD ......................................... 72
4.2.3 Substrate binding pocket of C. albicans Tps2PD ......................... 73
4.2.4 Tps2PD active site mutations ................................................... 76
4.2.5 Closed conformational structure of C. neoformans Tps2PD .......... 77
4.2.6 Tps2PD adopts a large conformational change upon substrate binding .... 79
4.2.7 Substrate specificity of Tps2PD ................................................. 81
4.3 Experimental Procedures .............................................................. 84
4.3.1 Protein purification and crystallization ....................................... 84
4.3.2 Data collection and structure determination .................................. 86
4.3.3 Steady-state kinetics and enzymatic activity assays ....................... 87
4.3.4 Cell survival assays and capsule induction assays ....................... 87
4.4 Acknowledgements .................................................................... 88

5. Structure of C. albicans Tps3PPD reveals a pseudo-phosphatase fold ................. 103
5.1 Introduction ............................................................................ 103
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Results and discussion</td>
<td>105</td>
</tr>
<tr>
<td>5.2.1 Tps3 protein expression and purification</td>
<td>105</td>
</tr>
<tr>
<td>5.2.2 Overall structure of <em>C. albicans</em> Tps3PPD</td>
<td>106</td>
</tr>
<tr>
<td>5.2.3 Potential ligand-binding pocket of Tps3PPD</td>
<td>108</td>
</tr>
<tr>
<td>5.3 Future directions</td>
<td>110</td>
</tr>
<tr>
<td>5.4 Experimental procedures</td>
<td>111</td>
</tr>
<tr>
<td>5.4.1 Protein expression and crystallization</td>
<td>111</td>
</tr>
<tr>
<td>5.4.2 Data collection</td>
<td>112</td>
</tr>
<tr>
<td>5.4.3 MD simulation and ligand docking</td>
<td>112</td>
</tr>
<tr>
<td>5.5 Acknowledgements</td>
<td>114</td>
</tr>
<tr>
<td>6. Future directions</td>
<td>120</td>
</tr>
<tr>
<td>6.1 Structure determinations of <em>C. neoformans</em> Tps1 and Tps2</td>
<td>120</td>
</tr>
<tr>
<td>6.2 Structure determinations of <em>A. fumigatus</em> Tps2NTD and Tps3</td>
<td>123</td>
</tr>
<tr>
<td>6.3 Inhibitor design against Tps1 and Tps2</td>
<td>124</td>
</tr>
<tr>
<td>6.4 Dissection and reconstitution of trehalose biosynthetic complex</td>
<td>126</td>
</tr>
<tr>
<td>6.5 Acknowledgements</td>
<td>128</td>
</tr>
<tr>
<td>Appendix A: PDB accession numbers</td>
<td>131</td>
</tr>
<tr>
<td>Appendix B: X-ray protein crystallography</td>
<td>132</td>
</tr>
<tr>
<td>X-ray crystallography, NMR and Cryo-EM</td>
<td>132</td>
</tr>
<tr>
<td>Diffraction basics</td>
<td>134</td>
</tr>
<tr>
<td>Protein crystallization and space group</td>
<td>139</td>
</tr>
<tr>
<td>Data collection and process</td>
<td>141</td>
</tr>
</tbody>
</table>
Structure determination by molecular replacement ................................................... 144
Structure determination by SAD .................................................................................. 147
Structure refinement and validation ............................................................................ 149
References ..................................................................................................................... 161
Biography ...................................................................................................................... 178
List of Tables

Table 1: Structures of proteins involved in trehalose biosynthesis .............................................. 22
Table 2: Selected *C. albicans* Tps1 data collection and refinement statistics ............................. 41
Table 3: Selected *A. fumigatus* Tps1 data collection and refinement statistics ........................... 59
Table 4: Selected *C. albicans* Tps2 data collection and refinement statistics ............................. 89
Table 5: Selected *C. neoformans* and *A. fumigatus* Tps2 data collection and refinement statistics ........................................................................................................................................ 90
Table 6: Selected *C. albicans* Tps3 data collection and refinement statistics ............................. 115
Table 7: Crystal systems and Bravais lattice types .......................................................................... 153
List of Figures

Figure 1: An enzymatic scheme of TPS pathway ............................................................. 23
Figure 2: Trehalose biosynthesis pathways .................................................................... 24
Figure 3: Overall folds of the GT-A and GT-B glycosyltransferases ............................. 25
Figure 4: Overall folds of the HADSF phosphatase subfamilies ................................. 26
Figure 5: Overall structure of C. albicans Tps1 ............................................................. 42
Figure 6: Sequence alignment of C. albicans Tps1 ....................................................... 43
Figure 7: Active site of C. albicans Tps1 ................................................................. 44
Figure 8: Electron density for substrates ...................................................................... 45
Figure 9: Relative enzyme activity of C. albicans Tps1 and selected mutants ............. 46
Figure 10: Tps1-VDM interactions .............................................................................. 47
Figure 11: C. albicans Tps1 dimer interactions ............................................................ 48
Figure 12: Structure of A. fumigatus Tps1A-UDP-VDM complex ............................... 60
Figure 13: Structure of A. fumigatus Tps1B-UDP-VDM complex ............................... 61
Figure 14: Sequence alignment of A. fumigatus Tps1B and Tps1A ............................... 62
Figure 15: Active site of A. fumigatus Tps1B ............................................................... 63
Figure 16: Size exclusion chromatography of Tps1A and Tps1B ................................. 64
Figure 17: Overall structure of Tps2NTD ................................................................. 91
Figure 18: Sequence alignment of C. albicans Tps2NTD ............................................. 92
Figure 19: Tps2PD activity assays ............................................................................. 93
Figure 20: Tps2PD overall structure ....................................................................... 94
Figure 21: Tps2PD sequence alignments .................................................................96
Figure 22: Tps2PD activity assays and closed conformation structure .........................97
Figure 23: C. neoformans Tps2PD overall structure ..................................................98
Figure 24: Capsule induction assays .........................................................................99
Figure 25: Superposition of C. albicans and B. malayi Tps2PD ..................................100
Figure 26: Structures of A. fumigatus Tps2PD in open conformation .........................101
Figure 27: Structure of C. albicans Tps2PD with docked sucrose ...............................102
Figure 28: Structure of C. albicans Tps3PPD ...........................................................116
Figure 29: Comparison of Tps3PPD to Tps2PD .......................................................117
Figure 30: Structure representation of MD simulations results ...................................118
Figure 31: Representative Tps3PPD-UDP-glucose complex structure from docking ....119
Figure 32: Crystallization of C. neoformans Tps1 .......................................................129
Figure 33: Purification of the C. neoformans Tps2NTD and full-length Tps2 .............130
Figure 34: Visualization of Bragg’s law .....................................................................154
Figure 35: Ewald sphere and the reciprocal lattice .....................................................155
Figure 36: Phase diagram of protein crystallization ....................................................156
Figure 37: Hanging-drop vapor diffusion method .......................................................157
Figure 38: Experimental phasing by isomorphous replacement ................................158
Figure 39: Se-Met scan result of A. fumigatus Tps2PD crystal .................................159
Figure 40: Introduction of anomalous scattering contributions to Friedel pairs .........160
List of Abbreviations

ABPA: allergic bronchopulmonary aspergillosis
ADP: adenosine 5-diphosphate
ALS: advanced light source
APS: advanced photon source
ASU: asymmetric unit
BSI: bloodstream infection
CPA: chronic pulmonary aspergillosis
Cryo-EM: cryo-electron microscopy
G6P: glucose-6-phosphate
HADSF: haloacid dehydrogenase superfamily
ITC: isothermal titration calorimetry
MAD: multi-wavelength anomalous diffraction
MD: molecular dynamics
MR: molecular replacement
NMR: nuclear magnetic resonance
PDB: protein databank
PEG: polyethylene glycol
rmsd: root mean square deviation
SAD: single-wavelength anomalous diffraction
SEC: size exclusion chromatography
SnRK1: sucrose non-fermenting kinase 1
TEV: tobacco etch virus
Tsl1: trehalose synthase long chain 1
Tps1: trehalose-6-phosphate synthase
Tps2: trehalose-6-phosphate phosphatase
Tps2NTD: trehalose-6-phosphate phosphatase N-terminal domain
Tps2PD: trehalose-6phosphate phosphatase phosphatase domain
Tps3: trehalose synthase regulatory protein
Tps3NTD: trehalose synthase regulatory protein N-terminal domain
Tps3PPD: trehalose synthase regulatory protein pseudo-phosphatase domain
TreA: trehalose-6-phosphate hydrolase
TreP: trehalose phosphorylase
TreS: trehalose synthase
TreT: trehalose glycosyltransferring synthase
TreY: maltooligosyl trehalose synthase
TreZ: maltooligosyl trehalose trehalohydrolase
T6P: trehalose-6-phosphate
UDP: uridine 5-diphosphate
VDM: validoxylamine A
WT: wide type

YMDB: yeast metabolite database

5-FC: 5-flucytosine

5-FU: 5-fluorouracil
Acknowledgements

I am sincerely thankful to several people and organizations for the completion of my dissertation, without whom the project would be impossible. I would specially like to thank my advisor Dr. Richard G. Brennan for his guidance throughout graduate school. His ideas and supervision are essential for the developments of this dissertation. In addition, critical thinking and problem solving skills learned from him significantly shape my future career. His optimistic attitude in the lab also provides mental support during difficult times.

I would like to thank Dr. John R. Perfect for his support and advice for this project. This project would have not started with John as the initiator. John has provided valuable suggestions from a clinician point of view and encouragements during this dissertation. I would also like to thank Drs. Jennifer L. Tenor and Dena L. Toffaletti for their fungal cellular work on this project, which implements my biochemical results and translates to cellular levels.

I would like to thank Dr. Maria A. Schumacher for her assistance with structure determination. Her knowledge and experience significantly accelerates experimental phasing of several proteins. Critical question arisen in lab group meetings and committee meetings are also contributing to the progress of this dissertation. I would like to than Dr. Jane S. Richardson for her suggestions during committee meetings.
I would like to thank Dr. Richard E. Lee, whose knowledge in synthetic chemistry and inhibitor design. I would also like to thank Drs. William Shadrick and Jiuyu Liu in Lee lab for synthesis of compounds and inhibitor screening.

I would like to thank Dr. Hashim M. Al-Hashimi for his assistance with molecular dynamics simulation and docking. These results from Dr. Yi Xue and Laura Ganser in Al-Hashimi lab significantly impacts part of this dissertation.

I would like to thank former and current Brennan lab and Schumacher lab members for constituting a supporting environment in the lab. I specially would like to thank Drs. Nam Tonthat and Ivan Birukou for their assistance with data collection and structure determination. I also would like to thank Tien-jui Yen for assistance with enzyme kinetics experiments.

I would like to thank staff at SER-CAT and BCSB for their assistance with data collection. Data collection at SER-CAT was organized by Drs. Charles Pemble and Nathan Nicely from Duke University Macromolecular X-ray Crystallography Shared Resource, to whom I owe thanks.

I would also like to thank National Institute of Health for financial support. The work was supported by grant 1P01AI104533-01A1 to Richard G. Brenan.
1. Introduction: Significance of trehalose biosynthesis in pathogenic fungi

Trehalose biosynthesis is a significant pathway in cellular survival and virulence for pathogenic fungi. Both the protective function of trehalose and regulatory function of this pathway are essential for maintaining cellular integrity. Absence of this pathway in mammalian systems confers this pathway as an ideal target for antifungal drug design. However, the lack of germane structural and biochemical information on fungal enzymes in the trehalose biosynthetic pathway hinders our understandings and delays inhibitor design.

1.1 Invasive mycoses and challenges

The fungal kingdom comprises of a large number of eukaryotic organisms ranging from unicellular to multicellular species. An estimated of 1.5 million species are included in this kingdom [1], though the majority have not been identified. Being an integral part of human activities, edible mushrooms are a well-known diet around the world. Fungi haven also long been used in food industries for fermentation of wine, soy sauce and vinegar. In addition, fungi are used for industrial production of antibiotics.

Despite widespread applications and significant impacts of fungi in multiple endeavors of human lives, infections by fungal pathogens threaten millions of lives worldwide annually. The most prevalent fungal infections are superficial infections of
the skin and nails, which affect 25% of the world’s population [2]. Invasive mycoses, though less predominant than superficial infections, are associated with high mortality rate, resulting in 1.5 million deaths worldwide [3]. Three major genera: Candida, Cryptococcus and Aspergillus contribute to the majority of invasive mycoses. In this dissertation, I focus on these genera and discuss the epidemiology of these infections.

1.1.1 Candidiasis, aspergillosis and cryptococcosis

Candidemia (Candida infections of the bloodstream and the most predominant form of invasive candidiasis) is a common infection in hospitalized patients, representing the fourth leading cause of nosocomial bloodstream infections (BSI) in the United States [4]. Candidemia significantly prolongs duration of hospitalization and results in high mortality in these patients (46 -75%) [3]. Five major species, C. albicans, C. glabrata, C. parapsilosis, C. tropicalis and C. krusei are responsible for the majority of Candida infections, with C. albicans contributing to more than half of Candida infections [5]. The worldwide annual incidence of candidemia is estimated to be approximately 400,000 cases [3,6] and result in a serious burden on healthcare systems, primarily in economically developed countries.

Invasive aspergillosis is a leading cause of death in patients with leukemia, bone marrow transplant, solid-organ transplants and increasingly diagnosed in AIDS patients [7]. A. fumigatus is the most frequent cause of aspergillosis. Invasive aspergillosis leads
to a mortality rate higher than 50% and delayed treatments lead to essentially complete fatality rate in patients [3]. In addition to invasive aspergillosis, allergic bronchopulmonary aspergillosis (ABPA) frequently develops in patients with underlying lung diseases, including asthma, with an estimated global incidence around 5 million cases [8]. ABPA complicates asthma and gradually cause chronic pulmonary aspergillosis (CPA) and death.

Cryptococcosis, primarily caused by *C. neoformans*, is one of the most serious opportunistic infections in immunocompromised patients. Cryptococcal meningitis is one of the most common causes of meningitis, with an estimated global incidence over one million cases per year, resulting in more than 600,000 deaths worldwide [9]. Recently, an outbreak of *Cryptococcus gattii* (*C. gattii*) infections on Vancouver Island, Canada and the northwestern United States showed hypervirulent isolates that could infect immunocompetent populations [10].

### 1.1.2 Treatments for invasive mycoses

Antifungal agents are administrated to patients with invasive fungal infections. In the clinics, candidemia is most frequently treated with fluconazole (67.7%), followed by caspofungin, micafungin and lipid formulations of amphotericin B [6]. The most common medical treatment for invasive aspergillosis is voriconazole [11]. *A. fumigatus*, the most predominant invasive aspergillosis species, is intrinsically resistant
to fluconazole [12]. Mild cryptococcal infections are primarily treated with fluconazole and cryptococcal meningitis is treated with lipid formulations of amphotericin B in combination with flucytosine [13].

Based on their mechanisms of action, current antifungal agents, except flucytosine, belong to one of the three major classes: azoles (more commonly triazole based antifungals), polyenes and echinocandins. Azoles (including fluconazole, itraconazole and voriconazole) target the synthesis of ergosterol, one major sterol present in fungal cell membranes [14]. Azoles function by inhibition of lanosterol 14α-demethylase, blocking synthesis of ergosterol and accumulating toxic intermediates.

Polyenes (primarily amphotericin B) interact with ergosterol in fungal cell membranes and form pores, resulting in leaking from fungal cells. Two forms of amphotericin B are prescribed in the clinics, amphotericin B deoxycholate and lipid formulations of amphotericin B, the latter more frequently utilized due to lower toxicity. Echinocandins (including caspofungin and micafungin) function by inhibiting 1,3-β-D-glucan synthase, an essential enzyme for fungal cell wall synthesis, thus interrupting cell wall integrity [15]. Echinocandins demonstrate antifungal activities against selected fungal species including Candida and Aspergillus.

Flucytosine (5-FC) utilizes a different mechanism from those described for the other major classes of antifungals [16]. 5-FC is converted to 5-fluorouracil (5-FU) by
cytosine deaminase in fungal cells. Two subsequent different inhibitory mechanisms have been proposed for the antifungal function of 5-FU. 5-FU can be converted to 5-fluorouracil triphosphate, altering aminoacylation of tRNA and inhibiting protein synthesis. Alternatively, 5-FU can be converted to 5-fluorouracil monophosphate, inhibiting the essential DNA synthesis enzyme thymidylate synthetase [16].

1.1.3 Antifungal drug resistance

Antifungal drug resistance is emerging as a problem in the clinics due to the continuous use of current limited arsenal of drugs. Two primary mechanisms, induction of drug efflux pumps and alternations of drug targets, result in drug resistance in pathogenic fungi [17]. Here, I briefly cover the resistance mechanisms of Candida, Aspergillus and Cryptococcus to clinically relevant drugs employed in the clinics.

Candida resistance to fluconazole is increasingly prevalent in AIDS patients and complicates the treatments of the underlying primary diseases. Fluconazole concentration could be reduced in fungal cells by induction of efflux pumps, encoded by Multi-Drug Resistance and Candida Drug Resistance genes. Upregulation of enzyme concentration achieves resistance in a similar fashion. In addition, either point mutations in lanosterol 14α-demethylase abolish azole binding or Candida bypasses the need of lanosterol 14α-demethylase with ergosterol synthesis [18].
Voriconazole has only recently been introduced for treatment of invasive aspergillosis, due to its superior pharmacokinetics when compared to amphotericin B. Reports on voriconazole resistance in *Aspergillus* remain rare, but have been observed in patients, who have been treated for long periods. Resistance mechanisms to azole, similar to those of *Candida*, have been proposed for *Aspergillus* [19].

Amphotericin B has been used in the clinics for over 50 years and repeated exposure to this antifungal results in resistance in major pathogenic fungal species. Two major causes of resistance include a change in membrane composition and reduced access of amphotericin B to its target. Cells with resistance to amphotericin B either change membrane composition, reducing amphotericin B binding affinity or alter their 1,3-β-D-glucan content in the cell wall, limiting access of amphotericin B to the plasma membrane [20]. Resistance to 5-FC is one major obstacle for treatment of cryptococcal meningitis [21]. Resistant strains show mutations in cytosine permease, resulting in decreased uptake of 5-FC. Alternatively, resistance strains show mutations in uridine-5-monophosphate pyrophosphorylase and uracil phosphoribosyl transferase, leading to defects in 5-FU conversion.

In addition to emerging antifungal drug resistance, drug toxicities and drug-drug interactions further complicate fungal infection treatments. Therefore, novel mechanisms of antifungal intervention are urgently required and I will describe how
disruption of the trehalose biosynthetic pathways of pathogenic could be employed for novel antifungal development in the following sections.

1.2 Discovery and protective function of trehalose

Trehalose, α-D-glucopyranosyl-(1→1)-α-D-glucopyranoside, is a non-reducing disaccharide first discovered in 1832 by H.A.L. Wiggers [22]. Trehalose gained the name from trehala manna, where trehalose was isolated in 1858. Trehalose has been found to be widely synthesized in bacteria, fungi, lower plants and invertebrates, but has never been identified in mammals [23]. Trehalose could be converted into two units of glucose and was first viewed as an energy storage molecule for these organisms [24]. Activation of ascospores of Neurospora tetrasperma from the dormant state through heat shock shows consumption of trehalose, confirming trehalose can be an energy source [25]. In addition, insect flight significantly decreases blood trehalose level in Phormia regina, suggesting trehalose was utilized as energy source in some insects as well [26].

Trehalose has gained tremendous attention as protective molecules through studies on anhydrobiotic organisms [27]. Trehalose is a significant component of organisms undergoing almost complete desiccation, termed as anhydrobiotic organisms. The most well-known anhydrobiotic organism is dry active baker’s yeast. A significant portion of dry weight of anhydrobiotic organisms is trehalose, suggesting a potential function in desiccation. Dehydration of a biological membrane normally results in
irreversible morphological damage. However, in the presence of trehalose, vesicles show similar morphological and functional phenotypes after a dehydration-rehydration cycle [28]. In addition, phospholipids undergo a transition from liquid crystalline state to gel phase upon dehydration. This damage to membranes could be alleviated with increasing concentration of trehalose [29]. In addition, trehalose has been shown to stabilize proteins in yeast during heat shock and suppress the aggregation of denatured protein [30]. Therefore, it has been proposed that trehalose protects proteins and membranes in “dry” states through direct interactions with polar residues and vitrification to stabilize dry materials [31]. Recently, the Koshland group found that intracellular trehalose is sufficient for desiccation tolerance in *Saccharomyces cerevisiae* (*S. cerevisiae*) by exogenously supplying trehalose in the media [32], thereby highlighting the significance of trehalose protective functions in desiccation.

Trehalose has gained widespread industrial use due to its protective properties and unreactive nature. Trehalose is an active stabilizer of enzymes, vaccines and has been widely used in the food industry [33]. Compared with expensive extraction from biomass, the unit price of trehalose has been significantly reduced due to novel developments in trehalose production through maltooligosyl trehalose synthase and maltooligosyl trehalose trehalohydrolase (TreY-TreZ). Large scale production of trehalose through maltose conversion significantly reduces the unit price of trehalose by
100 fold, thus rendering the usage of trehalose in industries economically applicable [33].

1.3 Emerging regulatory function of trehalose biosynthetic pathway

In addition to the protective function of trehalose in the stress response to dehydration and heat shock, emerging metabolic functions of the trehalose biosynthetic pathway is being realized. In pathogenic fungi, trehalose is synthesized in a two-step enzymatic reaction. The first step is condensation of UDP-Glucose and G6P to form trehalose-6-phosphate (T6P) and is catalyzed by Tps1. The second step, which is catalyzed by Tps2, is dephosphorylation of T6P to form trehalose (Fig. 1). One key obstacle to understand trehalose function in the stress response is to differentiate the function of the trehalose molecule from the metabolic function of trehalose biosynthetic pathway. Recently, the Botstein group developed a tool to understand the physiological function of trehalose by introducing AGT1 (the α-glucoside transporter) into S. cerevisiae [34]. One variant of this transporter imports trehalose into the cells from an exogenous supply. The Botstein group found that intracellular accumulation of trehalose is not sufficient to rescue the growth and sporulation defects of tps1Δ or tps2Δ, deletion mutants of the two trehalose biosynthetic genes. These results suggest trehalose accumulation is not sufficient for the stress response and highlight the metabolic function of the trehalose biosynthetic pathway in this response.
A metabolic function of the trehalose biosynthetic pathway has been proposed for a long time with the tps1Δ mutant unable to grow on glucose media [35]. T6P, the intermediate for trehalose biosynthesis was found to inhibit hexokinases in vitro [36], thereby inhibiting the flux of glucose into glycolysis. However, Schizosaccharomyces pombe expressing one T6P-insensitive form of hexokinase 2 was able to utilize glucose [37], rendering T6P inhibition of hexokinase not sufficient to explain the tps1Δ deficiency in glucose media. The molecular mechanism of growth defects of tps1Δ mutant remained elusive until recently [38]. Yeast cells transitioned into glucose media have a flux of glucose into the glycolysis pathway. High flux into the first steps of glycolysis but not subsequent steps leads to low levels of inorganic phosphate and high levels of G6P, resulting in cell death. Presence of trehalose biosynthetic pathway rescues these cells by releasing free inorganic phosphate and reducing G6P levels [38]. Studies on the metabolic function of the trehalose biosynthetic pathway have been focused on the model organism S. cerevisiae, which potentially is different from other fungal organisms. Trehalose biosynthesis has also been investigated in rice blast fungus Magnaporthe grisea (M. grisea) by the Talbot group. An M. grisea tps1Δ mutant failed to accumulate trehalose and showed attenuated pathogenicity [39]. However, the M. grisea tps1Δ mutant growth defects on glucose media could be rescued by addition of free amino acids. Further analyses by the Talbot group found Tps1 regulated NADPH levels via the pentose
phosphate pathway [40], further regulating NADP binding transcriptional corepressors [41], thus affecting fungus infections. Further detailed research is required to understand whether the *tps1* metabolic functions in *S. cerevisiae* and *M. grisea* are ubiquitous in other fungal species. Nevertheless, current research highlights that both the protective functions of trehalose and the metabolic functions of trehalose biosynthetic pathway are significant contributors in the stress response and virulence of pathogenic fungi.

Compared to the research on the metabolic function of the trehalose biosynthesis fungal systems, the significant impact of T6P in plant metabolism has been under intensive investigation. Trehalose was first discovered in the plant *Selaginella lepidophylla* in 1913 [42]. However, the effects of trehalose metabolism in plant biology have been overlooked for almost a century until the discoveries of *tps1* and *tps2* gene families in *Arabidopsis thaliana*. Meanwhile, due to the protective function of trehalose in stress response, bacterial and yeast trehalose biosynthetic genes have been engineered into plants cells. Indeed, introduction of yeast *tps1* gene into tobacco results in drought tolerance phenotypes [43] and overexpression of trehalose in rice confers tolerance to various abiotic stresses [44]. However, these transgenic plants also show developmental anomalies, which make them less favorable for agricultural applications. Nevertheless, these plant phenotypes of stress tolerance highlight the significance of trehalose
biosynthesis in plant metabolism, resulting in intensive investigations to understand the physiological functions of this pathway.

Similar to fungal cells, trehalose is synthesized through the intermediate T6P in plants. Recent research has shown the T6P concentration in plants has far-reaching effects on various plant developmental stages, including embryogenesis, leaf growth and flowering [42,45]. One model for T6P regulatory mechanism was proposed to involve its inhibition of the sucrose non-fermenting kinase 1 (SnRK1) [46]. The concentration of T6P was strongly correlated with the concentration of sucrose in plants [42]. A high concentration of T6P was found to correlate with a low concentration of sucrose in plant cells and *vice versa*. In addition, T6P inhibited SnRK1 with a $K_i$ around 5 μM [47], within the *in vivo* concentration range of T6P [48]. Therefore, the model proposes high concentration of T6P inhibits SnRK1, further repressing growth processes. This inhibition is activated by low sucrose concentrations and alleviated by high sucrose concentrations [46]. The significance of T6P concentration in regulating plant growth leads to scientists to engineer transgenic maize with overexpressed *tps2* as an effort to reduce intracellular T6P concentration. Interestingly, these transgenic maize showed phenotypes with increased yields [49]. Therefore, fine tuning of trehalose biosynthetic pathway could be exploited to enhance agricultural yields. These applications of
trehalose biosynthesis in plants will significantly impact multiple endeavors of human activities.

In summary, the protective function of trehalose is a key contributor to the stress response, particularly in severe stress conditions such as desiccation. The metabolic function of trehalose biosynthesis is also essential for maintaining cellular integrity and complements the protective function of trehalose. Fungal cellular stress response depends on both functions of the trehalose biosynthetic pathway. More research is required to gain a deeper understanding of trehalose biosynthetic pathways.

1.4 Significance of trehalose biosynthesis in fungal virulence

To date, at least five different pathways have been identified for the biosynthesis of trehalose, primarily in bacteria and archaeal systems (Fig. 2). The most predominant pathway in eukaryotic organisms is the TPS pathway, which involves the intermediate T6P. Another pathway present in eukaryotic organisms [50] (for instance, the edible mushroom *Agaricus bisporus*) is the TreP pathway, which convert glucose-1-phosphate and glucose to trehalose and inorganic phosphate. This reaction is catalyzed by trehalose phosphorylase (TreP). In prokaryotic systems (*Mycobacterium smegmatis*), trehalose synthase (TreS) catalyzes a reversible conversion of maltose to trehalose [51]. Trehalose glycosyltransferring synthase (TreT) from hyperthermophilic bacteria and archaea condenses ADP-glucose and glucose to form trehalose and ADP [52]. Last, the TreY-
TreZ pathway is commercially utilized for large scale trehalose production [33]. Emerging biochemical and structural data have been available for the enzymes involved in these trehalose biosynthesis pathways. A list of current available structures in Protein Data Bank (PDB) is shown in Table 1. Due to the length of this dissertation, detailed discussions of these enzymes are omitted.

In pathogenic fungi and plants, the TPS pathway (Fig. 1), first characterized by E. Cabib and L. F. Leloir [53,54], is the only pathway for trehalose biosynthesis (trehalose biosynthesis in subsequent chapter refer only to this pathway). The first step of TPS pathway, which is catalyzed by trehalose-6-phosphate synthase (Tps1), is the condensation of UDP-glucose and glucose-6-phosphate (G6P) to form trehalose-6-phosphate (T6P). The second step, which is catalyzed by trehalose-6-phosphate phosphatase (Tps2), is the de-phosphorylation of T6P to yield trehalose.

Disruption of tps1 in C. albicans inhibited growth at 42 °C and resulted in impaired the yeast-to-hyphae transition at 37 °C [55]. In addition, electron microscopy studies showed loss of materials of the outer cell wall in a C. albicans tps1Δ mutant. The C. albicans tps1Δ mutant is also more prone to macrophage killing compared with wild type cells [56] and consequently show diminished virulence. Disruption of tps2 in C. albicans showed increased sensitivity to heat shock and oxidative stress [57]. Similarly, growth defects at 37 °C were observed for tps1 and tps2 null mutants in both C.
*neoformans* [58] and *C. gattii* [59]. Furthermore, these null mutants also showed decreased virulence in rabbit models.

Interestingly, two copies of trehalose-6-phosphate synthase have been identified in *A. fumigatus* (annotated as Tps1A and Tps1B in this dissertation for consistency instead of TpsA and TpsB in some publications). Real time reverse transcription PCR (RT-PCR) revealed transcription of Tps1A and Tps1B mRNAs were increased with trehalose accumulation. Deletion of both genes is required to abolish trehalose accumulation and *A. fumigatus* thermotolerance. However, compared with *C. albicans* and *C. neoformans*, a tps1ABΔ mutant was hypervirulent instead of avirulent in murine models [60]. On the other hand, *A. fumigatus tps2Δ* showed defects in cell wall integrity and impaired virulence in invasive aspergillosis murine models [61]. Despite the contradictory virulence phenotypes of *tps1* and *tps2* deletion mutant, this pathway could still be targeted for intervention of *A. fumigatus* virulence either by inhibition of Tps2 or tuning Tps1 activity.

Clearly, trehalose biosynthesis is an integral pathway in pathogenic fungal survival under multiple stress conditions. The absence of this pathway in mammalian cells indicates that Tps1 and/or Tps2 are ideal targets for antifungal drug intervention. However, the lack of structural and biochemical characterizations of fungal Tps1 and Tps2 hinders our fuller understanding of their function and delays inhibitor design.
1.5 Outline of research

This dissertation proposes to utilize macromolecular crystallography and biochemical assays to deepen our understandings of Tps1, Tps2, Tps3 and overall regulatory function of this pathway in fungal stress response. These analyses focus on proteins from three major pathogenic fungal species C. albicans, A. fumigatus and C. neoformans.

1.5.1 Glycosyltransferases and Tps1

Tps1 is a nucleotide-sugar-dependent glycosyltransferase, which transfers glucose from UDP-glucose to G6P to synthesize trehalose-6-phosphate. Nucleotide-sugar-dependent glycosyltransferases are first discovered by Luis F. Leloir and catalyze glycosidic bond formation utilizing nucleotide-sugar donors [62]. Two different folds have been identified for nucleotide-sugar-dependent glycosyltransferases, namely the GT-A and GT-B folds (Fig. 3)[62]. Both the GT-A fold [63] and GT-B fold [64] glycosyltransferases contain two Rossmann-fold domains with the difference between GT-A and GT-B lying with the relative position of these two domains. The two Rossmann-fold domains in the GT-B fold face each other and form the active site within the interface created by them. Glycosyltransferases are further divided into different families based on amino acid sequence similarities. To date (10/28/2015), according to Carbohydrate-Active enZymes database (CAZy), 97 glycosyltransferase families have
been identified [65,66]. The Tps1 proteins from *A. fumigatus*, *C. neoformans* and *C. albicans* all belong to the GT-20 family and adopt the GT-B fold.

Glycosyltransferases can also be divided into two classes depending on the stereochemical outcome of their products. Either the anomic configuration of the donor substrate is retained or inverted, which allows classification of glycosyltransferases into either retaining or inverting glycosyltransferases [62]. The anomic configuration of trehalose thus classifies Tps1 into the retaining GT-B fold glycosyltransferase. The enzymatic mechanism of retaining glycosyltransferases still remains elusive. The earlier proposed double-displacement mechanism, resulting in a covalently-bound substrate-enzyme intermediate, has gained little support from an increasing amount of macromolecular structural data [62]. More recently, evidence for an S\(_{\text{ni}}\)-type mechanism has been provided by the Benjamin G. Davis group [67], who were studying the *E. coli* Tps1 homologue, OtsA. The Davis laboratory determined the structure of OtsA in complex with UDP and validoxylamine A-6’-O-phosphate, revealing a transition state mimic [68]. This transition state mimicry has been subsequently confirmed by linear free energy relationships. Further, kinetic isotope effects suggest an oxocarbenium ion-like transition state. Understanding the enzymatic mechanism of Tps1 further should facilitate novel inhibitor design.
Despite structures of *E.coli* OstA with substrates or substrate analogs [69,70,71], no structural information is available for homologues in Achaea and Eukaryotes. Sequence alignment of pathogenic fungal Tps1s with OtsA demonstrates that Tps1 shares the conserved UDP-Glucose and G6P binding residues, indicating that the active site interactions and enzymatic mechanism may be identical to that of OtsA. However, this hypothesis requires further confirmation from the *de novo* structure determination of eukaryotic Tps1 proteins and a detailed assessment of their catalytic activity. To underscore the need for such studies *C. neoformans* and *E. coli* Tps1 have significant size differences, with the former enzyme being 671 residues and the latter 474 residues. Structural and biochemical studies on *C. neoformans* Tps1 should provide insight into any function of these extra ~200 residues and might reveal additional allosteric sites for inhibitor targeting.

As part of this thesis and in order to fill in critical gaps in our knowledge of eukaryotic Tps1 proteins, structures of Tps1 from *C. albicans* (Chapter 2) and *A. fumigatus* (Chapter 3) were determined. Further, the Tps1 protein from *C. neoformans* was successfully purified and crystallized (Chapter 6). Thus, in this dissertation I report the structures of the first eukaryotic Tps1 enzymes. Together with biochemical and cellular assays, these results shed light on the enzymatic mechanism and provide a scaffold for rational inhibitor design.
1.5.2 HADSF phosphatases and Tps2

Tps2 is a trehalose-6-phosphate specific phosphatase with only the structure of the unliganded trehalose-6-phosphate phosphatase from *Brugia malayi* currently available [72]. Sequence alignments reveal Tps2 belongs to the haloacid dehalogenase superfamily (HADSF) phosphatases. HADSF phosphatases possess a Rossmann-fold core domain, distinguished by four conserved motifs necessary for phosphoryl group transfer [73, 74]. Motif I contains the nucleophilic aspartate, which forms a covalently bound aspartate-phosphoryl intermediate during catalysis. Motifs II and III contain conserved S/T and R/K residues, which bind and neutralize the charge of the substrate/transition state phosphoryl group. Motif IV includes two or three aspartate residues and coordinates a catalytical required Mg$^{2+}$ ion in the active site. Despite the conserved core domain, HADSF phosphatases demonstrate a wide range of substrate specificity, including phosphoserines, phosphothreonines, phosphotyrosines, nucleoside phosphates and phosphosugars. Substrate specificity is achieved primarily through an additional cap domain. The size and location of the cap domain (Fig. 4) further divides HADSF phosphatases into three subfamilies [75]. Subfamily I contains an α-helical bundle domain inserted between Motif I and II [76]. Subfamily II has an α/β domain inserted between Motif II and III [77]. Superfamily III lacks this additional cap domain [78]. Based on structure predictions, Tps2 from pathogenic fungi belong to superfamily
II of HADSF phosphatases. However, without structural information on Tps2 bound to substrates or substrate analogues, the mechanism of substrate specificity of Tps2 remains obscure.

Intriguingly sequences of fungal Tps2 proteins reveal each has a “pseudo-Tps1“ N-terminal domain (Tps2NTD) and a C-terminal phosphatase domain (Tps2PD). In this dissertation, the structures of *C. albicans* Tps2NTD is determined (Chapter 4) and indeed reveals a Tps1 overall fold. In addition, the structures of the Tps2PD from *C. albicans*, *A. fumigatus* and *C. neoformans* in multiple conformational states reveal the underlying mechanism of Tps2 substrate specificity, visualize the conformational changes effected by substrate binding and provide fuller insight into the catalytic mechanism of this phosphatase (Chapter 4).

1.5.3 Tps3 structure and function

Tps3 is an essential component of the trehalose biosynthetic complex [79,80], which is significant for trehalose accumulation. Deletion of *tps3* or *tsl1* in *S. cerevisiae* reduced or abolished trehalose production [81]. Tps3 has also been found to be essential for *M. grisea* pathogenicity [40]. However, the detailed structure and function of Tps3 remains unknown.

Compared to two copies of this gene in *S. cerevisiae* (Tps3 and Tsl1), *C. albicans* has a single copy of Tps3. Sequence alignment reveals *C. albicans* Tps3 (904 amino acid
residues) has a “pseudo-Tps1” domain at its N-terminus (Tps3NTD) and interestingly a pseudo-phosphatase domain at its C-terminus (Tps3PPD). A long disordered stretch of residues of unknown function is also present at the N-terminus. The structure of C. albicans Tps3PPD (Chapter 5) was determined and revealed a binding pocket for an uncharacterized ligand that clearly cannot be trehalose-6-phosphate. This potentiates inter-connection of trehalose biosynthesis with other metabolic pathways and novel mechanisms of fungal response to other cellular stresses.

1.5.4 Trehalose biosynthetic complex and regulatory function

In S. cerevisiae, Tps1 and Tps2 together with regulatory subunits, Tps3 and Tsl1, form the trehalose biosynthetic complex [79,80]. The interactions within the trehalose biosynthetic complex have been further confirmed by comprehensive protein-protein interaction studies in S. cerevisiae [82]. However, to date, the stoichiometry and function of complex formation is not understood. Part of this mystery is attributed to the large size of this complex, estimated to be 600 to 800 kDa [79,80]. I propose to reconstitute the trehalose biosynthetic complex in vitro with individual complex components that I have obtained and attempt to characterize its catalytic function (Chapter 6). Detailed studies of this complex might reveal a previously unrealized regulatory function of this complex and thus making it a novel target for antifungal intervention.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Organism</th>
<th>PDB accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tps1</td>
<td><em>Escherichia coli</em> [68,70,71]</td>
<td>1UQT, 1UQU, 1GZ5, 2WTX</td>
</tr>
<tr>
<td>Tps2</td>
<td><em>Brugia malayi</em> [83]</td>
<td>4OFZ, 5E0O</td>
</tr>
<tr>
<td>Tps2</td>
<td><em>Thermoplasma acidophilum</em> [77]</td>
<td>1U02</td>
</tr>
<tr>
<td>Trehalase</td>
<td><em>Escherichia coli</em> [84,85]</td>
<td>2JF4, 2JG0, 2JJB, 2WYN</td>
</tr>
<tr>
<td>TreS</td>
<td><em>Mycobacterium smegmatis</em> [51]</td>
<td>3ZO9, 3ZOA</td>
</tr>
<tr>
<td>TreS</td>
<td><em>Deinococcus radiodurans</em> [86]</td>
<td>4TVU, 4WF7</td>
</tr>
<tr>
<td>TreT</td>
<td><em>Pyrococcus horikoshii</em> [52]</td>
<td>2X6Q, 2X6R, 2XA1, 2XA2, 2XA9, 2XMP</td>
</tr>
<tr>
<td>TreY</td>
<td><em>Sulfolobus acidocaldarius</em> [87]</td>
<td>1IV8</td>
</tr>
<tr>
<td>TreZ</td>
<td><em>Sulfolobus solfataricus</em> [88,89]</td>
<td>1EH9, 1EHA, 3VGB, 3VGD, 3VGE, 3VGF, 3VGG, 3VGH</td>
</tr>
<tr>
<td>TreZ</td>
<td><em>Deinococcus radiodurans</em> [90]</td>
<td>2BHU</td>
</tr>
<tr>
<td>TreA</td>
<td><em>Bacillus licheniformis</em></td>
<td>5BRP, 5BRQ (unreleased)</td>
</tr>
</tbody>
</table>
Figure 1: An enzymatic scheme of TPS pathway
1. **TPS pathway**
   \[ \text{UDP-glucose} + \text{Glucose-6-phosphate} \xrightarrow{Tps1} \text{Trehalose-6-phosphate} + \text{UDP} \xrightarrow{Tps2} \text{Trehalose} \]

2. **TreP pathway**
   \[ \text{Glucose-1-phosphate} + \text{Glucose} \xrightarrow{\text{TreP}} \text{Trehalose} \]

3. **TreY-TreZ pathway**
   \[ \text{Maltooligosaccharide} \xrightarrow{\text{TreY}} \text{Maltooligosyltrehalose} \xrightarrow{\text{TreZ}} \text{Trehalose} + \text{Glucose} \]

4. **TreS pathway**
   \[ \text{Maltose} \xrightarrow{\text{TreS}} \text{Trehalose} \]

5. **TreT pathway**
   \[ \text{ADP-glucose} + \text{Glucose} \xrightarrow{\text{TreT}} \text{Trehalose} + \text{ADP} \]

**Figure 2: Trehalose biosynthesis pathways**
Figure 3: Overall folds of the GT-A and GT-B glycosyltransferases

(a) Cartoon diagram of a GT-A fold glycosyltransferase (PDB: 1QGQ). The N-terminal and C-terminal Rossmann-fold domains are depicted as ribbons and colored cyan and pink. (b) Cartoon diagram of GT-B fold glycosyltransferase (PDB: 1JG7). The N-terminal and C-terminal Rossmann-fold domains are depicted as ribbons and colored cyan and pink whilst the C-terminal helix that interacts with both domains is colored red. (Figure adapted from [62])
Figure 4: Overall folds of the HADSF phosphatase subfamilies

(a) Cartoon diagram of a member of the HADSF subfamily I (PDB: 1LVH). The core domain and cap domain depicted as ribbons and colored cyan and pink. (b) Cartoon diagram of a member of the HADSF subfamily II (PDB: 1U02). The core domain and cap domain are depicted as ribbons and colored cyan and pink. (c) Cartoon diagram of a member of the HADSF subfamily III (PDB: 1U7O). The core domain and cap domain are depicted as ribbons and colored cyan and pink. (Figure adapted from [74])
2. Structures of *C. albicans* Tps1 provide insight into antifungal drug design

Trehalose-6-phosphate (T6P) synthase (Tps1) catalyzes the first step of trehalose biosynthesis in fungi and serves as a potential antifungal target. Here, we report the first structures of a eukaryotic Tps1 from *Candida albicans* (*C. albicans*), in complex with substrates or substrate analogues. Tps1 adopts two Rossmann-fold domains, which form the UDP-Glucose and Glucose-6-phosphate (G6P) binding sites. Tps1 structures with substrates reveal those residues essential for catalysis. Subsequent enzymatic and cellular assays highlight the significance of essential Tps1 residues in enzyme function and fungal virulence. A Tps1 structure with a non-hydrolysable inhibitor, thereby mimicking the transition-state, demonstrates that Tps1 adopts an “internal return like” mechanism for catalysis. Interestingly, disruption of the trehalose biosynthetic complex formation through abolishing dimerization of *C. albicans* Tps1 reveals complex formation has a regulatory function in addition to trehalose production, both of which are significant for *C. albicans* stress response and virulence.

2.1 Introduction

During the past several decades, infectious diseases caused by fungi have tremendously increased in both plants and humans. [91,92]. *Candida* species represent one major pathogenic fungal species causing both mucosal and disseminated infections, among which *C. albicans* is the predominant species. Candidiasis leads to more than
400,000 cases annually worldwide and more than 50% mortality [3]. *C. albicans* is also a leading cause of nosocomial bloodstream infection in the United States [5]. The high mortality rate of candidiasis is partly attributed to limited classes of antifungal drugs, emerging cases of drug resistance in the clinics and toxic side effects of the drugs currently available. Furthermore, fungal cells have higher resemblance to mammalian cells compared to bacteria or virus, inevitably resulting in difficulty of novel target identification, thereby delaying antifungal drug development.

Regardless, fungi do have potential “nonhuman” targets and computer-aided antifungal target selection ranks the trehalose biosynthetic pathway as a top candidate for novel inhibitor/drug design [93]. Absence of this pathway in mammalian cells would likely alleviate inhibitor/drug toxicity. Trehalose is a non-reducing disaccharide with two glucose units linked by an α, α-1, 1-glycosidic linkage. Trehalose is only synthesized in bacteria, fungi, lower plants and invertebrates [23]. Fungal cells synthesize trehalose to protect proteins and membranes from external and internal stresses, including dehydration, heat shock and oxidation[29,94,95,96,97]. A single two-step trehalose biosynthetic pathway has been identified in pathogenic fungi (Fig. 1). The first step, which is catalyzed by Tps1, is the transfer of glucose from UDP-glucose to glucose-6-phosphate (G6P) to form trehalose-6-phosphate (T6P). The second step, which is carried out by the trehalose-6-phosphate phosphatase (Tps2), is the dephosphorylation of T6P to produce trehalose.
Disruption of either *tps1* or *tps2* decreases virulence in *C. albicans* [55,57,98]. Similar phenotypes are observed in pathogenic fungi including *Cryptococcus* [58,59] and *Aspergillus* [61] as well. To rationally design inhibitors against these enzymes, biochemical and structural characterization of Tps1 and Tps2 are required. However, the structure of only one homologue of Tps1, the *E.coli*, OtsA protein, has been determined [69,70,71] while no structural information is available for homologues in Achaea and Eukaryotes. Here, we describe the first eukaryotic Tps1 structures from *C. albicans* bound to substrates or substrate analogues. These structures, together with enzymatic activity assays and cell survival assays, identified key residues for substrate binding, shed light on the enzymatic mechanism and provided a structural scaffold for inhibitor design.

### 2.2 Results and discussion

#### 2.2.1 Structure of *C. albicans* Tps1

The structure of *C. albicans* Tps1 bound to UDP-Glucose was determined to 1.9 Å resolution by molecular replacement using *E. coli* OtsA (PDB: 1UQU) as the search model. The structure of Tps1 in complex with UDP and G6P was determined by molecular replacement using the Tps1 UDP-Glucose structure as the search model. Both structures take the P6₁ space group with two subunits in the asymmetric unit. Selected data collection and refinement statistics are listed in Table 2.
Tps1 adopts a typical GT-B fold for retaining glycosyltransferases with two modified Rossmann-fold domains [62]. The N-terminal Rossmann-fold domain contains a core six stranded parallel β sheet that is flanked by eight α helices (Fig. 5 and 6).

Different from a typical Rossmann-fold domain, β1 links to antiparallel β strands β2 and β3 instead of an α helix. In addition, antiparallel β strand β9 to β11 link the N-terminal Rossmann-fold domain to the C-terminal domain. The C-terminal domain adopts a β/α/β fold with a core of six parallel β strands flanked by eight α helices. A C-terminal helix at the end of the C-terminal domain extends into the N-terminal domain and interacts with structural elements of both domains. A kink around residue Y457 disrupts the integrity of this helix, which is characteristic of GT-B fold glycosyltransferases.

Electron density for UDP-Glucose and G6P is detected in the C-terminal and N-terminal domain, respectively. Both binding pockets are close to the interface, conferring that the interface between these two domains is the active site. Structure superposition of the Tps1-UDP-Glucose and Tps1-G6P-UDP complex structures reveals no detectable conformational change, indicating that UDP-Glucose induces Tps1 to adopt a closed conformation. Interestingly, no apoTps1 from any species has been crystallized suggesting that this form of the protein is highly flexible.

As expected a DALI search of Tps1-UDP-Glucose structure [99] reveals the structure with the highest structural homology is OtsA from E. coli with root mean square deviation (rmsd) of 1.8 Å for 452 corresponding Cα atoms. Further, the
structural superposition using the DALI pairwise comparison server [100] reveals a highly conserved active site. The most dramatic differences lie in the N-terminal domain of Tps1 in which the antiparallel β strands β2 and β3 are not present in OtsA (Fig. 5b).

2.2.2 Binding pocket of C. albicans Tps1

Each Rossmann-fold domain of Tps1 hosts one substrate binding site. UDP-Glucose binds to the C-terminal domain with the uracil base interacting directly with Tps1 only through a hydrogen bond between the exocyclic O4 atom and the backbone amide group of residue I357 backbone amide (Fig. 7a). The exocyclic O2 atom is interacting with a water molecule in the active site. Although there are not extensive interactions between uracil and Tps1, interaction between the uracil O4 with Tps1 discriminates UDP from CDP, as the CDP N4 amide group would not be able to form a hydrogen bond with the amide nitrogen of the I357 backbone. In addition, the size of the UDP-Glucose binding pocket would favor pyrimidine over a bulkier purine. Taken together, UDP-Glucose is the preferred substrate for Tps1. Adding to the substrate recognition and binding is the ribose ring of UDP as the O2 and O3 hydroxyl groups form hydrogen bonds with the side chain of residue E387. With regard to the phosphate moiety of UDP, one hydrogen bond is made between its α phosphate group and the backbone amide group of residue L383. A conserved Arg/Lys pair (residues R280 and K285) forms hydrogen bonds with the β phosphate. These interactions are significant for...
positioning the UDP-Glucose in the correct conformation for activity and the subsequent stabilization of the UDP leaving group. No cation is detected in the electron density, suggesting that retaining GT-B fold glycosyltransferases utilize a metal-independent catalytic mechanism.

Although the electron density is poor for the glucose moiety of the UDP-Glucose substrate, all of its hydroxyl groups are detected in electron density maps (Fig. 8). These hydroxyl groups participate in hydrogen bonds to Tps1 with the O2 hydroxyl hydrogen bonding to a water molecule, the O3 hydroxyl hydrogen bonding to the backbone of residues M381 and the side chain of residue D379, the O4 hydroxyl interacting with the backbone of residue N382 and the O6 hydroxyl group hydrogen bonding to the side chain of residue H171. These interactions also highlight the significance of the glucose unit conformation for enzymatic activity. However, without G6P bound, the other substrate, the glucose moiety of the UDP-Glucose likely adopts slightly different conformations, thereby contributing to the relatively poor electron density observed for this sugar. Finally, the Tps1 structure in complex with UDP-Glucose alone suggests Tps1 uses an ordered binding mechanism with UDP-Glucose binding first. No Tps1-G6P complexes have been crystallized despite multiple attempts.

In order to locate the G6P binding site, the structure of Tps1 complexed to UDP and G6P was solved. The most notable interactions between G6P and Tps1 are between the phosphate moiety and sidechains of residues Y89 and R318 (Fig. 7b). Another major
contributor is residue D143, the side chain of which interacts with the O-3 and O-4 hydroxyl groups. The Tps1 structure in complex with UDP and G6P also shows relatively weak electron density for the bound G6P glucose moiety (Fig. 8). This weaker density is probably due to the use of UDP as a substitute for UDP-Glucose hence allowing some additional flexibility in the binding pocket. Importantly, the structures of Tps1 bound to UDP-Glucose and bound to UDP and G6P indicates that the glucose moieties of the UDP-Glucose and G6P would be positioned with the appropriate stereochemistry to allow efficient catalysis.

To assess the universality of the substrate binding and catalytic mechanisms of Tps1 proteins, the sequence of C. albicans Tps1 was aligned (Fig. 6) with those from E. coli OtsA, Cryptococcus neoformans (C. neoformans) Tps1 and Aspergillus fumigatus (A. fumigatus) Tps1. Residues in substrate binding are conserved amongst these species except for I357, the side chain of which is not involved in UDP recognition. This conservation of all residues strongly supports the idea that all Tps1 proteins adopt the same binding pocket for both substrates and potentially utilize identical catalytic mechanisms. Intriguingly, these Tps1 homologues are of similar molecular weight with the exception of the Tps1 enzyme from C. neoformans, which has approximately 200 residues larger. These extra residues might compose additional domains in the C. neoformans Tps1. However, their structures and any possible function remain enigmatic.
2.2.3 C. albicans Tps1 mutant enzymatic activity

To assess the significance of the conserved substrate binding residues, four structure-guided mutations were made to Tps1. Coupled enzyme assays utilizing NADH were performed on Tps1 mutants Y89F, K285A, D379A and E387A (Fig. 9). Compared with the wild type protein, no enzymatic activity was detected for these mutants. Thus, the integrity of each substrate-binding residue is essential for Tps1 activity. The importance of each of these residues to the cellular fitness of both C. albicans is underway by their introduction into the chromosome and their subsequent ability to grow under heat stress.

2.2.4 Elucidation of the catalytic mechanism of Tps1

Tps1 is one member of the retaining GT-B fold glycosyltransferases, the catalytic mechanism of which remains unclear. Our Tps1 structures bound to either UDP-Glucose or UDP and G6P show poor electron density for glucose moieties of UDP-Glucose and G6P. The lack of accurate structural information on the stereochemistry of UDP-Glucose and G6P binding in the active site hinders our fullest understandings of the enzymatic mechanism of Tps1. In an attempt to overcome this obstacle, Tps1 has been crystallized with a transition-state inhibitor, validoxylamine A (VDM), plus UDP and the structure of this complex determined at 1.8 Å resolution. VDM represents a non-hydrolysable inhibitor with a structural scaffold that resembles trehalose. Good electron density is observed for the VDM backbone (Fig. 10a). Each of the VDM hydroxyl groups engage in
extensive interactions with Tps1 and solvent (Fig. 10b). Not surprisingly, the VDM hydroxyl groups share identical interactions to Tps1, when compared to its natural substrates. Interestingly, the absence of a phosphate moiety in VDM results in a water-mediated interaction between its 7’ hydroxyl group and the side chains of residues Y89 and R318. Furthermore, superposition three Tps1 protein structures (Tps1-UDP-Glucose, Tps1-UDP-G6P and Tps1-UDP-VDM) clearly reveal the retaining of anomeric configuration upon catalysis (Fig. 10d).

Two different enzymatic mechanisms, the double displacement mechanism and internal return mechanism, have been proposed for retaining glycosyltransferase [62]. The double displacement mechanism requires an active site residue to serve as nucleophile. However, none of Tps1 structures shows any nucleophile positioned in proximity to the substrate. Therefore, an internal return mechanism is favored. In the Tps1 structure bound to UDP and VDM, the β phosphate of UDP is 2.7 Å from the pseudo-glycosidic bond of VDM (Fig. 10c). This interaction and orientation favor a front-side, S\textsubscript{N}-i mechanism, which has been proposed for OtsA, the \textit{E. coli} homologue of Tps1 [67,68]. Therefore, both eukaryotic and prokaryotic Tps1 utilize an internal return like mechanism for catalysis. Clarification of the catalytic mechanism of the fungal Tps1 proteins by the structure determination of the Tps1-UDP-VDM complex provides insight into inhibitor design against Tps1 and deepens our understandings of the retaining glycosyltransferases in general.
2.2.5 Significance of the formation of trehalose biosynthetic complex

Tps1 adopts a dimer of dimers oligomeric state due to crystal packing with two different protomer-protomer interfaces (Fig 11). PDBePISA [101] reveals that the ΔG for each interface is -6.3 and 1.4 kcal/mol, respectively. This dramatic difference in ΔG indicates that C. albicans Tps1 tetramer formation would only occur at extremely high concentrations as found in the crystal. Indeed, Tps1 forms a dimer in solution as size exclusion chromatography (SEC) results demonstrate Tps1 is a dimer at concentrations 1-2 mg/mL (Fig. 11). The finding that Tps1 is a homotetramer in the crystal might have relevance on its function in vivo as Tps1 is one component of the hetero-oligomeric trehalose biosynthetic complex in other fungi. For instance, in S. cerevisiae, the trehalose biosynthetic complex complex is formed by Tps1, Tps2, Trehalose-6-phosphate regulatory protein (Tps3) and Trehalose synthase long chain 1 (Tsl1), with Tps3 and Tsl1 being interchangeable [79]. Research has also shown Tps1 adopts an oligomeric state within trehalose biosynthetic complex [79]. Therefore, the “stronger” Tps1 dimer that is observed in our crystal structure could well serve as a potential scaffold for recruitment of Tps2 and other proteins to allow formation of an analogous trehalose biosynthetic complex in C. albicans.

To dissect the significance of any trehalose biosynthetic complex formation in C. albicans, mutations were designed to disrupt the Tps1 stronger dimer interface. Key interactions mapped to this interface include E346-R261, E341-R328 and H354 ionic
contacts (Fig. 11b). Two sets of mutation were designed, the E341R/E346R (ER) and R261E/R328E (RE), with the charge swapped. Both mutations result in monomeric form of Tps1 in solution as determined via SEC (Fig. 11c). To further confirm the disruption of dimer interface, the structure of Tps1(ER) was determined in complex with UDP-Glucose. Indeed, the structure clearly reveals the disruption of hydrogen bonds within the dimer interface (Fig. 11c). Considering the dimerization interface is distant from the Tps1 active site, these mutations would appear to be unlikely to affect the enzymatic activity of Tps1. However, only the Tps1(ER) protein retained the majority of its catalytic activity whilst the Tps1(RE) protein is catalytically “dead” (Fig. 9). Regardless, the Tps1(ER) mutant demonstrates a form of catalytic active monomeric Tps1.

To assess the significance of this pair of mutations, further dissecting the significance of the formation of the trehalose biosynthetic complex, Tps1(ER) mutations are introduced to C. albicans through chromosome recombination. Phenotypes of these mutations are being undertaken in the Perfect lab at Duke University Medical Center.

Potential function of the trehalose biosynthetic complex can involve bringing Tps1 and Tps2 in proximity to each other and increasing trehalose biosynthesis efficiency. This rapid synthesis of trehalose is significant for response to stress conditions, enabling fungal cells to adapt to changing environment. In addition, formation of trehalose biosynthetic complex potentially traps T6P within the complex, preventing T6P from diffusing freely into cytosol [102]. Tight control of trehalose
metabolism regulates glucose flux into glycolysis [36] and rescues cells with unbalanced glycolysis [38]. Detailed research is still required to understand the trehalose biosynthetic complex, which may reveal novel interfacial antifungal inhibitor sites for intervention.

### 2.3 Experimental procedures

#### 2.3.1 Protein expression and crystallization

The Tps1 gene from *C. albicans* strain SC5314 was codon optimized for *E. coli* expression system (GenScript). This codon-optimized gene was cloned into a pET-28a kanamycin-resistant vector via restriction sites NdeI and SacI. This vector contained an N-terminal hexahistidine affinity tag followed by a thrombin cleavage site. It was transformed into BL21(DE3)plysS cells (Life Technologies), which were grown at 37 °C to an OD$_{600}$ of 0.6, followed by induction with 0.5 mM IPTG and continued incubation at 15 °C for 16 h. The Tps1 protein was purified by Ni$^{2+}$-NTA affinity chromatography followed by thrombin cleavage to remove the hexahistidine tag. Tps1 was further purified using size exclusion chromatography via a Superdex S200 column (GE Healthcare) in a buffer containing 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol, 5 mM MgCl$_2$, and 1 mM βME. Purified Tps1 was concentrated using an Amicon Ultra concentrator (30K MWCO, Millipore). To prepare single/double residue mutations of Tps1, a standard site directed mutagenesis protocol was utilized [103]. *C. albicans* Tps1 was concentrated to 20 mg/mL and crystallized with a series of different substrates or
substrate analogues. Crystals were grown at 25 °C using the hanging drop-vapor
diffusion method. Protein crystals appeared after one week from crystallization
solutions containing 0.2 M lithium sulfate, 0.1 M Tris pH 8.5 and 40% PEG 400.

2.3.2 X-ray data collection and structure determination

X-ray intensity data sets of the *C. albicans* Tps1-UDP-Glucose, Tps1-UDP-G6P
and Tps1-UDP-VDM complexes were collected at the Advanced Photon Source (APS)
ID-22 line. Data for the *C. albicans* Tps1(ER) mutant bound to UDP-Glucose were
collected at the Advanced Light Source (ALS) using beamline 5.0.1. Intensity data were
indexed and scaled using HKL 3000 [104] and HKL 2000[105], respectively. The
structure of the *C. albicans* Tps1-UDP-Glucose complex was determined by molecular
replacement (Phaser) [106] using the OtsA structure (PDB code: 1UQU) [70] as a search
model. Other structures were determined by molecular replacement using the *C.
albicans* Tps1 structure bound to UDP-Glucose with the UDP-Glucose removed as the
search model. All structures were improved by multiple rounds of refinement in Phenix
[107] and manual rebuilding in Coot [108]. Protein structure validation was performed
by MolProbity [109]. Selected of data collection and refinement statistics are presented at
Table 2.

2.3.3 Enzyme activity assays

Tps1 activity was measured utilizing a continuous enzyme coupled assay as
previously reported [68]. Briefly, a buffer comprised of 50 mM HEPES, pH 7.8, 100 mM
KCl, 5 mM MgCl$_2$, and 2 mM DTT was used for protein concentration and the coupled enzyme assays. A final concentration of 3 μM, 1 mM and 1mM were utilized for Tps1, UDP-Glucose and G6P, respectively. Activity assays were performed in Greiner 96-well plates and measured by a decrease in absorbance at A$_{340}$ nm that was recorded using a SpectraMax M5 (Molecular Devices).

2.4 Acknowledgements

Validoxylamine A (VDM) was kindly synthesized by Dr. Jiuyu Liu from the laboratory of Dr. Richard E. Lee at St. Jude Children’s Research Hospital. Cellular assays in C. neoformans and C. albicans are being undertaken by the Perfect laboratory at Duke University Medical Center.
Table 2: Selected *C. albicans* Tps1 data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>C. albicans Tps1</th>
<th>C. albicans Tps1</th>
<th>C. albicans Tps1</th>
<th>C. albicans Tps1 ER mutations +UDP-Glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UDP-Glucose</td>
<td>+UDP+G6P</td>
<td>+UDP+VDM</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P6₄</td>
<td>P6₄</td>
<td>P6₄,22</td>
<td>P6₄</td>
</tr>
<tr>
<td>Unit cell a, b, c (Å)</td>
<td>98.7, 98.7, 187.7</td>
<td>98.7, 98.7, 188.9</td>
<td>115.6, 115.6, 282.7</td>
<td>98.6, 98.6, 186.6</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90,120</td>
<td>90, 90,120</td>
<td>90, 90,120</td>
<td>90, 90,120</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-1.90 (1.93-1.90)</td>
<td>50.0-2.37 (2.41-2.37)</td>
<td>50.0-1.80 (1.83-1.80)</td>
<td>50.0-2.0 (2.03-2.0)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.97949</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.9)</td>
<td>99.6 (97.3)</td>
<td>99.6 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.8 (5.6)</td>
<td>5.7 (5.6)</td>
<td>10.3 (10.7)</td>
<td>7.5 (7.5)</td>
</tr>
<tr>
<td>I/Iσ</td>
<td>21.5 (2.5)</td>
<td>18.5 (2.0)</td>
<td>51.6 (4.9)</td>
<td>31.5 (4.9)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.07 (0.672)</td>
<td>0.075 (0.638)</td>
<td>0.092 (0.666)</td>
<td>0.091 (0.443)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rwork [%]</td>
<td>17.9</td>
<td>20.0</td>
<td>17.9</td>
<td>14.5</td>
</tr>
<tr>
<td>Rfree [%]</td>
<td>20.0</td>
<td>24.7</td>
<td>20.6</td>
<td>17.1</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds [Å]</td>
<td>0.008</td>
<td>0.003</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Angles [°]</td>
<td>1.18</td>
<td>0.679</td>
<td>1.113</td>
<td>1.107</td>
</tr>
<tr>
<td>Overall B factors (Å²)</td>
<td>35.9</td>
<td>62.5</td>
<td>32.0</td>
<td>26.9</td>
</tr>
<tr>
<td>Ramachandran [%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored, outliers</td>
<td>97.4, 0.3</td>
<td>95.9, 0.4</td>
<td>97.2, 0.2</td>
<td>97.6, 0.3</td>
</tr>
</tbody>
</table>

* statistics for the highest resolution shell are shown in parentheses.

* \( R_{merge} = \frac{\sum |I| - \langle |I| \rangle}{\sum |I|} \), where |I| is the observed intensity and \( \langle |I| \rangle \) is the average intensity of several symmetry-related observations.

* \( R_{work} = \frac{\sum|F_o| - |F_c|}{\sum|F_o|} \), where \( F_o \) and \( F_c \) are the observed and calculated structure factors, respectively.

* \( R_{free} = \frac{\sum|F_o| - |F_c|}{\sum|F_o|} \) for 5% of the data not used at any stage of the structural refinement.
Figure 5: Overall structure of *C. albicans* Tps1

(a) Cartoon diagram of *C. albicans* Tps1. The N-terminal domain is colored cyan and the C-terminal domain is colored pink. The C-terminal helix interacting with both domains is colored in red. The extra antiparallel β-sheets (β2-β3) and parallel β sheets (β9-β11) linking N-terminal and C-terminal domain are colored orange. (b) Superposition of *C. albicans* Tps1 and *E. coli* OtsA, which is colored grey. Differences between the N-termini of the two proteins are circled (red) and the parallel β sheets linking the two domains are also highlighted.
Figure 6: Sequence alignment of C. albicans Tps1

Tps1 sequences of C. albicans (Q92410), Cryptococcus neoformans (Q6IVK9), Aspergillus fumigatus (A0A084BNL5) and E. coli (P31677) are aligned. The sequences are annotated as Tps1_CA, Tps1_CN, Tps1_AF and Tps1_EC, respectively. The secondary structure of C. albicans Tps1 is also shown in the figure whereby α helices are depicted by arrow heads and β strands by rectangles. The color scheme of secondary structure is consistent with overall structure shown in Fig 5. Key residues are highlighted by red asterisks in the figure.
Figure 7: Active site of *C. albicans* Tps1

(a) View of residues and solvent involved in UDP-Glucose binding. UDP-Glucose is shown as atom-colored yellow sticks. Residues of Tps1 protein from N-terminal domain and C-terminal domain are shown as atom-colored cyan and pink sticks. Hydrogen bonds are shown by dashed lines. (b) View of residues and solvent involved in G6P binding. G6P is shown as atom-colored yellow sticks.
Figure 8: Electron density for substrates

(a, b) 2Fo-Fc electron density maps of the bound substrates of the described Tps1 structures. The electron density is shown as light blue mesh and contoured at 1.5 σ. Substrates are shown as atom-colored sticks.
Figure 9: Relative enzyme activity of *C. albicans* Tps1 and selected mutants
Figure 10: Tps1-VDM interactions

(a) A 2Fo-Fc electron density map of the Tps1-bound VDM, shown as light blue mesh and contoured at 1.5 σ. VDM is shown as atom-colored yellow sticks. The oxygen atoms are labelled where primed oxygens are found on the glucose moiety closest to N-terminal domain. (b) View of interactions involved in VDM binding. Residues of Tps1 protein from N-terminal domain and C-terminal domain are shown as atom-colored cyan and pink sticks. Water molecules are shown as red spheres. Hydrogen bonds are shown by dashed line. (c) Stick representation of VDM and UDP in a 2Fo-Fc electron density map contoured at 1.5 σ. (d) Superposition of UDP-Glucose from the Tps1 UDP-Glucose structure and G6P from Tps1-UDP-G6P structure onto UDP and VDM from the Tps1-UDP-VDM complex structure. UDP and VDM are shown by atom-colored yellow sticks. UDP-Glucose and G6P are shown by atom-colored grey sticks.
Figure 11: *C. albicans* Tps1 dimer interactions

(a) Cartoon diagram of the Tps1 tetramer due to crystal packing. Two subunits in the ASU are colored in wheat and light cyan, respectively. The tetramer is generated by crystallographic symmetry. The Tps1 dimer interface with the calculated stronger interactions is boxed in red. (b) Detailed interactions of the Tps1 dimer interface. The side chains of the interacting residues are atom-colored. Hydrogen bonds and ionic interactions between the two subunits are shown by dashed lines. The interface is formed between subunit B and A’, a symmetry generated mate. Residues of the symmetry mate are distinguished by a prime symbol. (c) Tps1 size exclusion chromatography using an S200 column (GE Healthcare). (d) Superimposition of the Tps1 ER mutant structure and wild type Tps1 structure. The Tps1ER mutant structure is shown as a cartoon and colored grey. The side chains of the Tps1ER mutant residues are shown by atom-colored grey sticks. Disruptions of key hydrogen bonds are highlighted in the figure using a red oval.
3. Structures of trehalose-6-phosphate synthase from *Aspergillus fumigatus* reveal the molecular mechanism of glucose transfer

Trehalose biosynthesis is an integral pathway in *Aspergillus fumigatus* (*A. fumigatus*) that is required for proper stress response and virulence. Trehalose is synthesized through the intermediate, trehalose-6-phosphate (T6P), by T6P synthase (Tps1) and T6P phosphatase (Tps2). Here, we report the transition-state structures of *A. fumigatus* Tps1A and Tps1B in complex with UDP and validoxylamine A. These structures reveal key residues involved in substrate binding and catalysis, which are essential for Tps1 enzymatic activity and trehalose accumulation in *A. fumigatus*. The stereochemistry of the transition-state active site also suggests that Tps1 adopts an internal-return mechanism for catalysis. Furthermore, the structures and structure-guided activity assays highlight key differences between the two Tps1 proteins in *A. fumigatus* and imply overlapping but different cellular functions.

3.1 *Introduction*

Invasive aspergillosis, primarily caused by *Aspergillus fumigatus*, is one of the most deadly infections (mortality rate higher than 50%) in immunocompromised patients. In addition, chronic pulmonary aspergillosis affects more than three million patients worldwide, especially patients with underlying lung diseases [3]. Limited therapeutic options [7] and high morbidity and mortality associated with *Aspergillus* infection require the identification of novel antifungal fungal targets for therapeutic
intervention. Trehalose biosynthesis, catalyzed by Trehalose-6-phosphate synthase (Tps1) and Trehalose-6-phosphate phosphatase (Tps2), has been ranked as one top novel antifungal target [93]. Deletion of tps1 or tps2 gene in pathogenic fungi, including Cryptococcus, Candida and Aspergillus, has been shown to decrease growth rate and diminish virulence [56,57,58,59,60,61]. In contrast to other pathogenic fungal species, deletion of tps1 genes in A. fumigatus has shown to enhance virulence [60]. Interestingly and different from Cryptococcus and Candida, A. fumigatus contains two copies of trehalose-6-phosphate synthase (annotated as Tps1A and Tps1B in this dissertation). However, detailed analyses on Tps1A and Tps1B are still lacking, hindering our understanding of the structures, enzymatic mechanisms and physiological functions of these two proteins.

Here we report the structures of both copies of Tps1 (Tps1A and Tps1B) in complex with UDP and VDM. These structures reveal key residues in enzymatic catalysis, shed additional light on the enzymatic mechanism of retaining glycosyltransferases, and highlight major differences between the two Tps1 proteins in A. fumigatus and suggest potential functional variations within cells.

### 3.2 Results and Discussion

#### 3.2.1 Structure of A. fumigatus Tps1A

The Tps1A structure was determined to 2.8 Å resolution by molecular replacement using the structure of C. albicans Tps1 as the search model. Selected
crystallographic data and refinement statistics are listed in Table 3. The structure was determined in the orthorhombic space group P2\(\_1\)2\(\_1\) with two molecules in the asymmetric unit (ASU). Electron density was not visible for residues 1-11, 35-39, 56-63, 68-70, and 480-515 (chain B). The presence of a disordered C-terminus (residue 480-515) is consistent with the secondary structure prediction by PSIPRED [110]. The overall structure of the Tps1A is similar to that of other enzymes belonging to the GT-B class of retaining glycosyltransferases [62] and contains N and C-terminal Rossmann-fold domains and a C-terminal helix that crosses and interacts with each Rossmann-fold domain (Fig. 12). The N-terminal and C-terminal Rossmann-fold domains bind to VDM and UDP, respectively. Structure superposition of A. fumigatus Tps1A with C. albicans Tps1 reveal a root mean square deviation (rmsd) of 0.48 Å for 452 corresponding Cα atoms.

### 3.2.2 Structure of A. fumigatus Tps1B

The Tps1B structure was determined to 2.46 Å resolution in P2\(\_1\)2\(\_1\) space group by molecular replacement using the structure of C. albicans Tps1 (Chapter 2) as a search model. Selected crystallographic data and refinement statistics are listed in Table 1. Electron density was clear for Tps1B protein with only 20 residues (residues 1-11, 29-30, 38-43, 479 from chain A) missing from the structure. The overall structure of the Tps1B is similar Tps1A [62] and contains two Rossmann-fold domains and a C-terminal helix.
(Fig. 13). There were four molecules of Tps1B in the ASU, each of which binds with one molecule of UDP and VDM.

The N-terminal Rossmann-fold of Tps1B is composed of six parallel $\beta$ sheets flanked by seven $\alpha$ helices in an $\alpha/\beta/\alpha$ fold. However, two pairs of additional antiparallel $\beta$ hairpins ($\beta2-\beta3$, $\beta9-\beta10$) are also present in the N-terminal domain. The first pair of antiparallel $\beta$ sheets is present at the N-terminus whilst $\beta9$ and $\beta10$ are located at the C-terminus, together with a downstream loop that links N and C-terminal Rossmann-fold domains. The C-terminal Rossmann fold was composed of core six parallel $\beta$ strands flanked by seven $\alpha$ helices in an $\alpha/\beta/\alpha$ arrangement. A C-terminal $\alpha$ helix ($\alpha14$) downstream of the C-terminal Rossmann-fold domain spans and interacts with both domains. This $\alpha14$ has a kink around Y461, which breaks the integrity of $\alpha14$ around the interface created by N and C-terminal Rossmann-fold domains.

Superpositions of A. fumigatus Tps1B with E. coli Tps1 homolog (OtsA), A. fumigatus Tps1A and C. albicans Tps1 reveals an rmsd of 0.83, 0.52 and 0.41 Å for 452, 452 and 459 corresponding C$\alpha$ atoms, respectively (Fig. 13). Thus, trehalose-6-phosphate synthases adopt highly conserved folds across both bacterial and fungal species.

**3.2.3 Substrate binding to Tps1A and Tps1B**

Tps1A and Tps1B were both crystallized with UDP and VDM, which allows them to be captured in their transition-states [67,68]. The electron density of these ligands was clear in the conserved catalytic pocket of both proteins. Sequence alignment
and structure superposition revealed highly conserved active site interactions between these two proteins (Fig. 14). To avoid a redundant description of the protein-ligand interactions, we focus on the higher resolution Tps1B interactions with UDP and VDM, in which solvent molecules are better resolved and hence provide a more comprehensive view of the active site.

The uracil base of UDP is positioned in the active site through only one hydrogen bond, from the O4 oxygen to the amide group of residue V364 (Fig. 15a). This contact selects uracil over cytosine. The positioning of the uracil is dependent on the interactions of the ribosyl 2 and 3 hydroxyl groups, which interact with the side chain of residue E394. Despite the seeming dearth of base-specifying contacts, the size of Tps1 substrate pocket would also favor a pyrimidine over larger purine and thus UDP-glucose is one of the preferred substrates of Tps1A and Tps1B. UDP binding is further effected by a loop between α12 and β12 that is essential for positioning the α phosphate through hydrogen bonds with the backbone amide groups of residues L390 and V391. The conserved K/R pair (K292 and R287) positions UDP for catalysis and shields the negative charge of the β phosphate leaving group. Both the α phosphate and β phosphate groups interact with a water molecule in the active site. One notable interaction is found between the β phosphate O3 atom and the VDM N-glycosidic bond, which are 2.3 Å apart. This interaction suggests that Tps1 utilizes an S_{Ni}-mechanism [67] involving the deprotonation of G6P by UDP and serving as the nucleophile.
VDM is a nonreactive molecular mimic of trehalose and T6P, the product of Tps1, in which the glycosydic bond joining the two glucose moieties of trehalose is replaced with an N-glycosydic bond. The hydroxyl groups on its two pseudo-glucose moieties are identical to glucose (Fig. 15b). The O2’ interacts with a water molecule, which is locked in the active site by hydrogen bonds with the side chain of residue H109 and the backbone amide group of residue G387. The O3’ and O4’ hydroxyl groups interact with the backbone amide groups of the loop between α12 and β12. This loop is of critical importance for positioning the α phosphate and glucose leaving groups of UDP-glucose. The O7’ hydroxyl group of VDM makes a hydrogen bond to the side chain of residue H179. Perhaps the key contributor to positioning the pseudo-glucose moiety of VDM is residue D150, which interacts with both the O2 and O3 hydroxyl groups. The O7 hydroxyl interacts with a water molecule, which takes the position of the phosphate moiety of G6P.

3.3 Future directions in understanding Tps1A and Tps1B

To date (12/05/2015), crystal structures of Tps1A and Tps1B in complex with UDP and VDM from A. fumigatus have been obtained. Tps1A and Tps1B share 78% sequence identity and the superposition of their structures revealed an rmsd of 0.52 Å for 433 corresponding Cα atoms. However, different expression levels of these proteins under stress conditions in A. fumigatus suggest potential functional differences between
these two proteins [60]. Detailed analyses are still required to understand the activity and any functional differences between Tps1A and Tps1B.

Two major differences arise during purification and structure determination of Tps1A and Tps1B, which shed light on potential functional differences. The most notable difference between Tps1A and Tps1B is the extra 34 residues at C-terminus of Tps1A, which is not visible in the structure. The other difference between Tps1A and Tps1B is the oligomeric state. Size exclusion chromatograph (SEC) revealed two peaks of Tps1A, corresponding to aggregated Tps1A and dimeric Tps1A (Fig. 16). In contrast, Tps1B eluted as a monomer in identical buffer conditions (Fig. 16). Taken these results into consideration, I propose the following experiments to gain deeper understandings of *A. fumigatus* Tps1A and Tps1B.

### 3.3.1 Activity difference between Tps1A and Tps1B

Previous research has shown different expression levels of Tps1A and Tps1B under stress conditions [60]. One potential explanation for this variation is that the optimal activities of Tps1A and Tps1B depend on environmental conditions. To test this hypothesis, Tps1A and Tps1B activity will be measured at different buffer conditions, with variations of salt, pH and temperature. Results from these experiments may be key to understand physiological functions of Tps1A and Tps1B in *A. fumigatus*. 
3.3.2 Contribution of C-terminal tail of Tps1A in activity and localization

The most notable difference between Tps1A and Tps1B is the extra C-terminal tail of Tps1A. To dissect whether the C-terminus residues affect Tps1A enzymatic activity, we generated Tps1A construct (residues 1-482, annotated as Tps1AC) with these residues deleted. Tps1AC will be purified and subject to enzymatic activity as described in Chapter 2. Activity differences of Tps1A, Tps1AC and Tps1B will reveal the contribution of Tps1A C-terminal tail. Interestingly, the C-terminal tail of Tps1A is abundant in serine and threonine, potential phosphorylation sites. Presence of post-translational modification site in this C-terminal tail suggest potential different regulatory mechanism of Tps1A. This C-terminal tail of Tps1 might suggest different localization of these two proteins. Confocal microscopy and other techniques will be employed to explore the localization of Tps1A/Tps1AC and Tps1B within cells.

3.3.3 Dissection of oligomeric state of Tps1A and Tps1B

Tps1A and Tps1B form different oligomers in solution (Fig. 16). Structure-based sequence alignment of these proteins reveals that one major difference between Tps1A and Tps1B is residue 353 (Fig. 14). In Tps1A, residue E353 is able to form hydrogen bonds and a salt bridge with residue arginine 268 of the dyadic mate. However, the presence of an aspartate in Tps1B precludes this interaction due to the shorter length of its side chain. To reveal the significance of this residue, E353 in Tps1A is mutated to aspartate and D353 in Tps1B is mutated to glutamate. SEC experiments are undergoing
to reveal the significance of this residue in oligomeric state of Tps1 proteins in *A. fumigatus*.

### 3.4 Experimental procedure

#### 3.4.1 Protein expression and crystallization

The *Tps1A* and *Tps1B* genes from *A. fumigatus* strain Af293 were codon optimized for *E. coli* expression system (GenScript). These codon-optimized genes were cloned into a pET-28a kanamycin-resistant vector using the restriction sites NdeI and SacI. This vector contained an N-terminal hexahistidine affinity tag followed by a thrombin cleavage site. The vectors were transformed into BL21(DE3)plysS cells (Life Technologies) and induced with 0.5 mM IPTG at 15 °C for 16 h. Each Tps1 protein was purified by Ni²⁺-NTA affinity column chromatography followed by thrombin cleavage. Each Tps1 was further purified using gel filtration chromatography via Superdex S200 column (GE Healthcare). The buffer was 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol, 5 mM MgCl₂, and 1 mM βME. Each purified Tps1 protein was concentrated to 12 mg/mL using an Amicon Ultra concentrator (30K MWCO, Millipore). The proteins were crystallized with 10 mM UDP and 10 mM VDM. Crystals were grown at 25 °C by the hanging drop-vapor diffusion method. The Tps1A crystals appeared after one week from solutions containing 0.1 M sodium malonate, pH 8.0, 0.1 M Tris, pH 8.0 and 30% PEG 1000. Tps1B crystals appeared after three weeks from solutions of 0.2 M lithium sulfate monohydrate, 0.1 M Tris, pH 8.5 and 20% PEG 3350.
3.4.2 Data collection and structure determination

X-ray intensity data of the *A. fumigatus* Tps1A and Tps1B crystals were collected at the Advanced Photon Source (APS) using ID-22 beamline. The data were indexed and scaled using HKL 2000 [105]. The structures were solved by molecular replacement using Phaser [106]. The search model utilized was the *C. albicans* Tps1 (PDB: 5HUT). The structures were manually built in Coot [108] and improved by multiple rounds of refinement in Phenix refine [107] and rebuilding. Protein structure validation was performed by MolProbity [109]. Selected data collection and structure refinement statistics are presented at Table 3.
Table 3: Selected *A. fumigatus* Tps1 data collection and refinement statistics

<table>
<thead>
<tr>
<th><strong>Data Collection</strong></th>
<th>A. <em>fumigatus</em> Tps1A</th>
<th>A. <em>fumigatus</em> Tps1B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P2,22,1</td>
<td>P22,2,1</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td>a, b, c (Å)</td>
<td>α, β, γ (°)</td>
</tr>
<tr>
<td></td>
<td>83.9, 101.2, 147.0</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-2.80 (2.85-2.80)</td>
<td>50.0-2.46 (2.50-2.46)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt; (%)</td>
<td>10.3 (88.2)</td>
<td>10.7 (68.2)</td>
</tr>
<tr>
<td>I/σI</td>
<td>16.3 (2.1)</td>
<td>17.3 (2.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100.0)</td>
<td>97.6 (98.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.8 (4.5)</td>
<td>5.0 (4.3)</td>
</tr>
</tbody>
</table>

| **Refinement**      |                     |                     |
| Resolution          | 43.3-2.82 (2.92-2.82) | 39.8-2.47 (2.56-2.47) |
| No. reflections     | 30852                | 74719                |
| R<sub>work</sub><sup>c</sup> / R<sub>free</sub><sup>d</sup> (%) | 17.6/23.1          | 20.0/23.6            |
| No. atoms           |                      |                      |
| Protein             | 7180                 | 14584                |
| Ligand/ion          | 96                   | 192                  |
| Water               | 10                   | 190                  |
| B-factor            |                      |                      |
| Protein             | 66.4                 | 51.6                 |
| Ligand/ion          | 58.3                 | 44.9                 |
| Water               | 55.1                 | 48.6                 |
| RMS deviations      |                      |                      |
| Bonds (Å)           | 0.008                | 0.008                |
| Angles (°)          | 1.21                 | 1.21                 |
| Ramachandran        |                      |                      |
| Favored (%)         | 98.0                 | 97.0                 |
| Outlier (%)         | 0.3                  | 0.5                  |

<sup>a</sup> statistics for the highest resolution shell are shown in parentheses.

<sup>b</sup> R<sub>merge</sub> = Σ|I - <I>|/ Σ|I|, where I is the observed intensity and <I> is the average intensity of several symmetry-related observations.

<sup>c</sup> R<sub>work</sub> = Σ||F<sub>o</sub>| - |F<sub>c</sub>||/ Σ|F<sub>o</sub>|, where F<sub>o</sub> and F<sub>c</sub> are the observed and calculated structure factors, respectively.

<sup>d</sup> R<sub>free</sub> = Σ||F<sub>o</sub>| - |F<sub>c</sub>||/ Σ|F<sub>o</sub>| for 5% of the data not used at any stage of the structural refinement.
Figure 12: Structure of *A. fumigatus* Tps1A-UDP-VDM complex

(a) Cartoon diagram of *A. fumigatus* Tps1A. The N-terminal and C-terminal Rossmann-fold domains depicted as ribbons and colored cyan and pink whilst the C-terminal helix that interacts with both domains is colored red. UDP and VDM are shown as yellow sticks and colored by atom type. (b) Superposition of *A. fumigatus* Tps1A with *C. albicans* Tps1-UDP-VDM structure, the latter of is shown in cartoon diagram colored in grey without bound substrates.
Figure 13: Structure of *A. fumigatus* Tps1B-UDP-VDM complex

(a) Cartoon diagrams of *A. fumigatus* Tps1B. The N-terminal and C-terminal Rossmann-fold domains are depicted as ribbons and colored cyan and pink whilst the C-terminal helix that interacts with both domains is colored red. (b) Superposition of *A. fumigatus* Tps1B with *A. fumigatus* Tps1A, the latter of is shown in cartoon diagram colored in grey. (c) Superposition of *A. fumigatus* Tps1B with *C. albicans* Tps1-UDP-VDM complex, the latter of is shown in cartoon diagram colored in grey without bound substrates.
Figure 14: Sequence alignment of *A. fumigatus* Tps1B and Tps1A

Primary sequences of *A. fumigatus* Tps1B and Tps1A are aligned. Identical sequences regions are shaded in the figure. The secondary structure of Tps1B is shown as arrows for α helices and rectangles for β strands. The secondary structure component is colored in identical scheme as Fig. 13. Note the C-terminal extension of Tps1A.
Figure 15: Active site of *A. fumigatus* Tps1B

(a, b) Active site interactions of Tps1B with UDP and VDM. A 2F\_o - F\_c electron density map of the bound UDP and VDM is shown as grey mesh and contoured at 1 \( \sigma \). UDP and VDM are shown as atom-colored sticks and solvent shown as red spheres. Protein-substrate and water-substrate interactions are indicated by dashed lines.
Size exclusion chromatography of Tps1A and Tps1B using an S200 column (GE Healthcare). The total column volume is 320 mL and the void volume is 112 mL. The two eluted peaks of Tps1A indicate a high molecular weight aggregated species and 160 kDa suggestive of an elongated dimer. The Tps1B peak corresponds to a calculated molecular weight of 70 kDa indicating a monomeric species. The peak in Tps1B eluted at 150 mL corresponds to contaminated chaperone during purification (SDS-PAGE data not shown).

Figure 16: Size exclusion chromatography of Tps1A and Tps1B
4. Structures of Tps2 from pathogenic fungi reveal the mechanisms of substrate specificity and catalysis

Trehalose is a disaccharide essential for the survival and virulence of pathogenic fungi. The biosynthesis of trehalose requires two critical enzymes, the trehalose-6-phosphate synthase, Tps1, and trehalose-6-phosphate phosphatase, Tps2. Here, we report the structures of the N-terminal domain of Tps2 (Tps2NTD) from *Candida albicans* and a transition state complex of its C-terminal trehalose-6-phosphate phosphatase domain (Tps2PD) bound to trehalose and BeF$_3$. The structure of the Tps2PD-trehalose-BeF$_3$ complex reveals extensive interactions between both glucose moieties of the trehalose involving all eight hydroxyl groups and multiple residues of both the cap and core domains of the protein. The Tps2PD-trehalose-BeF$_3$ structure also reveals that steric hindrance is the underlying factor for the exquisite substrate specificity of this phosphatase. In addition, Tps2PD structures in open and closed conformations provide visualization of the striking structural changes effected by substrate binding to the Tps2PD catalytic pocket.

4.1 Introduction

Fungal infections, both superficial and invasive, have enormous effects on human health. Superficial infections of skin and nails infect around 1.7 billion people. Invasive fungal infections, primarily opportunistic invasive mycoses, lead to substantial morbidity (more than 2 million cases) and mortality (around 50%) [3]. Amongst the
opportunistic invasive mycoses, cryptococcosis, candidiasis, and aspergillosis represent by far the most common human infections. The high mortality rate associated with opportunistic invasive mycoses is largely due to the availability of a limited set of rapidly fungicidal drugs, drug toxicities, drug-drug interactions and rapidly emerging drug resistance[14,111] combined with serious underlying diseases. Further hampering drug development efforts, is the problematic identification of novel antifungal drug targets due to the substantial overlap of essential fungal and mammalian biosynthetic pathways as eukaryotic organisms.

One promising novel antifungal target is the trehalose biosynthetic pathway, which does not have a mammalian counterpart. Computer aided target selection against fungal infections rank the trehalose biosynthetic pathway as a top candidate for antifungal intervention and targeting[93]. Trehalose, α-D-glucopyranosyl-(1→1)-α-D-glucopyranoside, is a non-reducing disaccharide synthesized in bacteria, fungi, lower plants and invertebrates, but has never been identified in mammals[23]. Trehalose levels significantly increase when fungal cells are exposed to multiple external and internal stresses, such as dehydration, heat shock or oxidative stress [94,95]. Increased production of trehalose has a significant role in protecting proteins and membranes against these environmental stresses [29,96,97], rendering trehalose essential for fungal cell stress response and survival during infections.
Trehalose is synthesized in most fungi by a two-step enzymatic reaction (Fig. 1). The first step, which is catalyzed by trehalose-6-phosphate synthase (Tps1), is the condensation of UDP-glucose and glucose-6-phosphate (G6P) to form trehalose-6-phosphate (T6P). The second step, which is catalyzed by trehalose-6-phosphate phosphatase (Tps2), is the dephosphorylation of T6P to yield trehalose. Disruption of the *tps1* gene in *C. albicans* results in impaired hyphae formation and decreased infectivity[55]. Furthermore, disruption of the *tps2* gene in *C. albicans* leads to the accumulation of T6P, which is cytotoxic in high concentrations, decreases yeast growth at higher temperatures and diminishes infectivity[57,98]. Similarly, the importance of trehalose production, and hence this biosynthetic pathway, in cell survival has been convincingly demonstrated in other important pathogenic fungi including *Cryptococcus*[58,59] and *Aspergillus*[60,61]. In addition to the protective role of trehalose in stress response, the trehalose biosynthetic pathway has been recognized as an integral part of basic fungal and plant cellular metabolism and energy homeostasis[38,42]. For instance, overexpression of *tps2* in maize decreases T6P concentration and increases maize yield in drought conditions [49]. These results further underscore Tps1 and Tps2 as important regulatory proteins for maintaining cellular integrity.

*C. albicans* Tps2 is an 888 amino acid residue protein that can be divided into two structural domains. The N-terminal domain (Tps2NTD) is 534 amino acid residues, and shares significant sequence similarity to the *C. albicans* Tps1 and *Escherichia coli* (*E. coli*)
OtsA (the bacterial Tps1 homologue) but is missing key catalytic residues and hence is posited to be a “pseudo-Tps1”. Such a large, two domain containing trehalose-6-phosphatase is unlike the bacterial Tps2 proteins, which are much smaller. For example, the *E. coli* Tps2 protein (OtsB) is only 266 amino acid residues and does not have a comparable Tps2NTD. Thus the structure and particularly the function of the Tps2NTD remain enigmatic. The C-terminal domain of the *C. albicans* Tps2 protein (Tps2PD) encompasses amino acid residues 535 through 888 and contains the putative phosphatase domain. Indeed, sequence comparisons reveal that Tps2PD is a member of the haloacid dehydrogenase superfamily (HADSF) phosphatases, enzymes which recognize a broad spectrum of substrates. To date only structures of substrate-free Tps2 from *Brugia malayi* [83,112] and Tps2-related protein from *Thermoplasma acidophilum* [77] have been determined, thus leaving the mechanisms of T6P binding specificity and catalysis by Tps2 enzymes unknown.

Our understanding of the substrate specificity and catalytic mechanism of any Tps2 has been hindered significantly by the lack of germane structural data, which consequently has delayed structure-guided design of novel inhibitors. Here we report the first structure of any fungal pathogen Tps2 protein and more important, the first transition-state structure of *C. albicans* Tps2PD in complex with trehalose and BeF₃. We also report here the structure of the “closed” state of the *Cryptococcus neoformans* (*C. neoformans*) Tps2PD, which has had the conserved nucleophilic aspartate mutated to
asparagine, in complex with T6P and the substrate-free “open” state of the Aspergillus fumigatus (A. fumigatus) Tps2PD. Our structural and biochemical analyses reveal the basis of substrate specificity, the conformational flexibility employed by the enzyme for catalysis and provide key insight into the catalytic mechanism of trehalose-6-phosphate phosphatases thus paving the way for the initiation of structure-guided inhibitor design.

4.2 Results

4.2.1 Structure of C. albicans Tps2NTD

The Tps2NTD structure (residues 1-534) was determined to 2.56 Å resolution by molecular replacement using the structure of E. coli OtsA, a Tps1 homologue, (PDB code: 1UQU) [70] as a search model. Selected crystallographic data and refinement statistics are listed in Table 1. Electron density was poor for N-terminal domain, with 58 residues (residues 1-17, 45-61, 91-114 from chain B) missing from the structure, and likely contributing to a relatively high R_free. These disordered regions are consistent with secondary structure prediction by PSIPRED [110]. However, deletion of these disordered residues renders Tps2NTD insoluble, and necessitated their presence during protein purification and for crystallization.

The overall structure of the Tps2NTD is similar to that of other enzymes belonging to the GT-B class of glycosyltransferases [62] and contains N and C-terminal Rossmann-fold domains and a C-terminal helix that crosses and interacts with each Rossmann-fold domain (Fig. 17a). As anticipated, a DALI search [99] revealed that the
structure most similar to the Tps2NTD is that of *E. coli* OtsA with a root mean square deviation (rmsd) of 2.2 Å for 433 corresponding residues. Structural superposition of the Tps2NTD and OtsA reveals the highly conserved Rossmann-fold of the C-terminal domain and C-terminal helix that spans both domains of the Tps2NTD (Fig. 17b).

Compared with *E. coli* OtsA, the most notable difference between the structures is the presence of an anti-parallel β strand in OtsA that is substituted by α2 in Tps2NTD (Fig. 17b).

Despite the overall structural and sequence similarity of Tps2NTD and OtsA, alignment (Fig. 18) of their sequences reveals that OtsA residues G22, involved in UDP binding, and R9, Y76 and R300, involved in G6P binding[71], are replaced with a tyrosine, serine, tryptophan and a deletion in Tps2NTD. This finding suggested strongly that the Tps2NTD could not synthesize T6P. Indeed, Tps2NTD has no glycosyltransferase activity. Furthermore, no T6P phosphatase activity was detected as well. However, at present, it is unclear whether Tps2NTD has any other enzymatic activity and its role in *C. albicans* and other pathogenic fungi thus remains enigmatic.

Despite the lack of trehaose biosynthesis function of Tps2NTD, deletion of this domain in *C. neoformans* resulted in a temperature sensitive phenotype at 39 °C (Fig. 17c), suggesting Tps2NTD is functionally essential for cell survival at elevated temperature.

*C. albicans* Tps2 is one component of the trehalose biosynthetic complex, which also consists of Tps1 and the Trehalose-6-phosphate regulatory protein (Tps3). The
formation of a multiprotein complex containing Tps1, Tps2 and a Tps3 homologue has been demonstrated in *Saccharomyces cerevisiae* (*S. cerevisiae*)[80] and such complex formation appears necessary for proper function in fungi since OtsA, the *E. coli* Tps1, which does not form a complex with OtsB, the *E. coli* Tps2, is unable to complement a *S. cerevisiae* Tps1 deletion strain[113]. The presence of the Tps2NTD, which is structurally very similar to OtsA, suggests its potential role in hetero-protein complex formation as the key residues in the conserved dimer interface of *E. coli* OtsA[71] are also present in the *C. albicans* Tps1, Tps2NTD and Tps3NTD (Fig. 18). Although direct interactions between the *C. albicans* Tps2NTD, Tps3NTD and Tps1 have not yet been reported, the presence of this conserved interface suggests the potential for the Tps2NTD and Tps3NTD to bind to each other as well as Tps1 and thus play a role in the formation of the trehalose biosynthetic complex in *Candida* and probably in *Aspergillus* and *Cryptococcal* species. One of the most likely roles for the formation of at least a Tps1-Tps2 heterodimer, but certainly not the only possibility, is substrate sequestration of T6P and its delivery to the latter enzyme. Beyond its function as an intermediate in the biosynthesis of trehalose, T6P has been demonstrated to be an important regulatory molecule in both yeast and plant [35,42]. Intriguingly the *K_m* of Tps2PD for T6P is 0.5 mM (Fig. 19), a surprisingly high concentration in comparison to the reported cytosolic concentration (100 μM) of T6P [48,98,102]. This difference in *K_m* and cytosolic T6P concentration requires a mechanism for controlling the free T6P concentration *in vivo.*
The formation of a trehalose biosynthetic complex, composed of at least Tps1 and Tps2, would reduce cytosolic T6P concentration and increase T6P concentrations in proximity to Tps2 by trapping T6P in the complex[102], thereby allowing the efficient dephosphorylation of this signaling molecule and the production of trehalose.

4.2.2 Structure of the C. albicans Tps2PD

In order to determine the structure of the C. albicans Tps2PD, a series of truncated proteins that encompassed the putative phosphatase domain were created and purified. Crystals of a Tps2PD construct comprising residues 1-299 were obtained in the presence of beryllium fluoride (BeF₃) and trehalose and diffracted to 2.0 Å resolution. This Tps2PD truncate is catalytically active and displays a $k_{cat}$ of $32.8 \pm 2.5 \text{ s}^{-1}$ and $K_m$ of $498.9 \pm 89.9 \mu\text{M}$ with specificity for T6P (Fig. 19), values which are similar to those reported for the Tps2 from Brugia malayi[83]. The structure was determined using single-wavelength anomalous dispersion (SAD) methods and selenomethionine (SeMet)-substituted Tps2PD. The model was later refined using a higher resolution native dataset (Table 4). There are two molecules in the asymmetric unit, which adopt nearly identical conformations (rmsd = 0.25 Å). The Tps2PD structure and size exclusion chromatography indicate that Tps2PD is monomeric in solution. All subsequent results and discussion focus on molecule A.

The Tps2PD structure consists of a core domain (residues 1-105 and 186-299), which is a modified Rossmann-fold, and a cap domain (residues 106-185), which is
inserted between the two halves of the core domain (Fig. 20 and 21). Residues 206-214 were not seen in electron density maps and thus are not included in the structure. The Rossmann-fold of the core domain is formed by six parallel β strands flanked by seven α helices. An additional β strand, β5, is antiparallel with respect to the core β sheets. A loop C-terminal to β5 links the core and cap domains. The cap domain is composed of four antiparallel β strands and two α helices. The cap domain reconnects to the core domain by a loop C-terminal of α7. The insertion position and topology of the cap domain is typically found in the subfamily 2B HADSF phosphatases [73]. Excellent electron density for trehalose, BeF₃, which is covalently linked to a conserved nucleophilic aspartate residue, D25, and a Mg²⁺ ion is observed (Fig. 20b) and demarcates the active site of this phosphatase. Thus, the structure reveals the first transition-state structure of any T6P-specific phosphatase.

4.2.3 Substrate binding pocket of C. albicans Tps2PD

Similar to other HADSF phosphatases, the C. albicans Tps2PD reveals four conserved motifs for nucleophilic catalysis [73,74,112]. Residue D25 (Motif I) is covalently bonded to the BeF₃ and therefore presents a visualization of the aspartyl-phosphate intermediate that is formed during the nucleophilic attack of its carboxylate side chain (Fig. 20c). Residues S65 (Motif II) and K188 (Motif III) make hydrogen bonds and electrostatic interactions with the BeF₃, contributing to the stabilization of the aspartyl-phosphate intermediate. A nearby Mg²⁺ ion also interacts with the BeF₃.
potentially to neutralize the charge of the covalent intermediate and facilitate the subsequent removal of the phosphate group from D25. The Mg\(^{2+}\) ion is positioned in the active site through interactions with the backbone carbonyl group of residue D27, a Motif I residue and the second aspartate of the characteristic DXD motif of the HADSF phosphatases, and the side chain of D230 (Motif IV). Two water molecules are also seen in the active site both of which interact with the Mg\(^{2+}\) ion. These water molecules also interact with side chains of residues D230 and D234 (Motif IV). Combined these interactions complete the octahedral coordination of the Mg\(^{2+}\) ion. In the current structure these water molecules also make protein-water-mediated hydrogen bonds to the trehalose, likely increasing the specificity and binding affinity of Tps2PD for T6P.

The two glucose moieties of trehalose engage in multiple interactions with residues of the extensive substrate binding pocket of Tps2PD, both directly and indirectly through water-mediated interactions (Fig. 20d and 20e). To adopt the correct conformation for nucleophilic attack by residue D25, the trehalose O6' atom, from which the phosphate group of T6P is cleaved, makes an extensive hydrogen bond network with Tps2PD core residues S65, G66 and R67. These interactions orient the substrate into an ideal position for nucleophile attack by D25. Additional interactions to this glucose molecule include hydrogen bonds from the side chain of cap residue K133 to both the O3' and O4' hydroxyl groups. O3' and O4' also interact with the side chains of cap residues E131 and E180, respectively. The O2' hydroxyl group is positioned in the active
site by its water-mediated interactions with the side chains of residues E131 and R67. The ring oxygen of this glucose moiety, O5’, interacts with carboxylate group of residue D27. Thus, all four hydroxyl groups of the glucose-6-phosphate moiety of the TP6 interact extensively with the catalytic pocket of Tsp2PD ensuring the proper chemical environment and orientation of the O6’-phosphate bond for nucleophilic attack by the D25 carboxylate side chain (Fig. 20c and 20d). Additionally, the proper alignment of the substrate is secured by cation-π stacking interactions between the side chains of residues R67 and F71 whereby the guanidinium side chain of R67 is locked into place over the glucose ring by interactions with the carboxyl group of D27 (Fig. 20d).

The other glucose moiety of trehalose also engages in extensive interactions with Tsp2PD (Fig. 20e). Its O2 hydroxyl group makes hydrogen bonds to the backbone carbonyl moiety of residue K176, the N178 sidechain and a water molecule. This water is involved in a hydrogen-bonding network that links this glucose to residue D231 (Motif IV) and a water molecule that coordinates the Mg^{2+} ion. O3 makes two hydrogen bonds to the side chain of residue R142 and also interacts with the K176 backbone. Both H140 and A177 coordinate one water molecule to bind to O3 as well. The O4 hydroxyl group forms a hydrogen bond with the V34 backbone carbonyl group whereas O6 forms hydrogen bonds to both the V34 backbone amide group and a water molecule. This water molecule interacts with P32 and D27 (Motif I), again demonstrating the strong link
between key catalytic residues and both glucose moieties of trehalose. Only the glucosyl ring oxygen atom O5 is not engaged in hydrogen bonding to either Tps2PD or solvent.

Thus, residues belonging to both the cap and core domains play key roles in substrate binding and specificity. However, the key catalytic residues, including the nucleophile D25 and those residues responsible for the direct or indirect stabilization of the transition state, appear to be found primarily in the core domain. Two additional contributors to trehalose binding are Loop 1 (residues 24-37) and Loop 2 (residues 140-146). As noted above, both loops host residues involved in trehalose binding whereby residues P32 and V34 from Loop 1 and H140 and R142 from Loop 2 make hydrogen bonds to the trehalose. Substrate binding brings the two loops into proximity, which effectively closes the active site from the bulk solvent (Figure 20a). Interestingly only one weak interaction, a D36-R143 electrostatic contact, is observed between the Loops. Such a minimal number of “pocket-closing” interactions might be facilitating active site reopening and substrate departure.

4.2.4 Tps2PD active site mutations

Tps2PD dephosphorylates T6P with a $k_{\text{cat}}$ of $32.8 \pm 2.5 \text{ s}^{-1}$ and $K_m$ of $498.9 \pm 89.9 \mu\text{M}$, values which are similar to those reported for Brugia malayi Tps2 [83] ($k_{\text{cat}}= 24 \pm 2 \text{ s}^{-1}$, $K_m= 360 \pm 60 \mu\text{M}$). To assess the significance of the structural identified nucleophilic catalytic and substrate recognition residues in Tps2PD phosphatase activity, a series of mutations were introduced to C. albicans Tps2PD. Mutation of the Tps2PD nucleophile
aspartate to asparagine completely abolishes Tps2PD enzymatic activity, highlighting the critical importance of the nucleophile in phosphatase activity (Fig. 22a). Mutation of each conserved motif (S65A, D230A, and D234A) severely impairs phosphatase activity as well (Fig. 22a). Disrupting trehalose-binding residues, R67, E131, K133 and E180 by separate alanine substitution showed similar effects on catalytic efficiency. Interestingly, the H140A mutant, whose sidechain interacts with trehalose through the E131 mediated hydrogen bond network instead of direct hydrogen bond showed 30% residual phosphatase activity. Therefore, second-shell residues in the binding pocket also contribute to enzyme efficiency. Together, our high-resolution structure of Tps2PD in the transition-state and Tps2PD phosphatase activity assays reveal residues critical for Tps2PD activity.

4.2.5 Closed conformational structure of C. neoformans Tps2PD

The C. neoformans Tps2PD (residues 1-306)-T6P complex structure was determined to 2.15 Å resolution by molecular replacement using the C. albicans Tps2PD structure as a search model. Selected crystallographic data and refinement statistics are listed in Table 2. To capture T6P in the binding pocket, the Tps2PD nucleophile aspartate, D24, was mutated to asparagine. As seen for the C. albicans Tps2PD-BeF₃-trehalose complex, the C. neoformans Tps2PD-T6P complex adopts a closed conformation (Fig. 23) with the two proteins taking essentially identical structures (rmsd = 1.0 Å). Furthermore, clear electron density was for T6P was detected in the binding pocket.
providing the first view of a Tps2PD-substrate complex (Fig. 22b). In addition to the overall structural similarity, the high sequence identity (46%) and the complete conservation of all active site residues indicate that *C. neoformans* Tps2PD utilizes the same nucleophilic catalytic mechanism and substrate binding mode as the *C. albicans* Tps2PD (Fig. 21). Detailed coordination of T6P in Tps2PD-T6P structure is not included due to this highly similar active site. Superposition of the *C. albicans* Tps2PD-BeF$_3$-trehalose and *C. neoformans* Tps2PD-T6P complexes reveals insight into the reaction coordinate of these T6P phosphatases (Fig. 22b). In the *C. albicans* Tps2PD transition-state structure, the BeF$_3$ phosphate mimic is covalently bonded to the carboxylate side chain of residue D25, whereas in the *C. neoformans* Tps2PD-T6P complex structure, the phosphate moiety of the T6P is 1.8 Å from the BeF$_3$ and hydrogen bonds with the side chain of N24. The two structures thus provide a visualization of phosphoryl group transfer from T6P to aspartate during nucleophilic attack, which appears to require the slight repositioning of the Mg$^{2+}$ ion in the structures by approximately 0.5 Å. This nucleophile mutation (D-N) was also introduced into *C. neoformans* (annotated as *tps2D705N*) and a similar temperature sensitive growth phenotype at 37 °C (Fig. 17c) comparable to a *tps2Δ* deletion mutant was detected. Interestingly, the *tps2D705N* cells had a pronounced defect in cell division and failed to produce large capsules under capsule-inducing conditions (Fig. 24). These findings highlight that Tps2 phosphatase activity is essential for stress responses, further underscoring the significance of T6P
concentration regulation. It also suggests that blocking the activity of the enzyme has more severe consequences compared to complete loss of the protein. Temperature in humans remains a constant host factor that critically affects fungal pathogen survival; therefore, a Tps2 specific inhibitor would be predicted to eliminate C. neoformans infections. This characteristic target of environmental stress serves as an innovative therapeutic approach, utilizing thermotolerance to rapidly kill fungal pathogens.

4.2.6 Tps2PD adopts a large conformational change upon substrate binding

Superpositions of the individual core and cap domains of B. malayi Tps2PD (residues 206-492, excluding the N-terminal domain) onto the corresponding core and cap domains of the C. albicans Tps2PD reveal rmsd of 3.9 Å and 1.9 Å, respectively (Fig. 25). Superposition of the entire B. malayi Tps2PD onto the C. albicans Tps2PD reveals an rmsd of 5.6 Å for 277 corresponding Cα atoms. This large rmsd is due to a substrate-binding induced conformational change within the protein. Indeed, such a conformational transition from a substrate-free “open” conformation to a substrate-bound “closed” conformation is critical for the exclusion of bulk solvent and catalysis. However, sequence alignment between C. albicans Tps2PD and B. malayi Tps2 reveal only an overall 24% sequence identity and more importantly no conservation of the residues involved in substrate recognition, suggesting a significant evolutionary difference between the two species. Therefore, we sought to solve the structure of a substrate-free Tps2PD from a more closely related species to verify the observed Tps2PD
conformational flexibility. Thus, we determined the substrate-free structure of Tps2PD from human pathogenic mold, *A. fumigatus*, which shares 58% sequence identity to the Tps2PD from *C. albicans* including all residues involved in T6P binding and catalysis (Fig. 21). We obtained three different crystal forms of the *A. fumigatus* Tps2PD (forms 1, 2 and 3). The structure of form 1 was solved by SAD using a SeMet-substituted protein crystal whilst the structures of forms 2 and 3 were determined by molecular replacement using the structure of crystal form 1 as the search model.

As expected the *A. fumigatus* Tps2PD structures exhibit the conserved overall fold of a HADSF phosphatase with the core Rossmann-fold domain and cap domain (Fig. 5a). But quite different from the transition-state Tps2PD and T6P-bound Tps2PD structures, all three *A. fumigatus* structures assume conformations in which the core and cap domains have opened significantly but to different degrees (Fig. 26a and 26b). Superposition of the structure of the three *A. fumigatus* core domains (residues 1-106, 184-270) reveals strong structure conservation (rmsd of 0.3 Å). However, a clear difference in the relative positions of their cap domains is observed and superposition of their core domains reveals the cap domain strand β6, as denoted by F129, is 3.4 Å closer to the core in form 1 than in form 2, hence revealing the significant conformational flexibility of Tps2PD in the absence of substrate (Fig. 26a). Superposition of the *C. albicans* Tps2PD transition-state structure with the open conformation that is observed for *A. fumigatus* Tps2PD form 2 visualizes the large conformational change between the
two states (Fig. 26c). Despite the high structural similarity of the core and cap domains (rmsd of 0.88 Å and 0.48 Å), respectively, when the core domains of the transition-state structure and form 2 substrate-free structure are overlaid, the tip of cap domain as denoted by E123 of helix 4 translocates 25.2 Å. Further analysis of this movement using DynDom[114] reveals a 53° rotation upon substrate binding, which is centered about C. albicans residues 102-103, 184-188 that are located on the linker between the core and cap domain and form a hinge (Fig. 26c). Thus, the structures of the highly related fungal Tps2PDs support the idea that type IIB HADSF phosphatases adopt a large conformational change upon substrate binding and substrate dephosphorylation (Fig. 26d), though interestingly the extent of the conformational change varies amongst different phosphatase. For example a 28° rotation is observed for substrate binding by the sucrose-6-phosphate phosphatase[115].

4.2.7 Substrate specificity of Tps2PD

Previous biochemical characterization in Mycobacterium smegmatis[116] and B. malayi Tps2 showed a strong preference for T6P as its substrate. C. albicans Tps2PD shows similar substrate specificity for T6P and indeed, does not dephosphorylate sucrose-6-phosphate (S6P), glucose-6-phosphate (G6P) or para-nitrophenylphosphate (pNPP, the “universal” phosphatase substrate) (Fig. 19). The structures of the Tps2PD transition-state and Tps2PD-T6P complexes reveal the underlying molecular basis of the specificity of this phosphatase for T6P and its discrimination against S6P. Superposition
of the glucose moiety of T6P, which does not contain the phosphate that is cleaved, and the glucose moiety of sucrose places the fructose ring of the latter sugar proximal to the nucleophilic aspartate (Fig. 27). However, the structural differences between the fructose of S6P and the glucose of T6P would result in the loss of several hydrogen bonds but more critically the steric clash of the fructose ring with the guanidinium side chain of residue R67. Residue R67 is unable to avoid this clash because of its stacking interaction with residue F71. It should be noted that this clash would be even more significant if the six position of the fructose sugar is phosphorylated and that this phosphate would not be positioned correctly for catalysis.

The strict conformational requirements of the Tps2PD binding pocket eliminate other potential sugar phosphate substrate for Tps2PD. Yet, Tps2PD demonstrates no activity against G6P, which can be positioned into the active site without steric clash since it could occupy the identical position of the glucose-6-phosphate moiety of T6P. However, the catalytic pocket is composed of two glucose-binding sites and the nonphosphorylated glucose moiety of T6P is critical for substrate affinity by interacting extensively with the protein (Fig. 20e). Loss of these interactions would decrease Tps2PD binding affinity for the G6P significantly and require extremely high G6P concentrations for enzymatic activity, which are not physiological relevant. Importantly, the glucose moiety of T6P binds to residues of Loop 1 and Loop 2 and
contributes to the conformational change necessary for the enzyme to take its closed, catalytically competent conformation.

The highly specific active site of Tps2PD distinguishes this enzyme from other mammalian phosphatases, and thus may alleviate potential off-target toxicity of Tps2PD specific inhibitors. Furthermore, sequence alignments of the *C. neoformans* and *A. fumigatus* Tps2PD (Fig. 21) reveal 46% and 58% sequence identity, respectively. Residues involved in nucleophilic catalysis and critically, substrate binding, are conserved amongst these pathogenic fungi species. The high conservation of these residues in the pathogenic fungal Tps2PDs suggests that this enzyme is a viable broad-spectrum antifungal drug target.

In summary, we report here the Tps2NTD structure from *C. albicans* and its potential function in the trehalose biosynthetic complex formation. The formation of this complex regulates T6P concentration intracellularly and increases trehalose production efficiency. Further biochemical and cellular experiments are required to understand the stereochemistry and function of this complex. We also report here the structures of Tps2PD in multiple conformational states. Combined with biochemical and cellular analyses, we underscore key residues in substrate recognition, reveal the basis of substrate specificity and provide insights into dephosphorylation mechanism. The conformational flexibility presented by Tps2PD structures deepens our understandings of Tps2 and other general HADSF phosphatases. Rational Tps2 inhibitor design is also
applicable with the structural information, paving the way for antifungal drug design targeting the trehalose biosynthetic pathway.

4.3 Experimental Procedures

4.3.1 Protein purification and crystallization

The Tps2NTD gene (residues 1-534) and Tps2PD gene (residues 1-299) from C. albicans strain SC5314 were cloned from a cDNA library into the pMCSG7 plasmid using a standard ligation-independent cloning protocol[117]. Tps2NTD residue 120 was mutated to serine since the CTG encodes a serine in C. albicans[118]. The pMCSG7 vector contains an N-terminal hexahistidine affinity tag followed by a TEV protease cleavage site. Tps2NTD and Tps2PD were expressed in the Rosetta 2(DE3) pLysS cell line (Novagen) and induced with 0.5 mM IPTG at 15 °C for 16 h. The Tps2NTD and Tps2PD proteins were purified by Ni²⁺-NTA affinity column chromatography followed by TEV protease cleavage of the His-tag. The cleaved proteins were reloaded onto the Ni²⁺-NTA column and eluted in our standard buffer without imidazole in order to eliminate noncleaved protein and TEV protease. The proteins were finally purified by size exclusion chromatography (SEC). To prepare our singly mutated Tps2PD proteins, a standard site directed mutagenesis protocol was utilized[103]. Mutated Tps2PD was purified as described for wild type protein.

C. albicans Tps2NTD was concentrated to 20 mg/mL and crystals were grown at 25 °C by hanging drop-vapor diffusion methods. Protein crystals appeared after one
week from solutions of 1.2 M sodium citrate and 0.1 M cacodylate pH 6.5. *C. albicans* Tps2PD was concentrated to 45 mg/mL and crystallized with 50 mM trehalose and 50 mM beryllium fluoride. Crystals were grown at 25 °C by hanging drop-vapor diffusion methods. Protein crystals appeared after one week from solutions of 0.2 M lithium sulfate, 0.1 M Tris pH 7.0 and 2.0 M ammonium sulfate.

Tps2PD genes from *A. fumigatus* and *C. neoformans* were codon optimized for *E. coli* expression system (GenScript). To prepare *C. neoformans* Tps2 D24N mutant, site directed mutagenesis protocol was utilized [103]. These codon optimized genes were cloned into a pET-28b kanamycin-resistant vector via restriction sites NcoI and XhoI. This vector contained a C-terminal hexahistidine affinity tag. They were transformed into BL21(DE3)plysS cells (Life Technologies) and were then induced with 0.5 mM IPTG at 15 °C for 16 h. Tps2PD was purified by Ni²⁺-NTA column and subsequent SEC via Superdex S75 column (GE Healthcare). Purified Tps2PD was concentrated in Buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 5% glycerol, 5 mM MgCl₂, and 1 mM βME) using Amicon Ultra concentrator (10K MWCO, Millipore). *C. neoformans* Tps2PD was concentrated to 12 mg/ml and co-crystallized with 5 mM T6P. Crystals were grown at 25 °C by hanging drop vapor diffusion method. Protein crystals appeared after two weeks from 0.1 M MgCl₂, 0.1 M Tris pH 8.5, 20% PEG 400 and 10% PEG 8000. *A. fumigatus* crystals were grown at 25 °C by hanging drop-vapor diffusion method using 20 mg/ml protein. Protein crystals in form 1 appeared after two weeks from 0.1 M sodium citrate,
pH 5.5, 16% PEG 8000. Protein crystals in form 2 appeared two weeks from 0.1 M HEPES, pH 7.0, 15% PEG 20000. Protein crystals in form 3 appeared after two weeks from 0.1 M HEPES, pH 7.5, 12% PEG 3350.

4.3.2 Data collection and structure determination

X-ray intensity data for the *C. albicans* Tps2NTD crystal were collected to 2.56 Å resolution at the Advanced Photon Source (APS) beamline ID-22. The Tps2NTD structure was determined by molecular replacement (Phaser[106]) using the OtsA structure (PDB code: 1UQU)[70] as a search model. X-ray intensity data for the *C. albicans* Tps2PD crystal and SeMet-substituted Tps2PD protein crystal were collected to 2.0 Å and 2.12 Å resolution using Advanced Light Source (ALS) beamlines 5.0.1 and 5.0.2. Experimental phases were calculated using Phenix Autosol[119] and the model was manually built in Coot[108]. X-ray intensity data for the *C. neoformans* Tps2PD crystal were collected to 2.15 Å resolution using ALS beamline 5.0.1. The structure was determined by molecular replacement using the *C. albicans* Tps2PD structure as the search model. X-ray intensity data for the *A. fumigatus* SeMet-substituted crystal and native crystals were collected at the APS BM-22 and ID-22 lines, respectively. Experimental phasing was calculated using Phenix Autosol [119] and the model was manually built in Coot[108]. All X-ray intensity data were indexed and scaled using HKL 2000[105]. All models were improved using multiple rounds of refinement in
Phenix refine [107] and manual rebuilding. Selected statistics of data collection and refinement are presented in Table 4 and Table 5.

4.3.3 Steady-state kinetics and enzymatic activity assays

Steady-state kinetic measurements on Tps2PD measured the initial enzymatic velocity utilizing 3 nM Tps2PD with a gradient concentration of T6P (0.05 mM to 1.2 mM). The assays were performed using the EnzChek Phosphate Assay Kit (Molecular Probes) and absorbance was measured using PerkinElmer Lambda 25 spectrophotometer at 360 nm. All measurements were performed in triplicate. Data were fit by the GraphPad Prism 6 Michaelis-Menten kinetics module. The relative activities of wild type Tps2PD and selected mutants were measured with the EnzChek Phosphate Assay Kit (Molecular Probes) using 5 nM protein and 500 μM T6P.

4.3.4 Cell survival assays and capsule induction assays

Cryptococcus neoformans var. grubii strain H99 was used to determine the viability and capsule induction of the wild type strain and selected mutants. PCR based site-directed mutagenesis in combination with the In-Fusion HD cloning system was used to generate two mutated genes: one that encodes a truncated Tps2 protein (residues 682-987) that is missing the NTD and a Tps2 protein containing the D705N mutation. Each DNA construct was transformed into a tps2Δ mutant strain by biolistic transformation [120].
To examine cells for capsule production, a capsule-inducing media was used as described with minor changes [121]. Briefly, a single colony of yeast was inoculated into 5 ml YPD broth and grown at 30 °C overnight on a shaking incubator (225 rpm). Cells were harvested by centrifugation at 3000 rpm for 3 minutes and washed twice with 10 mL PBS. The cell pellet was resuspended in 5 mL diluted Sabouraud (1/1) in MOPS buffer pH 7.3. The cells were grown overnight under the same conditions and stained the next day with India Ink for visualization of capsule. A Zeiss Axio Imager A1 fluorescence microscope equipped with an AxioCam MRM digital camera was used to capture differential interference contrast (DIC) images under oil immersion at 63X magnification. Images were acquired and analyzed by ZEN software (Zeiss).

4.4 Acknowledgements

Biolistic transformation of *C. neoformans*, temperature sensitive cellular assays and capsule induction assays were performed by Drs. Jennifer L. Tenor and Dena L. Toffaletti from the Perfect laboratory at Duke University Medical Center. For completeness of this chapter, these data were included.
**Table 4: Selected *C. albicans* Tps2 data collection and refinement statistics**

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>C. albicans Tps2NTD</th>
<th>C. albicans Tps2PD</th>
<th>C. albicans Se-Met</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P6;22</td>
<td>P2;2;2,1</td>
<td>P2;2;2,1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>98.3, 98.3, 452.8</td>
<td>58.8, 87.2, 139.3</td>
<td>59.2, 87.2, 137.5</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 120</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-2.56 (2.60-2.56)</td>
<td>50.0-2.0 (2.03-2.0)</td>
<td>50.0-2.12 (2.16-2.12)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.8 (52.5)</td>
<td>5.3 (52.7)</td>
<td>8.0 (31.5)</td>
</tr>
<tr>
<td>I/σI</td>
<td>21.4 (2.5)</td>
<td>31.5 (2.8)</td>
<td>33.0 (5.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.6 (92.2)</td>
<td>99.9 (100.0)</td>
<td>99.1 (93.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.1 (5.5)</td>
<td>4.8 (4.7)</td>
<td>6.9 (5.7)</td>
</tr>
<tr>
<td>Overall figure of merit (FOM)</td>
<td></td>
<td></td>
<td>0.458</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution</td>
<td>48.9-2.56 (2.65-2.56)</td>
<td>36.4-2.0 (2.07-2.0)</td>
<td>-</td>
</tr>
<tr>
<td>No. reflections</td>
<td>42019</td>
<td>49177</td>
<td>-</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; /R&lt;sub&gt;free&lt;/sub&gt; (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>26.6/29.8</td>
<td>20.0/24.1</td>
<td>-</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>7451</td>
<td>4602</td>
<td>-</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>-</td>
<td>56</td>
<td>-</td>
</tr>
<tr>
<td>Water</td>
<td>98</td>
<td>395</td>
<td>-</td>
</tr>
<tr>
<td>B-factor</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Protein</td>
<td>59.9</td>
<td>38.7</td>
<td>-</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>-</td>
<td>29.3</td>
<td>-</td>
</tr>
<tr>
<td>Water</td>
<td>47.5</td>
<td>42.8</td>
<td>-</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.006</td>
<td>0.008</td>
<td>-</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.09</td>
<td>1.10</td>
<td>-</td>
</tr>
<tr>
<td>Ramachandran (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favorable</td>
<td>94.7</td>
<td>97.4</td>
<td>-</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.9</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup> statistics for the highest resolution shell are shown in parentheses.

<sup>b</sup> R<sub>merge</sub> = Σ||I-<i>||/Σ||I||, where I is the observed intensity and <i>|| is the average intensity of several symmetry-related observations.

<sup>c</sup> R<sub>work</sub> = Σ||F<sub>o</sub>|-|F<sub>c</sub>||/Σ|F<sub>o</sub>, where F<sub>o</sub> and F<sub>c</sub> are the observed and calculated structure factors, respectively.

<sup>d</sup> R<sub>free</sub> = Σ||F<sub>o</sub>|-|F<sub>c</sub>||/Σ|F<sub>o</sub> for 5% of the data not used at any stage of the structural refinement.
Table 5: Selected *C. neoformans* and *A. fumigatus* Tps2 data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>C. neoformans Tps2PD</th>
<th>A. fumigatus Tps2PD 1</th>
<th>A. fumigatus Tps2PD 2</th>
<th>A. fumigatus Tps2PD 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P6;22</td>
<td>P22;2,1</td>
<td>P12,1</td>
<td>P12,1</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>63.1, 63.1, 284.3</td>
<td>45.4, 65.5, 118.5</td>
<td>39.8, 42.7, 80.2</td>
<td>45.0, 76.8, 83.7</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 90, 120</td>
<td>90, 90, 90</td>
<td>90, 99.5, 90</td>
<td>90, 94.4, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-2.15 (2.22-2.15)</td>
<td>50-1.57 (1.63-1.57)</td>
<td>50-1.65 (1.71-1.65)</td>
<td>50-1.89 (1.96-1.89)</td>
</tr>
<tr>
<td>Rmerge =&lt; I&gt; / &lt;I&gt;</td>
<td>0.088 (0.693)</td>
<td>0.108 (0.637)</td>
<td>0.077 (0.521)</td>
<td>0.07 (0.568)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.5 (99.8)</td>
<td>99.9 (99.0)</td>
<td>97.3 (89.0)</td>
<td>98.7 (84.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>9.9 (9.6)</td>
<td>5.7 (5.6)</td>
<td>3.5 (2.2)</td>
<td>3.7 (3.5)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>43.3-2.15 (2.23-2.15)</td>
<td>36.0-1.57 (1.63-1.57)</td>
<td>20.6-1.65 (1.71-1.65)</td>
<td>31.8-1.89 (1.96-1.89)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>19254</td>
<td>49835</td>
<td>31559</td>
<td>44688</td>
</tr>
<tr>
<td>Rwork / Rfree [%]</td>
<td>23.0/26.7</td>
<td>20.0/23.0</td>
<td>21.9/25.3</td>
<td>20.7/25.0</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>2412</td>
<td>2257</td>
<td>2122</td>
<td>4344</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>32</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Water</td>
<td>96</td>
<td>392</td>
<td>240</td>
<td>186</td>
</tr>
<tr>
<td>B-factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>53.0</td>
<td>19.9</td>
<td>28.6</td>
<td>45.2</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>47.0</td>
<td>-</td>
<td>19.2</td>
<td>27.4</td>
</tr>
<tr>
<td>Water</td>
<td>49.6</td>
<td>29.7</td>
<td>38.4</td>
<td>48.0</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.005</td>
<td>0.006</td>
<td>0.007</td>
<td>0.009</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>0.71</td>
<td>1.01</td>
<td>1.05</td>
<td>1.15</td>
</tr>
<tr>
<td>Ramachandran</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored [%]</td>
<td>98.4</td>
<td>98.6</td>
<td>97.3</td>
<td>97.0</td>
</tr>
<tr>
<td>Outliers [%]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* Statistics for the highest resolution shell are shown in parentheses.

b Rmerge = Σ|I-<I>| / Σ|I|, where I is the observed intensity and <I> is the average intensity of several symmetry-related observations.

c Rwork = Σ|Fo-Fc| / Σ|Fo|, where Fo and Fc are the observed and calculated structure factors, respectively.

d Rfree = Σ|Fo-Fc| / Σ|Fo| for 5% of the data not used at any stage of the structural refinement.
Figure 17: Overall structure of Tps2NTD.

(a) Cartoon diagram of Tps2NTD. The N-terminal and C-terminal Rossmann-fold domains depicted as ribbons and colored cyan and pink whilst the C-terminal helix that interacts with both domains is colored red. (b) Superposition of Tps2NTD with *E. coli* OtsA, the latter of which is shown as grey ribbons. Note the significant difference of the N-terminal Rossman fold domains.
Figure 18: Sequence alignment of C. albicans Tps2NTD

Primary sequences of C. albicans Tps2NTD (Q5AI14), E. coli OtsA (labelled as Tps1_EC) (P31677), C. albicans Tps1 (Q92410) and C. albicans Tps3NTD (residues 234-833) were aligned. Highly conserved regions are shaded in the figure. Residues involved in catalysis demonstrated in E. coli OtsA are highlighted by red asterisks. Two key residues involved in E. coli OtsA tetramer formation are highlighted by red pound sign.
(a) Tps2PD selectivity for different substrates. The error bar on the curve represents S.E. of three independent measurements. (b) Tps2PD protein Michalis-Menten kinetics fitting curve. The statistical data were analyzed and figures were generated by GraphPad Prism6. The error bar on the curve represents S.E. of three independent measurements.
Figure 20: Tps2PD overall structure

(a) Two views of the structure of the Tps2PD-trehalose-BeF$_3$-Mg$^{2+}$ transition-state complex. The core domain and cap domains are shown as cartoons and colored cyan and pink. The trehalose and BeF$_3$ are shown as sticks. (b) A 2F$_o$-F$_c$ electron density map of the bound trehalose and covalently bound BeF$_3$, shown as light blue mesh and contoured at 1.5 $\sigma$. Trehalose is shown as pale yellow sticks. The oxygen atoms are labelled where primed oxygens are found on the glucose ring close to the catalytic residue D25. The beryllium and fluoride are colored olive and light cyan, respectively. (c) View of cap-domain residues involved in transition-state stabilization, as depicted by the D25-BeF$_3$ covalent link and Mg$^{2+}$ ion coordination. The cap-domain residues are shown as atom-colored cyan sticks and the magnesium ion and waters are depicted as spheres and colored orange and red, respectively. Selected hydrogen bonds are shown by dashed lines. (d) View of residues and solvent involved in binding the catalytic-residue proximal glucose moiety of trehalose. Residues from the cap-domain are shown
as atom-colored pink sticks and from the core-domain as atom-colored cyan sticks. Hydrogen bonds are shown by dashed lines. (e) View of residues and solvent involved in binding the catalytic-residue distal glucose moiety of trehalose. Hydrogen bonds are shown by dashed lines.
Figure 21: Tps2PD sequence alignments

Primary sequences of *C. albicans* Tps2PD (Q5AI14 residues 535-833), *A. fumigatus* Tps2PD (A0A084BMG6 residues 673-949) and *C. neoformans* Tps2PD (Q059G6 residues 681-987) are aligned. Conserved regions are shaded and residues involved in catalysis and substrate binding are highlighted by red asterisk.
Figure 22: Tps2PD activity assays and closed conformation structure

(a) Relative phosphatase activity of *C. albicans* Tps2PD mutant compared with WT Tps2PD protein. Error bar represents S.E. of three independent measurements. (b) View of T6P in *C. neoformans* Tps2PD closed conformation structure superposed with trehalose and BeF₃ in *C. albicans* transition-state structure. T6P from *C. neoformans* Tps2PD structure is shown as atom-colored yellow sticks. A 2Fₒ-Fₑ electron density map of the bound T6P in *C. neoformans* Tps2PD closed conformation is shown as grey mesh and contoured at 1.5 σ. Trehalose and D25-BeF₃ covalent link are shown as atom-colored grey sticks. Mg²⁺ ions from *C. neoformans* and *C. albicans* Tps2PD structure are shown as orange and grey spheres, respectively.
Figure 23: *C. neoformans* Tps2PD overall structure

(a) Two views of the structure of the Tps2PD closed conformation structure. The core domain and cap domains are shown as cartoons and colored cyan and pink. T6P is shown as atom-colored yellow sticks and Mg$^{2+}$ is shown as orange sphere. (b) Superposition of *C. albicans* Tps2PD with *C. neoformans* Tps2PD. *C. albicans* Tps2PD is shown in cartoon and colored in grey.
Cells were grown initially in YPD broth, washed, and resuspended in diluted Sabouraud (1/10) in MOPS buffer pH 7.3. Cells were grown overnight and stained with India ink. Large capsules were observed for all strains except for tps2D705N, which had a small capsule size and defective in cell division. The tps2NTDΔ strain that retains the catalytic activity but has a deletion of the N-terminal domain had a phenotype similar to WT.
Figure 25: Superposition of *C. albicans* and *B. malayi* Tps2PD

(a) Superposition of *C. albicans* Tps2PD with *Brugia malayi* Tps2PD (residues 206-492) by the core domain. *C. albicans* Tps2PD core domain and cap domain are shown as cartoons and colored in cyan and pink, respectively. *Brugia malayi* Tps2PD is colored in grey cartoon. (b) Superposition of *C. albicans* Tps2PD with *Brugia malayi* Tps2PD (residues 206-492) by the cap domain.
Figure 26: Structures of *A. fumigatus* Tps2PD in open conformation

(a) Side view of *A. fumigatus* Tps2PD structures in open conformation from three different crystal forms. Tps2PD structures are depicted as cartoon and superpositioned together by the core domain. Crystal form 1, 2 and 3 is colored in cyan, pink and wheat, respectively. (b) Top view of superposition of *A. fumigatus* Tps2PD structures from three different crystal forms. (c) Structure superposition of *A. fumigatus* Tps2PD (crystal form 2) and *C. albicans* Tps2PD transition-state structure. The superposition is based on the core domain only. (d) Tps2PD conformational change (open-closed-transition state-open) effected by substrate binding and product release.
Figure 27: Structure of *C. albicans* Tps2PD with docked sucrose

Sucrose is superpositioned with trehalose at the active site in *C. albicans* Tps2PD-trehalose-BeF$_3$-Mg$^{2+}$ transition-state complex structure. Trehalose and sucrose are shown by atom-colored grey and yellow sticks. 6 position of phosphorylated glucose and fructose moiety of T6P and S6P is labelled as T6 and S6 in the figure.
5. Structure of *C. albicans* Tps3PPD reveals a pseudo-phosphatase fold

Trehalose biosynthesis is of critical importance for fungal cell stress response and virulence. Trehalose is synthesized by a multiprotein trehalose biosynthetic complex, including trehalose synthase regulatory protein, Tps3. Despite the necessity of Tps3 for trehalose biosynthesis, lack of structural and biochemical analyses hinders our understandings of Tps3 function in stress response. Here, we report the structure of the *Candida albicans* Tps3 C-terminal domain, which adopts a fold typical of other haloacid dehydrogenase superfamily (HADSF) phosphatases. Molecular dynamics simulations of this domain and subsequent ligand docking reveal a binding pocket for a yet identified ligand. Therefore, Tps3 may represent a novel cross-connection between trehalose biosynthesis and other aspects of cellular metabolism.

5.1 Introduction

Fungal infections, especially invasive mycoses, have increased significantly over past several decades and caused a severe burden to health care systems worldwide. Current antifungal drugs in the clinics often result in serious side effects to patients, primarily due to significant overlap of human and fungal metabolic and regulatory pathways [14,111]. The trehalose biosynthetic pathway, essential for fungal stress response and survival, thus serves as ideal target for antifungal intervention due to its absence in mammalian systems, thereby alleviating potential off-target drug toxicities.
Trehalose, α-D-glucopyranosyl-(1→1)-α-D-glucopyranoside, is a non-reducing disaccharide synthesized by fungal cells and acts as a protective agent for proteins and membranes against internal and external stresses [30,94,96]. Trehalose biosynthesis also rescues yeast cells from unbalanced glycolysis and maintains energy homeostasis [38]. These characteristics underscore the essentiality of trehalose biosynthesis for fungal cellular survival. Indeed, deletions of genes encoding trehalose biosynthesis genes abolish trehalose accumulation, reduce growth rates at high temperature and diminish virulence in murine models [55,56,57,58,59,61].

Trehalose is synthesized by transfer of glucose from UDP-glucose to glucose-6-phosphate to form trehalose-6-phosphate (T6P), followed by dephosphorylation of T6P (Fig. 1). These two reactions are catalyzed by trehalose-6-phosphate synthase (Tps1) and trehalose-6-phosphate phosphatase (Tps2). In *Saccharomyces cerevisiae* (*S. cerevisiae*), Tps1 and Tps2, together with two additional regulatory subunits Tps3 and Tsl1, comprise the trehalose biosynthetic complex. Tps3 and Tsl1 function interchangeably within the trehalose biosynthetic complex [79,80]. Deletion of both Tps3 and Tsl1 destabilized the trehalose biosynthetic complex and reduced trehalose concentration. More specifically, deletion of Tsl1 abolished trehalose accumulation whilst deletion of Tps3 reduced trehalose production [81]. However, except for the requirement for Tps3 for complex formation, the exact structure and function of Tps3 is unknown. Lack of this information hinders our understandings of the regulatory subunit and the trehalose biosynthetic
complex. Therefore, detailed research into Tps3 structure and function is required. Here, we report the structure of the C-terminal domain of Tps3, which we name Tps3 pseudo-phosphatase domain (Tps3PPD) due to its structural similarity to the Tps2PD. Molecular dynamics (MD) simulations and computational ligand docking reveals that Tps3PPD possesses a ligand-binding pocket that is not compatible with T6P binding.

5.2 Results and discussion

5.2.1 Tps3 protein expression and purification

*C. albicans* Tps3 (amino acid residues 1-904) is composed of a long disordered N-terminal peptide (residues 1-130), an N-terminal pseudo-Tps1 domain (residues 131-598) and a C-terminal pseudo-phosphatase domain (residues 599-904, hereafter named Tps3PPD). However, the exact origin of each domain is not known due to lack of previous research. Hence, a series of constructs were designed and cloned into expression vectors. However, none of these constructs led to soluble Tps3 N-terminal pseudo-Tps1 domain (Tps3NTD). Protein refolding also failed to rescue Tps3NTD from aggregates. Interestingly, *C. albicans* Tps2NTD shared 36% sequence identity with Tps3NTD (Fig. 18). Interestingly, Tps3NTD lack more than half of conserved residues in Tps1 enzyme whilst retaining conservation in dimeric interface residues. Previous research has proposed a complex formation function of Tps2NTD, which is essential for cell stress response (Chapter 4). This high sequence identity and the presence of Tps3...
within the trehalose biosynthetic complex suggested a similar complex formation function of Tps3NTD, although Tps3NTD structure has not yet been determined.

A long (130 residues) disordered N-terminal peptide predicted by PSIPRED [110] is present in Tps3. Interestingly, this peptide is rich in prolines, serines and threonines. A blast of the PEST motif (regions rich in proline, glutamic acid, serine and threonine) motif by EPESTFIND [122] reveals a PEST motif between positions 61 to 91 (Sequence: HGNSSATILLSLEDGTADSSPVILPSSLEDDDFDQNFH). Presence of this PEST motif suggests that Tps3 might be degraded rapidly within eukaryotic cells [122], thus implying one potential regulatory mechanism for trehalose biosynthesis. In addition, the presence of serines and threonines provides phosphorylation sites, indicating post-translational modifications might also play roles in trehalose biosynthesis. Indeed, Tps3 from S. cerevisiae has been shown to be phosphorylated [81]. Further cellular analyses are required to understand regulatory functions of Tps3 in yeast cells.

5.2.2 Overall structure of C. albicans Tps3PPD

Tps3PPD (residue 599-904) was expressed and purified to homogeneity as a monomer, shown by size exclusion chromatography. Tps3PPD crystal was obtained by hanging-drop vapor diffusion method. Tps3PPD structure was determined using single-wavelength anomalous dispersion (SAD) methods using selenomethionine (SeMet)-substituted Tps3PPD. The Tps3PPD structure was determined in P65 space group. Experimental phases were calculated by Phenix AutoSol [119]. Tps3PPD structure was
manually built in Coot [108] and the Tps3PPD structure was improved by multiple rounds of refinement against a higher resolution native dataset (1.75 Å resolution) and rebuilding. There is one molecule of Tps3PPD in the asymmetric unit. Electron density for residues 34-46 was not observed, thus not modelled.

The overall Tps3PPD (annotated as residues 1-306) structure consists of one Rossmann-fold core domain (residues 1-113 and 199-306) and one α/β cap domain (residues 114-198) (Fig. 28). The Rossmann-fold core domain is formed by seven parallel β strands flanked by seven α helices. An additional β strand, β5 is antiparallel to the core β sheets and links the core and cap domains through a C-terminal loop downstream of β5. Four antiparallel β strands topped by two α helices form the cap domain. The cap domain reconnects to the core domains through another loop downstream of β9.

The global fold of Tps3PPD resembles that of a typical HADSF type IIB phosphatase. Indeed, superposition of C. albicans Tps3PPD with C. albicans Tps2PD-trehalose-BeF₃ structure revealed a root mean square deviation (rmsd) of 2.4 Å for 290 Cα atoms, while only sharing 22% sequence identity (Fig. 29). DALI search of Tps3PPD structure [99] reveals the structure with the highest structural homology is Tps2-related protein from Thermoplasma acidophilum (PDB: 1U02) with root mean square deviation (rmsd) of 2.8 Å for 229 corresponding Cα atoms (Tps2PD structures from pathogenic fungi have not been released in PDB yet). Despite the high structural similarity, Tps3PPD lacks the four conserved Motifs of the HADSF phosphatases. Structure
alignment showed the aspartate nucleophile of *C. albicans* Tps2PD is an alanine in the Tps3PPD, thus unable to contribute to nucleophilic catalysis and as anticipated, Tps3PPD showed no phosphatase activity against T6P.

### 5.2.3 Potential ligand-binding pocket of Tps3PPD

The physiological function of Tps3PPD remains unknown. However, its structural similarity with Tps2 and lack of phosphatase activity renders Tps3PPD as a pseudo-phosphatase. Emerging data of human pseudophosphatome reveals that 17 human phosphatome members are pseudo-phosphatases based on lack of activity. These pseudo-phosphatases have been shown to be integral parts of signaling networks [123]. Therefore, we posited that Tps3PPD might be involved in binding to sugar substrates of trehalose biosynthetic pathway. Tps3PPD binding with trehalose, T6P and G6P were tested by isothermal titration calorimetry (ITC). However, no binding to Tps3PPD was detected with any of these ligands. Furthermore, structure determination of Tps3PPD with these ligands through *de novo* crystallization experiments failed to reveal any electron density at binding pocket. In an effort to expand our search for ligands that bind to the Tps3PPD, Molecular Dynamic (MD) simulations (Fig. 30) on this domain were carried out and computational docking experiments were employed using the Yeast Metabolite Database (YMDB).

This *in silico* search yielded a series of nucleotides as the top hits for some of the MD ensembles of Tps3PPD. One representative result, that of Tps3PPd binding to UDP-
galactose (YMDB00797) is shown (Fig. 31a). UDP-galactose binds to Tps3PPD at the binding pocket formed between the core and cap domain. The uracil moiety of the UDP-galactose engages in extensive interactions with Tps3PPD whereby its N1 hydrogen bonds to the side chain of D256 and O2 interacts with the side chain of residue Y76. Both the N3 and O4 atoms of uracil interact with the side chain of residue N98. The hydroxyl groups at the 2 and 3 positions of ribose from a hydrogen bond with the side chain of residue N49 and the G251 backbone amide group, respectively. The only notable interaction between the phosphate groups of the UDP nucleotide with Tps3PPD is a contact from residue R190 to the α phosphate. This interaction also marks the only contact between the UDP-galactose with the Tps3PPD cap domain.

In an effort to test the binding of Tps3PPD with nucleotides, a total of 12 potential ligands (including NAD⁺, CoA, UDP-galactose, CDP, ADP, GDP, UDP, UDP-glucose, dATP, dTTP, dCTP and dGTP) were co-crystallized with Tps3PPD. The structures of each were determined by molecular replacement using the Tps3PPD substrate-free structure as search model. However, electron density was not detected at the putative ligand-binding pocket in any of these structures. These results, though not eliminating the possibility of ligand binding potential of these ligands, suggested that these ligands are likely not actual Tps3PPD ligands. Further analyses are needed to identify the ligand, if any, for Tps3PPD and its potential role in the signaling function of Tps3PPD.
In summary, we report here the structure of the Tps3PPD, which adopts a pseudophosphatase fold. Tps3PPD contains a potential ligand/substrate-binding pocket for a yet unidentified ligand that might function in cellular metabolism or signaling. Further analyses are required to understand the physiological function of Tps3PPD and reveal connection between trehalose biosynthesis with other metabolic pathways.

5.3 Future directions

Detailed research is required to understand the function of C. albicans Tps3. Subsequent co-crystallization and isothermal titration calorimetry (ITC) experiments of Tps3PPD with the candidates from our in silico screens are ongoing. Another approach to identify this uncharacterized metabolite is through the use of metabolite screening cocktails [124]. In such an approach, a cocktail of available metabolites would be co-crystallized with Tps3PPD as a fortuitous effort to find physiologically relevant ligands.

Tps3PPD signaling via the binding of a small molecule is one, however, not the only possibility of a regulatory function for Tps3. Tps3 potentially has regulatory functions through interaction with or modulation of other proteins. Indeed, Tps3 has been found to interact with other proteins in Saccharomyces GENOME DATABASE. Clearly, more research is needed to confirm these interactions in pathogenic fungal species and understand the physiological function of these interactions.
5.4 Experimental procedures

5.4.1 Protein expression and crystallization

The Tps3PPD gene was codon optimized for expression in *E. coli* by Genscript. The gene was cloned into the pMCSG7 plasmid using a ligation independent cloning protocol with an N-terminal hexahistidine affinity tag followed by a Tobacco Etch Virus (TEV) protease cleavage site [117]. The vector was transformed into BL21(DE3)plysS cells (Life Technologies) and were then induced with 0.5 mM IPTG at 15 °C for 16 h. Tps3PPD was purified by Ni²⁺-NTA affinity column followed by TEV protease cleavage. Tag-cleaved Tps3PPD was separated by another round of Ni²⁺-NTA purification. Tps3PPD was further purified using size exclusion chromatography via a Superdex S75 column (GE Healthcare) in a buffer containing 20 mM Tris, pH 7.5, 100 mM NaCl, 5% glycerol, 5 mM MgCl₂, and 1 mM βME. Purified Tps3PPD was concentrated to 20 mg/mL using an Amicon Ultra concentrator (10K MWCO, Millipore). Crystals were grown at 25 °C by the hanging drop-vapor diffusion method. Protein crystals appeared after one week from 10% 2-Propanol, 0.1 M sodium citrate, pH 5.5, and 20% PEG 4000. Se-Met substituted protein was purified and crystallized using native protein crystallization conditions. Protein crystals appeared after one week from solutions containing 0.2 M ammonium chloride, 0.1 M MES, pH 6.0, and 20% PEG 6000.
5.4.2 Data collection

X-ray intensity data for the *C. albicans* Tps3PPD crystals and the Se-Met substituted crystals were collected at the Advanced Photon Source (APS) ID-22 and BM-22 line, respectively. Data were indexed and scaled using HKL 2000 [105]. Experimental phases were calculated using Phenix AutoSol [119]. The Tps3PPD structure was manually built with Coot [108] and improved by multiple rounds of refinement against a higher resolution dataset using Phenix refine [107]. Selected crystallographic statistics are presented at Table 6.

*C. albicans* Tps3PPD was co-crystallized with different potential ligands. Crystals were obtained by sparse matrix screening and include both original and *de novo* crystallization conditions. X-ray intensity data for these crystals were collected at the APS ID-22 and BM-22 lines. The data were indexed and scaled by HKL 2000 [105]. These structures were determined by molecular replacement using Phaser and wild type Tps3PPD as the search model [106] and improved by several rounds of refinement using Phenix refine [107]. However, electron density was not detected at the active site, thus the detailed statistics of data collection not reported here.

5.4.3 MD simulation and ligand docking

The initial coordinates of the MD simulation were based on Tps3PPD crystal structure. The missing loop (residue 34 through 46) was rebuilt according to the following procedure. First, 5000 random coil structures containing residue 33 through 46
were generated using the program FOLDTRAJ [125], and each random coil structure was glued to the crystal structure by superimposing residue 33. Among the resulting 5000 conformers, only those with distance between G46-C atom and K47-N atom less than 5Å were chosen for the further optimization. Next, the structures generated in the previous step (50 in total) were subjected to a structure optimization protocol using the GPU-accelerated Amber 14 MD simulation package [126] under ff14SB force field [127]. In this protocol, the structure was gradually heated to 1000 K in 40 ps and then gradually cooled to 0 K in 240 ps. During the optimization, harmonic restraints with a force constant of 500 kcal mol$^{-1}$ Å$^{-2}$ were applied to heavy atoms in all residues expect for those in the added loop. Finally, one structure with a closed loop conformation similar to that of the observed in *C. albicans* Tps2PD transition state structure was picked up for the MD simulation.

The MD simulation was conducted using Amber 14 under the same force field as described above. The protein was solvated using the explicit SPC/E water model [128]. A truncated octahedral water box was constructed such that the boundary of the water box was at least 10 Å away from any of the protein atoms. The hydrated system was neutralized by adding one Na$^+$ ion and subjected to two rounds of energy minimization with 1000 steps using harmonic restraints of the same force constant as described above and another 1000 steps using no restraints, then heated from 0 to 300 K, and equilibrated for 1 ns at 300 K under the Isothermal–isobaric (NPT) ensemble. The final 1 μs MD
simulation was conducted at 300 K using the NPT ensemble. During the simulation, all bonds involving hydrogen atoms were constrained using the SHAKE algorithm[129], and a nonbonded cutoff of 8 Å was used. The integration step was set to 2 fs. A total of 10000 snapshots were uniformly extracted from the MD trajectory. The clustering analysis was performed using ptraj facility in Amber. The rmsd of Cα atoms was used as similarity metric during clustering. Finally, 20 structures were output as representative conformations.

The Internal Coordinate Mechanics (ICM) (Molsoft) [130] docking program was used to dock the Yeast Metabolome Database (YMDB) library of 2323 compounds to 20 representative conformations of Tps3PPD generated by MD simulations. Binding pockets for each receptor were identified using the ICM PocketFinder module and only those near the cleft were selected as receptors for docking. Over the 20 structures, the library was docked against 28 binding pockets. Each compound was docked to each receptor one time with a thoroughness value of 1. Compounds for each binding pocket were ranked by ICM score.

5.5 Acknowledgements

Tps3PPD molecular dynamics simulation and docking experiments were performed by Dr. Yi Xue and Laura Ganser from Al-Hashimi laboratory at the Department of Biochemistry Duke University School of Medicine. These results are essential component of this chapter and thus included for completeness.
Table 6: Selected *C. albicans* Tps3 data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>C. albicans Tps3PPD</th>
<th>C. albicans Tps3PPD Se-Met</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P6₅</td>
<td>P6₅</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>106.9, 106.9, 41.9</td>
<td>107.8, 107.8, 42.1</td>
</tr>
<tr>
<td>α, β, γ (*)</td>
<td>90, 90, 120</td>
<td>90, 90, 120</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-1.75 (1.78-1.75)</td>
<td>50.0-2.11 (2.15-2.11)</td>
</tr>
<tr>
<td>I/σ</td>
<td>7.3 (36.0)</td>
<td>12.7 (71.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100.0)</td>
<td>100.0 (99.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.4 (3.6)</td>
<td>9.9 (6.3)</td>
</tr>
<tr>
<td>Overall figure of merit (FOM)</td>
<td>0.573</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>33.0-1.75 (1.81-1.75)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>27781</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>18.6/21.7</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>2300</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>0</td>
</tr>
<tr>
<td>Water</td>
<td>228</td>
</tr>
<tr>
<td>B-factor</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>33.2</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>-</td>
</tr>
<tr>
<td>Water</td>
<td>39.6</td>
</tr>
<tr>
<td>RMS deviations</td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.004</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>0.79</td>
</tr>
<tr>
<td>Ramachandran (%)</td>
<td>Favored</td>
</tr>
<tr>
<td>Favored</td>
<td>98.3</td>
</tr>
<tr>
<td>Outliers</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup> Statistics for the highest resolution shell are shown in parentheses.

<sup>b</sup> R<sub>merge</sub> = Σ||I - <I>|| / Σ||I||, where I is the observed intensity and <I>|| is the average intensity of several symmetry-related observations.

<sup>c</sup> R<sub>work</sub> = Σ||F<sub>o</sub>| - |F<sub>c</sub>| / Σ|F<sub>o</sub>|, where F<sub>o</sub> and F<sub>c</sub> are the observed and calculated structure factors, respectively.

<sup>d</sup> R<sub>free</sub> = Σ||F<sub>o</sub>| - |F<sub>c</sub>| / Σ|F<sub>o</sub>| for 5% of the data not used at any stage of the structural refinement.
Figure 28: Structure of C. albicans Tps3PPD

(a) Cartoon diagram of C. albicans Tps3PPD. The core Rossmann-fold domain and cap domain depicted as ribbons and colored cyan and pink. (b) The primary sequences of C. albicans Tps3PPD and Tps2PD were aligned. Identical residues are highlighted in blue shade. The secondary structure elements of Tps3PPD are labelled and shown as arrows for α helices and rectangles for β stands. The secondary structure component is also colored as cyan and pin for the core domain and cap domain, respectively.
Figure 29: Comparison of Tps3PPD to Tps2PD

(a) Structure superposition of *C. albicans* Tps3PPD and Tps2PD. The core Rossmann-fold domain and cap domain of Tps3PPD depicted as ribbons and colored cyan and pink. Tps2PD is depicted as ribbons and colored grey. (b) The putative active site of *C. albicans* Tps3PPD and the active site of Tps2PD were aligned. Side chains of Tps2PD Motif DXD, containing the aspartate nucleophile, are shown in atom-colored grey sticks and labelled D25 and D27. The location of corresponding to nucleophile (D25 in Tps2PD) in Tps3PPD is A26, clearly not a nucleophile, is shown as atom-colored cyan sticks.
Figure 30: Structure representation of MD simulations results

Structural superposition of *C. albicans* Tps3PPD MD simulation results. Three representative structures out of the ensemble of 20 are superposed. The structures are depicted as cartoons colored in cyan, pink and olive, respectively. Note the loops modelled onto the structure are in different conformations and were enclosed by a red oval.
Figure 31: Representative Tps3PPD-UDP-glucose complex structure from docking

(a) The structure of *C. albicans* Tps3PPD in complex with UDP-galactose as found by ligand docking program ICM. The core Rossmann-fold domain and cap domain of Tps3PPD depicted as cartoon diagram and colored cyan and pink. UDP-galactose is shown as atom-colored yellow sticks. (b) The potential ligand/substrate binding site of *C. albicans* Tps3PPD docked with UDP-galactose. The side chains of Tps3PPD are shown as atom-colored sticks and labelled. Calculated interactions of Tps3PPD residues with UDP-galactose are shown by dashed lines.
6. Future directions

6.1 Structure determinations of C. neoformans Tps1 and Tps2

The C. neoformans Tps1 is the largest Tps1 amongst the pathogens C. neoformans, A. fumigatus and C. albicans (Fig. 6). A sequence alignment reveals an additional 200 residues inserted within the N-domain domain. However, the function of these extra residues remains obscure. Hence, the full-length C. neoformans Tps1 gene was PCR amplified from a cDNA library and cloned into the pMCSG7 plasmid with an N-terminal hexahistidine affinity tag followed by Tobacco Etch Virus (TEV) protease cleavage site using a ligation independent cloning protocol [117]. The vector was transformed into Rosetta(DE3)plysS cells (Novagen) which were induced with 0.5 mM IPTG at 15 °C for 16 h. Full-length Tps1 was purified by Ni\textsuperscript{2+}-NTA affinity column chromatography followed by TEV protease cleavage. Non-tagged Tps1 was separated by another round of Ni\textsuperscript{2+}-NTA affinity column purification. Tps1 was further purified using size exclusion chromatography via Superdex S200 column (GE Healthcare) in a buffer containing 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol, 5 mM MgCl\textsubscript{2}, and 1 mM βME. Purified Tps1 was concentrated to 20 mg/mL using an Amicon Ultra concentrator (30K MWCO, Millipore) and crystallized with in the presence of either UDP-Glucose or UDP and G6P. Crystals were grown at 25 °C by the hanging drop-vapor diffusion method. Protein crystals appeared after one week from 8% Tacsimate, pH 6.5, 8% PEG3350 (Fig. 32a). Crystal was tested using a home-source Rigaku X-ray
machine. The crystal diffracted to approximately 8 Å resolution (Fig. 32b). Subsequent optimization, seeding and post-crystallization modification failed to improve the resolution. The structure of full-length *C. neoformans* Tps1 thus remains undetermined.

The *C. neoformans* Tps2NTD gene (encoding amino acid residues 1-681) was PCR amplified from cDNA library and cloned into pMCSG7 plasmid and contained an N-terminal hexahistidine affinity tag followed by Tobacco Etch Virus (TEV) protease cleavage site using ligation independent cloning protocol [117]. The vector was transformed into Rosetta(DE3)plysS cells (Novagen), which were then induced with 0.5 mM IPTG at 15 °C for 16 h. The Tps2NTD was purified by Ni²⁺-NTA affinity column chromatography followed by TEV protease cleavage (Fig. 33a). Non-tagged Tps2NTD was separated by another round of Ni²⁺-NTA affinity column purification. The Tps2NTD was further purified using size exclusion chromatography via Superdex S200 column (GE Healthcare) in buffer containing 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol, 5 mM MgCl₂, and 1 mM βME. The calculated molecular weight using a standard curve of representative globular proteins revealed that the Tps2NTD is a monomer in solution (Fig. 33b). Purified Tps2NTD was concentrated to 12 mg/mL using Amicon Ultra concentrator (30K MWCO, Millipore). Crystallization trials were performed at 25 °C by hanging drop-vapor diffusion methods. However, to date no crystals of the *C. neoformans* Tps2NTD have been obtained.
The *C. neoformans* full-length TPS2 gene was PCR amplified from a cDNA library and cloned into pMCSG7 plasmid using ligation independent cloning protocol [117]. The construct has an N-terminal hexahistidine affinity tag followed by Tobacco Etch Virus (TEV) protease cleavage site. The vector was transformed into Rosetta(DE3)plysS cells (Novagen) and induced with 0.5 mM IPTG at 15 °C for 16 h. Tps2 was purified by Ni²⁺-NTA affinity column chromatography followed by TEV protease cleavage (Fig. 33a). Non-tagged Tps2 was separated by another round of Ni²⁺-NTA purification. Tps2 was further purified using size exclusion chromatography via Superdex S200 column (GE Healthcare) in a buffer containing 20 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol, 5 mM MgCl₂, and 1 mM βME. Tps2 elutes as a monomer in solution (Fig. 33b). Purified Tps2 was concentrated to 12 mg/mL using an Amicon Ultra concentrator (30K MWCO, Millipore). Crystallization trials of Tps2 in complex with trehalose were performed at 25 °C using hanging drop-vapor diffusion methods in complex with trehalose. However, to date no crystals of the *C. neoformans* Tps2-trehalose complex have been grown.

Compared to the *C. albicans* and *A. fumigatus* Tps1 proteins, the extra residues of the *C. neoformans* Tps1 might result in poorly packed crystals. Secondary structure prediction reveals the majority of the extra residues (except for N-terminal and C-terminal tail of Tps1) is folded and suggests these extra residues might compose an additional domain. Despite continuous efforts, deletion of N-terminal and/or C-terminal tail of the *C. neoformans* Tps1 (N55Δ/ C42Δ/ N55Δ C42Δ) failed to increase the resolution.
To improve the resolution of *C. neoformans* Tps1, additional truncations will be applied. Compared to *C. albicans* and *A. fumigatus*, *C. neoformans* does not possess the additional Tps3 regulatory subunit. It is anticipated that the structures of the *C. neoformans* Tps1 and Tps2NTD will reveal insights into the underlying mechanism how *C. neoformans* circumvent the need of Tps3.

### 6.2 Structure determinations of *A. fumigatus* Tps2NTD and Tps3

Regardless of our current structures of Tps2NTD and Tps3PPD from *C. albicans*, *A. fumigatus* Tps2NTD and Tps3 structures might still reveal species-specific properties. The Tps2NTD gene from *A. fumigatus* strain Af293 was codon optimized for expression in *E. coli* (GenScript). This codon optimized gene was cloned into a pET-28a kanamycin-resistant vector via restriction sites Ndel and SacI. This vector contained an N-terminal hexahistidine affinity tag followed by a thrombin cleavage site. It was transformed into BL21(DE3)plysS cells (Life Technologies) and induced with 0.5 mM IPTG at 15 °C for 16 h. Purification of *A. fumigatus* Tps2NTD is ongoing.

The *A. fumigatus* Tps3 full length gene was codon optimized for expression in *E. coli* by Genscript and cloned into a standard pUC57 vector. Sequence alignment reveals that Tps3 consists of a disordered N-terminal peptide, a pseudo-Tps1 domain (Tps3NTD) and a C-terminal pseudo-phosphatase domain (Tps3PPD). A series of constructs are being generated for cloning TPS3NTD and TPS3PPD into the pMCSC7
vector. This vector contains an N-terminal hexahistidine affinity tag followed by a TEV protease cleavage site. Test expressions and purifications are underway.

**6.3 Inhibitor design against Tps1 and Tps2**

Accounting for approximately 1-2% of the genome of most bacterial or eukaryotic organisms [62], glycosyltransferase genes are major components of the whole genome. Inhibitors against glycosyltransferases emerged around 40 years ago with Paulsen’s innovative idea of using a nitrogen to replace the sugar ring oxygen [131]. In the past 20 years, this field has attracted tremendous attention due to its potential pharmaceutical use. The main method for inhibitor design is carbohydrate mimetics, which include iminosugars, carbasugars and bisubstrate or trisubstrate analogues [131,132,133]. These studies provide significant information and demonstrate a roadmap for promising future inhibitor designs against Tps1. In addition to substrate mimetics, high-throughput lead screening techniques using commercial libraries [134] and *in silico* docking [135] are also frequently used approaches to target inhibition. In collaboration with the Richard E. Lee laboratory at St. Jude Children’s Research Hospital and John R. Perfect laboratory at Duke University Medical Center, we propose to utilize high-throughput screening to search for potential inhibitors. High-throughput screening based on a Tps1 enzymatic activity assay or biophysical properties, e.g., thermal denaturation changes upon ligand binding, are underway. Initial enzymatic activity screening of *C. albicans* Tps1 has revealed a lead compound, closantel, a commercially
available anti-parasite drug used in sheep and cattle, can inhibit Tps1 [136].

Crystallization trials have begun using the C. albicans Tps1 and closantel. However, the resulting crystals have failed to yield the resolution required to resolve the structure. Crystal screening will continue. Furthermore, lead compounds from future screenings will be tested for their ability to crystallize with Tps1 in an effort to understand the atomic details of inhibitor binding to Tps1. Rational redesign of the lead compound will be based on this structural information in order to improve the inhibitory effects of the lead compound.

A significant concern with respect to Tps2 inhibitor design is nonspecific inhibition of other phosphatases, which can lead to toxicity. This is a primary underlying reason for the few successes of phosphatase inhibitor designs. However, one well-studied phosphatase is protein tyrosine phosphatase 1B, which functions as a regulator of insulin and serves as a potential drug target for type 2 diabetes. More than four general classes of inhibitors have been designed successfully for use against this specific phosphatase [137]. This information suggests that inhibitor design against Tps2 using either high throughput screening or in silico docking is feasible. As a proof of concept, trehalose-6-sulfate has been used as an inhibitor against the Brugia malayi Tps2 and showed a Ki of 82 μM [72]. Compound design with a scaffold of trehalose or VDM is ongoing and serves as one starting point for the creation of specific Tps2 inhibitors. These lead compounds will be co-crystallized with Tps2PD from multiple pathogenic
fungal species. Structural information from these protein-compound complexes will guide future Tps2-specific inhibitor optimization.

6.4 Dissection and reconstitution of trehalose biosynthetic complex

The trehalose biosynthetic complex is composed of Tps1, Tps2, Tps3 or Tsl1 in *S. cerevisiae* [79,80]. However, in *C. neoformans*, only Tps1 and Tps2 have been identified. The *C. neoformans* Tps1, Tps2NTD and Tps2PD proteins have been purified to homogeneity. However, ITC, co-expression and pull-down assays have all failed to detect a direct interaction between these proteins (data not shown). In *C. albicans*, only one copy of Tps3 has been identified. Tps1, Tps2NTD, Tps2PD and Tps3PPD have been purified to homogeneity. However, our preliminary pull-down and size exclusion chromatography assays have failed to show binding between the proteins. These results suggest the reconstitution of the trehalose biosynthetic complex might require additional proteins or post-translational modification or some other small molecule factors to fully assemble. Indeed, additional protein components have been identified to bind to trehalose biosynthetic complex proteins through yeast two-hybrid analysis (*Saccharomyces* GENOME DATABASE). In addition, Tps3 from *S. cerevisiae* has been shown to be phosphorylated, which is significant for Tps3 function [81].

To circumvent the need to identify uncharacterized component of the trehalose biosynthetic complex or ensure proper post-translational modification of each component, the most straightforward method is to purify the complex from the germane
cells. So far, the top candidate for extraction of the complex is *C. albicans*. The structures of Tps1, Tps2NTD, Tps2PD and Tps3PPD from *C. albicans* are available. Tps3NTD structure could be generated using structure homology-modelling server SWISS-MODEL [138]. Additionally there is only one copy of Tps3 in *C. albicans* and no identified Tsl1, which might allow a homogeneous trehalose biosynthetic complex to assembly as compared to the heterogeneous complexes that likely occur due to the interchangeability of Tps3 and Tsl1 in *S. cerevisiae*.

To extract the trehalose biosynthetic complex from *C. albicans*, a tandem affinity purification tag [139] be added to one of chromosomally encoded Tps1, Tps2, and Tps3. Cells will be exposed to heat shock at 42 °C or desiccation to induce overexpression of the trehalose biosynthetic complex proteins. The complex will be purified utilizing a standard tandem affinity purification protocol and subsequently subjected to mass spectrometry to identify each component within the complex. Ideally, if the trehalose biosynthetic complex is purified to homogeneity, the complex structure could be determined by small angle X-ray scattering methods [140] or single particle cryo-EM [141] utilizing current crystal structures of individual component. Less ideally, the proteins within the trehalose biosynthetic complex would be cross-linked and subject to enzymatic digestion and mass spectrometry to identify interacting residues/domains. Identification of the complex will shed light on the function of complex formation and identify inter-connection of trehalose biosynthesis with other metabolic pathways. In
addition, structure of the trehalose biosynthetic complex might reveal novel targets for interfacial inhibitor design.

6.5 Acknowledgements

Cloning of the C. neoformans genes from a cDNA library was performed by Dr. Jennifer L. Tenor from the Perfect laboratory at Duke University Medical Center. The initial Tps1 high-throughput inhibitor screening was performed by Dr. William Shadrick of the Richard E. Lee laboratory at St. Jude Children’s Research Hospital. These data were included for a fuller discussion of the future directions of this project.
Figure 32: Crystallization of *C. neoformans* Tps1

(a) Crystals of *C. neoformans* Tps1. (b) Diffraction of *C. neoformans* Tps1 crystal using home source X-ray (Rigaku) equipped with imaging plates.
Figure 33: Purification of the *C. neoformans* Tps2NTD and full-length Tps2

(a) SDS-PAGE gel of purified *C. neoformans* Tps2NTD and Tps2. Lane 4 and lane 8 shows highly purified Tps2NTD and Tps2 after Ni$^{2+}$-NTA column purification, respectively. Molecular weight markers are shown in lane M with selected molecular weights given on the side. (b) Tps2 size exclusion chromatography by S200 column (GE Healthcare).
**Appendix A: PDB accession numbers**

Accession number of crystal structures deposited in the Protein Data Band (PDB)

5DX9: Structure of *C. neoformans* Tps2PD D24N in complex with T6P

5DXF: Structure of *C. albicans* Tps2NTD

5DXI: Structure of *C. albicans* Tps2PD in complex with trehalose and beryllium fluoride

5DXL: Structure of *A. fumigatus* Tps2PD crystal form 1

5DXN: Structure of *A. fumigatus* Tps2PD crystal form 2

5DXO: Structure of *A. fumigatus* Tps2PD crystal form 3

5HUT: Structure of *C. albicans* Tps1 in complex with UDP-Glucose

5HUU: Structure of *C. albicans* Tps1 in complex with UDP and G6P

5HVL: Structure of *C. albicans* Tps1 in complex with UDP and VDM

5HUV: Structure of *C. albicans* Tps1 E341R/E346R in complex with UDP-Glucose

5HVM: Structure of *A. fumigatus* Tps1A in complex with UDP and VDM

5HUO: Structure of *A. fumigatus* Tps1B in complex with UDP and VDM

5HUS: Structure of *C. albicans* Tps3PPD
Appendix B: X-ray protein crystallography

Macromolecular X-ray crystallography was employed to determine structures of multiple proteins in this dissertation. These results laid the foundation for this dissertation and represented the majority work of my graduate studies. Here, I briefly describe the general principles of X-ray protein crystallography. I also share here the experience of working with proteins involved in trehalose biosynthesis.

**X-ray crystallography, NMR and Cryo-EM**

Proteins are macromolecules in cellular organisms that are composed of long chains of amino acid residues, which perform various functions in cells. The function of a protein is primarily determined by its three-dimensional structure. Therefore, there have been continuous efforts to determine structures of various proteins. To solve the secondary structure of proteins and side chain identities of amino acids, atomic resolution of protein structures, though in some cases, subatomic resolution, is required. To date, three major methods, X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM) are utilized to achieve atomic resolution protein structures, the latter two methods were not utilized in this dissertation due to their inherent limitations.

NMR is a method for determination of atomic resolution protein structures, gaining information of dynamics of proteins as well. However, protein sizes (typically under 35 kDa) often limit the practical use of NMR in structure determination [142] of
large proteins. In my dissertation, the molecular weights of Tps1 to Tps3 range from 55 kDa to over 100 kDa. The sizes of these proteins limit the potential of NMR to determine these structures. However, NMR could be complementary method to probe the dynamics of specific domain (for example, Tps2PD from *A. fumigatus*).

Cryo-EM has long been used [143] to determine structures of large macromolecules, especially viral capsids. Recently, cryo-EM, particularly, single particle cryo-EM, has gained tremendous attention as a result of the substantial progress of data-collection software and hardware [141]. These advances permit the determination of atomic resolution structures. Compared with X-ray crystallography, cryo-EM does not need protein crystals for data collection, thus circumventing the “trial and error” step of protein crystallization. In addition, electron micrographs contain phase information required to calculate electron density map. However, to date (12/01/2015), only two cryo-EM structures below 2.5 Å have been deposited in the EMDataBank. The higher resolution 1.9 Å structure of lens-specific aquaporin was determined by electron crystallography of a double layered two-dimensional crystal [144]. The other 2.2 Å resolution structure of β-galactosidase in complex with the cell-permeant inhibitor thus represents the highest resolution structure achieved with single particle cryo-EM so far [145]. Low success rates associated with atomic resolution protein structures using cryo-EM renders this method less favorable in structure determination of single proteins in this dissertation. However, cryo-EM could be a complementary method for solving the
structure of the trehalose biosynthetic complex or any component resisting crystallization.

In summary, X-ray crystallography served as the more suitable methodology and was employed in this dissertation to determine the structures of Tps1, Tps2 and Tps3 involved in the trehalose biosynthetic pathway. In the following sections, I address general principles of protein X-ray crystallography.

**Diffraction basics**

To resolve the atomic resolution structures of proteins, electromagnetic radiation of similar wavelength as atomic bond distances is required. Such wavelength or radiation belongs to X-rays (0.1 – 100 Å), first discovered by Wilhelm Röntgen. In 1912, Max von Laue recorded X-ray diffraction of a copper sulfate crystal, which was marked as a milestone for X-ray crystallography [146]. The diffraction of X-ray by a salt crystal established the electromagnetic wave nature of X-rays. Because of the wave nature of X-rays, defined by amplitude and phase, X-ray could be expressed as a vector $F$ in a complex plane. Superposition of waves then could then be simplified as addition of vectors in the complex plane. These significantly facilitate mathematical calculations and understanding of X-ray diffraction. Vector $F$ in the complex plane could further be expressed as a real component as $F \cos \Psi$ and an imaginary component $F \sin \Psi$, where $i$ is root square of -1. A mathematical superposition of multiple waves could be shown as:
An atom is composed of a positively charged, dense nucleus and negatively charged electrons moving around the nucleus in orbits. The distribution of electrons around the nucleus is defined as ρ(r). The diffraction of x-ray by a single atom is defined as atomic scattering factor F and equation is shown below, where S is the difference vector between reflective wave vector and incident wave vector.

\[
F = \sum_{j} F_j = \sum_{j} F_j (\cos \varphi + i \sin \varphi) = \sum_{j} F_j e^{i\varphi}
\]

From the atomic scattering factor of a single atom, diffraction from two atoms is a combination of each wave. However, this contribution can be constructive or destructive depending on the phase of individual waves. If we hypothesize the amplitudes of these two waves are identical, the final combined amplitude could be twice the amplitude or totally cancel each other. Sir William Lawrence Bragg derived a simplified interpretation of interference construction, leading to the famous Bragg’s law (Fig. 34). The distance between the parallel planes that dissect two atoms is defined as d while the incident X-ray is reflected by an angle of θ. In order for two reflective waves to be constructive and show maximum amplitude, their phases must be identical. The travelling distance between two reflective waves is 2dsin θ. This must equal the integer of the wavelength (λ) of the X-ray, thus the Bragg’s law could be written as:

\[
n\lambda = 2d \sin \theta
\]
Similarly to diffraction of two atoms, scattering from a molecule is the summation of diffraction of each atom within this molecule. Therefore, the total scattering factor can be written as:

\[ F_s = \sum_j^n f_{s,j} \exp(2\pi i S r_j) \]

In practice, focusing X-rays on a single molecule and detecting single molecule diffraction is impractical. The earliest diffraction patterns were observed on salt crystal, a three-dimensional periodic lattice of molecules. The basic unit of a crystal is called the unit cell, which can be translated in all directions and form the three-dimensional lattices. Scattering from a crystal, therefore, can be considered the addition of the scattering of each unit cell. Scattering of a unit cell is the addition of scattering of each atom within the unit cell. However, as established by Bragg’s law, constructive diffraction needs to meet certain conditions.

To ease the understanding of diffraction and simplify the diffraction conditions, one new concept, reciprocal lattice, has to be introduced. As stated above, three-dimensional lattices could be considered a repetition of unit cells. A three-dimensional unit cell could be defined by six parameters, three vectors \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \) and three angles \( \alpha, \beta \) and \( \gamma \). \( \alpha, \beta \) and \( \gamma \) are the angles between vector \( \mathbf{bc}, \mathbf{ac} \) and \( \mathbf{ab} \). A reciprocal space is also defined by three vectors, \( \mathbf{a}^*, \mathbf{b}^* \) and \( \mathbf{c}^* \). \( \mathbf{a}^*, \mathbf{b}^* \) and \( \mathbf{c}^* \) are normal to real space planes \( \langle \mathbf{b}, \mathbf{c} \rangle, \langle \mathbf{a}, \mathbf{c} \rangle \) and \( \langle \mathbf{a}, \mathbf{b} \rangle \). The amplitudes of reciprocal lattice vectors follow equation:

\[ \mathbf{a} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{b}^* = \mathbf{c} \cdot \mathbf{c}^* = 1 \]
A real space lattice could be sliced in multiple pieces by Miller indices \((h, k, l)\), which \(h, k\) and \(l\) are all integers. The distance between these real space lattice planes are thus defined as \(d_{hkl}\). In a reciprocal space, these Miller indices planes \((h, k, l)\) correspond to the reciprocal lattice points \((h, k, l)\).

\[
d'_{hkl} = \frac{1}{d_{hkl}}
\]

With the introduction of reciprocal lattice and Miller indices, revisit the scattering will simplify the mathematical calculations and diffraction conditions. To visualize the diffraction, a new concept, Ewald sphere, is introduced (Fig. 35). Ewald sphere is a sphere centered at crystal center with a radius of \(1/\lambda\). The interconnection (one closer to detector) of Ewald sphere and incoming X-ray is defined as the origin of reciprocal lattice. To have constructive inflection, Bragg’s law conditions have to be met. Therefore, the following equations apply.

\[
\frac{n\lambda}{2 \sin \theta} = d_{hkl} = \frac{1}{d'_{hkl}}
\]

As defined in the reciprocal lattice, \(d'_{hkl}\) is the distance between origin and a reciprocal lattice point \((h, k, l)\). As shown in Figure 35, the distance of \(d'_{hkl}\) is:

\[
d'_{hkl} = \frac{2 \sin \theta}{\lambda}
\]

Therefore, any reciprocal lattice point on Ewald sphere meets the conditions of Bragg’s law. In addition, any orders of this \(hkl\) point also meet Bragg’s law since the integer \(n\) is omitted in this equation. From this point, X-ray crystallography diffraction is
simplified as Ewald sphere and reciprocal lattice point. With several parameters known, diffraction pattern could be predicted. Wavelength of X-ray, unit cell parameters (reciprocal lattice parameters), incident X-ray angle and crystal position are basics to estimate diffraction. To have more reciprocal lattice points intersecting with Ewald sphere, we have to rotate the reciprocal lattice. However, in practice, X-ray diffraction is also limited by other parameters; for example, detector size and the inherent packing of the macromolecules in the crystal, which can limit crystal diffraction. These issues are not covered in detail here in this dissertation.

At this point, we have established the conditions for protein crystal diffraction. In a typical X-ray crystallography experiment, we obtain the diffraction pattern and intensity of each reciprocal lattice point. However, to determine protein structures, this information has to be translated into more interpretable results for model building. In crystallography, this interpretable information is electron density map, which accurately reflect atomic positions within the asymmetric unit. Here, I will describe how the initial recorded data can be translated into electron density information.

To define a position of an atom within the unit cell, we use \((x, y, z)\) to describe the position. Within the unit cell parameter, any atom position relative to unit cell origin can be written as:

\[ r = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} \]
Also as mentioned above, reciprocal lattice point (h, k, l) has to be on Ewald sphere to have constructive reflection. The S vector (Fig. 35) in this case, is identical to d\^{*}{hkl}. Therefore, in the structure factor equation, the following component could be simplified as:

\[ \mathbf{S}_r = (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*)(xa + yb + zc) = (hx + ky + lz) \]

Therefore, structure factors could be expressed as:

\[ F_{hkl} = \int \rho(x,y,z)\exp(2\pi i(hx + ky + lz)) \]

Furthermore, the electron density is the Fourier transform of the structure factors (with phases). The back Fourier transform of this equation:

\[ \rho(x,y,z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| \exp(-2\pi i(hx + ky + lz) + i\alpha'_{hkl}) \]

These two equations thus represent the fundamentals of X-ray crystallography, correlating structure factors with electron density in real space. In X-ray crystallography experiments, the intensity of diffraction is recorded, which is proportional to |F\^{*}{hkl}|^2. One more factor, \(\alpha_{hkl}\) (the phase angle) is required to calculate the electron density map.

**Protein crystallization and space group**

As stated above, X-ray diffraction on a single protein molecule is not feasible, thus requiring proteins to be arranged into a periodic three-dimensional lattice, or crystal, in order to obtain measurable X-ray diffraction data. In order for protein crystals to form, homogenous protein (95% purity or higher) is driven to supersaturated state
(Fig. 36), which in turn can drive spontaneous nucleation of protein molecules.

Nucleation is subsequently followed by crystal growth and results in a protein crystal with reasonable volume. To drive protein into the supersaturated state, crystallization cocktails, composed of different precipitants, are frequently utilized to reduce protein solubility. One of the most common precipitants is different molecular weight polyethylene glycols (PEGs). PEGs (HO-(O-CH₂-CH₂)ₙ-OH) compete with protein for water molecules in solution, thus reducing protein solubility. Other common variable factors in crystallization include pH and salt concentration. Successes with protein crystallization have led to commercially available screening kits, which contain conditions that have generated protein crystals. These frequently serve as starting point for protein crystallization screening.

Several different crystallization techniques are commonly utilized, including vapor-diffusion, batch crystallization and dialysis. In this dissertation, hanging-drop vapor diffusion methods were performed (Fig. 37) using 48-well plates with sealant and siliconized glass cover slides (Hampton Research). Initial crystal screenings were performed using sparse matrix sampling methods utilizing commercial crystallization kits, including Wizard (Rigaku Reagents), Wizard Cryo (Rigaku Reagents), Index (Hampton Research), PEG-ION (Hampton Research) and PEG-Rx (Hampton Research). Proteins were purified to homogeneity (95% pure by SDS-PAGE) and mixed with an equal volume of well solution on siliconized glass cover slides. These cover slips were
flipped and sealed over one of 48 wells, each of which was filled with 200 µL of a
different crystallization solution. After obtaining initial crystal hits, crystals were
optimized by varying well solution conditions (including pH and precipitant
concentration) and protein concentration. Micro-seeding was also utilized to facilitate
crystallization. To tolerate radiation damage in data collection, protein crystals in this
dissertation are at least 50 µm in dimension. These crystals were tested using a home-
source X-ray data collection instrument in order to assess the diffraction limit. Well-
diffracting crystals were flash frozen and sent to synchrotron sources.

**Data collection and process**

In order to reduce radiation damage to protein crystals that is caused by X-rays,
all data sets were collected at cryogenic temperatures (100 K). Protein crystals were flash
frozen in liquid nitrogen. To prevent ice formation in the crystals during flash freezing,
additional cryo-protectants were normally needed. On some occasions, protein crystals
appeared from (*C. albicans* Tps1 in 40% PEG 400) the crystallization solution with
enough cryo-protectant already present and thus circumvented the need for additional
cryo-protectant. In this dissertation, additional ethylene glycol or glycerol was added to
their crystallization buffers as their additional cryo reagents in the original
crystallization buffers failed to include enough cryo-protection. The concentration of the
additional ethylene glycol or glycerol ranged from 5% to 25%. Before sending crystals to
synchrotron sources, initial exposure and testing of the cryo conditions were performed
using a Rigaku FRE+ X-ray instrument located in the Brennan and Schumacher laboratories. Two key criteria for cryo-condition selection are the prevention of ice ring formation and minimized influence on the crystal diffraction resolution and crystal mosaicity. The former is achieved by naked eye observation while the latter requires comparison between room temperature snapshots and cryo-condition snapshots. After characterizing the cryo-protectant, crystals were harvested by cryo-loops, flash frozen, stored in crystal pucks and a Dewar (Hampton Research).

In theory, we could collect 360° of data on every single crystal and calculate electron density information based on the diffraction results. However, in practice, radiation damage poses significant problems for this strategy and it is improvident not to utilize the symmetry present within the crystal. Protein crystal symmetry reduces data collection time and simplifies subsequent calculations during refinement. Three-dimensional lattices could be considered translations of unit cells. Seven different crystal systems generate three-dimensional repeats only by translation (Table 7). Depending on whether lattice points are present only on the corners of unit cell or include other origins, the lattice can be described as primitive (P), face-centered (F or C) or interior (I). Together, these lead to 14 Bravais lattices (Table 7). In addition, 1, 2, 3, 4 and 6 fold rotational symmetry, can be present within the unit cell. Screw axes are also allowed in the unit cell with rotational symmetry followed by translation along an axis or diagonal
direction. Together, 230 space groups are allowed for crystals with only 65 space groups for protein crystal, due to the stereochemical nature of their amino acids residues.

Diffraction data in this dissertation were collected at the Advanced Photon Source (APS) or the Advanced Light Source (ALS). Here, I focus on data collection using synchrotron sources, more frequently utilized for *de novo* structure determination due to tunable wavelengths. Crystals were shipped to these synchrotron sources and loaded in front of the x-ray beam with the assistance of beamline scientists and mounting robot. X-ray diffraction was recorded on charge coupled device (CCD) detectors. HKL2000 [105] was utilized to index the first several frames collected. Indexing of the reflections assigns the parameters for the unit cell and further ranked Bravais lattices by a penalty function. Based on this information, a collection strategy was generated for a complete dataset. More data were collected if the space group of the crystal was not known for sure.

Typically, data were collected with exposure times of 1 second and an oscillation angle of 0.5 degree with a fixed axis at synchrotron. Intensity information was recorded and is proportional to the square of structure factors. After data collection, these oscillation frames were integrated and scaled together to a final scaled file. This file contained space group information, the six parameters for unit cell, the intensity (I_{hkl}) of each reflection and their attendant intensity errors.

A few parameters have been proposed to assess the quality of a crystallographic dataset. The most common quality indicator [147] is $R_{merge}$. 

143
\[ R_{\text{merge}} = \frac{\sum_{hkl} \sum_{j} |I_{hkl,j} - <I_{hkl}>|}{\sum_{hkl} \sum_{j} I_{hkl,j}} \]

In this equation, \(<I_{hkl}>\) represents the average of intensity of N observations for a unique (h, k, l). \(R_{\text{merge}}\) value is a representation of merge of symmetry related reflections, typically below 0.1 for reasonable dataset. Due to the nature of \(R_{\text{merge}}\), it also indicates the correct symmetry. If the scaled symmetry is higher than the actual symmetry, \(R_{\text{merge}}\) values are much higher. Therefore, this value is significant to differentiate correct symmetry, especially ambiguity of Bravais lattice type during index. Another important quality indicator is \(I/\sigma I\) (recorded intensity over its standard deviation). This parameter is normally utilized to determine the cutoff of the highest resolution shell (normally demonstrating \(I/\sigma I\) value higher than 2.0).

To this point, the obtained scale file contains intensity information of each (h, k, l), from which structure factor \(F_{hkl}\) is available. However, to calculate electron density, one more parameter \(\alpha\) (phase) for each structure factor is still required. Currently, molecular replacement and experimental phasing (more commonly SAD or MAD) are utilized for phase determination. In the following sections, I will discuss these two methods separately.

**Structure determination by molecular replacement**

Molecular replacement (MR) is generally the more convenient method for phase determination as compared to experimental phasing, as there is no need of additional experimental data. Currently, with the increasing number of structures deposited in the
RCSB Protein Data Bank, the success of molecular replacement has significantly increased. MR utilizes phase information directly from a search model for the unknown dataset. In order for MR to succeed, the search model has to share almost identical structure. However, structural identity information is unknown before MR, thus amino acid sequence identity serves as a starting point for MR. MR frequently utilizes current PDB structures with the highest sequence identities to the unknown protein structure, normally 30% or higher. In cases when structural similarity information is available, search models with lower sequence identities but high structural similarity could also result in reasonable solutions. This requirement of structural similarity rather than sequence identity also suggests structures with large conformational change could fail to serve as search model despite high sequence identity. For example, *A. fumigatus* Tps2PD shares 58% sequence identity with *C. albicans* Tps2PD. The MR determination of the *A. fumigatus* Tps2PD open-conformation structure utilizing *C. albicans* Tps2PD transition-state structure as the search model failed.

To describe the principles of MR in detail, we have to introduce one new concept, the Patterson function, $P(u, v, w)$:

$$P(u, v, w) = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 \exp(-2\pi i (hu + kv + lw))$$

The Patterson function contains the interatomic distances (vectors) of all atoms within the unit cell. For a unit cell of N atoms, the Patterson map will contain N peaks at 0 position and N(N-1) interatomic distance peaks (vectors). As shown in the equation,
the Patterson function can be calculated from the structure factors and is thus directly available from the scaled data file.

The Patterson function maps for unknown structures can be calculated directly from the scale file. Similarly, the search model Patterson function map can be derived from coordinates. In theory, a 6-dimensional (three rotation angles and three translational degrees) method can be utilized to correlate these two Patterson functions. However, in practice, such method requires enormous computational loads. Therefore, a three-dimensional rotation function followed by another three-dimensional translation function is typically utilized in MR search programs. The rotation function $R(R)$ aims to determine the rotation angles that are necessary to best match the unknown structure to the search model. Self-Patterson vectors (intramolecular Patterson vectors) $P_s(r)$ are utilized for maximum overlap of the unknown structure and the search mode.

$$R(R) = \frac{1}{V} \int_{r_{\text{min}}}^{r_{\text{max}}} P_T(r) \cdot P_S(R^T r) d^3r$$

After the orientation of the structure has been determined, a translation function $T(R)$ based on intermolecular Patterson $P_I(r)$ functions are utilized for determining actual position of the unknown model within the unit cell.

$$T(R) = \int P_I(r)P_S(r - R) d^3r$$

MR further utilizes the phase information of the search model for calculating the electron density map for this unknown structure. Additional structure refinement is required since the structure is heavily biased towards the search model. In Phaser [106]
within the Phenix suite [148], a fast rotation function with a fast translation function is incorporated for MR calculation. The translational function Z-score (TFZ) is primary indicator whether MR solution is successful, typified by values > 8 as a confident solution. However, structure refinement followed by manual inspection is more appropriate for final decision.

**Structure determination by SAD**

The *de novo* determination of structures of proteins and their complexes often requires additional experimental data for phase determination. Multiple isomorphous replacement (MIR) utilizing heavy atom derivatized crystals are utilized for phase determination and requires both a native protein dataset and the heavy-atom derivative dataset. After obtaining the scaled file for both datasets, the structure factors $|F_{PH}|$ (heavy atom derivative) and $|F_P|$ (protein dataset) are available. In addition, Patterson functions from these data can be calculated. MIR utilizes the difference Patterson function:

$$P(u, v, w) = \frac{1}{V} \sum_{hkl} (|F_{PH}| - |F_P|)^2 \cos(2\pi(hu + kv + lw))$$

The difference Patterson function calculated here will only contain vectors contributed by heavy atoms alone. Therefore, the heavy atom positions will be readily known. After obtaining the heavy atom positions, structure factors for heavy atoms $F_H$ are known. Combined with the structure factor amplitudes $|F_{PH}|$ and $|F_P|$, protein phases can be calculated from the Harker construction (Fig. 38). Two solutions are
obtained from the Harker construction, thus there is phase ambiguity, which is typically resolved by at least one additional heavy atom derivative.

However, MIR faces a few drawbacks. First, obtaining heavy atom derivatized crystals without affecting crystal diffraction quality is difficult. Second, the presence of heavy atoms results in more radiation damage to the crystal during data collection. More important, these dataset are required to be isomorphous, which requires all dataset to have the same space group, cell dimensions and etc. This can be difficult to achieve in practice. Anomalous scattering methods, especially Seleno-Methionine (Se-Met) -based SAD or MAD is frequently utilized nowadays. In this dissertation, all experimental phasing was performed by SAD methods. Anomalous scattering is caused by X-ray absorption by elements at specific wavelengths, resulting in additional components to the scattering factor: a real component $f'(\lambda)$ of absorbed photons emitted as lower energy and an imaginary component $f''(\lambda)$ of absorbed photons reemitted at the same wavelength. Therefore, the atomic scattering factor is expressed in a new form:

$$f(\lambda) = f^0 + f'(\lambda) + f''(\lambda)$$

A crystal of a Se-Met substituted protein is frequently utilized for anomalous scattering experimental phasing. A fluorescence scan resulting from a Se-Met protein crystal (A. fumigatus Tps2PD) is shown in Figure 39. With the introduction of anomalous contributions, Friedel pair amplitudes are no longer identical (Fig. 40). Similar to isomorphous replacement, the differences between $|F_{PA}|$ and $|F_{PA}|$ can be employed to
locate the heavy atom positions, the selenium sites, allowing the downstream calculation of *de novo* phases. Compared to MIR, data is collected on a single crystal, thus circumventing the need of isomorphism. Alternatively, instead of MAD methods, SAD data are frequently collected using inverse beam mode, reducing radiation damage.

Similar to MIR, the phase ambiguity of SAD can be resolved by additional data, thus named MAD. Further, density modification methods are gaining popularity in resolving phase ambiguities. In this dissertation, all *de novo* phasing were performed by SAD and the phase ambiguity was resolved by solvent flattening. Solvent flattening takes advantage that, in a reasonable electron density map, protein electron density can be distinguished from solvent electron density, or solvent channels. These solvent regions are distinguished by computer programs and its electron density is set to a constant value. New phases are calculated combining the original phases with constant solvent region density to generate an improved solution.

**Structure refinement and validation**

With a scaled data file and calculated phase information, an electron density map can be calculated. The remaining part of structure determination is to build a model of the protein into the calculated electron density map. Structures in this dissertation were manually built in Coot and subjected to multiple rounds of refinements using Phenix-refine [107]. The input files for structure refinement include a PDB file (including cell dimensions, space group, atom x, y, z positions (coordinates), occupancies and B-factors)
and the mtz file (the structure factor file with $R_{\text{free}}$ flagged-reflections included). The overall goal of structure refinement is to improve the agreement between the structural model (PDB file with parameters stated above) and the measured reflections from data collection (mtz file). Structure factors can be calculated from the structure and a least square algorithm is utilized to maximize their overlap to the observed structure factors. During refinement, to reduce the load of computation, several parameters are varied in one optimization step instead of all parameters. Therefore, the users have the options to define the varied parameters and optimize the refinement procedures. In this section, I discuss several refinement options.

Rigid body refinement is commonly utilized immediately after MR. As described, MR utilizes phase information directly from the search model, thus resulting in a heavily biased electron density map towards the search model. Rigid body refinement constrains the structure to move as a rigid body to optimize the position within the unit cell. Rigid bodies can contain the entire protein at first and then broken into domains followed by another round of rigid body refinement.

In protein structures, we can use one parameter, the B-factor, [149] to describe the isotropic atomic displacement of each atom (equation shown below), where $\mu$ represents the projection of atomic displacement. In a protein structure, B-factors of adjacent atoms are very similar in value. In the refinement configuration employed in this dissertation, individual B-factor refinement was utilized. Structures in this
dissertation were solved from medium to high resolution, thus a B-factor could be assigned to each atom. At low resolution, group B-factor refinement is preferred, which assign a shared B-factor to groups of atoms.

$$B = 8\pi^2 \mu^2$$

Several other common options for refinements include applying additional restraints. For example, secondary structure restraints, and non-crystallographic symmetry related B-factor restraints. Structure refinement represents the final step for protein structure determination. R-factors (including R\textsubscript{work}, R\textsubscript{free}) have been introduced to assess the agreement of the final global model compared to the experimental data. In this equation, R\textsubscript{free} utilizes a “free” data set (5% -10% of the data) that was not utilized in any step of refinement for cross-validation.

$$R = \frac{\sum_{hkl} |F_{obs}| - |F_{calc}|}{\sum_{hkl} |F_{obs}|}$$

In addition to assessing the agreement between the calculated data with the global model, other parameters are also introduced to assess the final protein model. These parameters include the root means square deviations on bond lengths and bond angles and Ramachandran statistics. These statistics are reported during PDB deposition. MolProbity [109] has been developed to validate the structure and assist the crystallographer to fix model errors, such as atom clashes, rotamer choices and improperly positioned side chains. Nevertheless, the structure still has to be carefully built by a crystallographer and inspected during the refinement process. These
structures are finally deposited to the RCSB PDB for sharing, significantly deepening our understandings about these particular proteins.
<table>
<thead>
<tr>
<th>Crystal system</th>
<th>Unit cell parameters</th>
<th>Bravais lattice types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>$a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90^\circ$</td>
<td>$P$</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>$a \neq b \neq c$, $\alpha = \gamma = 90^\circ$, $\beta \neq 90^\circ$</td>
<td>$P$, $C$</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>$a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$</td>
<td>$P$, $I$, $C$, $F$</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>$a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$</td>
<td>$P$, $I$</td>
</tr>
<tr>
<td>Trigonal</td>
<td>$a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$</td>
<td>$P$, $R$</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>$a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$</td>
<td>$P$</td>
</tr>
<tr>
<td>Cubic</td>
<td>$a = b = c$, $\alpha = \beta = \gamma = 90^\circ$</td>
<td>$P$, $I$, $F$</td>
</tr>
</tbody>
</table>
Figure 34: Visualization of Bragg's law

Two incident waves (1 and 2) and reflected waves (1’ and 2’) are shown in the figure. The perpendicular distance between the Miller planes is defined as $d_{hkl}$. The path difference between two waves therefore equals to $A_2B + A_2C = 2d \sin \theta$. The figure was adapted from Public Domain figure (http://en.wikipedia.org/wiki/File:Bragg_diffraction.png).
The Ewald sphere has a radius of 1/λ and is centered at the crystal. A reciprocal lattice point on the Ewald sphere will have identical vector $d_{hkl}^*$ with the scattering vector $S$. In addition, $d_{hkl}^*$ has the amplitude of $2 \sin \theta / \lambda$, thus meeting the requirements of Bragg’s law.
Figure 36: Phase diagram of protein crystallization
Figure 37: Hanging-drop vapor diffusion method

(a) Plates with sealant and siliconized glass cover slides (Hampton Research) (b) Schematic sketch of hanging-drop vapor diffusion method.
Figure 38: Experimental phasing by isomorphous replacement

Experimental phasing by single wavelength isomorphous replacement method with phase information calculated. Note the two values that are obtained, thus leading to phase ambiguity.
Figure 39: Se-Met scan result of *A. fumigatus* Tps2PD crystal

Scan result of Se-Met *A. fumigatus* Tps2PD crystal. Peak, inflection and high remote wavelengths are highlighted in the figure.
Figure 40: Introduction of anomalous scattering contributions to Friedel pairs

(a) A Friedel pair has the same amplitude when there is no contribution from anomalous scattering. (b) The introduction of anomalous scattering contributions to Friedel pair (real component: magenta; imaginary component: green) result in different amplitudes.
References


ADPglucose pyrophosphorylase and higher rates of starch synthesis in Arabidopsis thaliana. Biochemical Journal 397: 139-148.


170


decreases infectivity without affecting hypha formation. Infection and Immunity 70: 1772-1782.


Biography

Yi was born in Anhui, China. He attended Tsinghua University in 2006, where he graduated with Bachelor of Science in Biological Sciences in July 2010. In August 2010, he attended Duke University to obtain the PhD degree under the supervision of Dr. Richard G. Brennan in Biochemistry Department. During graduate school, he focused on studying trehalose biosynthesis in pathogenic fungi and co-authored a few publications.

Publications


Miao Y, Tenor JL, Toffaletti DL, Perfect JR, Brennan RG, Structures of trehalose-6-phosphate synthase from pathogenic fungi provide insights into antifungal drug design. (in preparation for PLOS Pathogens)

Miao Y, Tenor JL, Toffaletti DL, Perfect JR, Brennan RG, Structures of trehalose-6-phosphate phosphatase from pathogenic fungi reveal the molecular mechanism of substrate specificity and catalysis. (under revision at PNAS)