

**INVESTIGATION OF *CANDIDA ALBICANS* PRENYLTRANSFERASES:  
SUBSTRATE RECOGNITION AND ENZYME INHIBITION**

by

Elaina A. Zverina

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Chemical Biology)  
in the University of Michigan  
2012

Doctoral Committee:

Professor Carol A. Fierke, Chair  
Professor Anna K. Mapp  
Associate Professor Bruce A. Palfey  
Professor Ronald W. Woodard

© Elaina A. Zverina  
2012

## **DEDICATION**

To my parents and grandparents

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Carol Fierke, for her support, guidance, and encouragement throughout my time in graduate school. I would also like to thank my committee, Prof. Anna Mapp, Prof. Bruce Palfey, and Prof. Ron Woodard, for their helpful comments and suggestions. I would like to thank the Pharmacological Sciences Training Program for funding and training.

I sincerely thank all my labmates, both past and present, who have made been great colleagues over the years. I would especially like to thank Drs. Corissa Lamphear, Terry Watt, and James Hougland for their help and discussions about my project. I would like to thank Noah Wolfson, Elia Wright, Xin Liu, Carol Ann Pitcairn, and Andrea Stoddard for all of their help along the way. I would like to thank the Chemical Biology Program for providing an excellent academic environment, and all the members of my class.

I would like to thank Prof. Kenneth Merz and Dr. Mark Benson at the University of Florida for collaboration on prenyltransferase inhibitors.

I would like to thank all the people who have encouraged me to continue my education, especially Drs. Dave Ehmann, Tom Keating, Adam Shapiro, and Grant Walkup.

I would like to thank my family for their unconditional love, support and encouragement. I would like to express my deepest love and gratitude to my parents Alex and Irina.

## TABLE OF CONTENTS

DEDICATION.....	ii
ACKNOWLEDGEMENTS.....	iii
LIST OF FIGURES.....	vi
LIST OF TABLES.....	ix
ABSTRACT.....	x
<b>CHAPTER ONE.....</b>	<b>1</b>
INTRODUCTION.....	1
Post-translational modifications.....	1
Protein lipidation.....	2
Protein prenylation and processing.....	5
Prenylation and disease implications.....	7
Mammalian FTase and GGTase-I.....	8
Prenyltransferase substrate identification.....	11
Computational work.....	13
<i>Candida albicans</i> prenyltransferases.....	15
Objectives of this work.....	19
REFERENCES.....	21
<b>CHAPTER TWO.....</b>	<b>27</b>
INVESTIGATION OF SUBSTRATE RECOGNITION BY <i>CANDIDA ALBICANS</i> PRENYLTRANSFERASES USING A PEPTIDE LIBRARY APPROACH.....	27
INTRODUCTION.....	27
EXPERIMENTAL PROCEDURES.....	34
Cloning of <i>C. albicans</i> FTase and GGTase-I genes.....	34
<i>C. albicans</i> FTase and GGTase-I expression and purification.....	36
Peptide libraries.....	38
Multiple turnover FTase and GGTase-I Activity Assay.....	42
Single turnover assay.....	43
Statistical analysis of amino acid sequence in FTase and GGTase-I peptide substrates.....	44
RESULTS.....	46
Peptide substrates of FTase and GGTase-I.....	46
Sequence analysis of MTO and STO <i>C. albicans</i> FTase and GGTase-I substrates.....	51
Statistical analysis of individual amino acid distribution in <i>C. albicans</i> FTase and GGTase-I substrates.....	57
Predictions for Cxxx sequences found in <i>C. albicans</i> genome.....	61
DISCUSSION.....	65
Comparison of mammalian and <i>C. albicans</i> prenyltransferase substrate pools.....	65
Comparison of mammalian and <i>C. albicans</i> prenyltransferase substrate recognition.....	67
REFERENCES.....	71

<b>CHAPTER THREE</b> .....	<b>73</b>
MOLECULAR BASIS OF Ca <sub>1</sub> a <sub>2</sub> X RECOGNITION IN MAMMALIAN AND <i>CANDIDA ALBICANS</i> PRENYLTRANSFERASES .....	73
INTRODUCTION .....	73
EXPERIMENTAL PROCEDURES.....	80
Multiple Turnover Prenylation Activity Assay.....	80
Data analysis .....	81
RESULTS .....	81
<i>C. albicans</i> FTase reactivity with GCVa <sub>2</sub> X where X=A, S, Q, M, L and F .....	81
<i>C. albicans</i> GGTase-I reactivity with GCVa <sub>2</sub> X where X=A, S, Q, M, L and F.....	87
Comparison of mammalian and <i>C. albicans</i> GGTase-I reactivity with GCVa <sub>2</sub> L.....	90
Recognition of the X group by mammalian and <i>C. albicans</i> prenyltransferases.....	92
DISCUSSION.....	94
Mammalian and <i>C. albicans</i> FTase recognition of -a <sub>2</sub> X.....	94
Mammalian and <i>C. albicans</i> GGTase-I recognition of -a <sub>2</sub> X.....	98
Recognition of X group by mammalian and <i>C. albicans</i> prenyltransferases .....	101
REFERENCES .....	105
<b>CHAPTER FOUR</b> .....	<b>107</b>
INVESTIGATION OF SMALL MOLECULE INHIBITORS OF MAMMALIAN AND <i>CANDIDA ALBICANS</i> PRENYLTRANSFERASES .....	107
INTRODUCTION .....	107
EXPERIMENTAL PROCEDURES.....	111
Selection of small molecule prenyltransferase inhibitors .....	111
Prenyltransferase inhibition assays .....	112
Mode of inhibition of <i>C. albicans</i> FTase by compound 36745 .....	113
RESULTS .....	114
FTase and GGTase-I inhibition.....	114
Mode of inhibition of <i>C. albicans</i> FTase by 36745 .....	116
Analog of 36745 identified in ZINC database .....	120
<i>C. albicans</i> FTase inhibition by close analogs of compound 36745 .....	122
Inhibition of <i>C. albicans</i> FTase by Fmoc-substituted dipeptides .....	123
DISCUSSION.....	128
Prenyltransferase inhibition .....	128
SAR of Fmoc-modified natural dipeptides .....	129
REFERENCES .....	132
<b>CHAPTER FIVE</b> .....	<b>134</b>
SUMMARY, CONCLUSIONS, AND FUTURE DIRECTIONS.....	134
SUMMARY AND CONCLUSIONS .....	134
Substrate recognition by <i>C. albicans</i> FTase and GGTase-I.....	135
Molecular basis of Ca <sub>1</sub> a <sub>2</sub> X recognition by <i>C. albicans</i> FTase and GGTase-I .....	136
Inhibition of <i>C. albicans</i> FTase and GGTase-I.....	138
FUTURE DIRECTIONS .....	139
Role of the exit groove in <i>C. albicans</i> GGTase-I product release .....	139
Prediction of <i>C. albicans</i> FTase and GGTase-I substrates .....	140
<i>C. albicans</i> inhibitor design .....	141
REFERENCES .....	142

## LIST OF FIGURES

Figure 1.1. Types of lipidation.....	3
Figure 1.2. Enzymes in the prenylation pathway.....	6
Figure 1.3. Structures of mammalian FTase and GGTase-I.....	8
Figure 1.4. Inactive FTase•FPP•KCVIM ternary complex.....	10
Figure 1.5. Minimal FTase kinetic mechanism.....	12
Figure 1.6. <i>C. albicans</i> GGTase-I superimposed on mammalian GGTase-I.....	17
Figure 2.1. Structure comparison of mammalian and yeast FTase and GGTase-I.....	29
Figure 2.2. The CaaX peptide-binding site in mammalian and <i>C. albicans</i> FTase and GGTase-I. .....	30
Figure 2.3. Kinetic mechanism of mammalian FTase.....	31
Figure 2.4. Conformational change of the product in mammalian GGTase-I.....	32
Figure 2.5. Mammalian and <i>C. albicans</i> GGTase-I exit grooves.....	33
Figure 2.6. Plasmid maps for <i>C. albicans</i> FTase and GGTase-I.....	36
Figure 2.7. Analysis of molecular mass of <i>C. albicans</i> GGTase-I by LC-MS.....	38
Figure 2.8. Continuous fluorescence assay to monitor prenyltransferase reaction.....	42
Figure 2.9. The percentages of canonical and non-canonical residues for peptides in the overall library and MTO, STO and non-substrate pools for <i>C. albicans</i> FTase.....	51
Figure 2.10. Distribution of amino acids in peptides at the a <sub>1</sub> , a <sub>2</sub> and X positions that are substrates for <i>C. albicans</i> FTase under MTO and STO conditions or are not reactive. ....	54
Figure 2.11. The percentages of canonical and non-canonical residues for peptides in the overall library and in the MTO substrate pool for <i>C. albicans</i> GGTase-I.....	55
Figure 2.12. Distribution of amino acids in peptides at the a <sub>1</sub> , a <sub>2</sub> and X positions that are substrates for <i>C. albicans</i> GGTase-I under MTO and STO conditions or are not reactive. .....	56
Figure 2.13. Prenyltransferase MTO substrate pool composition.....	66

Figure 3.1. Structure of a peptide substrate bound to FTase illustrating the a <sub>2</sub> residue binding site. ....	74
Figure 3.2. a <sub>2</sub> reactivity of mammalian FTase W102Aβ and W106βA mutants. ....	75
Figure 3.3. Model of <i>C. albicans</i> FTase with a bound peptide substrate illustrating the proposed a <sub>2</sub> residue binding site. ....	77
Figure 3.4. Structure of a peptide substrate bound to GGTase-I illustrating the a <sub>2</sub> residue binding site. ....	78
Figure 3.5. Mammalian and <i>C. albicans</i> GGTase-I a <sub>2</sub> binding pocket. ....	79
Figure 3.6. <i>C. albicans</i> FTase-I reactivity with GCVa <sub>2</sub> X peptides. ....	84
Figure 3.7. <i>C. albicans</i> GGTase-I reactivity with GCVa <sub>2</sub> X peptides. ....	89
Figure 3.8. Comparison of mammalian and <i>C. albicans</i> GGTase-I reactivity with dns-GCVa <sub>2</sub> L peptides. ....	91
Figure 3.9. Differential reactivities of <i>C. albicans</i> FTase and GGTase-I as a function of hydrophobicity of the X residue. ....	93
Figure 3.10. Differential reactivities of mammalian FTase and GGTase-I as a function of hydrophobicity of the X residue. ....	93
Figure 3.11. Structure of mammalian, <i>C. albicans</i> , and <i>C. neoformans</i> FTase a <sub>2</sub> and X binding pockets. ....	95
Figure 3.12. Mammalian and <i>C. albicans</i> FTase linear correlations between a <sub>2</sub> residue volume and peptide reactivity for peptides with varying X residues. ....	96
Figure 3.13. Comparison of FTase interactions with peptides that have X = S, Q, M, and F. ....	97
Figure 3.14. Comparison of a <sub>2</sub> binding pockets of mammalian and <i>C. albicans</i> GGTase-I. ....	100
Figure 3.15. FTase X binding pocket. ....	101
Figure 3.16. <i>C. albicans</i> FTase and GGTase-I reactivity as a function of a <sub>2</sub> and X residue volumes. ....	102
Figure 3.17. Mammalian and <i>C. albicans</i> GGTase-I X residue binding pocket. ....	102
Figure 3.18. X group contacts in <i>C. albicans</i> GGTase-I. ....	103
Figure 4.1. Selected FTase and GGTase-I inhibitors. ....	108
Figure 4.2. <i>C. albicans</i> FTase and GGTase-I inhibitors. ....	109



Figure 4.3. Small molecules selected in the initial virtual screen against mammalian FTase....	111
Figure 4.4. Types of reversible inhibition.....	114
Figure 4.5. IC <sub>50</sub> curve of compound 35253 with mammalian FTase.....	115
Figure 4.6. Time dependence of <i>C. albicans</i> FTase inhibition by compound 36745.....	116
Figure 4.7. Mode of inhibition of 36745 with respect to dns-GCVLS peptide substrate. ....	117
Figure 4.8. Mode of inhibition of 36745 with respect to FPP prenyl donor.....	119
Figure 4.9. Activity of analogs of compound 36745.....	121
Figure 4.10. IC <sub>50</sub> curve of Fmoc-Asp-Phe analog with <i>C. albicans</i> FTase. ....	122
Figure 4.11. Benzylthioether analogs of compound 36745.....	123
Figure 4.12. Activity of Fmoc-protected dipeptides against <i>C. albicans</i> FTase.....	124
Figure 4.13. <i>C. albicans</i> FTase inhibition results for Fmoc-substituted dipeptide compounds.	130
Figure 5.1. Structure of the mammalian GGTase-I exit groove. ....	140

## LIST OF TABLES

Table 2.1. Molecular mass of <i>C. albicans</i> FTase and GGTase-I.....	38
Table 2.2. Unique Cxxx> sequences found in the <i>C. albicans</i> genome. ....	39
Table 2.3. List of all TKCxxx peptides tested with <i>C. albicans</i> FTase and GGTase-I. ....	41
Table 2.4. MTO, STO, and non-substrates of <i>C. albicans</i> FTase.....	48
Table 2.5. MTO, STO, and non-substrates of <i>C. albicans</i> GGTase-I.....	50
Table 2.6. <i>C. albicans</i> FTase and GGTase-I substrate pools.....	50
Table 2.7. Amino acids that are overrepresented or underrepresented in <i>C. albicans</i> FTase MTO substrate, STO substrate, and non-substrate pools as compared to the overall library.....	58
Table 2.8 Amino acids that are overrepresented or underrepresented in <i>C. albicans</i> GGTase-I MTO substrate pool as compared to the overall library. ....	59
Table 2.9. Cxxx sequences found in <i>C. albicans</i> genome predicted to be <i>C. albicans</i> FTase, GGTase-I, dual enzyme, or non-substrates based on peptide library studies. ....	61
Table 2.10. Experimental results, PrePS substrate predictions, and predictions based on new algorithm.....	63
Table 3.1. Steady-state kinetic parameters measured for <i>C. albicans</i> FTase. ....	83
Table 3.2. Residue volumes used in analysis of peptide reactivity. ....	85
Table 3.3. Steady-state kinetic parameters measured for <i>C. albicans</i> GGTase-I. ....	88
Table 3.4. Steady-state kinetic parameters measured for mammalian GGTase-I.....	91
Table 3.5. Steady-state kinetic parameters measured for <i>C. albicans</i> FTase and GGTase-I.....	92
Table 4.1. Summary of mammalian and <i>C. albicans</i> FTase and GGTase-I inhibition.....	115
Table 4.2. Global fit parameters for inhibition of <i>C. albicans</i> FTase by 36745 with respect to dns-GCVLS.....	118
Table 4.3. Apparent $K_M$ , $k_{cat}$ , and $k_{cat}/K_M$ values for inhibition of <i>C. albicans</i> FTase by 36745 with respect to dns-GCVLS.....	119
Table 4.4. Global fit parameters for 36745 inhibition of <i>C. albicans</i> FTase wrt FPP.....	120

## ABSTRACT

### INVESTIGATION OF *CANDIDA ALBICANS* PROTEIN PRENYLTRANSFERASES: SUBSTRATE RECOGNITION AND ENZYME INHIBITION

by

Elaina A. Zverina

**Chair: Carol A. Fierke**

Prenylation is a post-translational modification that is essential for the proper membrane localization of many cellular proteins. Protein prenylation is carried out by protein farnesyltransferase (FTase) and protein geranylgeranyltransferase-I (GGTase-I), zinc-dependent sulfur alkyltransferases that catalyze attachment of either a 15-carbon farnesyl or a 20-carbon geranylgeranyl isoprenoid moiety to an invariant cysteine residue near the C-terminus of the substrate protein. Inhibitors targeting human FTase and GGTase-I are being developed as therapeutics for various diseases. Prenyltransferases from *Candida albicans* pathogen have been investigated as targets for anti-mycotic agents. Early studies with mammalian FTase and GGTase-I proposed a “CaaX” recognition motif where C is the modified cysteine, ‘a’ is often an aliphatic amino acid, and X determines enzyme specificity. Recent work indicates that this substrate recognition paradigm is too restrictive, and many “non-canonical” CaaX sequences are efficiently prenylated by both FTase and GGTase-I. In this study substrate specificities of *C. albicans* FTase and GGTase-I were evaluated to assess their substrate recognition overlap with

each other and were compared to mammalian enzymes to determine whether small molecule inhibitors could be specifically targeted to these. Peptide library studies showed that *C. albicans* FTase has similar substrate specificity to the mammalian enzyme; however, *C. albicans* GGTase-I has a significantly broader substrate specificity, including recognition of substrates with large amino acids. *C. albicans* GGTase-I structure shows that it has a larger  $\alpha_2$  binding pocket, providing rationale for the observed substrate preferences. *C. albicans* GGTase-I also lacks a product exit groove and appears to have a faster product dissociation rate. Docking studies identified a dipeptide derivative as a potent and selective *C. albicans* FTase inhibitor with an  $IC_{50} = 200$  nM. This compound shows competitive mode of inhibition with FPP and mixed mode of inhibition with peptide substrate, indicating that it does not behave as a peptidomimetic. SAR studies with related compounds showed that addition of zinc-coordinating ligands enhances *in vitro* potency. Together, these studies propose a substrate recognition model for *C. albicans* prenyltransferases, identify similarities and differences between mammalian and yeast substrate interactions, and suggest possible lead compounds for the treatment of *C. albicans* infection.

## CHAPTER ONE

### INTRODUCTION<sup>1</sup>

#### *Post-translational modifications*

Protein post-translational modification (PTM) is defined as the covalent addition of various functional groups or proteolytic cleavage of a protein after its translation. Although the human genome sequence indicates that there are somewhere between 20,000 and 25,000 genes [1], the number of different proteins is estimated to be over 1 million [2]. Clearly, the ‘one gene, one protein’ central dogma of molecular biology, first introduced in the 1940’s, is an oversimplification that does not hold up to the rigor of modern discovery [3]. Proteomic diversity is the result of a combination of both transcriptional and post-translational diversification steps. Genomic recombination events, alternative promoter sites that modulate transcription initiation, differential transcription termination, and alternative splicing processes generate over 100,000 transcripts [4]. Additionally, PTMs exponentially increase proteome complexity and functional diversity [5]. It is estimated that there are over 200 types of PTMs which include phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation, and proteolysis. The majority of these modifications are reversible and catalyzed by enzymes in response to both intra- and extracellular stimuli. These enzymes include kinases, phosphatases,

---

<sup>1</sup> Portions of Chapter 1 were adapted from Lamphear, C. L., Zverina, E. A., Hougland, J. L., and Fierke, C. A. (2011) Global Identification of Protein Prenyltransferase Substrates: Defining the Prenylated Proteome, in *The Enzymes* (Tamanoi, F., Hrycyna, C. A., and Bergo, M. O., Eds.) pp 207-234, Academic Press.

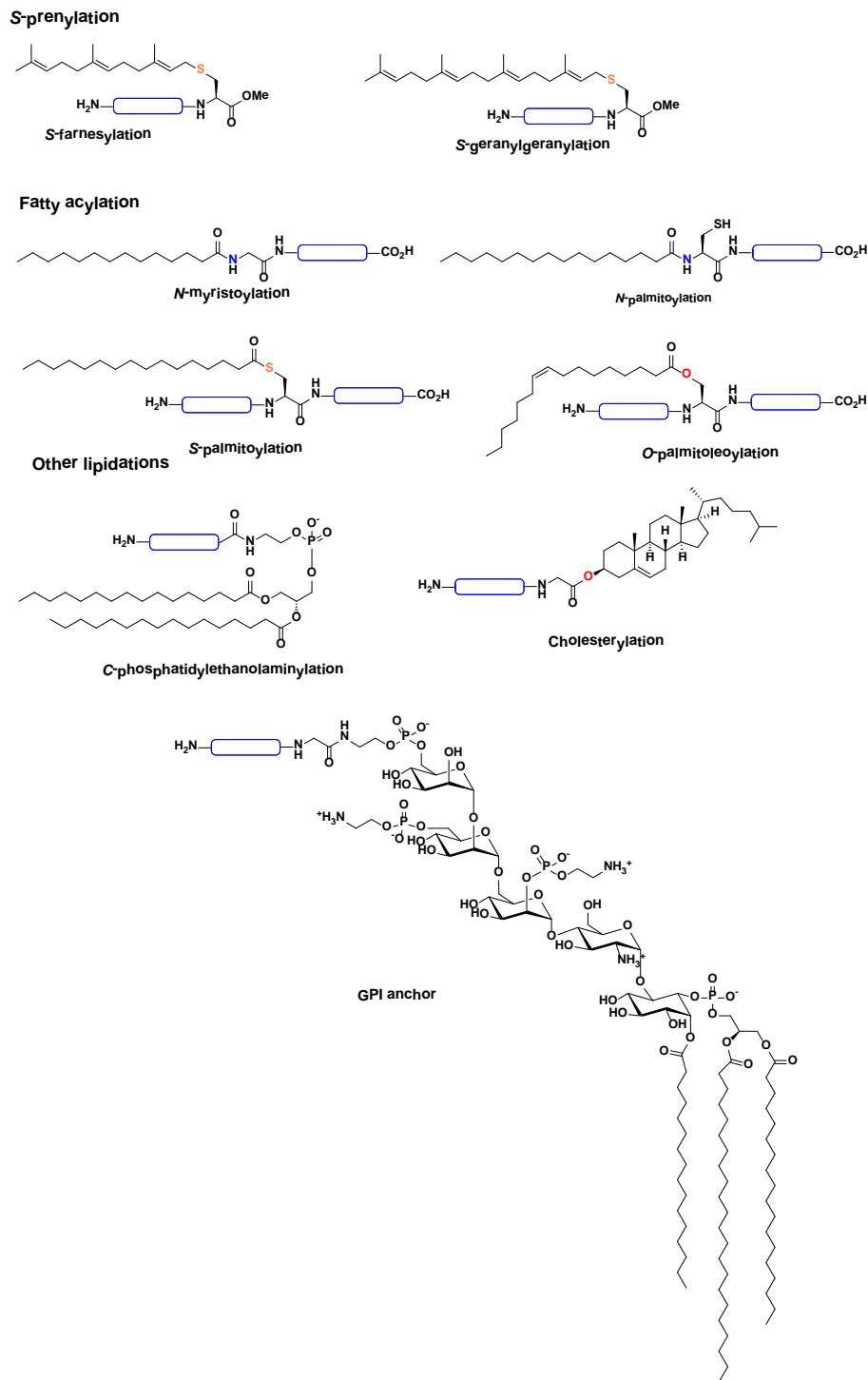
transferases, ligases, and proteases, and together constitute 5% of the proteome, or over 50,000 proteins.

PTMs can occur at any point during the 'life cycle' of a protein. Immediately after translation, PTMs mediate proper protein folding and stability or engage in cellular localization. Over the course of the functional lifetime of a protein, PTMs activate or inactivate its catalytic activity, regulate protein-protein interactions (PPIs) or alter localization. Finally, at the end PTMs tag proteins for proteasomal degradation. Naturally, many proteins undergo several sequential PTMs that differentially regulate their function and enhance or abrogate their cellular activity. As proper functioning of both modified proteins and enzymes that modify them is crucial for maintaining healthy cellular processes, significant deviations from the norm result in pathogenesis. Thus, understanding PTMs and enzymes that regulate them is critical for disease diagnosis, treatment, prevention, and prognostication. In particular, treatment of cancer, neurodegenerative, autoimmune, inflammatory, cardiovascular, and infectious diseases have all benefited from advances in understanding the role of PTMs. Various therapies have been developed based on this knowledge that are more efficacious, less toxic and more tailored for individual patients.

### *Protein lipidation*

Lipidation is a method whereby a lipid molecule is covalently attached to a protein and is generally used to target proteins to membranes in organelles (e.g. endoplasmic reticulum (ER), Golgi compartment, and mitochondria), vesicles (e.g. endosomes and lysosomes) and the plasma membrane (PM). A wide range of proteins, including many that are involved in human diseases, are modified by covalent linkage of lipids, including fatty acids, isoprenoids, and cholesterol [6]. Lipidation can be subdivided into five major groups: 1) C-terminal glycosyl phosphatidylinositol

(GPI) anchor attachment, 2) myristoylation, 3) palmitoylation, 4) prenylation [7], and 5) cholesterol attachment (Figure 1.1).



**Figure 1.1. Types of lipidation.** The targeting of proteins to membranes by lipidation plays key roles in many physiological processes. *N*-, *S*- and *O*- prefixes describe the linkage of the lipid attachment to proteins.

GPI is a complex glycolipid that consists of phosphatidylinositol, glucosamine, mannose, and ethanolaminephosphate that is synthesized by a sequential ten step ER-localized biosynthetic pathway. GPI attachment via a carbohydrate linker is catalyzed by ER-localized GPI transamidase, and it can be removed by phosphoinositol-specific phospholipase C [8]. GPI anchors tether cell surface proteins to the PM, and such proteins are often localized to cholesterol- and sphingolipid-rich lipid rafts where they act in signaling platforms and play important roles in host's self-defense, immune response and signal transduction [9]. Cholesterylation regulates secretion of the Sonic hedgehog protein and its misregulation can lead to abnormal tissue development and promote carcinogenesis [10].

Fatty-acylated proteins are synthesized by discrete families of acyltransferases that use fatty acyl coenzyme A (CoA) substrates to make cytoplasmic *N*-myristoylated [11] or *S*-palmitoylated [12] proteins as well as *S*-, *N*-, or *O*-acylated proteins that are secreted. Myristoylation is carried out by *N*-myristoyltransferases (NMTs), which catalyze the transfer of myristic acid from myristoyl-CoA to N-terminal glycine residues contained within the GXXXS/T consensus sequence. DHHC-protein acetyltransferases (DHHC-PATs) catalyze transfer of a palmitate group from palmitoyl-CoA onto cysteine residues. There is no defined amino acid consensus sequence for palmitoylation sites. In contrast to other lipidations, *S*-palmitoylation is a readily reversible thioester linkage which can be cleaved both non-enzymatically or by acyl protein thioesterases (APTs) [13]. Due to the reversible nature of *S*-palmitoylation, it is believed that the *S*-acylation/deacylation cycle is particularly important for regulating dynamic cellular processes. In particular, *S*-palmitoylation of *N*-myristoylated or *S*-prenylated proteins is typically required for stable membrane targeting of many proteins [14].

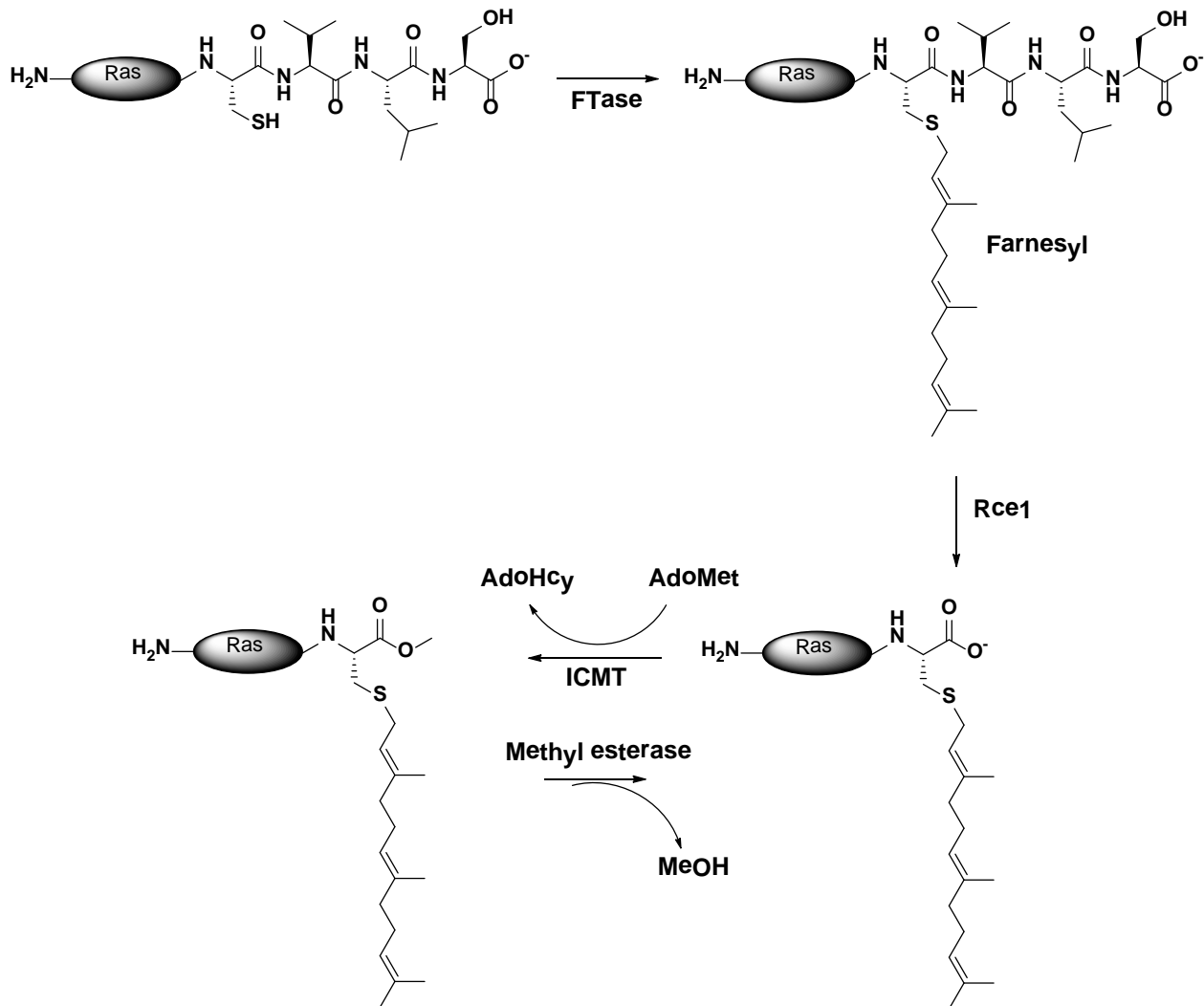


### *Protein prenylation and processing*

Protein prenylation encompasses formation of the thioethers, *S*-farnesylation and *S*-geranylgeranylation, catalyzed by protein farnesyltransferase (FTase) and protein geranylgeranyltransferase I and II (GGTase-I and GGTase-II) [15]. These enzymes attach either a 15-carbon farnesyl or a 20-carbon geranylgeranyl isoprenoid group to an invariant cysteine residue near the C-terminus of the substrate protein [16]. These hydrophobic modifications help to localize proteins to cellular membranes to carry out their function and facilitate protein-protein interactions [17,18]. FTase and GGTase-I are proposed to recognize a C-terminal “CaaX” motif on their substrates where C is the modified cysteine, ‘a’ is often an aliphatic amino acid, and X represents an amino acid that frequently determines specificity for the prenyltransferase. Known CaaX proteins include the family of Ras and Rho small GTPases, G protein  $\gamma$  subunits, protein phosphatases, phosphodiesterases, and nuclear lamins, many of which are associated with tumorigenesis [19]. Unlike FTase and GGTase-I, GGTase-II (also called RabGGTase) adds two *S*-geranylgeranyl groups to Rab proteins at the C-terminal cysteine residues, usually found in CC or CXC motifs [20-22]. RabGGTase substrate specificity is not dictated by a consensus sequence but by Rab escort proteins (REPs) that bind Rab proteins and target them for enzymatic modification [23,24].

As shown in Figure 1.2, the next step in the prenylation pathway is the proteolysis of the last three amino acids (-aaX) by the ER-bound protease RAS-converting enzyme 1 (*Rce1*) or zinc metalloproteinase sterile-24 homolog (*ZMPSTE24*) [25-28]. The final step is the methyl esterification of the newly exposed isoprenylcysteine residue catalyzed by isoprenylcysteine carboxyl methyltransferase (*Icmt*), another integral ER protein [29]. The importance of these processing steps is highlighted by mouse studies where germline knockouts of *Rce1* or *Icmt* were

embryonically lethal [30,31]. In general, farnesylated proteins frequently require such post-prenylation processing for proper localization, whereas the more hydrophobic geranylgeranylated proteins do not [32].



**Figure 1.2. Enzymes in the prenylation pathway.** CaaX processing is catalyzed by three enzymes that work sequentially: protein farnesyltransferase (FTase), RAS-converting enzyme 1 (Rce1), and isoprenylcysteine carboxymethyltransferase (Icmt). FTase catalyzes prenylation, the first rate-limiting reaction in the sequence. Rce1 proteolyzes the last three amino acids, and ICMT methylates the  $\alpha$ -carboxyl group of the farnesylated cysteine using S-adenosylmethionine (AdoMet) as the methyl donor. Reactions catalyzed by FTase and Rce1 are irreversible, while carboxyl methylation is reversible, catalyzed by a methyl esterase. Processing of H-Ras protein with the CVLS CaaX sequence is shown.

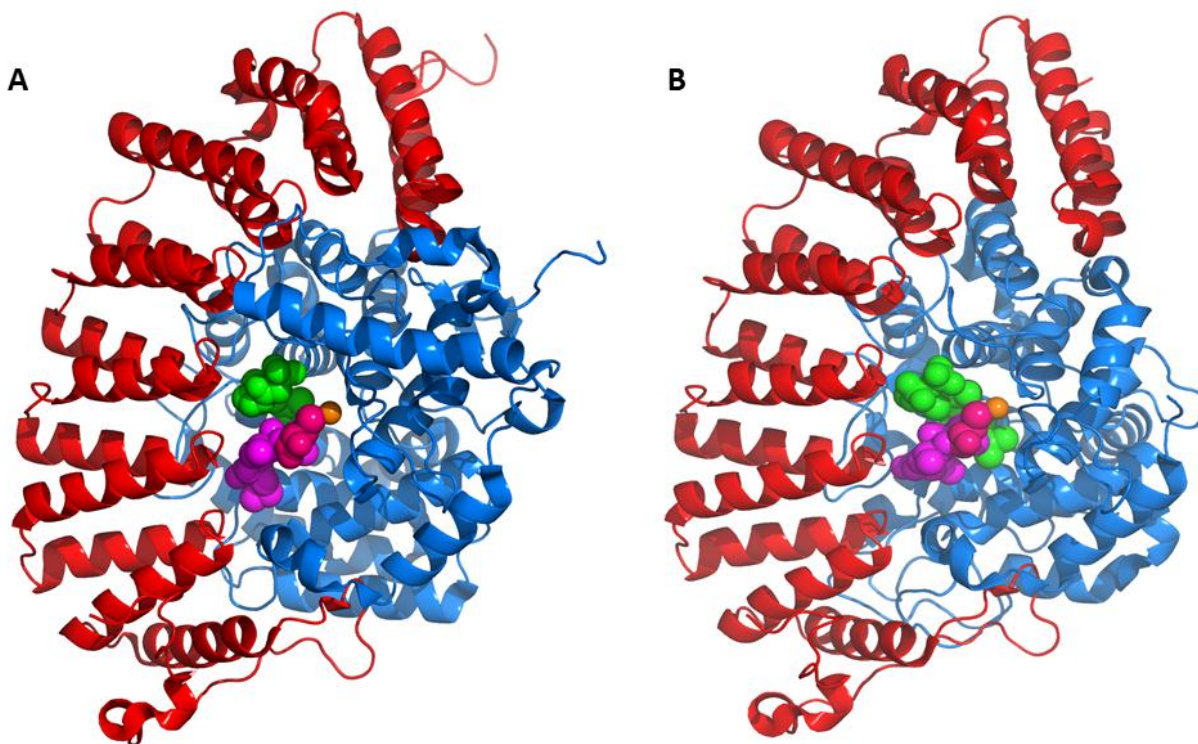
### *Prenylation and disease implications*

Many proteins are modified by the prenyltransferases, including the small Ras and Rho GTPase superfamilies. The discovery that a sizable portion of some, but not all, human cancers harbor activating oncogenic mutations in the Ras genes, and that Ras GTPases require prenylation for their transforming activity, led to the development of FTase inhibitors (FTIs) for the treatment of cancer [33]. Despite initial promise demonstrated in *in vitro* studies and preclinical animal models where these compounds showed good efficacy and limited toxicity, FTIs have not been as successful as hoped in clinical trials, with little to no efficacy observed in most patients [34-36]. However, a subset of patients appears to respond well to FTI monotherapy, prompting researchers to look for biomarkers that can be used to select patients whose tumors are sensitive to FTI treatment [37-39].

One of the issues with FTase inhibitors is that GGTase-I catalyzes geranylgeranylation of K-Ras and N-Ras allowing full functionality of these proteins in the cells [40]. This ability of GGTase-I to compensate for the inhibited FTase stems from their overlapping substrate specificities. Therefore, a better delineation of FTase and GGTase-I substrate selectivity will provide insight into proteins that are modified in the presence of FTIs and GGTase-I inhibitors (GGTIs), and indicate when dual-specificity FTase/GGTase-I inhibitors (DPIs) might be required. In addition, determining which downstream pathways of prenylated proteins are relevant to oncogenesis can help in patient selection, and provide insight for combination therapy with other targeted agents to produce synergistic effects [41,42]. One major challenge in the field remains the identification of prenyltransferase substrates and their role in proliferation and survival of different cancer types.

FTIs are currently being evaluated for treatment of the premature aging Hutchinson-Gilford progeria syndrome (HGPS) [43,44]. Normally, prelamin A is farnesylated, transferred to the nucleus and then the C-terminus is proteolyzed by ZMPSTE24 to release mature lamin A. In HGPS, a genetic mutation removes the protease recognition site, resulting in prenylated prelamin A that is permanently anchored to the nuclear membrane. Prevention of farnesylation by FTIs improves disease symptoms in a mouse model [45,46] and could provide a novel therapy for this disease in humans. Additionally, inhibitors of non-mammalian prenyltransferases are subject of investigation for treatment of infectious diseases such as malaria [47-49], African sleeping sickness [50], and infections caused by the pathogenic yeast *Candida albicans* [51,52].

#### *Mammalian FTase and GGTase-I*

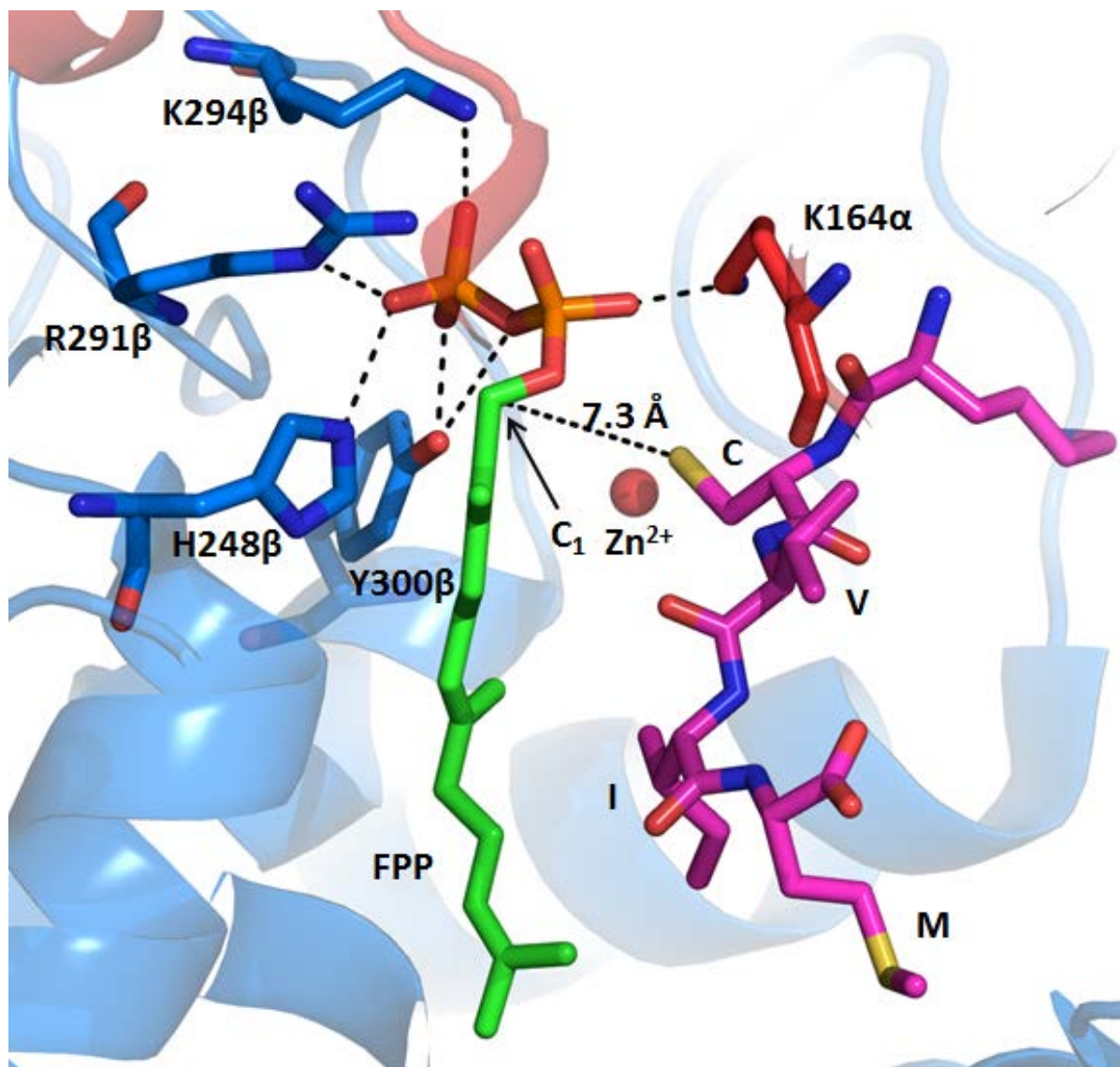


**Figure 1.3. Structures of mammalian FTase and GGTase-I.** A) Mammalian FTase complexed with FPP analog and CVIM peptide (PDB ID 1D8D) B) Mammalian GGTase-I complexed with GGPP analog and CVIL peptide (PDB ID 1N4Q).  $\alpha$  subunit is red,  $\beta$  subunit is blue, peptide is magenta with cysteine highlighted in pink, prenyl donor analog is green, and  $Zn^{2+}$  ion is orange.

FTase and GGTase-I are heterodimeric zinc-metalloenzymes composed of  $\alpha$  and  $\beta$  subunits with molecular masses around 48 and 43 kDa, respectively (Figure 1.3). FTase and GGTase-I share  $\alpha$  subunits, and the  $\beta$  subunits show 25% identity [53]. Previous studies have indicated that the  $\beta$  subunit of FTase contains the binding sites for farnesyl diphosphate (FPP), peptide substrate, and  $Zn^{2+}$ , which is required for binding of the peptide substrate and enzyme activity [54,55]. Portions of the  $\alpha$  subunit of FTase are near the peptide binding site and this subunit is required for stabilization of the  $\beta$  subunit [56,57]. The zinc ion coordinates the cysteine sulfur of the protein substrate and lowers the thiol  $pK_a$  from 8.1 to 6.4, creating a reactive thiolate at neutral pH [58]. In FTase, the  $Zn^{+2}$  stabilized nucleophile reacts with the C1 of FPP via a proposed concerted transition state with dissociative character where the peptide thiolate and the bridging oxygen of FPP both bear partial negative charges and the C1 of FPP bears a partial positive charge [59]. Recent work using quantum mechanical molecular mechanical studies (QM/MM) of the FTase-catalyzed reaction suggests that the transition state structure varies with the peptide structure; the transition state for farnesylation of CVLS is more  $S_N2$ -like (dissociative) while that of CVIM is associative [60]. This theoretical analysis is consistent with previously determined peptide-dependent alterations in secondary kinetic isotope effects [61]. In FTase a magnesium ion coordinates a side chain carboxylate in the enzyme and the diphosphate of the FPP substrate and is proposed to stabilize the negative charge accumulation on the diphosphate group in the transition state [62,63]. Mammalian GGTase-I, which is Mg-independent, contains a positively charged lysine residue whose side chain partially replaces the role of the magnesium ion [64].

In crystal structures of inactive FTase or GGTase-I E•peptide•prenyl donor complexes, the two reacting atoms, the C1 of the prenyl donor and the cysteine sulfur of the protein

substrate, are over 7 Å apart as shown in Figure 1.4 [65,66]. Therefore, to enable efficient transfer, a conformational change of the farnesyl diphosphate substrate is proposed to occur where C1 moves toward the zinc-activated thiolate. Kinetic, mutagenesis, and crystallographic data all support this mechanism [67].

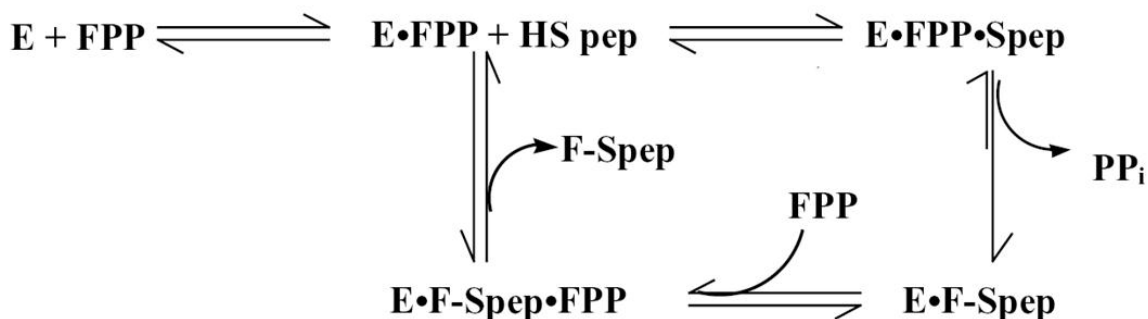


**Figure 1.4. Inactive FTase•FPP•KCVIM ternary complex.** C1 of FPP is 7.3 Å away from peptide thiolate. The complex was created by overlaying the two crystal structures FTase•FPP•L-739750 (PDB ID 1JCQ) and FTase•I2•K-Ras peptide (PDB ID 1D8D). Side chains that are proposed to stabilize diphosphates of FPP are highlighted.

### *Prenyltransferase substrate identification*

Identification of FTase and GGTase-I substrates has been an important goal for understanding the biological function of these enzymes. Traditional radiological methods where cells are treated with a radiolabeled analog of FPP, GGPP, or a precursor of these molecules have not been generally useful for identifying prenyltransferase substrates mainly due to the low signal of the radiolabel [68-70]. Furthermore, antibodies raised against the farnesyl or geranylgeranyl moieties cross-react and cannot unequivocally differentiate between these modifications [71-73]. Despite these challenges, great progress has been made in identifying potential and substantiated *in vivo* substrates of the prenyltransferases using peptide library studies, computational techniques, and lipid analogs.

Analysis of the reactivity of prenyltransferases with diverse libraries of peptides has provided significant insight into the substrate selectivity of FTase and GGTase-I, as these enzymes react efficiently with short peptides [74]. The substrate selectivity of prenyltransferases can be tested quickly and efficiently using a library of dansylated peptide substrates [74,75]. Krzysiak and colleagues demonstrated that FTase readily catalyzes farnesylation of a large number of peptides terminating in Leu (-CaaL), contrary to the CaaX paradigm which describes these as canonical GGTase-I substrate sequences [76]. This result expands the pool of FTase substrates and demonstrates additional cross-reactivity with GGTase-I. Hougland and colleagues probed the molecular recognition of FTase for amino acids at the  $a_2$  and the X positions of  $Ca_1a_2X$  peptides revealing that the identity of the amino acid at the X position affects the selectivity for the amino acid at the  $a_2$  position [77]. Therefore, FTase recognizes the entire  $Ca_1a_2X$  sequence cooperatively rather than each amino acid individually.



**Figure 1.5. Minimal FTase kinetic mechanism.** The kinetic pathway has a preferred binding order with FPP binding before the protein substrates, followed by a rapid prenylation step and slow product dissociation. Product release is facilitated by binding of a second FPP molecule. Multiple turnover (MTO) substrates are proposed to undergo all the steps in this mechanism while single turnover (STO) substrates are proposed to only undergo farnesylation, but not product release.

Two kinetic parameters,  $k_{\text{cat}}/K_M^{\text{peptide}}$  and  $k_{\text{farnesylation}}$ , can be measured to describe peptide reactivity with FTase. The value of  $k_{\text{cat}}/K_M$ , measured under multiple turnover (MTO) conditions, is also termed the “specificity constant” and is most representative of the reactivity of a particular substrate in a biological context where multiple substrates are present [78]. The relative rate of prenylation of a given substrate *in vivo* depends on both the concentration of the substrate and the value of  $k_{\text{cat}}/K_M$ . Previous kinetic studies with FTase suggest the basic kinetic pathway shown in Figure 1.5. Substrate binding is functionally ordered, with FPP binding before peptide. The chemical step is followed by rapid dissociation of diphosphate. Dissociation of prenylated product from E·F-Spép complex is enhanced by binding of a second FPP molecule, and subsequent regeneration of E·FPP. For FTase, the  $k_{\text{cat}}/K_M$  parameter includes the rate constant for peptide binding to E·FPP through the formation of the prenylated peptide and diphosphate products, where  $\text{PP}_i$  dissociation is the first irreversible step; therefore, dissociation of the prenylated peptide product does not contribute to the observed value of  $k_{\text{cat}}/K_M$ . A second parameter,  $k_{\text{farn}}$ , measures the formation of the prenylated peptides from the ternary E·FPP·Spép complex. This kinetic constant is measured under single turnover conditions with limiting FPP



and saturating concentrations of FTase and peptide. For the single turnover (STO) substrates, FTase binds the substrate and catalyzes farnesylation but product dissociation is so slow that multiple turnovers are not readily observed.

To further evaluate FTase substrate selectivity, a large scale study of the reactivity of a library of 301 dansyl-TKCxxx peptide sequences taken from the human proteome [79] revealed two classes of peptide reactivity with FTase: 1) peptides farnesylated under multiple turnover (MTO) conditions (where  $[E] \ll [S]$ ); and 2) peptides farnesylated only under single turnover (STO) (where  $[FPP] < [E]$ ) conditions. Statistical analysis of these two classes of substrates revealed significantly different sequence preferences. The MTO peptides typically display canonical  $Ca_1a_2X$  sequences, including Ile and Leu at the  $a_2$  position and Phe, Met, and Gln at the X position. STO FTase substrates are more diverse in sequence, often containing a Ser at the  $a_2$  position and little sequence preference at the X position. The biological role of the STO substrates is not yet clear. Together, these studies show that FTase catalyzes farnesylation of a wide range of substrates, suggesting that a large cohort of proteins is farnesylated *in vivo*.

### *Computational work*

In addition to biochemical methods used to identify FTase and GGTase-I substrates, which rely on various *in vitro* and *in vivo* approaches, *in silico* approaches have also been developed to address the challenge of substrate identification. Various computational approaches have been introduced over the past several years to aid in large-scale prediction of prenylated proteins [80,81], with features that enable the user to rank the likelihood of a particular sequence being a substrate for FTase, GGTase-I, or both enzymes. The methods are iteratively improved by continuously incorporating new biochemical data as it becomes available to further refine the predictive power of the computational analyses.

One of the first prenylation prediction tools was Prosite protocol PS00294 which used the consensus pattern C-(DENQ)-[LIVM]-x> ([www.expasy.org/prosite/](http://www.expasy.org/prosite/)) [82]. However, this tool does not distinguish between FTase and GGTase-I substrates, nor does it predict prenylation by GGTase-II. Crystallographic analysis of FTase and GGTase-I complexed with eight cross-reactive substrates used interactions with the binding pocket in the structures of the enzyme-substrate complexes to draw inferences about FTase and GGTase-I substrate recognition elements [83]. The most significant drawback of this approach is that it only identifies a subset of verified substrates, missing key substrate-protein interactions that are not covered by the peptide diversity in the available crystal structures.

PrePS is a more recent algorithm developed to predict prenyltransferase substrates ([mendel.imp.ac.at/PrePS/index.html](http://mendel.imp.ac.at/PrePS/index.html)) [80]. To define this algorithm, the authors built a learning set of known and homologous substrates as defined by specific rules which resulted in a set of 692 FTase and 486 GGTase-I substrates ([mendel.imp.ac.at/sat/PrePS/PRENbase/](http://mendel.imp.ac.at/sat/PrePS/PRENbase/)). They also expanded the list of acceptable amino acids in the Ca<sub>1</sub>a<sub>2</sub>X motif. One of the difficulties in predicting prenylation substrates is the inherent complexity of substrate recognition motifs, which may extend beyond the Ca<sub>1</sub>a<sub>2</sub>X box to include the upstream region of the protein. To address this additional complexity, the authors of PrePS included an 11-amino acid upstream region of the Ca<sub>1</sub>a<sub>2</sub>X motif to refine their algorithm. This upstream region typically consists of a flexible linker region that often has a compositional bias toward small or hydrophilic amino acids. Using this learning set, the PrePS algorithm defines a set of rules that is used to predict the likelihood of a 15-amino acid sequence being an FTase, GGTase-I, and/or GGTase-II substrate. In cross-validation experiments, the authors were able to establish 92.6% and 98.6% true positive and 0.11% and 0.02% false positive rates for FTase and GGTase-I substrates, respectively.

Consistent with this, analysis of the substrates identified in the peptide library screen described above using the PrePS algorithm yielded a very low number of false positive results [79]. However, the PrePS analysis led to a large number of false negative predictions, around 40%, indicating that the PrePS algorithm potentially misses a large number of prenyltransferase substrates.

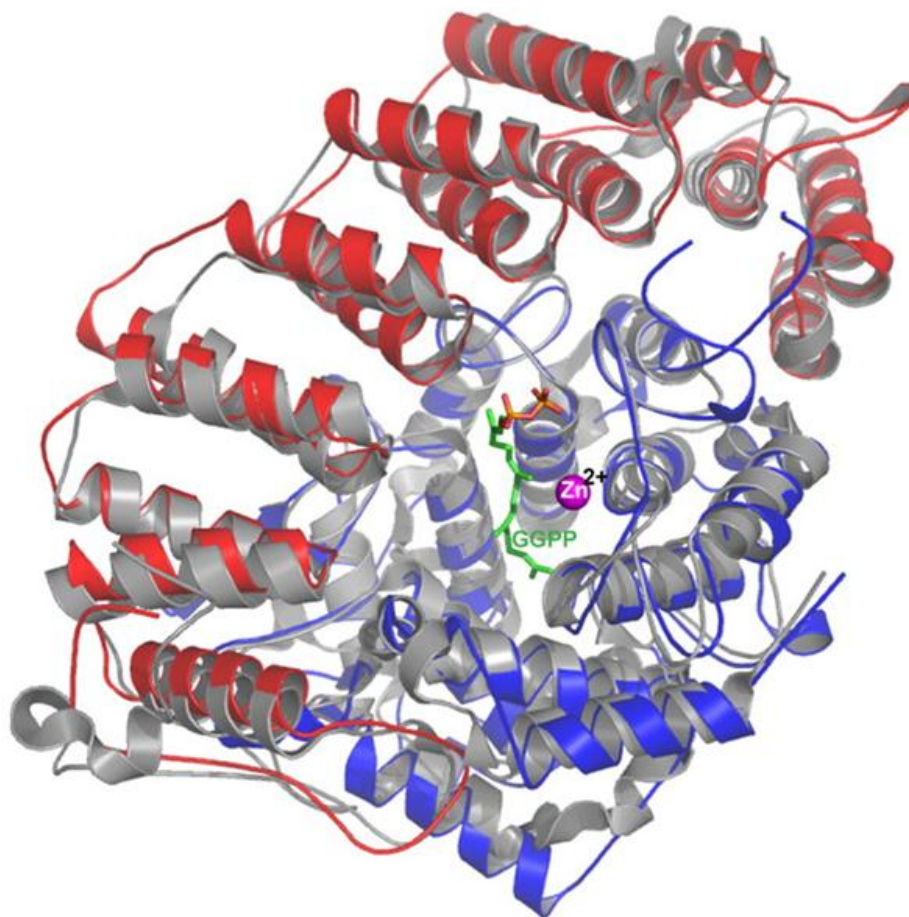
Recently, London and coworkers used the FlexPepBind algorithm to distinguish FTase substrates based on a calculated binding energy using the Hougland peptide library as a training set [84]. Analysis of the binding energy of all 8,000 possible CaaX sequences identified potential novel FTase substrates. In a test of the validity of these predictions, FTase catalyzed prenylation of 26 out of 29 proposed peptides, including a subgroup containing Asp or Glu at the X residue. These data confirm that FTase catalyzes prenylation of a variety of diverse sequences.

#### *Candida albicans prenyltransferases*

Although most studies to date have focused on mammalian prenyltransferases, lower eukaryotic organisms, such as yeast, also contain FTase and GGTase-I and inhibition of these enzymes has been proposed as a possible treatment for pathogenic yeast infection [51,85-89]. *Candida albicans* is a dimorphic yeast that is a major opportunistic human fungal pathogen which causes life-threatening infection in many immunocompromised patients. Although there are currently therapies on the market that combat *C. albicans* infections, resistance is always a concern, and thus novel therapies with unique mechanisms of action are highly sought after. Identifying *C. albicans* prenylation substrates and mining the differences between the substrate recognition of the human and *C. albicans* prenyltransferases could pave a way for novel drug candidates against this infection.

The first report of sequencing, cloning, and purification of *C. albicans* GGTase-I was published in 1999 by the Mazur group [85]. The  $\alpha$  and  $\beta$  subunits of GGTase-I are 28% and 25% identical to their mammalian homologs, respectively, with conservation of the zinc ligands and substrate-binding residues of the  $\beta$  subunit. Based on sequence alignment and comparison to mammalian FTase crystal structure, H231 $\beta$ , R296 $\beta$ , and K299 $\beta$  were identified as interacting with the diphosphate of FPP; and R160 $\beta$ , A165 $\beta$ , and G344 $\beta$  were implicated in protein substrate binding in *C. albicans* GGTase-I. In 2008, the Beese group solved the crystal structure of *C. albicans* GGTase-I complexed with GGPP at 1.8 Å resolution [90]. Despite the relatively low overall sequence identity, the structure of *C. albicans* GGTase-I is quite similar in overall architecture to mammalian GGTase-I, as shown in Figure 1.6, with 1.58 Å RMSD calculated over all aligned  $\alpha$ -carbon atoms. A noted difference between the mammalian and *C. albicans* GGTase-I structures is the size of the “exit groove”. In the structure of mammalian GGTase-I complexed with both the prenylated peptide product and GGPP, reflecting an intermediate in the dissociation of the prenylated product during the catalytic cycle, the prenyl group of the product flips into the exit groove when a second GGPP molecule binds in the active site pocket [66]. In the crystal structure of *C. albicans* GGTase-I, the exit groove is narrowed, suggesting that the geranylgeranyl group of the product could not be accommodated in the same binding mode as observed for the mammalian GGTase-I. This difference could alter the kinetic mechanism of *C. albicans* GGTase-I, particularly the steps involved in the release of prenylated product. In addition, although no peptide substrate was bound in the solved crystal structure of *C. albicans* GGTase-I, comparison of the putative Ca<sub>1</sub>a<sub>2</sub>X binding site of *C. albicans* GGTase-I to that of the mammalian enzyme revealed significant variation in the identities of amino acids, and most of

these were non-conservative changes. This alteration in *C. albicans* GGTase-I could change the substrate recognition pattern compared to the mammalian enzyme.



**Figure 1.6.** *C. albicans* GGTase-I superimposed on mammalian GGTase-I. *C. albicans* GGTase-I is shown in red and blue ( $\alpha$  and  $\beta$  subunits) and rat GGTase-I is in grey. (PDB ID 3DRA and 1N4P).

Biochemical studies of a small number of peptide substrates have been carried out to investigate fungal prenyltransferase substrate specificity. The Mazur group initially measured the reactivity of partially purified *C. albicans* GGTase-I and FTase with seven peptides containing a Ca<sub>1</sub>a<sub>2</sub>X sequences (CVIL, CVVL, CTIL, CAIL, CVLM, CVLS, and CVIA) from prenylated proteins in either *C. albicans* or *Saccharomyces cerevisiae* with the upstream sequence from *S. cerevisiae* Ras [85]. Five of these substrates were geranylgeranylated upon incubation with *C. albicans* GGTase-I, although the  $K_M$  values were considerably higher than those measured for

mammalian GGTase-I [85]. The two inactive sequences, CVLS and CVIA, are prototypical mammalian FTase substrates. *C. albicans* FTase catalyzed farnesylation of all seven Ca<sub>1</sub>a<sub>2</sub>X sequences. Unexpectedly, the kinetic parameters for FTase-catalyzed prenylation of –CaaL substrates was comparable to that of GGTase-I, suggesting both that *C. albicans* FTase has broad substrate specificity and that there is significant overlap in substrate specificity between the two enzymes. In a follow-up study, Mazur and coworkers explored the function of a polybasic sequence upstream of the Ca<sub>1</sub>a<sub>2</sub>X using biotinylated peptide substrates [52]. Overall, the upstream polybasic region has little impact on  $k_{cat}$  values, while the  $K_M$  values decreased; for example, addition of multiple lysine residues lowered the values of  $K_M$  ~15-fold for the CVIL and CAIL peptides. This is similar to the role of the polybasic region in rat FTase [91]. In addition, the polybasic region broadened the substrate pool for GGTase-I, enhancing reactivity for peptides that contain an amino acid other than L as the terminal amino acid (e.g. CVIM). The general conclusion from this work was that peptides ending in residues other than leucine are poor substrates for *C. albicans* GGTase-I unless they contain an upstream polybasic region.

Substrate specificity of the fungal prenyltransferases was also investigated *in vivo*. FTase and GGTase-I were shown to be essential for *C. albicans* growth as determined by knockout of the shared  $\alpha$  subunit [51]. The  $\beta$  subunits of both FTase and GGTase-I are essential for growth of *S. cerevisiae* but not *C. albicans*, possibly due to cross-prenylation [92,93]. In *S. cerevisiae*, loss of FTase can be overcome by growth at lower temperatures, and lack of GGTase-I activity can be circumvented by overexpression of the GGTase-I substrates Rho1p and Cdc42p [92,94]. To address the issue of *in vivo* cross-prenylation in *C. albicans*, the prenylation status of the GGTase-I substrates Rho1p (CVVL) and Cdc42p (CTIL) was assessed in the absence of GGTase-I protein expression [86]. Under such conditions, Rho1p and Cdc42p were prenylated,

as determined by localization of the proteins in the membrane fraction of the cells. This reaction was presumably catalyzed by FTase, although the prenyl donor group attached to these proteins was not verified. These data suggest that a prenyltransferase inhibitor with dual specificity for FTase and GGTase-I would be necessary to obtain *in vivo* antifungal activity against *C. albicans* pathogen.

### *Objectives of this work*

The main goals of this work are to determine the specificity and substrate recognition parameters of *C. albicans* FTase and GGTase-I. Detailed substrate specificity studies for mammalian FTase and GGTase-I have been carried out using peptide library studies, but there has not been a rigorous study performed on *C. albicans* prenyltransferases to define their molecular recognition. In this study, *C. albicans* FTase and GGTase-I were first cloned, heterologously overexpressed in *Escherichia coli*, and purified. The reactivity of *C. albicans* FTase and GGTase-I with a 328 d ansyl-TKCxxx peptide library was tested using a high throughput fluorescence assay. This screen demonstrated that 127 and 100 peptides were active with *C. albicans* FTase under MTO and STO conditions, respectively, and 207 and 36 peptides were active with *C. albicans* GGTase-I under MTO and STO conditions, respectively. Statistical analysis of these data revealed amino acid preferences for *C. albicans* FTase and GGTase-I in each pool of peptides (MTO, STO and non-substrates). This analysis showed that *C. albicans* FTase has similar substrate preferences to mammalian FTase, whereas *C. albicans* GGTase-I has an expanded substrate reactivity profile where more amino acid variability is tolerated at all positions of the Ca<sub>1</sub>a<sub>2</sub>X sequence, and bulky amino acids such as F and W, normally sterically discriminated against by mammalian FTase, are enriched in the *C. albicans* GGTase-I MTO substrate pool. Detailed substrate recognition studies of -a<sub>2</sub>X residues of the CaaX sequence

showed that *C. albicans* FTase exhibits cooperativity similar to that observed with mammalian FTase [77], where amino acids in the CaaX sequence are recognized not as independent moieties but in a context-dependent fashion. In contrast, *C. albicans* GGTase-I reactivity appears to be largely insensitive to the size or hydrophobicity of the a<sub>2</sub> and X residues.

In collaboration with Prof. Kenneth Merz and Dr. Mark Benson at the University of Florida, potent and selective inhibitors of *C. albicans* FTase were identified using virtual screening. The most potent fluorenylmethyloxycarbonyl-capped (Fmoc) dipeptide inhibitor showed competitive behavior with the FPP ( $K_i = 20$  nM) and uncompetitive behavior with the peptide substrate. Structure-activity relationship of a series of Fmoc-substituted dipeptides was used to identify potentially important features of peptidomimetic *C. albicans* FTase inhibitors, and suggests that zinc-coordinating ligands such as carboxylates, imidazoles, or thiolates enhance inhibitor potency. All together this work has advanced our understanding of substrate recognition by *C. albicans* FTase and GGTase-I, and provided insight into the development of inhibitors that could serve as possible lead compounds for treatment of *C. albicans* infections.



## REFERENCES

1. International Human Genome Sequencing Consortium: **Finishing the euchromatic sequence of the human genome.** *Nature* 2004, **431**:931-945.
2. Jensen ON: **Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry.** *Curr Opin Chem Biol* 2004, **8**:33-41.
3. Bussard AE: **A scientific revolution? The prion anomaly may challenge the central dogma of molecular biology.** *EMBO Rep* 2005, **6**:691-694.
4. Ayoubi TA, Van De Ven WJ: **Regulation of gene expression by alternative promoters.** *Faseb J* 1996, **10**:453-460.
5. Walsh CT: *Posttranslational Modification of Proteins: Expanding Nature's Inventory.* Englewood, Colorado: Roberts and Company Publishers; 2005.
6. Resh MD: **Targeting protein lipidation in disease.** *Trends Mol Med* 2012, **18**:206-214.
7. Nadolski MJ, Linder ME: **Protein lipidation.** *FEBS J* 2007, **274**:5202-5210.
8. Mayor S, Riezman H: **Sorting GPI-anchored proteins.** *Nat Rev Mol Cell Biol* 2004, **5**:110-120.
9. Triantafilou M, Triantafilou K: **Lipopolysaccharide recognition: CD14, TLRs and the LPS-activation cluster.** *Trends Immunol* 2002, **23**:301-304.
10. Mann RK, Beachy PA: **Novel lipid modifications of secreted protein signals.** *Annu Rev Biochem* 2004, **73**:891-923.
11. Farazi TA, Waksman G, Gordon JI: **The biology and enzymology of protein N-myristoylation.** *J Biol Chem* 2001, **276**:39501-39504.
12. Linder ME, Deschenes RJ: **Palmitoylation: policing protein stability and traffic.** *Nat Rev Mol Cell Biol* 2007, **8**:74-84.
13. Bijlmakers MJ, Marsh M: **The on-off story of protein palmitoylation.** *Trends Cell Biol* 2003, **13**:32-42.
14. Resh MD: **Trafficking and signaling by fatty-acylated and prenylated proteins.** *Nat Chem Biol* 2006, **2**:584-590.
15. Casey PJ, Seabra MC: **Protein prenyltransferases.** *J Biol Chem* 1996, **271**:5289-5292.
16. Zhang FL, Casey PJ: **Protein prenylation: molecular mechanisms and functional consequences.** *Annu Rev Biochem* 1996, **65**:241-269.
17. Marshall CJ: **Protein prenylation: a mediator of protein-protein interactions.** *Science* 1993, **259**:1865-1866.
18. Casey PJ, Moomaw JF, Zhang FL, Higgins YB, Thissen JA: **Prenylation and G protein signaling.** *Recent Prog Horm Res* 1994, **49**:215-238.
19. Winter-Vann AM, Casey PJ: **Post-prenylation-processing enzymes as new targets in oncogenesis.** *Nat Rev Cancer* 2005, **5**:405-412.
20. Armstrong SA, Seabra MC, Sudhof TC, Goldstein JL, Brown MS: **cDNA cloning and expression of the alpha and beta subunits of rat Rab geranylgeranyl transferase.** *J Biol Chem* 1993, **268**:12221-12229.
21. Horiuchi H, Kawata M, Katayama M, Yoshida Y, Musha T, Ando S, Takai Y: **A novel prenyltransferase for a small GTP-binding protein having a C-terminal Cys-Ala-Cys structure.** *J Biol Chem* 1991, **266**:16981-16984.
22. Seabra MC, Goldstein JL, Sudhof TC, Brown MS: **Rab geranylgeranyl transferase. A multisubunit enzyme that prenylates GTP-binding proteins terminating in Cys-X-Cys or Cys-Cys.** *J Biol Chem* 1992, **267**:14497-14503.

23. Anant JS, Desnoyers L, Machius M, Demeler B, Hansen JC, Westover KD, Deisenhofer J, Seabra MC: **Mechanism of Rab geranylgeranylation: formation of the catalytic ternary complex.** *Biochemistry* 1998, **37**:12559-12568.
24. Gelb MH, Brunsveld L, Hrycyna CA, Michaelis S, Tamanoi F, Van Voorhis WC, Waldmann H: **Therapeutic intervention based on protein prenylation and associated modifications.** *Nat Chem Biol* 2006, **2**:518-528.
25. Wright LP, Philips MR: **Thematic review series: lipid posttranslational modifications. CAAX modification and membrane targeting of Ras.** *J Lipid Res* 2006, **47**:883-891.
26. Barrowman J, Michaelis S: **ZMPSTE24, an integral membrane zinc metalloprotease with a connection to progeroid disorders.** *Biol Chem* 2009, **390**:761-773.
27. Boyartchuk VL, Ashby MN, Rine J: **Modulation of Ras and a-factor function by carboxyl-terminal proteolysis.** *Science* 1997, **275**:1796-1800.
28. Ahearn IM, Haigis K, Bar-Sagi D, Philips MR: **Regulating the regulator: Post-translational modification of RAS.** *Nature Reviews Molecular Cell Biology* 2012, **13**:39-51.
29. Dai Q, Choy E, Chiu V, Romano J, Slivka SR, Steitz SA, Michaelis S, Philips MR: **Mammalian prenylcysteine carboxyl methyltransferase is in the endoplasmic reticulum.** *J Biol Chem* 1998, **273**:15030-15034.
30. Kim E, Ambroziak P, Otto JC, Taylor B, Ashby M, Shannon K, Casey PJ, Young SG: **Disruption of the mouse Rce1 gene results in defective Ras processing and mislocalization of Ras within cells.** *J Biol Chem* 1999, **274**:8383-8390.
31. Bergo MO, Leung GK, Ambroziak P, Otto JC, Casey PJ, Gomes AQ, Seabra MC, Young SG: **Isoprenylcysteine carboxyl methyltransferase deficiency in mice.** *J Biol Chem* 2001, **276**:5841-5845.
32. Michaelson D, Ali W, Chiu VK, Bergo M, Silletti J, Wright L, Young SG, Philips M: **Postprenylation CAAX processing is required for proper localization of Ras but not Rho GTPases.** *Mol Biol Cell* 2005, **16**:1606-1616.
33. Berndt N, Hamilton AD, Sebti SM: **Targeting protein prenylation for cancer therapy.** *Nature Reviews Cancer* 2011, **11**:775-791.
34. Harousseau JL, Martinelli G, Jedrzejczak WW, Brandwein JM, Bordessoule D, Masszi T, Ossenkoppele GJ, Alexeeva JA, Beutel G, Maertens J, et al.: **A randomized phase 3 study of tipifarnib compared with best supportive care, including hydroxyurea, in the treatment of newly diagnosed acute myeloid leukemia in patients 70 years or older.** *Blood* 2009, **114**:1166-1173.
35. Rao S, Cunningham D, de Gramont A, Scheithauer W, Smakal M, Humblet Y, Kourteva G, Iveson T, Andre T, Dostalova J, et al.: **Phase III double-blind placebo-controlled study of farnesyl transferase inhibitor R115777 in patients with refractory advanced colorectal cancer.** *J Clin Oncol* 2004, **22**:3950-3957.
36. Blumenschein G, Ludwig C, Thomas G, Tan E, Fanucchi M, Santoro A, Crawford J, Breton J, O'Brien M, Khuri F: **A randomized phase III trial comparing lonafarnib/carboplatin/paclitaxel versus carboplatin/paclitaxel (CP) in chemotherapy-naive patients with advanced or metastatic non-small cell lung cancer (NSCLC).** *Lung Cancer* 2005, **49**:S30.
37. Raponi M, Lancet JE, Fan H, Dossey L, Lee G, Gojo I, Feldman EJ, Gotlib J, Morris LE, Greenberg PL, et al.: **A 2-gene classifier for predicting response to the**

- farnesyltransferase inhibitor tipifarnib in acute myeloid leukemia.** *Blood* 2008, **111**:2589-2596.
38. Rolland D, Ribrag V, Haioun C, Ghesquieres H, Jardin F, Bouabdallah R, Franchi P, Briere J, De Kerviler E, Chassagne-Clement C, et al.: **Phase II trial and prediction of response of single agent tipifarnib in patients with relapsed/refractory mantle cell lymphoma: a Groupe d'Etude des Lymphomes de l'Adulte trial.** *Cancer Chemother Pharmacol* 2010, **65**:781-790.
  39. Kauh J, Chanel-Vos C, Escuin D, Fanucchi MP, Harvey RD, Saba N, Shin DM, Gal A, Pan L, Kutner M, et al.: **Farnesyl transferase expression determines clinical response to the docetaxel-lonafarnib combination in patients with advanced malignancies.** *Cancer* 2011, **117**:4049-4059.
  40. Whyte DB, Kirschmeier P, Hockenberry TN, Nunez-Oliva I, James L, Catino JJ, Bishop WR, Pai JK: **K- and N-Ras are geranylgeranylated in cells treated with farnesyl protein transferase inhibitors.** *J Biol Chem* 1997, **272**:14459-14464.
  41. Balasis ME, Forinash KD, Chen YA, Fulp WJ, Coppola D, Hamilton AD, Cheng JQ, Sebti SM: **Combination of farnesyltransferase and Akt inhibitors is synergistic in breast cancer cells and causes significant breast tumor regression in ErbB2 transgenic mice.** *Clin Cancer Res* 2011, **17**:2852-2862.
  42. Garcia-Ruiz C, Morales A, Fernandez-Checa JC: **Statins and protein prenylation in cancer cell biology and therapy.** *Anticancer Agents Med Chem* 2012, **12**:303-315.
  43. Kieran MW, Gordon L, Kleinman M: **New approaches to progeria.** *Pediatrics* 2007, **120**:834-841.
  44. Wong NS, Morse MA: **Lonafarnib for cancer and progeria.** *Expert Opin Investig Drugs* 2012, **21**:1043-1055.
  45. Fong LG, Frost D, Meta M, Qiao X, Yang SH, Coffinier C, Young SG: **A protein farnesyltransferase inhibitor ameliorates disease in a mouse model of progeria.** *Science* 2006, **311**:1621-1623.
  46. Yang SH, Meta M, Qiao X, Frost D, Bauch J, Coffinier C, Majumdar S, Bergo MO, Young SG, Fong LG: **A farnesyltransferase inhibitor improves disease phenotypes in mice with a Hutchinson-Gilford progeria syndrome mutation.** *J Clin Invest* 2006, **116**:2115-2121.
  47. Nallan L, Bauer KD, Bendale P, Rivas K, Yokoyama K, Hornéy CP, Pendyala PR, Floyd D, Lombardo LJ, Williams DK, et al.: **Protein farnesyltransferase inhibitors exhibit potent antimalarial activity.** *Journal of Medicinal Chemistry* 2005, **48**:3704-3713.
  48. Bulbule VJ, Rivas K, Verlinde CL, Van Voorhis WC, Gelb MH: **2-Oxotetrahydroquinoline-based antimalarials with high potency and metabolic stability.** *J Med Chem* 2008, **51**:384-387.
  49. Fletcher S, Cummings CG, Rivas K, Katt WP, Horney C, Buckner FS, Chakrabarti D, Sebti SM, Gelb MH, Van Voorhis WC, et al.: **Potent, Plasmodium-selective farnesyltransferase inhibitors that arrest the growth of malaria parasites: structure-activity relationships of ethylenediamine-analogue scaffolds and homology model validation.** *J Med Chem* 2008, **51**:5176-5197.
  50. Gelb MH, Van Voorhis WC, Buckner FS, Yokoyama K, Eastman R, Carpenter EP, Panethymitaki C, Brown KA, Smith DF: **Protein farnesyl and N-myristoyl transferases: Piggy-back medicinal chemistry targets for the development of**

- antitrypanosomatid and antimalarial therapeutics.** *Molecular and Biochemical Parasitology* 2003, **126**:155-163.
51. Song JL, White TC: **RAM2: an essential gene in the prenylation pathway of *Candida albicans*.** *Microbiology* 2003, **149**:249-259.
  52. Smalera I, Williamson JM, Baginsky W, Leiting B, Mazur P: **Expression and characterization of protein geranylgeranyltransferase type I from the pathogenic yeast *Candida albicans* and identification of yeast selective enzyme inhibitors.** *Biochim Biophys Acta* 2000, **1480**:132-144.
  53. Seabra MC, Reiss Y, Casey PJ, Brown MS, Goldstein JL: **Protein farnesyltransferase and geranylgeranyltransferase share a common alpha subunit.** *Cell* 1991, **65**:429-434.
  54. Reiss Y, Seabra MC, Armstrong SA, Slaughter CA, Goldstein JL, Brown MS: **Nonidentical subunits of p21H-ras farnesyltransferase. Peptide binding and farnesyl pyrophosphate carrier functions.** *J Biol Chem* 1991, **266**:10672-10677.
  55. Lane KT, Beese LS: **Thematic review series: lipid posttranslational modifications. Structural biology of protein farnesyltransferase and geranylgeranyltransferase type I.** *J Lipid Res* 2006, **47**:681-699.
  56. Andres DA, Seabra MC, Brown MS, Armstrong SA, Smeland TE, Cremers FP, Goldstein JL: **cDNA cloning of component A of Rab geranylgeranyl transferase and demonstration of its role as a Rab escort protein.** *Cell* 1993, **73**:1091-1099.
  57. Andres DA, Goldstein JL, Ho YK, Brown MS: **Mutational analysis of alpha-subunit of protein farnesyltransferase. Evidence for a catalytic role.** *J Biol Chem* 1993, **268**:1383-1390.
  58. Hightower KE, Huang CC, Casey PJ, Fierke CA: **H-Ras peptide and protein substrates bind protein farnesyltransferase as an ionized thiolate.** *Biochemistry* 1998, **37**:15555-15562.
  59. Huang C, Hightower KE, Fierke CA: **Mechanistic studies of rat protein farnesyltransferase indicate an associative transition state.** *Biochemistry* 2000, **39**:2593-2602.
  60. Yang Y, Wang B, Ucisik MN, Cui G, Fierke CA, Merz KM, Jr.: **Insights into the mechanistic dichotomy of the protein farnesyltransferase peptide substrates CVIM and CVLS.** *J Am Chem Soc* 2012, **134**:820-823.
  61. Pais JE, Bowers KE, Fierke CA: **Measurement of the alpha-secondary kinetic isotope effect for the reaction catalyzed by mammalian protein farnesyltransferase.** *J Am Chem Soc* 2006, **128**:15086-15087.
  62. Pickett JS, Bowers KE, Fierke CA: **Mutagenesis studies of protein farnesyltransferase implicate aspartate beta 352 as a magnesium ligand.** *J Biol Chem* 2003, **278**:51243-51250.
  63. Saderholm MJ, Hightower KE, Fierke CA: **Role of metals in the reaction catalyzed by protein farnesyltransferase.** *Biochemistry* 2000, **39**:12398-12405.
  64. Hartman HL, Bowers KE, Fierke CA: **Lysine beta311 of protein geranylgeranyltransferase type I partially replaces magnesium.** *J Biol Chem* 2004, **279**:30546-30553.
  65. Long SB, Casey PJ, Beese LS: **Reaction path of protein farnesyltransferase at atomic resolution.** *Nature* 2002, **419**:645-650.
  66. Taylor JS, Reid TS, Terry KL, Casey PJ, Beese LS: **Structure of mammalian protein geranylgeranyltransferase type-I.** *Embo J* 2003, **22**:5963-5974.

67. Pickett JS, Bowers KE, Hartman HL, Fu HW, Embry AC, Casey PJ, Fierke CA: **Kinetic studies of protein farnesyltransferase mutants establish active substrate conformation.** *Biochemistry* 2003, **42**:9741-9748.
68. Hannoush RN, Sun J: **The chemical toolbox for monitoring protein fatty acylation and prenylation.** *Nat Chem Biol* 2010, **6**:498-506.
69. Benetka W, Koranda M, Maurer-Stroh S, Pittner F, Eisenhaber F: **Farnesylation or geranylgeranylation? Efficient assays for testing protein prenylation in vitro and in vivo.** *BMC Biochem* 2006, **7**:6.
70. Corsini A, Farnsworth CC, McGeady P, Gelb MH, Glomset JA: **Incorporation of radiolabeled prenyl alcohols and their analogs into mammalian cell proteins. A useful tool for studying protein prenylation.** *Methods Mol Biol* 1999, **116**:125-144.
71. Baron R, Fourcade E, Lajoie-Mazenc I, Allal C, Couderc B, Barbaras R, Favre G, Faye JC, Pradines A: **RhoB prenylation is driven by the three carboxyl-terminal amino acids of the protein: evidenced in vivo by an anti-farnesyl cysteine antibody.** *Proc Natl Acad Sci U S A* 2000, **97**:11626-11631.
72. Lin HP, Hsu SC, Wu JC, Sheen IJ, Yan BS, Syu WJ: **Localization of isoprenylated antigen of hepatitis delta virus by anti-farnesyl antibodies.** *J Gen Virol* 1999, **80 ( Pt 1)**:91-96.
73. Liu XH, Suh DY, Call J, Prestwich GD: **Antigenic prenylated peptide conjugates and polyclonal antibodies to detect protein prenylation.** *Bioconjug Chem* 2004, **15**:270-277.
74. Pompliano DL, Gomez RP, Anthony NJ: **Intramolecular fluorescence enhancement: A continuous assay of Ras farnesyl:protein transferase.** *Journal of the American Chemical Society* 1992, **114**:7946-7948.
75. Cassidy PB, Dolence JM, Poulter CD: **Continuous fluorescence assay for protein prenyltransferases.** *Methods Enzymol* 1995, **250**:30-43.
76. Krzysiak AJ, Aditya AV, Hougland JL, Fierke CA, Gibbs RA: **Synthesis and screening of a CaaL peptide library versus FTase reveals a surprising number of substrates.** *Bioorg Med Chem Lett* 2010, **20**:767-770.
77. Hougland JL, Lamphear CL, Scott SA, Gibbs RA, Fierke CA: **Context-dependent substrate recognition by protein farnesyltransferase.** *Biochemistry* 2009, **48**:1691-1701.
78. Fersht A: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W. H. Freeman; 1998.
79. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA: **Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities.** *J Mol Biol* 2010, **395**:176-190.
80. Maurer-Stroh S, Eisenhaber F: **Refinement and prediction of protein prenylation motifs.** *Genome Biol* 2005, **6**:R55.
81. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F: **Towards complete sets of farnesylated and geranylgeranylated proteins.** *PLoS Comput Biol* 2007, **3**:e66.
82. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJA, Hofmann K, Bairoch A: **The PROSITE database, its status in 2002.** *Nucleic Acids Research* 2002, **30**:235-238.
83. Reid TS, Terry KL, Casey PJ, Beese LS: **Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity.** *J Mol Biol* 2004, **343**:417-433.

84. London N, Lamphear CL, Hougland JL, Fierke CA, Schueler-Furman O: **Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity.** *PLoS Comput Biol* 2011, **7**:e1002170.
85. Mazur P, Register E, Bonfiglio CA, Yuan X, Kurtz MB, Williamson JM, Kelly R: **Purification of geranylgeranyltransferase I from *Candida albicans* and cloning of the CaRAM2 and CaCDC43 genes encoding its subunits.** *Microbiology* 1999, **145** ( Pt 5):1123-1135.
86. Kelly R, Card D, Register E, Mazur P, Kelly T, Tanaka KI, Onishi J, Williamson JM, Fan H, Satoh T, et al.: **Geranylgeranyltransferase I of *Candida albicans*: null mutants or enzyme inhibitors produce unexpected phenotypes.** *J Bacteriol* 2000, **182**:704-713.
87. McGeady P, Logan DA, Wansley DL: **A protein-farnesyl transferase inhibitor interferes with the serum-induced conversion of *Candida albicans* from a cellular yeast form to a filamentous form.** *FEMS Microbiol Lett* 2002, **213**:41-44.
88. Murthi KK, Smith SE, Kluge AF, Bergnes G, Bureau P, Berlin V: **Antifungal activity of a *Candida albicans* GGTase I inhibitor-Alanine conjugate. inhibition of Rho1p prenylation in *C. albicans*.** *Bioorganic & Medicinal Chemistry Letters* 2003, **13**:1935-1937.
89. Nishimura S, Matsunaga S, Shibazaki M, Suzuki K, Furihata K, van Soest RW, Fusetani N: **Massadine, a novel geranylgeranyltransferase type I inhibitor from the marine sponge *Stylissa aff. massa*.** *Org Lett* 2003, **5**:2255-2257.
90. Hast MA, Beese LS: **Structure of Protein Geranylgeranyltransferase-I from the Human Pathogen *Candida albicans* Complexed with a Lipid Substrate.** *Journal of Biological Chemistry* 2008, **283**:31933-31940.
91. Hicks KA, Hartman HL, Fierke CA: **Upstream polybasic region in peptides enhances dual specificity for prenylation by both farnesyltransferase and geranylgeranyltransferase type I.** *Biochemistry* 2005, **44**:15325-15333.
92. Trueblood CE, Ohya Y, Rine J: **Genetic evidence for in vivo cross-specificity of the CaaX-box protein prenyltransferases farnesyltransferase and geranylgeranyltransferase-I in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1993, **13**:4260-4275.
93. He B, Chen P, Chen SY, Vancura KL, Michaelis S, Powers S: **RAM2, an essential gene of yeast, and RAM1 encode the two polypeptide components of the farnesyltransferase that prenylates a-factor and Ras proteins.** *Proc Natl Acad Sci U S A* 1991, **88**:11373-11377.
94. Ohya Y, Qadota H, Anraku Y, Pringle JR, Botstein D: **Suppression of yeast geranylgeranyl transferase I defect by alternative prenylation of two target GTPases, Rho1p and Cdc42p.** *Mol Biol Cell* 1993, **4**:1017-1025.

## CHAPTER TWO

### INVESTIGATION OF SUBSTRATE RECOGNITION BY *CANDIDA ALBICANS* PRENYLTRANSFERASES USING A PEPTIDE LIBRARY APPROACH

#### INTRODUCTION

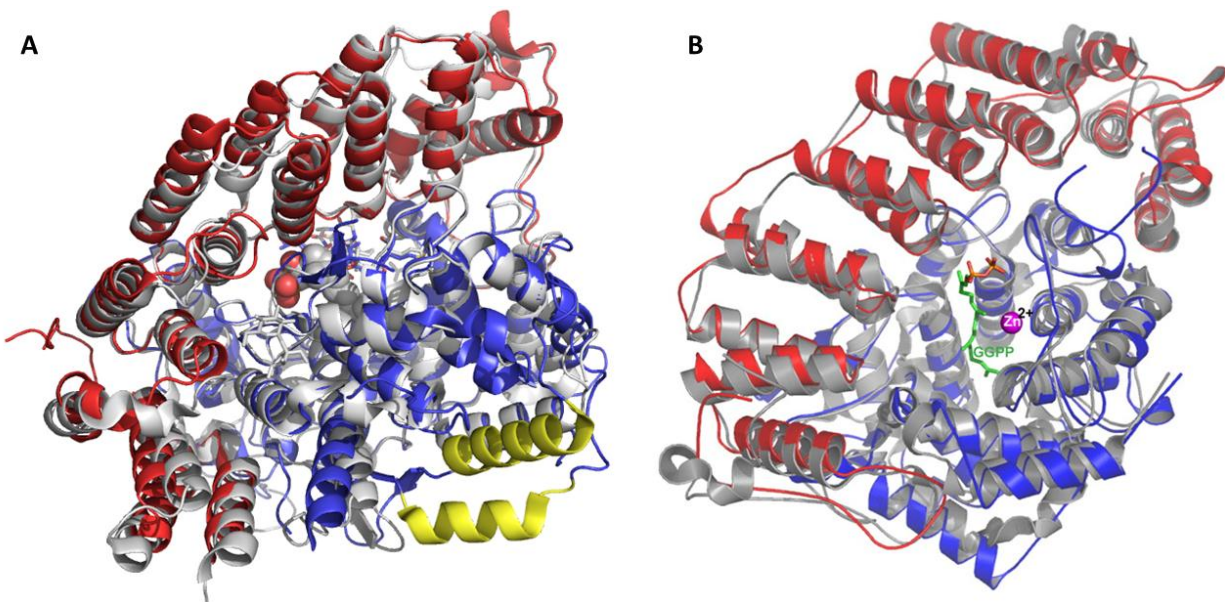
Protein farnesyltransferase (FTase) and geranylgeranyltransferase (GGTase-I) belong to a family of prenyltransferases that are zinc-dependent sulfur alkyltransferases [1,2]. FTase and GGTase-I are cytosolic heterodimeric proteins composed of  $\alpha$  and  $\beta$  subunits where these enzymes share the  $\alpha$  subunit and the  $\beta$  subunits have 25% identity [3]. FTase and GGTase-I catalyze the transfer of a 15-carbon farnesyl group from farnesyl diphosphate (FPP) and a 20-carbon geranylgeranyl group from geranylgeranyl diphosphate (GGPP), respectively, onto the cysteine sulfur four amino acids from the C-terminus [4]. Substrate proteins include the Ras and Rho small GTPase families, G protein  $\gamma$  subunits, protein phosphatases, phosphodiesterases, and nuclear lamins [5]. The attached lipid moiety aids in localization of proteins to cellular membranes and enhances protein-protein interactions [6]. The FTase substrate Ras, a small GTPase, was one of the first oncogenes to be identified as its deregulation or constitutive activation often leads to hyperplasia and tumorigenesis. Inhibitors of human protein prenyltransferases have been explored as treatments for cancer [7], laminopathies [8], and malaria [9].

Although historically most of the research on prenyltransferases has been concentrated on mammalian enzymes due to their downstream targets being implicated in cancer development and progression, genetic studies of protein prenylation in yeast have suggested prenyltransferases as potential targets for the development of new anti-mycotic agents [10]. In *Saccharomyces cerevisiae* genes encoding both the  $\alpha$  and  $\beta$  subunits of GGTase-I are essential, and disruption of the gene encoding the  $\beta$  subunit of FTase results in a temperature sensitive phenotype. Genetic studies also showed that the gene encoding the  $\alpha$  subunit of prenyltransferases in pathogenic yeast *Candida albicans* is essential for *C. albicans* viability [11]. Genetic and pharmacological interventions that specifically target the *C. albicans* GGTase-I resulted in a morphologically abnormal phenotype but failed to inhibit yeast growth [12]. To explain this, the authors hypothesized that in the absence of GGTase-I, FTase prenylates two essential GGTase-I substrates, Cdc42p and Rho1p, and this compensates for the lack of a functional GGTase-I. Such cross-reactivity has been observed with mammalian prenyltransferases and is one of the reasons proposed to limit clinical utility of FTase and GGTase-I inhibitors as cancer therapeutics [7]. Thus, a better understanding of substrate recognition by *C. albicans* FTase and GGTase-I can provide important information about these enzymes as drug targets for anti-fungal therapeutics.

FTase and GGTase-I are proposed to recognize their substrate proteins by the C-terminal “CaaX” motif, where “C” is the modified cysteine residue, “a” are aliphatic amino acids, and “X” represents an amino acid that determines specificity for either FTase (methionine, serine, glutamine, and alanine) and GGTase-I (leucine and phenylalanine) enzymes [13]. This model was derived from studies with mammalian prenyltransferases and identifies only the best substrates while leaving out many other CaaX sequences that biochemical studies have shown to be good prenyltransferase substrates. To date, most of the rules for substrate recognition of *C.*



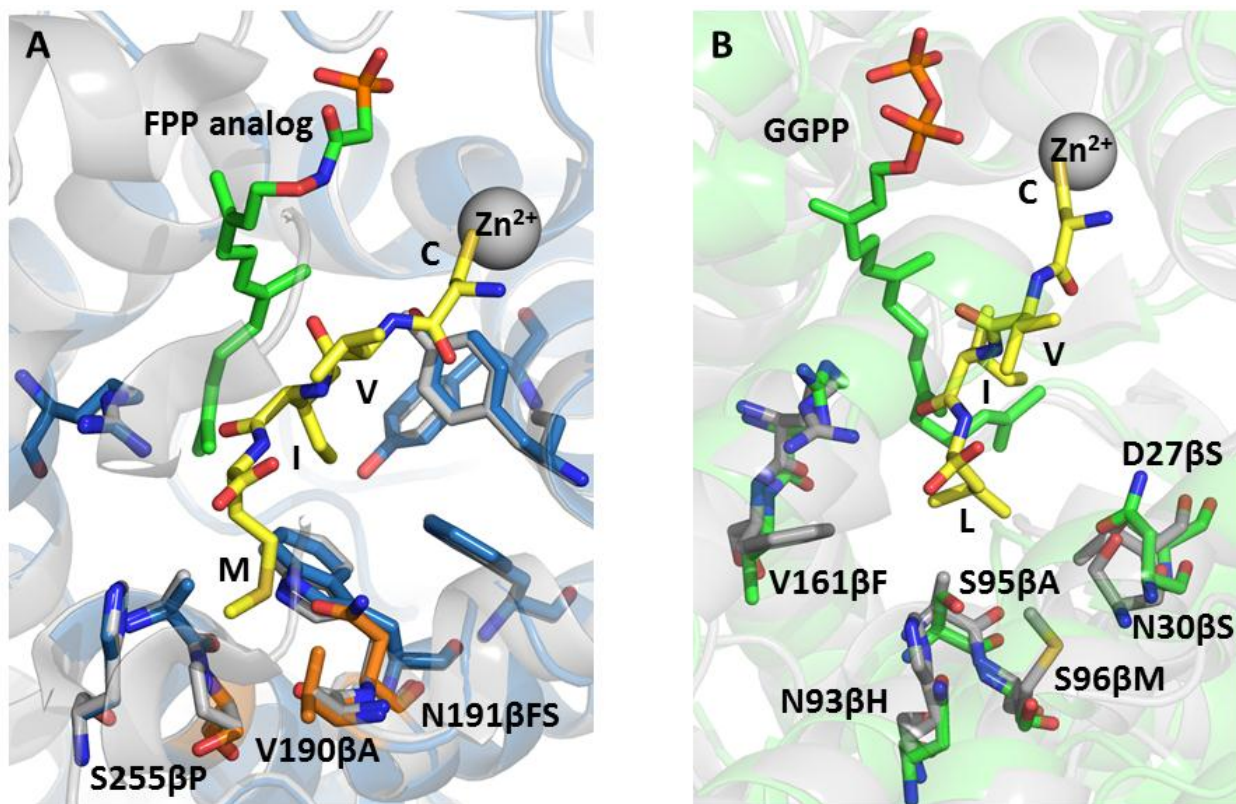
*albicans* prenyltransferases have been based on our understanding of mammalian prenylation. However, the sequence identity between the human and *C. albicans* prenyltransferases is ~30%, providing a possibility that substrate recognition varies between these species [14].



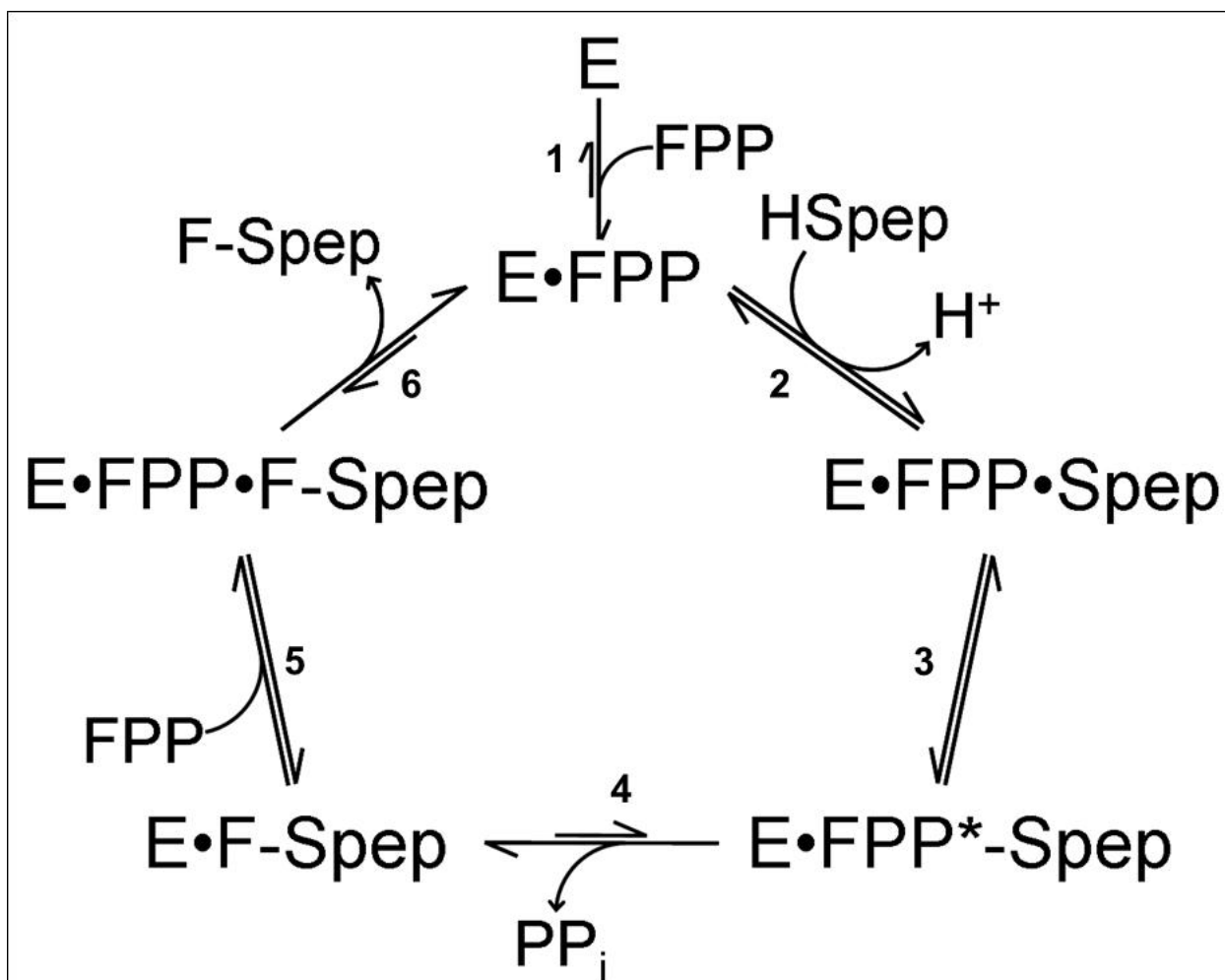
**Figure 2.1. Structure comparison of mammalian and yeast FTase and GGTase-I.** A) Mammalian FTase (1NT6) in grey and *C. neoformans* FTase (3Q75) in blue/red/yellow; B) Mammalian GGTase-I (3DRA) in grey and *C. albicans* GGTase-I (1N4P) in blue/red. Despite low overall sequence identity, the structures of mammalian and yeast prenyltransferases are quite similar in overall architecture. The  $\alpha$  subunits (red and grey) consist of  $\alpha$ -helices arranged in antiparallel pairs to form a crescent that envelopes part of the  $\beta$  subunit (blue and grey).

Crystal structures of mammalian FTase and GGTase-I complexed with numerous peptide substrates, lipid donors, products, and inhibitors have been solved [15,16]. Structures of non-mammalian prenyltransferases include that of FTase from the pathogenic yeast *Cryptococcus neoformans* [17] and GGTase-I from *C. albicans* [18]. Overlays of the mammalian and yeast prenyltransferases are shown in Figure 2.1 and, despite low overall sequence identity, the overall architectures of mammalian and yeast enzymes are quite similar, as can be seen by the excellent overall structural alignment of the two proteins. However, both FTase and GGTase-I have sequence alterations in the binding pockets for both the lipid and CaaX substrates, leading to

potential differences in substrate specificity (Figure 2.2). Based on structural and sequence data, FTase substrate binding sites contain more conserved residues than GGTase-I.



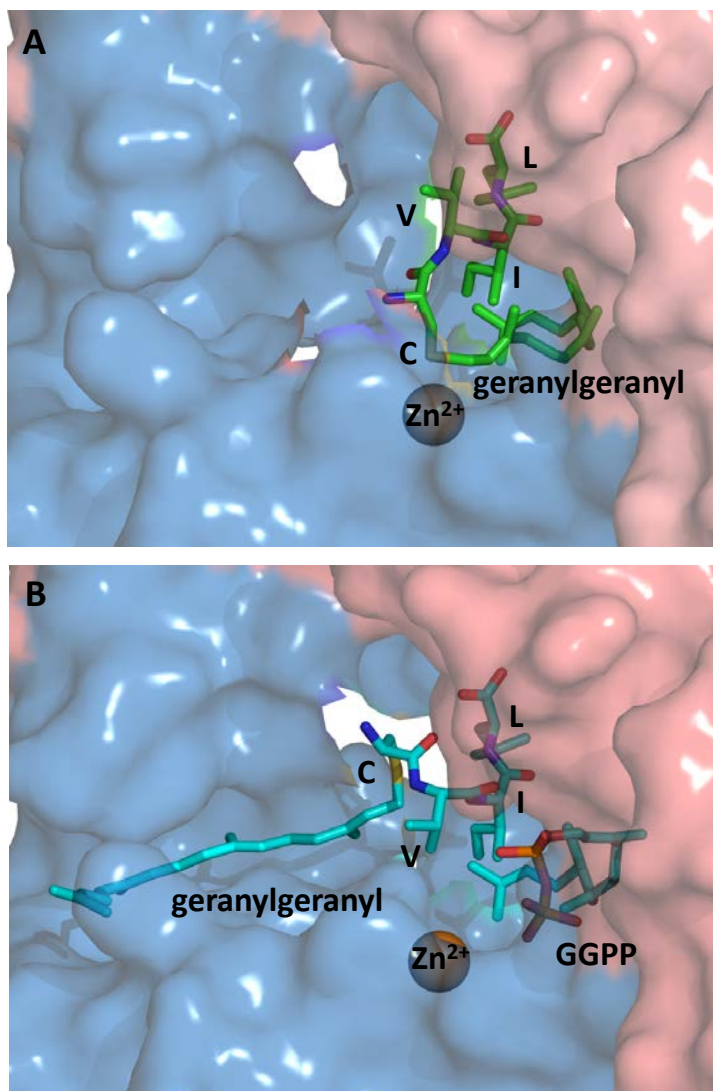
**Figure 2.2. The CaaX peptide-binding site in mammalian and *C. albicans* FTase and GGTase-I.** A) Superposition of FTase substrate binding pockets with CVIM peptide. Non-conserved residues are labelled and orange. B) Superposition of GGTase-I substrate binding pockets with CVIL peptide. Non-conserved residues are labelled. (PDB ID 1D8D, 1N4Q, and 3DRA). Mammalian enzymes are in grey and *C. albicans* are in blue or green.



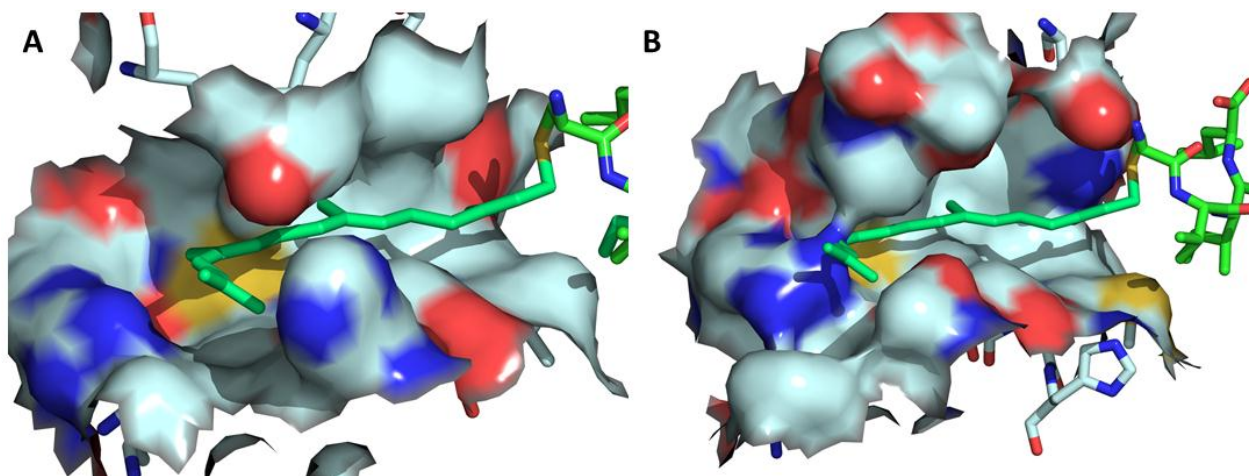
**Figure 2.3. Kinetic mechanism of mammalian FTase.** This is a simplified diagram representing the preferred kinetic pathway. 1) FTase binds the FPP lipid donor. 2) FTase•FPP binds the peptide substrate. 3) Conformational change occurs whereby C1 of FPP moves closer to the peptide thiolate. 4) Chemistry step is followed by rapid release of PP<sub>i</sub>. 5) Second molecule of FPP binds to the E•F-Spep complex to facilitate product release. 6) Dissociation of prenylated product.

To study substrate recognition in mammalian prenyltransferases, the reactivity of rat FTase and rat GGTase-I with peptide libraries consisting of different CaaX sequences have been measured. Unexpectedly, peptides that are substrates for these prenyltransferases fall into two different pools: 1) multiple turnover substrates (MTO) and 2) single turnover substrates (STO), where peptides only undergo stoichiometric prenylation and do not dissociate from the enzyme [19]. FTase MTO substrates are able to progress through the entire catalytic cycle as depicted in Figure 2.3. Product release from the FTase•F-Spep complex requires an additional

conformational change whereby the lipid group of the product swings into the product exit groove (Figure 2.4). This step is facilitated by binding of a second FPP molecule in the substrate binding pocket (step 5) which traps the product in the conformation where the prenyl group is bound in the exit groove. In contrast, the STO substrates bind and proceed through the chemical step but do not dissociate. It is not yet clear why the dissociation kinetics of the STO substrates are so slow or what is the biological role of these substrates.



**Figure 2.4. Conformational change of the product in mammalian GGTase-I .A)** Geranylgeranylated CVIL peptide (green) immediately following catalysis. The lipid group is in the substrate binding pocket. **B)** Binding of a second GGPP (blue) in the substrate pocket traps the geranylgeranyl group of the product (blue) in the exit groove. The GGTase-I  $\alpha$  subunit is in light pink and  $\beta$  subunit is in light blue in both panels. (PDB ID 1N4R and 1N4S)



**Figure 2.5. Mammalian and *C. albicans* GGTase-I exit grooves.** A) Mammalian GGTase-I (1N4S) exit groove shows the isoprenoid moiety (teal) fits into the primarily hydrophobic cavity of the enzyme. B) *C. albicans* GGTase-I does not fit the isoprenoid group when it is docked into the enzyme, as a steric clash is observed in the region of the 3<sup>rd</sup> isoprenoid unit of the geranylgeranyl group.

One of the most significant differences between the structures of mammalian and *C. albicans* GGTase-I is that the yeast enzyme has a very narrow exit groove, a mere 6.4 Å wide at its narrowest point. The ~4 Å narrowing of the *C. albicans* GGTase-I exit groove compared to the mammalian exit groove is caused by Pro21β and Pro347β projecting into the channel. Comparison of the mammalian and *C. albicans* GGTase-I structures in Figure 2.5 shows that the prenyl group of the product cannot fit into this narrow tunnel, preventing the interaction between the lipid and the exit groove that occurs upon the conformational change (step 5 in Figure 2.3). Although the effect of this structural change on product dissociation is currently unknown, product dissociation rate could increase due to the loss of the lipid interactions with the exit groove. This hypothesis is supported by transient kinetic studies with *S. cerevisiae* FTase that show the product release rate constant is  $3.5 \pm 0.2 \text{ s}^{-1}$  for this enzyme compared to  $0.061 \pm 0.004 \text{ s}^{-1}$  for rat FTase, a rate that is ~60 times faster [20,21]. Sequence alignment of *S. cerevisiae*, *C. neoformans* and rat FTase shows that the *S. cerevisiae* enzyme lacks an exit groove, possibly accounting for the large difference in product release kinetics.

In this chapter, I describe cloning, heterologous expression, and purification of active *C. albicans* FTase and GGTase-I. To obtain information about substrate recognition determinants of *C. albicans* FTase and GGTase-I, the reactivity of these enzymes with a library of 328 peptides was measured. This study showed that while *C. albicans* FTase has similar substrate preferences to the mammalian enzyme, *C. albicans* GGTase-I has a relaxed substrate specificity. In addition, *C. albicans* GGTase-I prenylates a significantly larger pool of peptides under MTO conditions than the other enzymes. These studies provide a better understanding of substrate recognition by *C. albicans* prenyltransferases, and differential substrate recognition between mammalian and *C. albicans* enzymes. The latter information can help decide whether *C. albicans* FTase and/ or GGTase-I should be targeted in anti-mycotic drug discovery efforts and provide clues about the structures of peptide mimetic inhibitors that specifically target the yeast enzymes.

## EXPERIMENTAL PROCEDURES

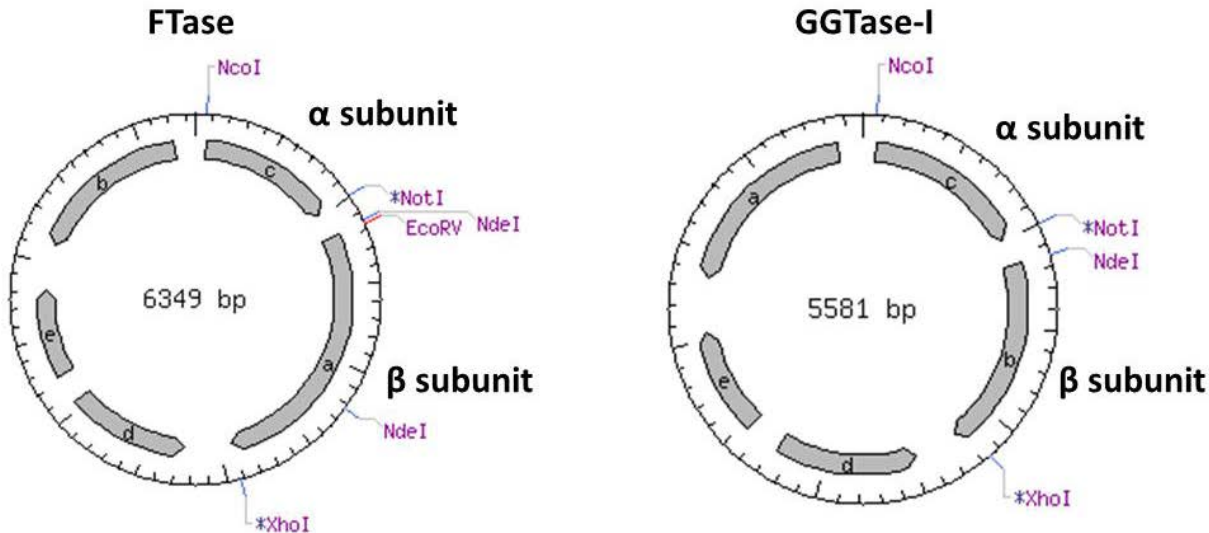
### *Cloning of C. albicans FTase and GGTase-I genes*

The complete reading frames of *C. albicans* *RAM2* ( $\alpha$  subunit), *RAM1* (FTase  $\beta$  subunit), and *CDC43* (GGTase-I  $\beta$  subunit) were amplified by PCR from *C. albicans* genomic DNA strain SC5314 (American Type Culture Collection, ATCC). Complete cDNA sequences for these genes were obtained from NCBI GenBank using accession numbers AF110691 for *RAM2*, AF110690 for *CDC43*, and XP\_710548 for *RAM1*. The *RAM2* gene encoding the  $\alpha$  subunit open reading frame was amplified using forward primer 5'-CTG ACG CCA TGG ATG ACA GAC TCC AAA TAT GAC-3' with an NcoI restriction site (underlined), and reverse primer 5'-CAT TAT GCG GCC GCT TAC ACC GAT GTG AGT TTG G-3' with a NotI restriction site (underlined). The *CDC43* gene encoding the GGTase-I  $\beta$  subunit open reading frame was

amplified using forward primer 5'-GCG CTG CAT ATG AAC CAA **CTG CTG** ATT **AAC** AAA CAT GAG AAA T TT TT-3' with an NdeI restriction site (underlined). In the *CDC43* forward primer, codons for leucine and asparagine were changed from TTA to CTG and AAT to AAC, respectively (bold), to optimize for bacterial codon usage. The *CDC43* reverse primer was 5'-GCG CTG CTC GAG TTA ATA CTT TAT TTC TTC TTT AAA AAA TTG-3' with an XhoI restriction site (underlined). The *RAM1* gene encoding the FTase  $\beta$  subunit was amplified using forward primer 5'-GCG CTG GAT ATC ATG AGT CAA GAT TCT **AAC** GCT AAA ATT AA-3' with an EcoRV restriction site (underlined) and an optimized codon for asparagine (bold) and reverse primer 5'-GTA ATA CTC GAG TTA ACG TTT TTG TTC TGC TCT-3' with an XhoI restriction site (underlined). All PCR amplifications were carried out using *PfuTurbo* DNA polymerase (Agilent Technologies), 10 mM dNTP mix, and 10X buffer supplied with the enzyme containing 10 mM MgCl<sub>2</sub>, with temperature cycling setup as follows: 1 cycle of 95°C for 2 min; 30 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 2 min; 1 cycle of 72°C for 10 min. The PCR products were digested with appropriate restriction enzymes (New England Biolabs) and fragments were gel purified (Promega Wizard SV Gel and PCR Clean-up System).

The pCDFDuet-1 vector (Novagen) was chosen to co-express both subunits of FTase and GGTase-I in *Escherichia coli* (*E. coli*); this vector contains two multiple cloning sites (MCS) under the control of separate isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG)-inducible T7 promoters for robust co-expression of two gene products. To construct the GGTase-I expression plasmid, the digested *RAM2* gene encoding the  $\alpha$  subunit was subcloned into the NcoI and NotI restriction sites in MCS1 of the pCDFDuet-1 vector. The digested *CDC43* gene encoding the  $\beta$  subunit was subcloned into the NdeI and XhoI restriction sites of MCS2 of the pCDFDuet-1 vector to achieve the final GGTase-I expression plasmid (Figure 2.6). To construct the FTase

expression plasmid, the *RAM1* gene encoding the FTase  $\beta$  subunit was subcloned into the EcoRV and XhoI sites of the pCDFDuet-1 vector with the *RAM2* gene already inserted (Figure 2.6). The University of Michigan DNA Sequencing Core performed the sequence analysis to confirm error-free construction of the expression vectors. Since the *RAM1* gene contains 1749 nucleotides, and the sequencing methodology used only provides accurate results for up to 800 nucleotides, additional primers located in the middle of the *RAM1* gene were used to obtain accurate sequencing information, with forward primer 5'-CGA AGT AGT TGT GGT AGT GGG GCA T-3' and reverse primer 5'-CGA GCA TCC ATC TCT CCA TTT TCA T-3'.



**Figure 2.6. Plasmid maps for *C. albicans* FTase and GGTase-I.** The  $\alpha$  and  $\beta$  subunits of FTase and GGTase-I are under separate T7 promoters.

#### *C. albicans* FTase and GGTase-I expression and purification

The FTase expression plasmid was transformed into BL21 (DE3) chemically competent *E. coli* cells (Novagen) for expression. A single colony was picked from an LB plate supplemented with 50  $\mu$ g/mL streptomycin (Sigma) and grown for 8 hours in a 50 mL LB culture with the same antibiotic. The 50 mL LB culture was used to inoculate 2 L of LB/strep media and incubated at 37°C until  $A_{600} = 0.6$ . At this point the temperature was reduced to 25°C



and protein expression was induced by addition of 1 mM IPTG and 0.5 mM ZnCl<sub>2</sub>. Cells were incubated for 18 hours, harvested at 6000 x g, and cell paste was stored at -80°C.

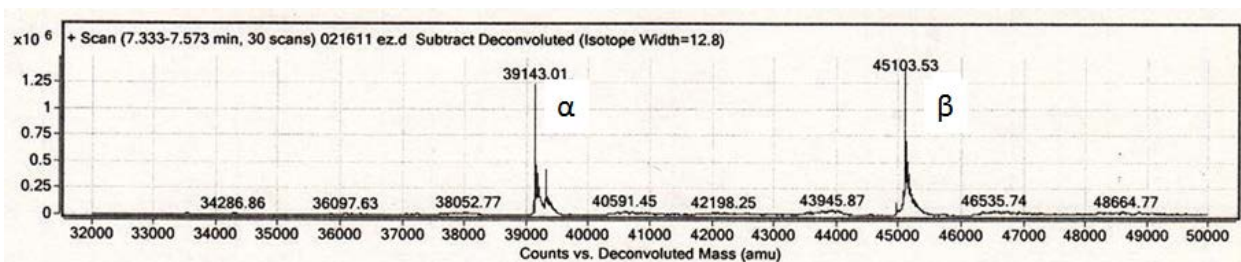
The cell paste from a 2 L culture was resuspended in 50 – 100 mL of buffer A (50 mM Hepes pH 7.8, 10 μM ZnCl<sub>2</sub>, 2 mM Tris(2-carboxyethyl)phosphine (TCEP)), complete EDTA-free Protease Inhibitor Cocktail (Roche), 4-(2-Aminoethyl)-benzenesulfonyl fluoride hydrochloride Pefabloc SC (Roche), and 1 μL Benzonase® nuclease (Sigma). To ensure complete resuspension, cells were shaken for 45 min at room temperature. Cells were lysed using a pressurized microfluidizer (Microfluidics Corp) and the resulting crude lysate was clarified by centrifugation at 12,000 rpm for 45 min at 4°C. The supernatant was decanted and 10% w/v streptomycin sulfate was added drop-wise at 4°C, and allowed to stir for additional 30 min at 4°C. The lysate was centrifuged again at 12,000 rpm for 45 min at 4°C, and filtered through a 0.45 μm hydrophilic mixed cellulose esters filter (Millipore).

The filtered lysate was applied to a D EAE Sephacel column (GE Healthcare Life Sciences) and fractionated in Buffer A using a linear gradient of 0.1 M NaCl to 0.5 M NaCl over 10 column volumes. Fractions containing FTase, as determined by SDS-PAGE, were pooled and concentrated. The protein was applied to a Talon column containing Cobalt resin (Clontech) and fractionated using a 2 – 100 mM imidazole gradient in buffer containing 50 mM Hepes pH 7.8, 2 mM TCEP, and 300 mM NaCl. Fractions containing FTase were pooled, concentrated, and dialyzed against 50 mM Hepes pH 7.8, 2 mM TCEP, and 10% glycerol to a final concentration of 5.5 mg/mL, aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C.

GGTase-I expression and purification was performed similarly to that of FTase, with minor adjustments. Expression was induced with 0.5 mM IPTG and 0.5 mM ZnSO<sub>4</sub>, and cells were grown for 6 hours at 25°C. Cells were pelleted, resuspended, and lysed following the same

protocol at *C. albicans* FTase. The purification procedure for GGTase-I was identical to that for FTase, except that the imidazole gradient for the Talon column was 2 – 200 mM.

The molecular mass of *C. albicans* FTase and GGTase-I was determined using Agilent Q-TOF HPLC-MS Mass Spectrometer (University of Michigan Chemistry Department Analytical Core). A deconvoluted spectrum of *C. albicans* GGTase-I is shown in Figure 2.7 and predicted and measured molecular masses are in Table 2.1.



**Figure 2.7. Analysis of molecular mass of *C. albicans* GGTase-I by LC-MS.** *C. albicans* GGTase-I was diluted to 0.1 mg/ml in 0.1% formic acid and injected onto an Agilent Q-TOF HPLC-MS. LC separation was performed on ZORBAX Poroshell 300SB-C8 column (Agilent) using a 5 – 95% acetonitrile linear gradient, followed by ESI+ MS. GGTase-I  $\alpha$  and  $\beta$  subunits are labeled accordingly.

Protein	Measured mass (Da)	Predicted mass (Da)
FTase/GGTase-I $\alpha$ subunit	39,143	39,142
FTase $\beta$ subunit	66,726	66,725
GGTase-I $\beta$ subunit	45,103	45,102

**Table 2.1. Molecular mass of *C. albicans* FTase and GGTase-I.** Experimentally determined molecular masses of FTase and GGTase-I exactly match their calculated values.

### Peptide libraries

To identify potential *C. albicans* FTase and GGTase-I substrates, the open reading frames within the *C. albicans* genome were searched using the Swiss-Prot database to identify all proteins that contain a cysteine residue at the fourth amino acid position from the C-terminus (-xxxxxxxxxxCxxx>). This search produced 137 unique 15 amino acid sequences and 120 unique Cxxx sequences, as shown in Table 2.2.

QFLFLFQNHFYCKLA	NQOKTQQQQDHCNQI	SLSMMMITVIECIIN
ANNKFGKVKENCHYA	NTCSSSSIQRKCVYI	FVHEVLHNCPCGENN
NADVCCGCCCCCCC	AEIAKELSQLVCKEK	GSECGFPAFASHCIRN
HPHPLPGAARFCWFC	NKYKTMRTNLLCSEK	VGSFYSDDYQACLNR
MATFIGNMLFACCKC	EREVIMEVIKTCQFK	VHGVVTDTKKCCFSN
<b>RKLFQNFKLNACIQC</b>	HKYFPNISIIPCIGK	IVDNDYLWELYCHSN
CQLLYLYRVGICWYC	PPIRKSEDSIIICRGK	IYDQLPECAKPCMFQ
ACGGDGGDGGGCGGD	RPNIYINNVQSCIKI	RKYEQQGYEQSCAQQ
<b>IKNTSSDGGDHCVID</b>	KSSRLAFDDGSCCKK	<b>PPEFGKDLQEGCKQQ</b>
EYIKEHGLGNACLKD	ERLGVNDYCVICGKK	<b>DEEAHGGPGVQCASQ</b>
ICFSYYQDEEDCFSD	AANLDIVSVEECMKK	SPSLHVLKALKCFWQ
TEFIVGFVLGVCYSD	NVCTILNFSKNCGSK	MNEIQSNILDSCNLR
LELVSSFRKDWCDYD	HKYFPNISIIPCISK	QMESWKSHRNWCLAS
YGVSKRIPMDPCTYD	YVCPVPCNCRNSK	KRSGTWLSWIFCCCS
DWAKVYFFGRDCSKE	SGAFTVPVAEQCADL	CTKFYCFVSLECCFS
RYTMVAINVGVCTLE	DPVSPSALKNRCYFL	DLKRLEEGSVTCIKS
GLLELSNRLELCFPE	<b>PGVDINILNSECLIL</b>	SCRYLVIKSSICSQS
DWIKVYLFGRDCSRE	<b>LEPPVIKSKKCTIL</b>	<b>KQKKYLRNQSGCVVS</b>
MGFNKYVKWAGCKCF	<b>AFTREKSKKCVIL</b>	<b>KSSQIHAENKCCIIT</b>
LVVKLGLDNGSCSDF	EEGEDVEVPYICVKL	QGLFTLMIPIRCST
YNSFYLYILLFCIIF	<u>MNEIQSNILDSCNLL</u>	EKGVFGNEIGCPVT
VEHLSKNISLVCKNF	KGTKCFPFLAICSML	VNSIARRSIGSCDFV
PYCESRPEYLQCNVF	YKNWRKEALEKCDQL	LEKFMNNKLKDCSFV
SRDIIICVIDLSCVFG	NKEIIFYKALLCESL	<b>NKDEKSKKLNCTIV</b>
FLYFLSDDNINCQKG	NRNLLQSVHFECNSL	<b>KQSSSKSKNGCCVIV</b>
ISTFHSYLIYLCHLG	NKEIIFYQPLLCEYL	LTGTVKLGTPVCDMV
QQWRFDHVSQCYLG	<u>IWYGKIDTRACFVL</u>	LESSASSMTVKCENV
DLDTQLFTIGFCKMG	FSFDPLMISSCCNVL	TWTPQPGWICQVRV
KIYEFFVHMGWCSQG	<b>RVKEKKEKSKCVVL</b>	NCVNILKTSFNCHWV
WDVETITFYVVCISG	TIYFIDIWNEKCLYL	VENLCQHYIGWCSFW
WDVETITFYVVCISG	PHTGLFQLFLCSYL	HLVLIQCRYLLCGHW
FTDVTEMFCVQCAVG	SDSDLLEVFTHCSFM	FDRLITKNFRYCIKW
QPVTDRQFPNQCGFH	<b>PETAEKSSCKCIIM</b>	SCVCRRCRCRYCLLW
<b>YDYSLDMWSCGCIH</b>	<b>DHPKSSSGSKCCTIM</b>	GTVMELEEVFKCIRW
RLRLRYLKDLKACEGI	<b>FYKQAKKTNSCCVIM</b>	EHNNGESTYHQCPYW
<b>LVKRPWRDLYNCKII</b>	<b>SKRQDDYKKEHCIMM</b>	KAKAKAKYSIQCFDY
<b>DHPKSSSGSKFCTII</b>	<b>ELRKNANSGEKCIVM</b>	RLDRINENRCPCIDY
IFVLLINLSLRCKLI	<b>KFSIRKSKKENCVVM</b>	RLDRINENRCPCIEY
RDIESKFVPDVCINI	QSDKTLKYSIDCKFN	LLYIVMTGVEICLEY
VSQTLTSLISLCHQI	WLRNLASAALVCHHN	HAAFLKMRERRCHNY

**Table 2.2. Unique Cxxx> sequences found in the *C. albicans* genome.** Predicted FTase substrates are bolded and GGTase-I substrates are underlined, and dual substrates are bolded and underlined (predictions based on PrePS algorithm). Peptides are ordered alphabetically by the X residue.

Out of the 120 unique Cxxx sequences, only fourteen of them overlap with peptides that correspond to mammalian proteins where reactivity with rat FTase and GGTase-I has been previously tested in our laboratory [19]. Additionally, only eleven of these sequences match known human FTase substrates. Running these sequences through PrePS prenylation prediction software (refer to Chapter 1 for details) [22,23], 21 of them were predicted to be FTase substrates and 12 were predicted to be GGTase-I substrates (bolded and underlined in Table 2.2, respectively). Since the overlap between Cxxx sequences found in the mammalian and *C. albicans* genomes was so small, a good comparison could not be performed if only Cxxx sequences found in the *C. albicans* genome were tested. In order to compare substrate specificity between mammalian and *C. albicans* prenyltransferases, a peptide library chosen to test *C. albicans* prenyltransferases consisted primarily of Cxxx sequences whose reactivity with mammalian prenyltransferases had been measured previously. As a result, 328 unique peptide sequences (Table 2.3) were tested against *C. albicans* prenyltransferases, only eleven of which had not been tested with mammalian enzymes.

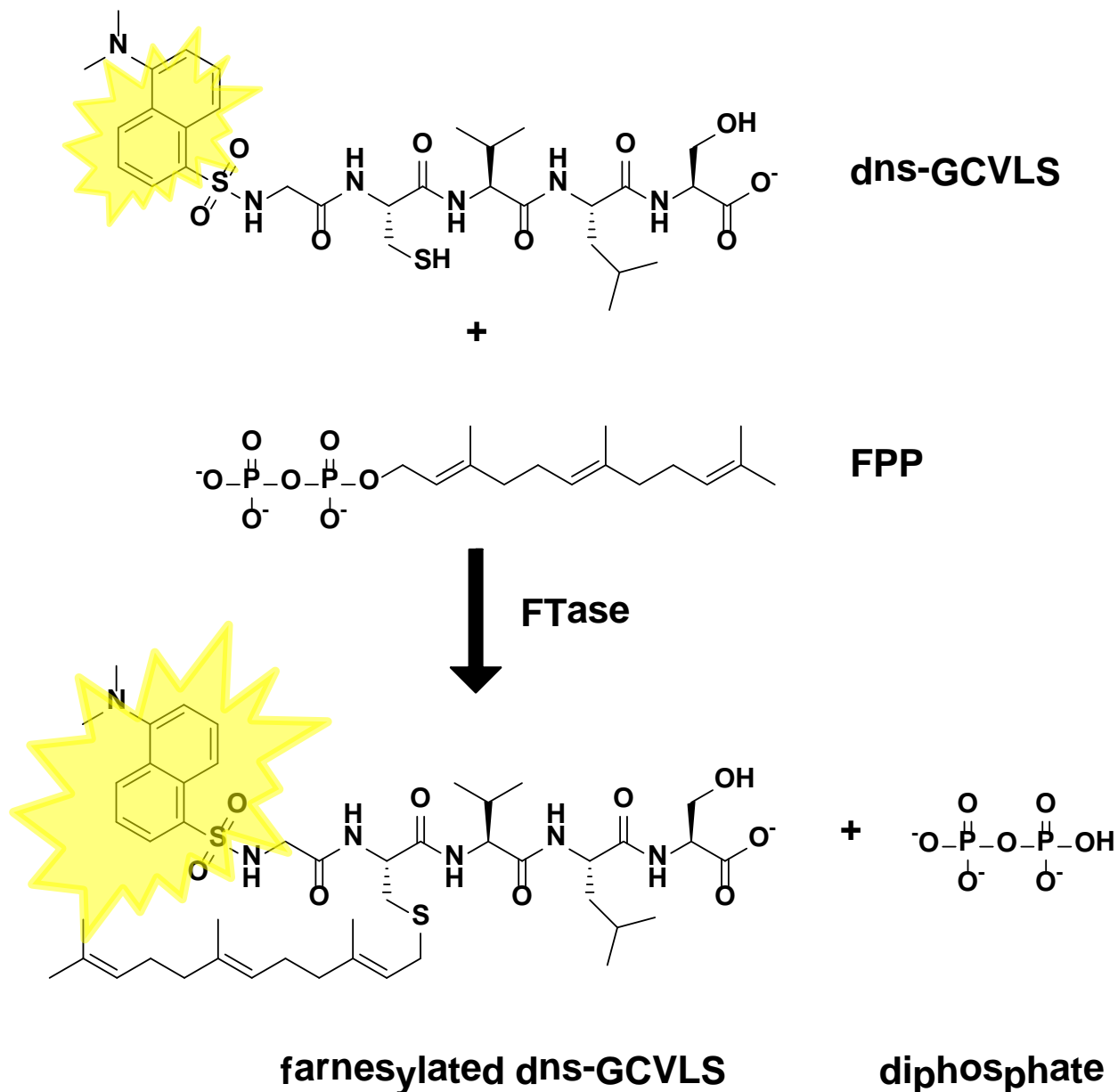
CKAA	CTVC	CNLF	CIII	CGGL	CISL	<b>CKFN</b>	CVTQ	CFPS	CCFV
CGCA	CPED	CGQF	CSII	CWGL	CLSL	<b>CNFN</b>	CAVQ	CLPS	CTHV
CEDA	CQED	<b>CNVF</b>	<b>CTII</b>	CIHL	CVSL	CGGN	CSVQ	CMPS	<b>CVIV</b>
CQDA	CLFD	CTVF	CVII	CLHL	CTTL	CIGN	CACR	CQPS	CTKV
CQEA	CEGD	CADG	CMKI	CAIL	CAVL	CVIN	CGCR	CNRS	CTMV
CSEA	CNHD	CSKG	CNKI	CCIL	CCVL	CRKN	CTCR	CFSS	CGNV
CGIA	<b>CVID</b>	CFLG	CQKI	CGIL	<b>CNVL</b>	CKLN	CFDR	CMSS	CWQV
CVIA	CALD	CPLG	CTKI	CIIL	CSVL	CLLN	CVIR	CSSS	CAWV
CKKA	CYPD	CSLG	CQNI	<b>CLIL</b>	<b>CVVL</b>	CYQN	CMKR	CVSS	CLYV
CTKA	CPRD	CFMG	CSPI	CNIL	CAYL	CCSN	CTKR	CLTS	CPFW
CCLA	CSSD	CPMG	CGQI	CPIL	CGEM	CYSN	CAVR	CNTS	CIGW
<b>CKLA</b>	CSVD	CCNG	CNTI	CSIL	CWEM	CVVN	CMVR	CSVS	CQLW
CTLA	CRGE	CSSG	CHVI	<b>CTIL</b>	<b>CSFM</b>	CQWN	<b>CLAS</b>	CLWS	CRLW
CVLA	CCIE	CATG	CSVI	<b>CVIL</b>	CVFM	CSIP	CPAS	CQWS	CGTW
CWLA	CFLE	CCTG	CCWI	CEKL	CLHM	CKKP	CGCS	CEYS	CEVW
CCNA	CLME	CWVG	CHWI	CFKL	CGIM	CLKP	CKCS	CLET	CQVW
CTQA	CENE	CGYG	CKWI	<b>CVKL</b>	<b>CIIM</b>	CLMP	CTDS	CLIT	CRVW
CCTA	CKSE	CRYG	CRWI	CWKL	CNIM	CAQP	CAES	CVIT	CSVW
CQTA	CYSE	CAAH	CTAK	CALL	CSIM	CCRP	CGHS	CCKT	CWDY
CEVA	CGTE	CVAH	CEEK	CCLL	<b>CTIM</b>	CWRP	CVHS	CLKT	CDIY
CIVA	CAVE	CGCH	CWHK	CILL	<b>CVIM</b>	CAGQ	CKIS	CALT	CVIY
CNWA	CKYE	CLEH	<b>CKIK</b>	<b>CNLL</b>	CKKM	CSGQ	CRIS	CHLT	CTKY
CYYA	CHDF	<b>CGFH</b>	CVKK	CRLI	CSKM	CVIQ	CSIS	CLLT	CGLY
CSAC	CMGF	CCHH	CMMK	CSLL	CLLM	CVKQ	CVIS	CGNT	CVLY
CRGC	CIHF	CGKH	CWNK	CTLL	CNLM	CNLQ	CAKS	CSPT	CARY
CTGC	CSHF	CCLH	CIVK	CVLL	CSLM	CTLQ	CQKS	CRRT	CNSY
CRKC	CCIF	CVLH	CVYK	CLML	CTLM	CVLQ	CFLS	<b>CSRT</b>	CSVY
CSLC	CKIF	CQRH	CHDL	<b>CSML</b>	<b>CIMM</b>	CWLQ	CQLS	CKTT	CTVY
CDRC	CVIF	CACI	CSDL	<b>CLNL</b>	CVQM	CAMQ	CRLS	CVTT	CIYY
CGRC	CSKF	CGCI	CTDL	CQNL	<b>CIVM</b>	CNNQ	CVLS	CNVT	
CGSC	CTKF	CKFI	CFFL	CVPL	CLVM	<b>CKQQ</b>	CDMS	CSVT	
CPSC	CGLF	CIHI	CSFL	CGQL	CYVM	CITQ	CTMS	CVVT	
CMTC	CLLF	CGII	CDGL	CSQL	CKWM	CQTQ	CKNS	CACV	

**Table 2.3. List of all TKCxxx peptides tested with *C. albicans* FTase and GGase-I.** CaaX sequences that are found in the *C. albicans* genome are bolded. CaaX sequences that were added based on *C. albicans* genome are underlined. Peptides are ordered alphabetically by the X residue.

Peptides used in the screening were of the form dansyl-TKCxxx and were purchased from Sigma-Aldrich in a PEPscreen® custom peptide library format. The peptides were shipped lyophilized and upon receipt were dissolved in a DMSO/ethanol mix and stored at -80°C. Peptide concentrations were measured using Ellman's reagent to detect free sulfhydryl groups [24]. Peptides were spot-checked for mass accuracy and stability using Micromass LCT Time-of-Flight mass spectrometer (University of Michigan Chemistry Department Analytical Core).

### Multiple turnover FTase and GGTase-I Activity Assay

The prenylation activity of *C. albicans* FTase and GGTase-I with all of the peptides was measured using a continuous fluorescent assay, as shown in Figure 2.8 [25].



**Figure 2.8. Continuous fluorescence assay to monitor prenyltransferase reaction.** Short dansylated peptides can be used as prenyltransferase substrates, and the dansyl group increases in fluorescence upon modification due to the enhanced hydrophobic environment of the dansyl group in the prenylated product [25]. The farnesyltransferase reaction is used here as an example.

Initial measurements were performed at a single peptide and enzyme concentration, and the conditions were further optimized as necessary. For GGTase-I, 3  $\mu\text{M}$  peptide was incubated in assay buffer (50 mM Hepes pH 7.8, 5 mM TCEP, and 5 mM  $\text{MgCl}_2$ ) for 30 min. Reactions were initiated by adding 500 nM GGTase-I and 10  $\mu\text{M}$  GGPP in assay buffer, which were pre-incubated for 5 – 10 minutes. Prenylation of a dansyl peptide lead to an enhancement in fluorescence that was monitored continuously ( $\lambda_{\text{ex}} = 340 \text{ nm}$ ,  $\lambda_{\text{em}} = 520 \text{ nm}$ ) in a 96-well plate (Costar, black, non-binding surface) using a POLARstar Galaxy plate reader (BMG Technologies). The fluorescence was measured at 25°C for 3 hours, and a no-enzyme control was included for each peptide to correct for background changes in fluorescence, such as photobleaching. *C. albicans* FTase was assayed in a similar fashion, except that 10 nM enzyme was used with 10  $\mu\text{M}$  FPP and 3  $\mu\text{M}$  peptide. For peptides that showed multiple turnover activity,  $k_{\text{cat}}/K_{\text{M}}$  values were approximated from initial velocities under the assumptions that  $[\text{peptide}] < K_{\text{M}}^{\text{peptide}}$  and that [FPP] or [GGPP] was saturating. The rate of product formation was calculated from the rate of fluorescence intensity increase using a conversion factor. The lower limit for MTO substrate detection using these assay conditions was  $\geq 150 \text{ M}^{-1} \text{ s}^{-1}$ , similar to other work with mammalian prenyltransferases [19].

#### *Single turnover assay*

Peptides that did not show activity under multiple turnover conditions were assayed under single turnover conditions. The same fluorescence assay was used, but measurements were made under  $[\text{E}] > [\text{S}]$  conditions: 1  $\mu\text{M}$  enzyme, 3  $\mu\text{M}$  peptide, and 0.8  $\mu\text{M}$  FPP/GGPP. Reactions were initiated by addition of enzyme and reaction progress was monitored continuously for 3 hours at 25°C. No-enzyme controls were run side-by-side to correct for background fluorescence changes. Single turnover rate constants were obtained by a single

exponential fit of the change in fluorescence intensity. The lower limit of detection for STO substrates was  $\geq 0.002 \text{ s}^{-1}$ .

*Statistical analysis of amino acid sequence in FTase and GGTase-I peptide substrates*

In collaboration with Dr. Terry Watt (Xavier University of Louisiana), *C. albicans* FTase and GGTase-I substrate preferences were analyzed using statistical analysis. A  $\chi^2$  test was used to determine whether MTO, STO and non-substrate pools had distributions of amino acids at the  $a_1$ ,  $a_2$ , and X position that were significantly different from distributions in the peptide library. A hypergeometric square model analysis [19] was used to determine if any amino acids were over-represented or under-represented compared to the overall library at the  $a_1$ ,  $a_2$  and X positions. At each position, the probability ( $p$ ) of a given amino acid occurring with the observed frequency at the position of interest in the substrate pool by random chance was calculated using equation (1) where  $N$  is the total library size,  $P$  is the number of peptides in the total library with a given amino acid at the position,  $S$  is the total number of peptide substrates, and  $R$  is the number of substrate peptides in the pool that have a particular amino acid at the position of interest.

$$p = \frac{\frac{P!(N-P)!}{(R!(P-R)!(S-R)!(N-P-S+R)!}}{N!}}{\frac{1}{(S!(N-S)!)}} \quad (\text{Eq. 1})$$

The basis for the equation used in this analysis comes from calculating the number of ways of drawing  $r$  items from a pool of  $n$  items without repetition, where order of drawing does not matter (i.e. combinations):

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (\text{Eq. 2})$$

To determine enhancement or depletion of a particular amino acid at a particular position, there are three relevant combinations: (1) the number of ways to draw the  $S$  peptides from the total



library of  $N$  peptides, (2) the number of ways to draw the  $R$  peptides from the total set of  $P$  peptides, and (3) the number of ways to draw the  $S - R$  peptides, which are peptides from the  $S$  peptide pool that do not contain the amino acid(s) of interest at the particular position, from the total pool of  $N - P$  peptides. The probability of drawing the  $R$  peptides from the total library of  $N$  peptides is expressed as:

$$p = \frac{\binom{P}{R} \binom{N-P}{S-R}}{\binom{N}{S}} \quad (\text{Eq. 3})$$

Equation (1) is an expansion of (3) with substitution of (2) for all three combinations. The hypergeometric distribution is analogous to the binomial distribution where items are drawn without replacement, such that a single peptide sequence can only be drawn and classified as MTO, STO, or non-substrate once.

## RESULTS

### *Peptide substrates of FTase and GGTase-I*

Overall, the reactivity of purified *C. albicans* FTase and GGTase-I with peptides of varying sequences was tested under both MTO and STO conditions (where  $[S] > [E]$  and  $[E] > [FPP]$ , respectively). Results for *C. albicans* FTase are presented in Table 2.4 and for *C. albicans* GGTase-I in Table 2.5. For each enzyme, peptides were classified as MTO substrates, STO substrates, or non-substrates. STO activity indicates that the peptides bind to FTase or GGTase-I and are prenylated, but the enzyme does not react with a second peptide, presumably due to slow product dissociation. To ensure that very slow MTO substrates are not being falsely classified as STO substrates, it was calculated that based on the cut-off value set for  $k_{STO}$ , the slowest peptide classified as STO, with  $k_{STO} = 0.002 \text{ s}^{-1}$ , would have a  $k_{cat}/K_M$  of  $\sim 650 \text{ M}^{-1} \text{ s}^{-1}$ , a rate that is well within the limit of detection of the MTO assay. In addition, for *C. albicans* FTase STO rate constants vary over 40-fold, with the fastest rate being  $0.082 \text{ s}^{-1}$ , giving a MTO rate constant of  $2.7 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ . This suggests that STO behavior is a sequence-dependent property of peptide substrates.

<i>C. albicans</i> FTase MTO substrate sequences and MTO rate constants ( $M^{-1} s^{-1}$ )									
CAMQ	$3.3 \times 10^5$	CLLF	$3.3 \times 10^4$	CVIL	$9.0 \times 10^3$	CKIS	$3.7 \times 10^3$	CVID	$1.9 \times 10^3$
CTVC	$1.9 \times 10^5$	CITQ	$3.0 \times 10^4$	CLIL	$8.1 \times 10^3$	CGIL	$3.6 \times 10^3$	CQTQ	$1.9 \times 10^3$
CSAC	$1.3 \times 10^5$	CTLM	$2.9 \times 10^4$	CRIS	$8.0 \times 10^3$	CALT	$3.5 \times 10^3$	CIGW	$1.8 \times 10^3$
CVIT	$9.6 \times 10^4$	CSFM	$2.7 \times 10^4$	CVTT	$7.6 \times 10^3$	CLHM	$3.2 \times 10^3$	CVHS	$1.7 \times 10^3$
CAVQ	$8.9 \times 10^4$	CVIV	$2.5 \times 10^4$	CMPS	$6.9 \times 10^3$	CSVS	$3.1 \times 10^3$	CIVK	$1.7 \times 10^3$
CVIM	$8.9 \times 10^4$	CLIT	$2.3 \times 10^4$	CIIL	$6.8 \times 10^3$	CVVN	$3.1 \times 10^3$	CTKF	$1.7 \times 10^3$
CSIM	$8.8 \times 10^4$	CVLA	$2.3 \times 10^4$	CVIN	$6.5 \times 10^3$	CSVY	$3.0 \times 10^3$	CSLG	$1.7 \times 10^3$
CCIF	$8.5 \times 10^4$	CIIM	$2.1 \times 10^4$	CNIL	$6.5 \times 10^3$	CTVF	$3.0 \times 10^3$	CCFV	$1.6 \times 10^3$
CVIM	$8.4 \times 10^4$	CSLM	$2.0 \times 10^4$	CWLA	$6.2 \times 10^3$	CQLS	$2.9 \times 10^3$	CWDY	$1.6 \times 10^3$
CVFM	$7.3 \times 10^4$	CNLM	$1.9 \times 10^4$	CTII	$6.2 \times 10^3$	CSLL	$2.9 \times 10^3$	CLTS	$1.6 \times 10^3$
CVIQ	$6.9 \times 10^4$	CTLQ	$1.8 \times 10^4$	CRVW	$5.9 \times 10^3$	CSVW	$2.7 \times 10^3$	CRLS	$1.5 \times 10^3$
CIII	$6.7 \times 10^4$	CTLA	$1.6 \times 10^4$	CTLL	$5.6 \times 10^3$	CTMS	$2.7 \times 10^3$	CACV	$1.4 \times 10^3$
CNIM	$6.3 \times 10^4$	CCIE	$1.6 \times 10^4$	CALD	$5.3 \times 10^3$	CCIL	$2.7 \times 10^3$	CPFW	$1.4 \times 10^3$
CLVM	$6.2 \times 10^4$	CSII	$1.6 \times 10^4$	CEVA	$5.1 \times 10^3$	CSIS	$2.6 \times 10^3$	CALL	$1.3 \times 10^3$
CVIF	$5.8 \times 10^4$	CVIY	$1.5 \times 10^4$	CISL	$5.0 \times 10^3$	CVII	$2.5 \times 10^3$	CSLC	$1.3 \times 10^3$
CVVT	$5.2 \times 10^4$	CIMM	$1.4 \times 10^4$	CILL	$4.8 \times 10^3$	CCTG	$2.5 \times 10^3$	CVSS	$1.3 \times 10^3$
CIVA	$5.0 \times 10^4$	CTIM	$1.4 \times 10^4$	CSVI	$4.7 \times 10^3$	CVLS	$2.4 \times 10^3$	CVVL	$1.3 \times 10^3$
CWLQ	$4.6 \times 10^4$	CAVE	$1.3 \times 10^4$	CFLS	$4.5 \times 10^3$	CLLN	$2.4 \times 10^3$	CSVD	$1.1 \times 10^3$
CLLM	$4.6 \times 10^4$	CSVQ	$1.3 \times 10^4$	CQTA	$4.3 \times 10^3$	CNTI	$2.3 \times 10^3$	CNVF	$9.0 \times 10^2$
CVIS	$4.4 \times 10^4$	CSVT	$1.2 \times 10^4$	CVQM	$4.2 \times 10^3$	CQVW	$2.2 \times 10^3$	CVLY	$8.6 \times 10^2$
CYVM	$4.3 \times 10^4$	CCLA	$1.1 \times 10^4$	CLKT	$4.1 \times 10^3$	CTMV	$2.2 \times 10^3$	CTIL	$7.6 \times 10^2$
CIVM	$3.9 \times 10^4$	CNLQ	$1.1 \times 10^4$	CGII	$4.1 \times 10^3$	CWQV	$2.2 \times 10^3$	CNVL	$5.7 \times 10^2$
CVLQ	$3.8 \times 10^4$	CGIA	$1.0 \times 10^4$	CSIL	$4.1 \times 10^3$	CHLT	$2.1 \times 10^3$	CSPT	$4.2 \times 10^2$
CGIM	$3.6 \times 10^4$	CVTQ	$9.7 \times 10^3$	CVSL	$4.0 \times 10^3$	CTVY	$2.1 \times 10^3$		
CNVT	$3.3 \times 10^4$	CKIF	$9.1 \times 10^3$	CNNQ	$3.9 \times 10^3$	CVLH	$1.9 \times 10^3$		
CVIA	$3.3 \times 10^4$	CLLT	$9.0 \times 10^3$	CAVL	$3.9 \times 10^3$	CGFH	$1.9 \times 10^3$		
<i>C. albicans</i> FTase STO substrate sequences and STO rate constants ( $s^{-1}$ )									
CRGC	0.0822	CGNV	0.0147	CKKM	0.0066	CKKP	0.0040	CWHK	0.0029
CCNG	0.0636	CQNI	0.0137	CSKM	0.0065	CGLF	0.0040	CGQI	0.0028
CFDR	0.0442	CSSG	0.0129	CFPS	0.0064	CSQL	0.0039	CGNT	0.0028
CMGF	0.0413	CPMG	0.0119	CLHL	0.0063	CVKQ	0.0039	CWKL	0.0027
CLYV	0.0380	CSSD	0.0119	CKFI	0.0061	CKNS	0.0039	CTCR	0.0027
CPSC	0.0367	CSHF	0.0118	CTKI	0.0054	CCNA	0.0038	CLWS	0.0026
CYQN	0.0367	CLKP	0.0114	CIHF	0.0053	CVPL	0.0038	CNLL	0.0026
CMKI	0.0279	CMSS	0.0112	CQKS	0.0053	CLMP	0.0037	CGCS	0.0026
CQWS	0.0274	CWNK	0.0105	CLSL	0.0051	CTTL	0.0036	CMKR	0.0025
CVYK	0.0257	CWRP	0.0096	CATG	0.0051	CGHS	0.0036	CAQP	0.0025
CARY	0.0245	CKYE	0.0090	CHDF	0.0050	CPED	0.0035	CQKI	0.0024
CYSE	0.0208	CAES	0.0086	CGKH	0.0050	CHVI	0.0034	CPLG	0.0024
CYYA	0.0203	CKSE	0.0083	CTHV	0.0047	CHDL	0.0034	CRWI	0.0024
CRGE	0.0201	CQPS	0.0081	CNTS	0.0047	CLPS	0.0034	CVKL	0.0023
CNLF	0.0173	CPRD	0.0077	CGYG	0.0046	CNHD	0.0033	CGTE	0.0023
CGEM	0.0165	CRLW	0.0073	CTDS	0.0044	CTDL	0.0032	CNSY	0.0023
CMMK	0.0159	CNRS	0.0072	CGGN	0.0043	CRYG	0.0032	CKLA	0.0022
CQED	0.0159	CIYY	0.0071	CAGQ	0.0042	CTKR	0.0031	CKLN	0.0022
CIGN	0.0154	CTQA	0.0068	CSSS	0.0042	CRKN	0.0030	CAWV	0.0022
CWVG	0.0150	CGLY	0.0067	CKIK	0.0041	CPIL	0.0030	CLAS	0.0021

<i>C. albicans</i> FTase Non-substrate peptides									
CKAA	CRKC	CENE	CLEH	CHWI	CWGL	CLNL	CCSN	CAVR	CCKT
CGCA	CDRC	CSKF	CCHH	CKWI	CIHL	CQNL	CYSN	CMVR	CRRT
CEDA	CGRC	CGQF	CCLH	CTAK	CAIL	CGQL	CQWN	CPAS	CSRT
CQDA	CGSC	CADG	CQRH	CEEK	CEKL	CCVL	CSIP	CKCS	CKTT
CQEA	CMTC	CSKG	CACI	CVKK	CFKL	CSVL	CCRP	CVIS	CTKV
CSEA	CLFD	CFLG	CGCI	CSDL	CCLL	CAYL	CSGQ	CAKS	CQLW
CKKA	CEGD	CFMG	CIHI	CFFL	CRLl	CWEM	CKQQ	CDMS	CGTW
CTKA	CYPD	CCAH	CNKI	CSFL	CVLL	CKWM	CACR	CFSS	CEVW
CCTA	CFLE	CVAH	CSPI	CDGL	CLML	CKFN	CGCR	CEYS	CDIY
CNWA	CLME	CGCH	CCWI	CGGL	CSML	CNFN	CVIR	CLET	CTKY
CTGC									

**Table 2.4. MTO, STO, and non-substrates of *C. albicans* FTase.** Peptides were classified as MTO substrates if their  $k_{\text{cat}}/K_M \geq 150 \text{ M}^{-1} \text{ s}^{-1}$ ; peptides were classified as STO substrates if their  $k_{\text{STO}} \geq 0.002 \text{ s}^{-1}$ . Within each pool, peptides are ordered by rate constant in descending order.

<i>C. albicans</i> GGTase-I MTO substrate sequences and MTO rate constants ( $M^{-1} s^{-1}$ )									
CPIL	$4.3 \times 10^4$	CNVF	$5.0 \times 10^3$	CCLL	$1.2 \times 10^3$	CYVM	$6.7 \times 10^2$	CSFM	$3.4 \times 10^2$
CTIM	$3.6 \times 10^4$	CWEM	$4.6 \times 10^3$	CLIT	$1.2 \times 10^3$	CISL	$6.7 \times 10^2$	CSQL	$3.3 \times 10^2$
CSIL	$3.3 \times 10^4$	CCWI	$4.5 \times 10^3$	CKIS	$1.2 \times 10^3$	CLHM	$6.5 \times 10^2$	CGSC	$3.2 \times 10^2$
CAIL	$3.3 \times 10^4$	CVVL	$4.3 \times 10^3$	CLEH	$1.2 \times 10^3$	CMVR	$6.4 \times 10^2$	CMPS	$3.2 \times 10^2$
CRLL	$3.2 \times 10^4$	CIVM	$4.3 \times 10^3$	CYYA	$1.1 \times 10^3$	CAMQ	$6.3 \times 10^2$	CEVW	$3.2 \times 10^2$
CTIL	$2.5 \times 10^4$	CLLF	$4.2 \times 10^3$	CWKL	$1.1 \times 10^3$	CWVG	$6.3 \times 10^2$	CSLM	$3.2 \times 10^2$
CVLL	$2.1 \times 10^4$	CRLW	$4.2 \times 10^3$	CTLM	$1.1 \times 10^3$	CVLA	$6.2 \times 10^2$	CWNK	$3.1 \times 10^2$
CKIF	$2.1 \times 10^4$	CTVF	$3.8 \times 10^3$	CLLM	$1.1 \times 10^3$	CCNG	$6.2 \times 10^2$	CFKL	$3.1 \times 10^2$
CCIF	$2.0 \times 10^4$	CQWN	$3.7 \times 10^3$	CGLF	$1.1 \times 10^3$	CQWS	$6.1 \times 10^2$	CITQ	$2.0 \times 10^2$
CIIM	$2.0 \times 10^4$	CAYL	$3.6 \times 10^3$	CSVS	$1.1 \times 10^3$	CNLM	$6.1 \times 10^2$	CVLQ	$2.9 \times 10^2$
CTLL	$1.8 \times 10^4$	CVFM	$3.5 \times 10^3$	CRLS	$1.1 \times 10^3$	CWHK	$6.1 \times 10^2$	CDRC	$2.8 \times 10^2$
CILL	$1.7 \times 10^4$	CTMV	$3.4 \times 10^3$	CGCH	$1.1 \times 10^3$	CWQV	$5.9 \times 10^2$	CCLA	$2.8 \times 10^2$
CKWI	$1.6 \times 10^4$	CSLL	$3.4 \times 10^3$	CRIS	$1.1 \times 10^3$	CCAH	$5.8 \times 10^2$	CQTA	$2.8 \times 10^2$
CSML	$1.6 \times 10^4$	CLME	$3.4 \times 10^3$	CIYY	$1.0 \times 10^3$	CSVI	$5.2 \times 10^2$	CEDA	$2.8 \times 10^2$
CNLL	$1.5 \times 10^4$	CQLW	$3.3 \times 10^3$	CTTL	$1.0 \times 10^3$	CVSL	$5.2 \times 10^2$	CIGN	$2.7 \times 10^2$
CNIL	$1.5 \times 10^4$	CHWI	$3.2 \times 10^3$	CLPS	$1.0 \times 10^3$	CGLY	$5.2 \times 10^2$	CNFN	$2.7 \times 10^2$
CLIL	$1.4 \times 10^4$	CVIN	$3.2 \times 10^3$	CWDY	$9.9 \times 10^2$	CSSS	$5.1 \times 10^2$	CSLC	$2.6 \times 10^2$
CNLF	$1.4 \times 10^4$	CLET	$3.1 \times 10^3$	CWRP	$9.7 \times 10^2$	CNKI	$5.1 \times 10^2$	CCHH	$2.6 \times 10^2$
CCIL	$1.3 \times 10^4$	CSVL	$3.0 \times 10^3$	CFDR	$9.6 \times 10^2$	CCNA	$5.0 \times 10^2$	CLTS	$2.6 \times 10^2$
CVIM	$1.2 \times 10^4$	CSII	$3.0 \times 10^3$	CIHL	$9.4 \times 10^2$	CVVT	$4.9 \times 10^2$	CYSN	$2.6 \times 10^2$
CRWI	$1.1 \times 10^4$	CLAS	$2.2 \times 10^3$	CWLQ	$9.1 \times 10^2$	CATG	$4.9 \times 10^2$	CLNL	$2.5 \times 10^2$
CVIV	$9.8 \times 10^3$	CCVL	$2.1 \times 10^3$	CIHI	$8.5 \times 10^2$	CKTT	$4.9 \times 10^2$	CNVT	$2.5 \times 10^2$
CVIT	$8.9 \times 10^3$	CTKF	$2.1 \times 10^3$	CLFD	$8.2 \times 10^2$	CGIM	$4.7 \times 10^2$	CCSN	$2.5 \times 10^2$
CAVL	$8.8 \times 10^3$	CNVL	$2.0 \times 10^3$	CQVW	$8.2 \times 10^2$	CFSS	$4.6 \times 10^2$	CFMG	$2.4 \times 10^2$
CKFN	$8.3 \times 10^3$	CLML	$2.0 \times 10^3$	CLHL	$8.0 \times 10^2$	CGIA	$4.5 \times 10^2$	CSKF	$2.4 \times 10^2$
CKFI	$8.2 \times 10^3$	CSIM	$1.9 \times 10^3$	CHDF	$7.9 \times 10^2$	CLWS	$4.5 \times 10^2$	CALT	$2.4 \times 10^2$
CIIL	$8.2 \times 10^3$	CAWV	$1.9 \times 10^3$	CLLT	$7.7 \times 10^2$	CLKT	$4.3 \times 10^2$	CTDL	$2.2 \times 10^2$
CIII	$8.0 \times 10^3$	CFLE	$1.8 \times 10^3$	CFLG	$7.7 \times 10^2$	CCLH	$4.3 \times 10^2$	CSLG	$2.1 \times 10^2$
CVIA	$7.8 \times 10^3$	CMTC	$1.8 \times 10^3$	CCIE	$7.6 \times 10^2$	CGII	$4.3 \times 10^2$	CYQN	$2.0 \times 10^2$
CALL	$7.5 \times 10^3$	CRRT	$1.8 \times 10^3$	CHDL	$7.5 \times 10^2$	CCFV	$4.3 \times 10^2$	CTHV	$2.0 \times 10^2$
CVIY	$7.4 \times 10^3$	CKLA	$1.7 \times 10^3$	CIHF	$7.3 \times 10^2$	CLMP	$4.1 \times 10^2$	CPFW	$1.9 \times 10^2$
CKWM	$7.2 \times 10^3$	CKIK	$1.6 \times 10^3$	CCTG	$7.3 \times 10^2$	CHVI	$4.1 \times 10^2$	CNNQ	$1.9 \times 10^2$
CVIQ	$7.2 \times 10^3$	CLKP	$1.5 \times 10^3$	CKLN	$7.2 \times 10^2$	CMSS	$3.9 \times 10^2$	CDMS	$1.9 \times 10^2$
CGIL	$7.2 \times 10^3$	CVLY	$1.5 \times 10^3$	CIVA	$7.2 \times 10^2$	CSHF	$3.9 \times 10^2$	CQNI	$1.7 \times 10^2$
CVIL	$7.1 \times 10^3$	CLVM	$1.4 \times 10^3$	CLLN	$7.1 \times 10^2$	CIGW	$3.9 \times 10^2$	CDIY	$1.7 \times 10^2$
CVII	$6.6 \times 10^3$	CRVW	$1.4 \times 10^3$	CNIM	$7.1 \times 10^2$	CVTT	$3.8 \times 10^2$	CRYG	$1.5 \times 10^2$
CTII	$6.0 \times 10^3$	CHLT	$1.4 \times 10^3$	CSSG	$7.0 \times 10^2$	CTVC	$3.8 \times 10^2$	CFPS	$1.5 \times 10^2$
CFFL	$5.4 \times 10^3$	CWLA	$1.3 \times 10^3$	CMGF	$7.0 \times 10^2$	CNWA	$3.7 \times 10^2$	CPLG	$1.5 \times 10^2$
CIMM	$5.4 \times 10^3$	CWGL	$1.3 \times 10^3$	CVAH	$6.9 \times 10^2$	CVLH	$3.7 \times 10^2$	CEVA	$1.4 \times 10^2$
CVIF	$5.1 \times 10^3$	CVKL	$1.3 \times 10^3$	CTLA	$6.9 \times 10^2$	CMMK	$3.7 \times 10^2$	CGFH	$1.4 \times 10^2$
<i>C. albicans</i> GGTase-I STO substrate sequences and STO rate constants ( $s^{-1}$ )									
CSVW	0.0406	CVLS	0.0111	CNLQ	0.0062	CACV	0.0040	CTKI	0.0029
CFLS	0.0256	CGCS	0.0097	CVID	0.0060	CSIS	0.0039	CSIP	0.0028
CVVN	0.0220	CVHS	0.0091	CNTI	0.0058	CCTA	0.0038	CGQI	0.0024
CTVY	0.0144	CGTW	0.0088	CTMS	0.0056	CEYS	0.0038	CTCR	0.0024
CSVY	0.0139	CNHD	0.0086	CKQQ	0.0055	CYPD	0.0035		
CQLS	0.0116	CSPI	0.0080	CSVD	0.0053	CGQF	0.0034		
CNSY	0.0113	CVTQ	0.0076	CCRP	0.0050	CGGN	0.0031		
CTLQ	0.0112	CGHS	0.0064	CSKG	0.0041	CVKK	0.0030		

<i>C. albicans</i> GGTase-I Non-substrate peptides									
CKAA	CRGC	CPRD	CADG	CTAK	CVPL	CAQP	CMKR	CQKS	CSRT
CGCA	CTGC	CSSD	CPMG	CEEK	CGQL	CAGQ	CTKR	CKNS	CSVT
CQDA	CRKC	CRGE	CGYG	CIVK	CLSL	CSGQ	CAVR	CQPS	CTKV
CQEA	CGRC	CENE	CGKH	CVYK	CGEM	CVKQ	CPAS	CNRS	CGNV
CSEA	CPSC	CKSE	CQRH	CSDL	CKKM	CQTQ	CKCS	CVSS	CLYV
CKKA	CPED	CYSE	CACI	CDGL	CSKM	CSVQ	CTDS	CNTS	CTKY
CTKA	CQED	CGTE	CGCI	CGGL	CVQM	CACR	CAES	CCKT	CARY
CTQA	CEGD	CAVE	CMKI	CEKL	CRKN	CGCR	CVIS	CGNT	
CSAC	CALD	CKYE	CQKI	CQNL	CKKP	CVIR	CAKS	CSPT	

**Table 2.5. MTO, STO, and non-substrates of *C. albicans* GGTase-I.** Peptides were classified as MTO substrates if their  $k_{\text{cat}}/K_M \geq 150 \text{ M}^{-1} \text{ s}^{-1}$ ; peptides were classified as STO substrates if their  $k_{\text{STO}} \geq 0.002 \text{ s}^{-1}$ . Within each pool, peptides are ordered by rate constant in descending order.

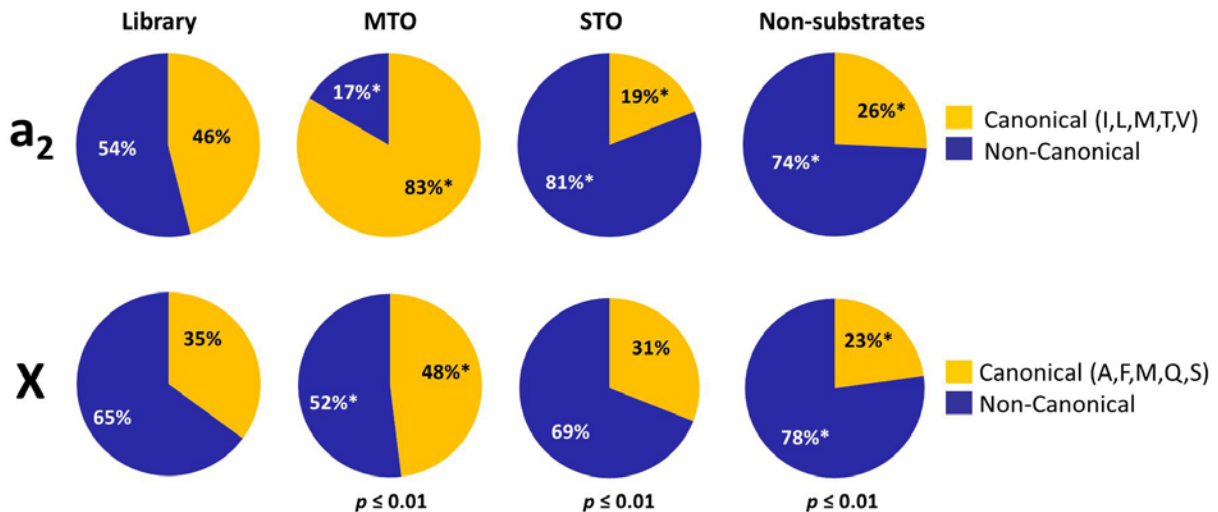
As shown in Table 2.6, *C. albicans* FTase catalyzes farnesylation of 127 or 39% of the peptides under MTO conditions. Additionally, under STO conditions, FTase catalyzes farnesylation of an additional 100 peptides, or 30% of the library. Overall, *C. albicans* FTase catalyzes prenylation of 227 peptides, or 69% of the library. *C. albicans* GGTase-I catalyzes geranylgeranylation of 207 or 63% of peptides under MTO conditions, and additional 36, or 11% of the library, under STO conditions. Overall, *C. albicans* GGTase-I catalyzes prenylation of 243 peptides, or 73% of the library.

<b>Pool</b>		<b><i>C. albicans</i> FTase Number of peptides</b>	<b><i>C. albicans</i> GGTase-I Number of peptides</b>
Library		328	331
Multiple turnover (MTO)		127 (39%)	207 (63%)
$k_{\text{cat}}/K_M$ ( $\text{M}^{-1} \text{ s}^{-1}$ )	$\geq 10^5$	3	0
	$\geq 10^4$	46	21
	$\geq 10^3$	73	82
	$\geq 10^2$	5	104
Single turnover (STO)		100 (30%)	36 (11%)
$k_{\text{STO}}$ ( $\text{s}^{-1}$ )	$\geq 0.010$	29	9
	$\geq 0.002$	71	27
Total substrates		227 (69%)	243 (74%)

**Table 2.6. *C. albicans* FTase and GGTase-I substrate pools.** Dns-TKCxxx peptides were screened for activity with *C. albicans* FTase and GGTase-I. Number of MTO and STO substrates for each enzyme are given and expressed as percentages of total library in parentheses. MTO and STO substrates are binned by reactivity.

*Sequence analysis of MTO and STO C. albicans FTase and GGTase-I substrates*

With the help of Dr. Terry Watt, overall patterns in the peptide sequences were analyzed for the MTO, STO, and non-substrate pools of *C. albicans* FTase and GGTase-I substrates using a  $\chi^2$  test. The peptides were separated into canonical or non-canonical Ca<sub>1</sub>a<sub>2</sub>X sequences and the sequence preferences were analyzed at the a<sub>2</sub> and X positions. As no information is currently available on yeast prenyltransferase substrate recognition, canonical residues defined for the mammalian enzymes were used in this analysis. Canonical amino acids were defined based on amino acids most commonly found in mammalian prenyltransferase substrates at each position of the Ca<sub>1</sub>a<sub>2</sub>X motif, as well as important structural features of FTase and GGTase-I highlighted by ternary complexes of enzymes with prenyl donor analogs and a variety of peptide substrates [26]. Figure 2.9 shows the percentages of canonical and non-canonical amino acids contained in the library, the MTO pool, and the STO pool at the a<sub>2</sub> and X positions for FTase.



**Figure 2.9.** The percentages of canonical and non-canonical residues for peptides in the overall library and MTO, STO and non-substrate pools for *C. albicans* FTase. At the a<sub>2</sub> position, canonical residues are defined as I, L, M, T and V whereas at the X position, canonical residues are defined as A, F, M, Q and S. The sequences of the peptides on the MTO and STO pools are significantly different from the library ( $p \leq 0.01$ ). The sequences of the MTO substrates generally follow typical CaaX predictions, but the sequences of the STO peptides are not well described by the canonical CaaX paradigm.

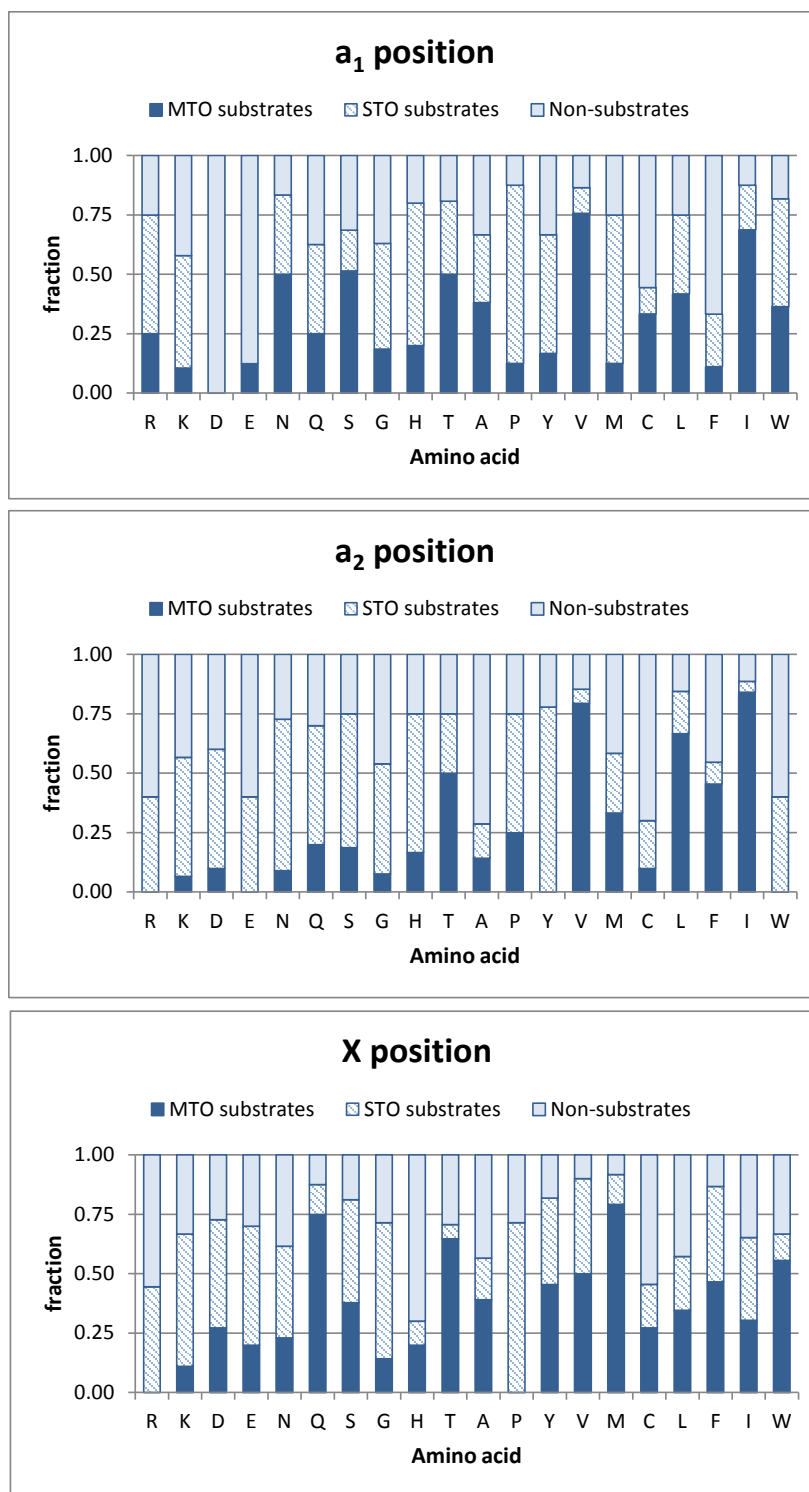
The MTO and STO pools of *C. albicans* FTase substrates reveal unique sequence patterns. For the MTO substrates, there is a significant increase in the percentage of canonical sequences at both the  $a_2$  and X positions as compared to the overall library. This is especially true of residues at the  $a_2$  position, where the percentage of peptides with a canonical residue (I, L, M, T or V) increases from 46% in the library to 83% in MTO substrates. This shows that peptides that are active with *C. albicans* FTase under MTO conditions are very likely to have a canonical residue at the  $a_2$  position. While canonical residues are also significantly enriched at the X position, this enrichment is weaker than the rat FTase, 48% in MTO pool *versus* 35% in the library for *C. albicans* FTase compared to 61% in MTO pool *versus* 36% in the library, suggesting that this enzyme has a slightly modified substrate specificity at the X position.

Figure 2.10 shows the fraction of each amino acid in each type of FTase substrate at the  $a_1$ ,  $a_2$ , and X positions of the peptides. At the  $a_1$  position, *C. albicans* FTase shows a strong preference for amino acids V and I in the MTO substrates. While the canonical model for mammalian prenyltransferase substrate specificity consists primarily of amino acid preferences at the  $a_2$  and X positions within the  $Ca_1a_2X$  sequence [13,26], recent data from our lab suggests that mammalian FTase also has sequence preferences at the  $a_1$  position [19], and it appears that *C. albicans* FTase is also selective at the  $a_1$  position. At the  $a_2$  position, the canonical residues I, V and L are significantly enriched and at the X position amino acids M, Q, and T are favored. This result is consistent with the observation that recognition at the  $a_2$  position is more similar between the yeast and mammalian enzymes than recognition at the X position.

The sequences of the STO substrates for FTase are more varied. Comparing the STO substrate pool to the library peptides indicates that the non-canonical sequences are significantly enhanced at the  $a_2$  (81% *versus* 54%) and X (61% *versus* 52%) positions, although the statistical

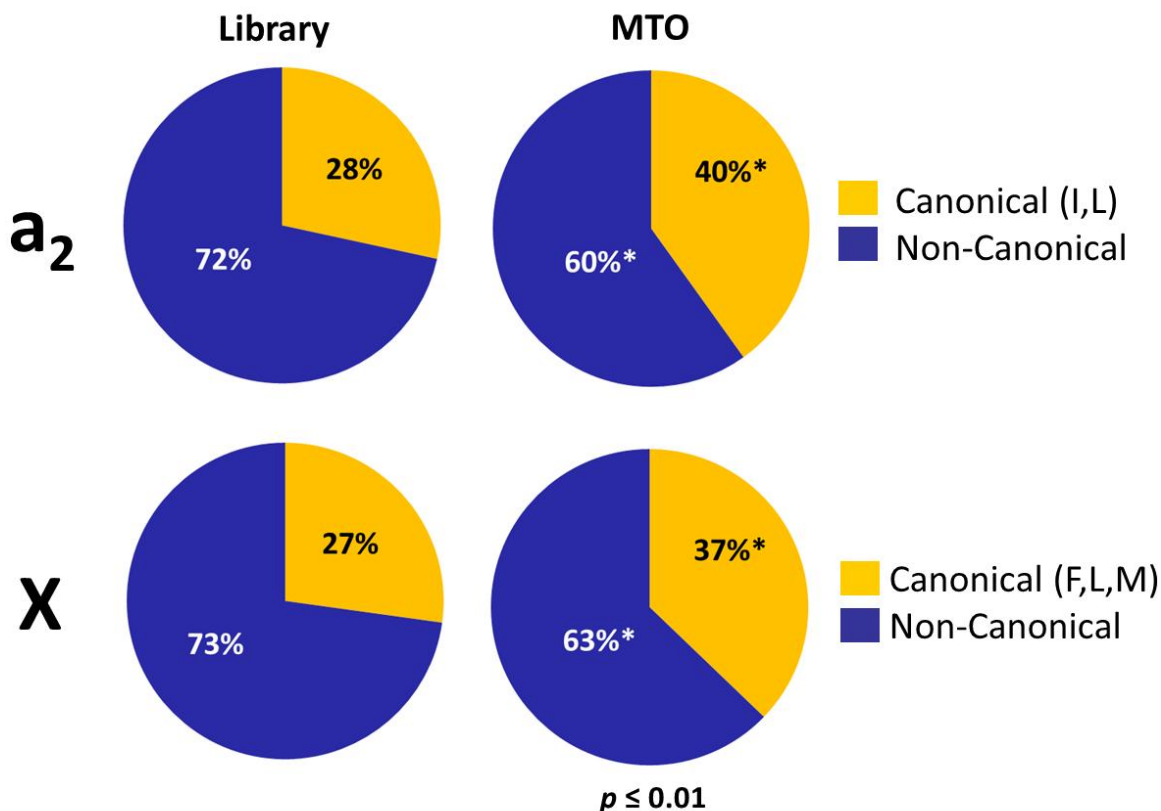


significance of this enrichment at the X position is not as strong ( $p = 0.026$ ). This trend is illustrated in Figure 2.10, where at the  $a_2$  position non-canonical residues Y and N, and to a lesser extent H and S, are enriched in STO substrates, and at the X position STO substrates have a higher fraction of P and G amino acids.

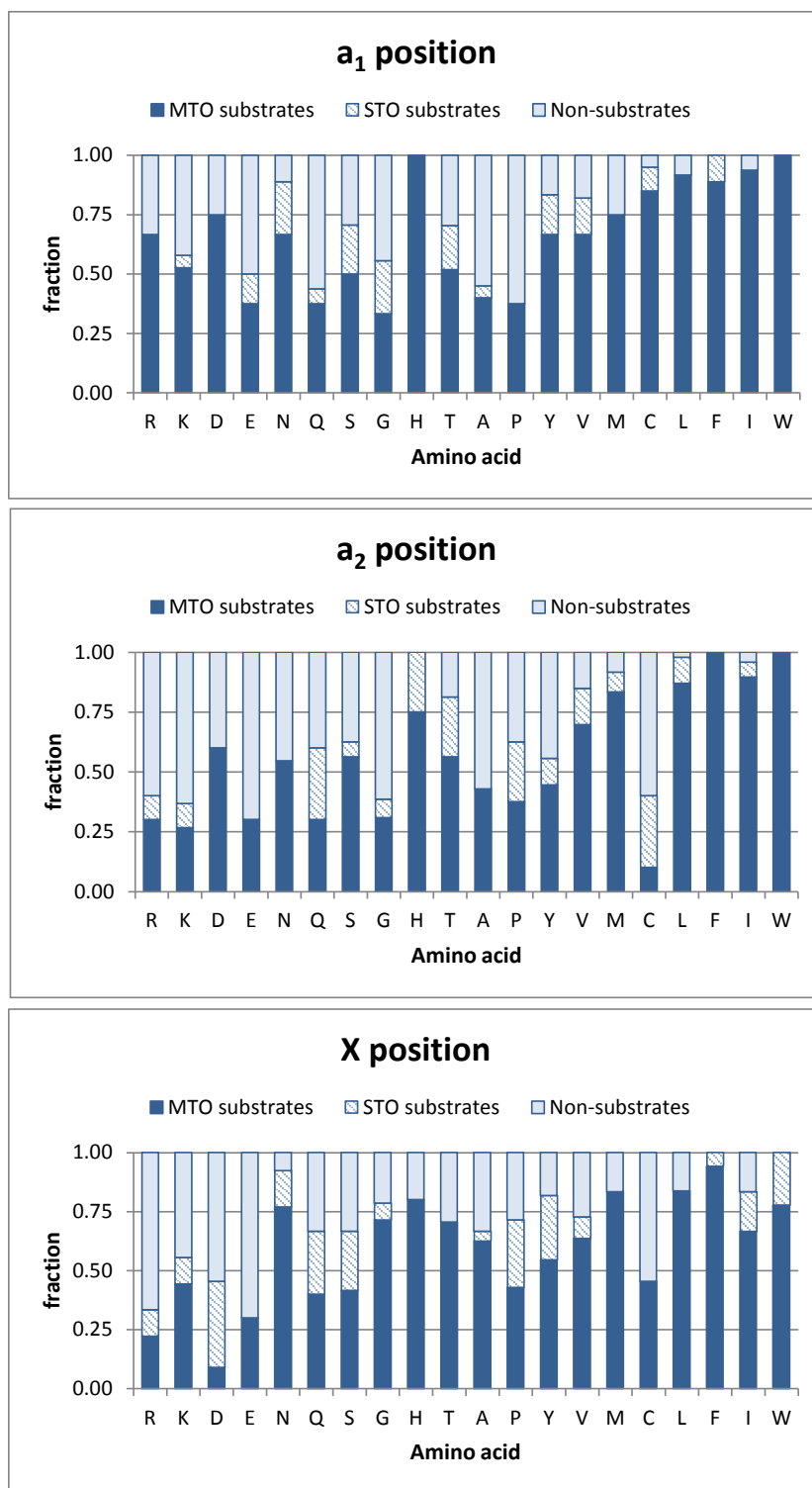


**Figure 2.10. Distribution of amino acids in peptides at the a<sub>1</sub>, a<sub>2</sub> and X positions that are substrates for *C. albicans* FTase under MTO and STO conditions or are not reactive. Each amino acid was tallied at the a<sub>1</sub> (top), a<sub>2</sub> (middle) and X (bottom) position for each pool of substrates and graphed by % amino acid vs. amino acid identity. MTO peptides are dark, STO peptides are hatched, and non-substrates are light. Amino acids are organized on the X-axis by increasing hydrophobicity based on  $\Delta G_{\text{transfer}}$  values (kcal/mol) [27].**

Similar to FTase, analysis of *C. albicans* GGTase-I shows that canonical mammalian amino acids are enriched for the MTO substrates compared to the peptide library as shown in Figure 2.11. However, this enrichment is not as pronounced as for *C. albicans* FTase, with only 40% versus 28% and 37% versus 27% enrichment for canonical amino acids at the a<sub>2</sub> and X positions, respectively. This enrichment for canonical amino acids is also smaller than that observed with the rat GGTase-I, where canonical amino acids at the a<sub>2</sub> position represent 62% of the MTO substrate pool compared to 30% of the library, and there is an enrichment of 65% compared to 28% at the X position. These data suggest that *C. albicans* GGTase-I is able to catalyze prenylation of a wider and more diverse set of substrates than is described by the canonical Ca<sub>1</sub>a<sub>2</sub>X sequences.



**Figure 2.11.** The percentages of canonical and non-canonical residues for peptides in the overall library and in the MTO substrate pool for *C. albicans* GGTase-I. At the a<sub>2</sub> position, canonical residues are defined as I and L, and at the X position, canonical residues are defined as F, L and M.



**Figure 2.12. Distribution of amino acids in peptides at the a<sub>1</sub>, a<sub>2</sub> and X positions that are substrates for *C. albicans* GGTase-I under MTO and STO conditions or are not reactive.** Each amino acid was tallied at the a<sub>1</sub> (top), a<sub>2</sub> (middle) and X (bottom) position for each pool of substrates and graphed by % amino acid vs. amino acid identity. MTO peptides are dark, STO peptides are hatched, and non-substrates are light. Amino acids are organized on the X-axis by increasing hydrophobicity based on  $\Delta G_{\text{transfer}}$  values (kcal/mol) [27].

*C. albicans* GGTase-I was able to catalyze prenylation under MTO conditions of 63% of all of the substrates in this library, a significantly higher percentage than mammalian prenyltransferases (35% for FTase and 28% for GGTase-I) [19] or *C. albicans* FTase (39%). Figure 2.12 shows the frequency of each amino acid in each pool of *C. albicans* GGTase-I substrates at the a<sub>1</sub>, a<sub>2</sub> and X positions. These graphs illustrate that *C. albicans* GGTase-I has a larger number of MTO substrates than *C. albicans* FTase, as evident from the higher fractions of MTO substrates for practically all amino acids, as well as much more variability in residues in the MTO substrate pool. At the a<sub>1</sub> position, in addition to amino acids V and I that are enriched in the MTO pool for FTase, residues H, W, L, F, C, M and D are also enriched for *C. albicans* GGTase-I. While canonical residues are enriched at the a<sub>2</sub> (I, L) and X (F, L, M) positions, W, F, M, V and T are also favored at the a<sub>2</sub> position and H, W, and N are enriched at the X position. This represents a much broader repertoire of peptides that are prenylated by *C. albicans* GGTase-I under MTO conditions, reinforcing that this enzyme accepts a wide range of CaaX sequences. Since there was only a total of 36 STO substrates with *C. albicans* GGTase-I, only MTO substrates were considered in this analysis.

#### *Statistical analysis of individual amino acid distribution in C. albicans FTase and GGTase-I substrates*

To further analyze substrate preferences of *C. albicans* FTase and GGTase-I, the sequences of the MTO, STO, and non-substrate peptides were analyzed using hypergeometric probabilities (with the help of Dr. Terry Watt). This analysis determines whether individual amino acids are enriched (preferred) or depleted (selected against) at the a<sub>1</sub>, a<sub>2</sub>, and X positions in a statistically significant manner. In this analysis, enrichments and depletions occurring with *p*

$\leq 0.02$  are considered statistically significant, and those occurring with  $0.02 < p \leq 0.05$  are suggestive.

Position	$a_1$		$a_2$		X	
	over-represented	under-represented	over-represented	under-represented	over-represented	under-represented
<b>MTO</b>	<b><i>I V T</i></b>	<b><i>G K M</i></b>	<b><i>I L V</i></b>	<b><i>D E G K R W</i></b> <b><i>Y N</i></b>	<b><i>M Q T</i></b>	<b><i>G H R K P</i></b>
<b>STO</b>	<b><i>M Y P</i></b>	<b><i>V S</i></b>	<b><i>N S K Y G H</i></b>	<b><i>I V L</i></b>	<b><i>G P</i></b>	<b><i>L T M</i></b>
<b>NON</b>	<b><i>G K D E C F</i></b>	<b><i>V</i></b>	<b><i>K C A D R W</i></b>	<b><i>I L V</i></b>	<b><i>H L R</i></b>	<b><i>M Q F</i></b>
<b>MTO</b>	<i>L</i>	<i>C</i>	<b><i>I L V</i></b>	<b><i>C K D</i></b>	<b><i>F M Q</i></b>	None
<b>STO</b>	<b><i>C A</i></b>	<b><i>L</i></b>	<b><i>S A</i></b>	<b><i>I K</i></b>	None	<b><i>M P Q</i></b>
<b>NON</b>	<i>K</i>	<i>I</i>	<b><i>D K R E</i></b>	<b><i>V I L T</i></b>	<b><i>P R</i></b>	<i>F Q</i>

**Table 2.7. Amino acids that are overrepresented or underrepresented in *C. albicans* FTase MTO substrate, STO substrate, and non-substrate pools as compared to the overall library.** For amino acids in bold,  $p \leq 0.02$  and for amino acids in italics,  $0.02 < p \leq 0.05$ . Substrate preferences for mammalian FTase are given for comparison (last three rows) [19].

The final results of analysis of *C. albicans* FTase substrates presented in Table 2.7 shows a preference for MTO reactivity with peptides containing some of the canonical CaaX sequences at the  $a_2$  (I, L, V) and X (M, Q, T) positions, but it also shows that there is a significant preference for V and I ( $p \leq 0.02$ ) and to a lesser extent T ( $p \leq 0.05$ ) at the  $a_1$  position for MTO substrates. The preference for V at the  $a_1$  position is so significant that this amino acid is depleted in both the STO and non-substrate pools ( $p \leq 0.02$ ). Such strong anti-correlation trends validate the preferences and confirm the usefulness of this approach in understanding substrate selectivity and recognition parameters. Similarly, residues that are selected against at the  $a_1$  position in the MTO substrate pool, G, K ( $p \leq 0.02$ ) and M ( $p \leq 0.05$ ) are overrepresented in non-substrate (G and K) and STO substrate (M) pools ( $p \leq 0.02$ ). In addition to K and G, charged (D, E), small (C, P) and aromatic (F, Y) amino acids are commonly found in STO and non-substrate pools. The different amino acid preferences for MTO and STO substrate pools suggest that recognition of different residues may affect step(s) in the prenylation reaction cycle that impact product release, as slow product release is hypothesized to contribute to STO substrate reactivity.

At the  $a_2$  position, canonical residues I, L and V are preferred ( $p \leq 0.02$ ) while charged (D, E, K, R), aromatic (W, Y) and small (G, N) residues are depleted in MTO substrates. Following the anti-correlation trend, I, L and V are under-represented in STO and non-substrate pools, while charged (K, D, R), aromatic (H, W, Y), and small (N, S, G, C, A) are all enriched in STO and non-substrate pools at the  $a_2$  position. At the X position M, Q and T side chains are significantly enriched ( $p \leq 0.02$ ) in the MTO substrate pool while charged (K, R, H) and small (G, P) residue are depleted. This result is confirmed by the enrichment of charged (H, R) and small (G, P) amino acids in the STO and non-substrate pools and depletion of M, Q and T in the STO and non-substrate pools.

Position	$a_1$		$a_2$		X	
	over-represented	under-represented	over-represented	under-represented	over-represented	under-represented
<i>C. alb</i> GGTase-I MTO Mamm	<b>ILWC</b>	<b>GAQ</b>	<b>ILFW</b>	<b>CKGEQR</b>	<b>FLM</b>	<b>DSRE</b>
GGTase-I MTO Mamm	<i>V</i>	<i>E</i>	<b>ILV</b>	<b>CHKSTD</b> <i>ENQR</i>	<b>FLMVI</b>	<b>AGKRS</b> <i>DHP</i>
GGTase-I STO	<i>Y</i>	<b>F</b>	<i>H</i>	<b>I</b>	<b>PSH</b>	<b>LVW</b>

**Table 2.8 Amino acids that are overrepresented or underrepresented in *C. albicans* GGTase-I MTO substrate pool as compared to the overall library.** For amino acids in bold,  $p \leq 0.02$  and for amino acids in italics,  $0.02 < p \leq 0.05$ . For comparison, mammalian GGTase-I substrate preferences are included.<sup>2</sup>

The analysis of *C. albicans* GGTase-I substrates, presented in Table 2.8, shows a preference for MTO reactivity with peptides containing canonical rat CaaX sequences at the  $a_2$  (I, L) and X (F, L, M) positions. Analysis of residues at the  $a_1$  position shows a preference for hydrophobic side chains I and L ( $p \leq 0.02$ ). However, a preference for larger aromatic amino acid W is also observed at this position, contrasting with the preferences for smaller amino acids (L, V) for the rat enzymes. This result is quite interesting as it not only shows that *C. albicans*

<sup>2</sup> Data for mammalian GGTase-I enzyme is from Corissa Lamphear's dissertation.

GGTase-I has a strong preference for particular residues at the a<sub>1</sub> position, but it also contradicts the ‘small aliphatic residue’ doctrine that is commonly used to describe amino acids at the a<sub>1</sub> and a<sub>2</sub> positions. A strong preference for W suggests that *C. albicans* GGTase-I not only has enough space to accommodate such a large residue, but that it forms favorable interactions with the a<sub>1</sub> residues that contribute to overall substrate reactivity. While the observation that W is over-represented at the a<sub>1</sub> position is interesting, structural data with mammalian GGTase-I indicate that the a<sub>1</sub> residue is exposed to solvent and the a<sub>1</sub> binding pocket could accommodate larger residues. Given that C is enriched at the a<sub>1</sub> position, it is somewhat counterintuitive that small (G, A) and polar (Q) amino acids at the a<sub>1</sub> position are depleted in *C. albicans* MTO substrates; this enrichment/depletion may be context dependent.

The most significant difference in *C. albicans* GGTase-I substrate recognition is observed at the a<sub>2</sub> position. While residues at this position show enrichment in canonical I and L residues, they also show statistically significant enrichment in large hydrophobic F and W residues ( $p \leq 0.02$ ). In contrast to the a<sub>1</sub> binding pocket, the a<sub>2</sub> side chain bound to mammalian FTase and GGTase-I is significantly more buried in a well defined pocket, and in many cases unable to accommodate large hydrophobic groups without a significant penalty in reactivity [28]. While *C. albicans* GGTase-I MTO substrates differ in enriched residues at the a<sub>2</sub> position, the depleted residues are similar to those observed for other prenyltransferases, including charged (K, E, R), small (C, G) and polar (Q) amino acids [19]. At the X position, the *C. albicans* GGTase-I MTO substrates are enriched in the canonical residues F, L and M, while charged (D, R, E) and small (S) residues are depleted. However, closer examination of the X position revealed that canonical X residues are enriched mostly when they are coupled with a canonical a<sub>2</sub> residue (I or L), and not very much when a<sub>2</sub> is different. Thus, it appears that *C. albicans* GGTase-I substrate



recognition is affected by several positive and negative factors at all positions, and context-based recognition plays a key role in *C. albicans* GGTase-I substrate selection.

*Predictions for Cxxx sequences found in C. albicans genome*

Based on the substrate recognition profile derived from the peptide library statistical analysis, a simple prediction algorithm was designed to predict whether proteins containing Cxxx> sequences found in the *C. albicans* genome are MTO substrates for *C. albicans* FTase, GGTase-I, both enzymes, or neither. In this algorithm, if that amino acid found at each position in the Cxxx sequence was enriched in the MTO substrates, a “+1” score was assigned, and a “-1” score was assigned for every amino acid that was depleted. If the sum was > 0, the Cxxx sequence was considered a substrate, and the results of this analysis are given in Table 2.9.

FTase		GGTase-I			Dual		Non-substrates		
CAQQ	CNVL	CCFS	<b>CKII</b>	<u>CTIL</u>	CFVL	<b><u>CTIM</u></b>	CADL	CGSK	CMKK
<b>CASQ</b>	CPVT	CDFV	CKIK	<u>CTIM</u>	CIIF	<u>CTIV</u>	CAVG	CHHN	CNLR
CFVL	<b>CTII</b>	CESL	CKLA	<u>CTIV</u>	CIIH	<b><u>CVIL</u></b>	CCCC	CHNY	CNQI
CIIF	<b>CTIL</b>	CETL	CKLI	CVFG	<b><u>CIIM</u></b>	<b><u>CVIM</u></b>	CCCS	CHQI	CNSK
<b>CIIH</b>	<b>CTIM</b>	<u>CFVL</u>	CKNF	<u>CVIL</u>	CIIN	CVIV	CCKC	CHSN	CPYW
<b>CIIM</b>	<b>CTIV</b>	CFWQ	<u>CLIL</u>	<u>CVIM</u>	CIIT	<b><u>CVVL</u></b>	CCKK	CHYA	CQFK
CIIN	CTLE	CHLG	CLLW	CVIV	CIMM	<b><u>CVVM</u></b>	CDMV	CIEY	CQKG
<b>CIIT</b>	<b>CVID</b>	CHWV	CLYL	<u>CVVL</u>	<b><u>CIIM</u></b>	<b><u>CLIL</u></b>	CDQL	CIGK	CRGK
<b>CIMM</b>	<b>CVIL</b>	CIDY	CMFQ	<u>CVVM</u>	<b><u>CLIL</u></b>	CLLW	CDYD	CIKS	CSEK
<b>CIQC</b>	<b>CVIM</b>	CIIF	<u>CNLL</u>	CWFC	CLLW	CNLL	CEGI	CIKW	CSKE
<b>CIVM</b>	<b>CVIV</b>	CIIH	CNSL	CWYC	CNLL	CNLL	CENN	CIRN	CSQG
<b>CLIL</b>	<b>CVVL</b>	<u>CIIM</u>	CNVF	CYFL	CNVF	CNVF	CENV	CIRW	CSQS
CLLW	<b>CVVM</b>	CIIN	CNVL	CYLG	CNVL	CNVL	CFDY	CKCF	CSRE
CNLL	<b>CVVS</b>	CIIT	CSDF		<b><u>CTII</u></b>	<b><u>CTIL</u></b>	CFPE	CKEK	CSRT
CNVF		CIMM	CSFM				CFSD	CKMG	CTSG
		CINI	CSFV				CFSN	<b>CKQQ</b>	CTYD
		CISG	CSFW				CGFH	CLAS	CVKL
		CISK	CSML				CGGD	CLEY	CVRV
		<u>CIVM</u>	CSYL				CGHW	CLKD	CVYI
		CKFN	<u>CTII</u>				CGKK	CLRN	CYSD

**Table 2.9. Cxxx sequences found in *C. albicans* genome predicted to be *C. albicans* FTase, GGTase-I, dual enzyme, or non-substrates based on peptide library studies.** Sequences that were predicted to be FTase substrates by PrePS are bolded, GGTase-I substrates are underlined, and dual substrates are bolded and underlined.

Out of the 120 unique Cxxx sequences found in the *C. albicans* genome, the PrePS algorithm predicted: 21 FTase substrates, 12 GGTase-I substrates, 10 dual FTase and GGTase-I substrates, and 97 non-substrates. However, the reactivity of these enzymes with the peptide library suggested a much larger potential pool of substrates. This new simple algorithm predicts: 29 *C. albicans* FTase substrates, 53 GGTase-I substrates, 22 dual FTase and GGTase-I substrates, and 60 non-substrates. With the exception of two sequences (CKII was predicted to be a *C. albicans* GGTase-I substrate and CKQQ was not predicted to be a substrate), all PrePS sequences that were predicted to be FTase and GGTase-I substrate were also correctly predicted by the new algorithm. However, 10 and 41 additional *C. albicans* FTase and GGTase-I substrates, respectively, were added to the list of substrates using the new algorithm.

To test the validity of the new algorithm at predicting *C. albicans* prenyltransferase substrates, we compared the predictions with the reactivity of the 25 CaaX sequences in the peptide library that correspond to *C. albicans* Cxxx sequences found in the genome, as shown in Table 2.10. Although this is not a complete set of all genomic sequences, it provides a useful test case for comparing the two algorithms.

Sequence	FTase			GGTase-I		
	Substrate?	Substrate by PrePS?	Substrate by new algorithm?	Substrate?	Substrate by PrePS?	Substrate by new algorithm?
CGFH	Y	N	N	Y	N	N
CIIM	Y	Y	Y	Y	Y	Y
CIMM	Y	Y	Y	Y	N	Y
CIVM	Y	Y	Y	Y	Y	Y
CKFN	N	N	N	Y	N	Y
CKIK	N	N	N	Y	N	Y
CKLA	N	N	N	Y	N	Y
CKQQ	N	Y	N	N	N	N
CLAS	N	N	N	Y	N	N
CLIL	Y	Y	Y	Y	Y	Y
CNLL	N	N	Y	Y	Y	Y
CNVF	Y	N	Y	Y	N	Y
CNVL	Y	N	Y	Y	N	Y
CSFM	Y	N	N	Y	N	Y
CSML	N	N	N	Y	N	Y
CSRT	N	N	N	N	N	N
CTII	Y	Y	Y	Y	Y	Y
CTIL	Y	Y	Y	Y	Y	Y
CTIM	Y	Y	Y	Y	Y	Y
CVID	Y	Y	Y	N	N	N
CVIL	Y	Y	Y	Y	Y	Y
CVIM	Y	Y	Y	Y	Y	Y
CVIV	Y	Y	Y	Y	N	Y
CVKL	N	N	N	Y	N	N
CVVL	Y	Y	Y	Y	Y	Y

**Table 2.10. Experimental results, PrePS substrate predictions, and predictions based on new algorithm.** 25 CaaX sequences that are found in the *C. albicans* genome were included in the peptide library. Experimentally determined activity is compared with PrePS and new algorithm predictions for MTO substrates. Correct prediction by PrePS and incorrect prediction by new algorithm are highlighted in yellow, and correct predictions using new algorithm and incorrect prediction by PrePS are highlighted in green.

Out of the 25 C aaX sequences, PrePS correctly classified 20 (80%) and 13 (52%) peptides as substrates and non-substrates for *C. albicans* FTase and GGTase-I, respectively, while the new algorithm correctly classified 22 (88%) and 22 (88%) peptides for the respective enzymes. For FTase, the PrePS false positive and false negative rates were 4% and 16% respectively, and for the new algorithm they were 4% and 8%, respectively. For GGTase-I, the PrePS false positive rate was 0% and false negative rate was 48%, while for the new algorithm the rates were 0% and 12%, respectively. This suggests that while PrePS and the new algorithm

perform similarly well for FTase, the new algorithm is a significantly better predictor for *C. albicans* GGTase-I. This is likely due to the differential recognition between rat and *C. albicans* GGTase-I, and as PrePS algorithm was made to predict mammalian substrates, this could account for the discrepancy. In particular, 9 peptides including CIMM, CKFN, CKIK, CKLA, CNVF, CNVL, CSFM, CSML, and CVIV were incorrectly predicted not to be GGTase-I substrates by PrePS, while the new algorithm correctly identified them as substrates. Neither PrePS nor the new algorithm correctly predicted 3 peptides CGFH, CLAS, and CVKL as substrates. For *C. albicans* FTase, peptides CKQQ, CNVF, and CNLV were correctly predicted by the new algorithm but not by PrePS. Neither algorithm predicted CGFH and CSFM substrates correctly, and the new algorithm identified CNLL as a substrate while it is not, resulting in the only false positive prediction.

Although the above analysis suggests that the new algorithm has good ability to predict *C. albicans* prenyltransferase substrates, the test data set was quite small and further investigation is warranted with other Cxxx sequences found in the *C. albicans* genome. Specifically, for *C. albicans* FTase, predicted substrates that are particularly interesting are CLLW and CTLE, the first one because it would be interesting if *C. albicans* FTase can accept such a large residue at the X position, and the second because of the negatively charged X group. For *C. albicans* GGTase-I, there are 15 peptides that have a large residue at a<sub>2</sub>, 12 of them which have not yet been tested (CCFS, CDFV, CFWQ, CHWV, CMFQ, CSFV, CSFW, CSYL, CVFG, CWFC, CWYC, CYFL), while the three that have been tested (CGFH, CKFN, CSFM) are all *C. albicans* GGTase-I substrates. In the peptide library, over 80% of peptides that contain F, W or Y at the a<sub>2</sub> position are MTO substrates for this enzyme.

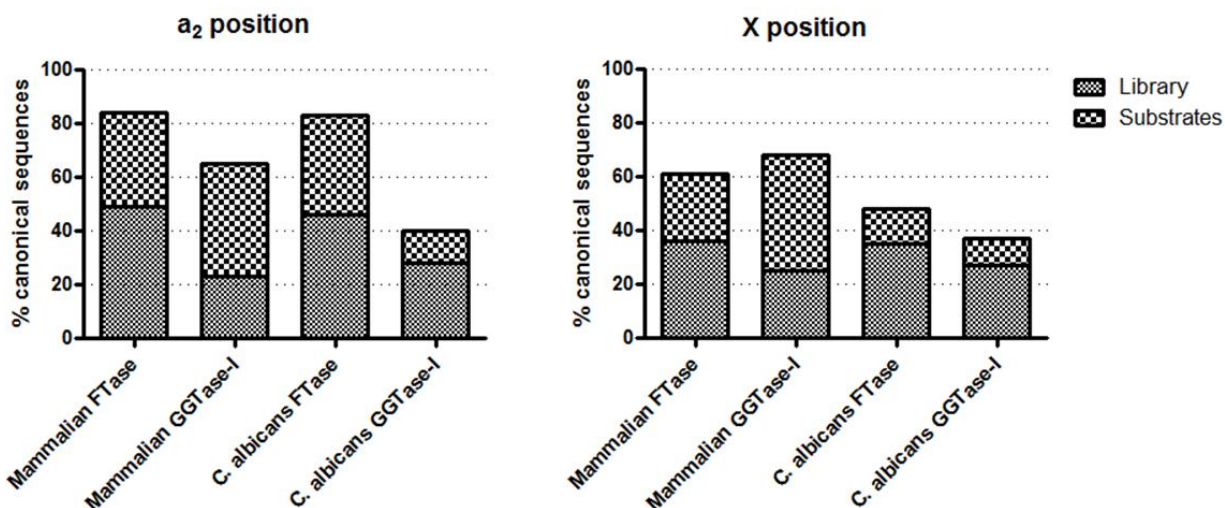
## DISCUSSION

### *Comparison of mammalian and C. albicans prenyltransferase substrate pools*

Understanding *C. albicans* prenyltransferase substrate recognition motifs provides basic knowledge about these important enzymes. This information could potentially be useful in designing inhibitors for these enzymes, and thus paving the way to finding new drugs to combat infections caused by *C. albicans* pathogen. However, in order to develop a successful drug against a pathogenic organism, this compound must not only kill the infectious agent, but it must also be safe and well-tolerated by the human body. Since prenyltransferases are not unique to pathogenic organisms, care must be taken to ensure that inhibitors of the pathogenic enzyme do not significantly affect the human enzymes. To determine the likelihood of a drug discovery effort to find *C. albicans* prenyltransferase-specific inhibitors, a comparison of the mammalian and *C. albicans* prenyltransferase substrate pools and recognition parameters was performed.

Both *C. albicans* FTase and GGTase-I show significant enrichment in canonical residues in their MTO substrate pools compared to the peptide library, but mammalian FTase and GGTase-I follow canonical predictions for the  $\alpha_2$  and X amino acids more closely than *C. albicans* enzymes. In Figure 2.13 compositions of MTO substrate pools for mammalian and *C. albicans* prenyltransferases are compared relative to the peptide library. It is important to point out that although slightly different libraries were used for these studies, relative compositions of the libraries were very similar in terms of their representation of canonical *versus* non-canonical residues, where the library used for mammalian FTase contained 49% and 53% of canonical amino acids at the  $\alpha_2$  and X positions, compared to 46% and 52% in the library used for *C. albicans* FTase. The library to test mammalian GGTase-I contained 30% and 28% of canonical

sequences at the  $a_2$  and X positions, respectively, compared to 28% and 27% in the *C. albicans* library.



**Figure 2.13. Prenyltransferase MTO substrate pool composition.** FTase and GGTase-I MTO substrate pools show significant enrichment in canonical amino acids at the  $a_2$  and X positions. However, *C. albicans* FTase shows more diversity at the X position, and *C. albicans* GGTase-I shows more diversity at both the  $a_2$  and X positions.

Figure 2.13 shows that only 40% and 37% of MTO substrates of the *C. albicans* GGTase-I contain canonical amino acids at the  $a_2$  and X positions, respectively, while 65% and 68% of mammalian GGTase-I MTO substrates have canonical amino acids at these positions. *C. albicans* FTase  $a_2$  position is described by canonical residues better than GGTase-I, but at the X position only 48% of the substrates have a canonical residue. This is in contrast to mammalian FTase, where canonical X residues account for 61% of the MTO substrate pool. This analysis suggests that both *C. albicans* prenyltransferases, and especially GGTase-I, while following general rules of substrate selectivity observed with mammalian enzymes, nonetheless have differences in their substrate preferences.

The difference between mammalian and *C. albicans* prenyltransferases is highlighted when relative MTO and STO substrate pools are compared across different organisms. In the analysis of mammalian FTase with a similarly composed peptide library containing 301 unique

CaaX sequences, 35% of peptides are MTO substrates, and 43% of peptides are STO substrates [19]. These percentages are quite similar to *C. albicans* FTase which suggests that these enzymes appear to have similar modes of substrate recognition. Library screening of mammalian GGTase-I showed 28% of peptides are MTO substrates, and another 44% are STO substrates. In the case of geranylgeranyltransferases, *C. albicans* GGTase-I appears to have a much larger MTO substrate pool, and thus suggests that this enzyme is able to accept a wider variety of CaaX sequences. However, the combined MTO+STO substrate pools are similar between mammalian and *C. albicans* GGTase-I (72% versus 73% respectively), as *C. albicans* GGTase-I has a much smaller pool of STO substrates. If the hypothesis that STO substrates have very slow product dissociation rates is valid, then this would suggest that *C. albicans* GGTase-I in fact recognizes (i.e. is able to bind to and catalyze prenylation of) approximately the same number (and potentially sequence) of peptides as mammalian GGTase-I, but it is much better able to release prenylated products, possibly via a different product dissociation mechanism that does not utilize the product exit groove.

#### *Comparison of mammalian and C. albicans prenyltransferase substrate recognition*

A pairwise comparison of substrate selectivity between the four enzymes reveals that the enzymes with the most similarity to each other are mammalian and *C. albicans* FTases. The only significant difference between these two enzymes is that *C. albicans* FTase appears to be more selective, with more amino acids under-represented at each position. Furthermore, the substrate preferences at the  $a_1$  position are different. At the  $a_2$  position, *C. albicans* FTase, like rat FTase, discriminates against charged residues (D, E, K and R) and large aromatic side chains. The similar recognition is not unexpected as the  $a_2$  binding pocket in mammalian and *C. albicans* FTases are identical. At the X position, both mammalian and *C. albicans* FTases prefer M and Q,

although *C. albicans* enzyme also favors T while mammalian FTase prefers F. The substitution of N191 $\beta$  for S99 $\beta$  and S255 $\beta$  for P152 $\beta$  in *C. albicans* FTase compared to rat FTase may slightly increase the polar nature of the X binding pocket. Overall, these data suggest that it will be difficult to design a *C. albicans* FTase specific peptidomimetic inhibitor, although a smaller substituent at the a<sub>2</sub> position and a more polar moiety at the X position might provide some selectivity.

The mammalian and *C. albicans* GGTase-I have more pronounced differences in substrate selectivity. First, *C. albicans* GGTase-I shows a strong preference for I, L and W (and weaker for C) at the a<sub>1</sub> position, in contrast to mammalian GGTase-I where only V is somewhat enriched. Second, at the a<sub>2</sub> position *C. albicans* GGTase-I shows a preference for large aromatic residues F and W, as well as the canonical I and L. This difference can be exploited in specific inhibitor design, where *C. albicans* GGTase-I specific inhibitors should have large substituents at both the a<sub>1</sub> and a<sub>2</sub> binding sites.

One important question about the selectivity of the *C. albicans* prenyltransferases is whether the lack of toxicity of FTIs is due to cross-prenylation by GGTase-I. Mammalian FTase and GGTase-I have similar substrate preferences at the a<sub>2</sub> position and overlapping preferences at the X position. The X residue is considered to define selectivity where peptides with X residues; however, several studies have shown that this selectivity is much more flexible, and this is evidenced by the F and M being enriched for both enzymes, with FTase additionally preferring more polar residue Q while GGTase-I prefers the more hydrophobic L, V or I side chains. In contrast, *C. albicans* prenyltransferases show a greater overlap in substrate preferences, having at least one amino acid in common that is favored and depleted at each position. However, there are also significant differences between the two *C. albicans* enzymes. At the a<sub>1</sub> and a<sub>2</sub> positions,



GGTase-I prefers larger residues (F, W) while *C. albicans* FTase discriminates against W and Y at the  $a_2$  position. Just as with mammalian enzymes, the two enzymes have overlapping preferences at the X position with FTase preferring more polar side chains than GGTase-I. A comparison of *C. albicans* FTase and GGTase-I substrate preferences shows that *C. albicans* GGTase-I is in general more permissive than FTase and shows greater variability at all positions of the CaaX sequence (Figure 2.12 *versus* Figure 2.10). Out of all the CaaX sequences that were measured for reactivity with *C. albicans* prenyltransferases, 9.4% of FTase substrates had preferred residues at all positions, and 96% of substrates had a preferred residue at at least one position. In contrast, when a similar analysis was carried out for GGTase-I, only 5.8% of all substrates had preferred residues at all positions, and only 79.7% had a preferred residue at at least one position. This suggests that *C. albicans* GGTase-I has a more relaxed substrate composition requirement and is able to accept more sequences that may not be “preferred” substrates. These data could help explain the lack of efficacy of FTIs against *C. albicans* organism as GGTase-I would be able to cross-prenylate FTase substrates. From previous observations that mammalian and *C. albicans* FTases are quite similar, and the differences discussed in mammalian and *C. albicans* GGTase-I substrate preferences, it follows that *C. albicans* GGTase-I should be the enzyme with altered substrate preferences that leads to mammalian *versus* *C. albicans* prenyltransferase substrate preference differences.

In conclusion, based on the molecular substrate recognition determinants elucidated in peptide library studies, it is quite conceivable to design a *C. albicans* GGTase-I-specific peptidomimetic inhibitor by having a large substituent at the  $a_1$  and especially at the  $a_2$  position. For a *C. albicans* FTase specific inhibitor, it is possible that more hydrophilic residues at the X position could offer some selectivity. However, from the data gathered on substrate selectivity, it

is not possible to say whether a dual *C. albicans* FTase/GGTase-I inhibitor is possible that does not also have some activity against mammalian enzymes, at least if only the CaaX binding site is considered. There may be more subtle differences that could potentially be exploited, but these are not readily apparent from this analysis.

## REFERENCES

1. Zhang FL, Casey PJ: **Protein prenylation: molecular mechanisms and functional consequences.** *Annu Rev Biochem* 1996, **65**:241-269.
2. Benetka W, Koranda M, Eisenhaber F: **Protein prenylation: An (almost) comprehensive overview on discovery history, enzymology, and significance in physiology and disease.** *Monatshefte fur Chemie* 2006, **137**:1241-1281.
3. Seabra MC, Reiss Y, Casey PJ, Brown MS, Goldstein JL: **Protein farnesyltransferase and geranylgeranyltransferase share a common alpha subunit.** *Cell* 1991, **65**:429-434.
4. Casey PJ, Seabra MC: **Protein prenyltransferases.** *J Biol Chem* 1996, **271**:5289-5292.
5. Wright LP, Philips MR: **Thematic review series: lipid posttranslational modifications. CAAX modification and membrane targeting of Ras.** *J Lipid Res* 2006, **47**:883-891.
6. Marshall CJ: **Protein prenylation: a mediator of protein-protein interactions.** *Science* 1993, **259**:1865-1866.
7. Berndt N, Hamilton AD, Sebt SM: **Targeting protein prenylation for cancer therapy.** *Nature Reviews Cancer* 2011, **11**:775-791.
8. Wong NS, Morse MA: **Lonafarnib for cancer and progeria.** *Expert Opin Investig Drugs* 2012, **21**:1043-1055.
9. Eastman RT, Buckner FS, Yokoyama K, Gelb MH, Van Voorhis WC: **Fighting parasitic disease by blocking protein farnesylation.** *Journal of Lipid Research* 2006, **47**:233-240.
10. Finegold AA, Johnson DI, Farnsworth CC, Gelb MH, Judd SR, Glomset JA, Tamanoi F: **Protein geranylgeranyltransferase of *Saccharomyces cerevisiae* is specific for Cys-Xaa-Xaa-Leu motif proteins and requires the CDC43 gene product but not the DPR1 gene product.** *Proc Natl Acad Sci U S A* 1991, **88**:4448-4452.
11. Song JL, White TC: **RAM2: an essential gene in the prenylation pathway of *Candida albicans*.** *Microbiology* 2003, **149**:249-259.
12. Kelly R, Card D, Register E, Mazur P, Kelly T, Tanaka KI, Onishi J, Williamson JM, Fan H, Satoh T, et al.: **Geranylgeranyltransferase I of *Candida albicans*: null mutants or enzyme inhibitors produce unexpected phenotypes.** *J Bacteriol* 2000, **182**:704-713.
13. Fu HW, Casey PJ: **Enzymology and biology of CaaX protein prenylation.** *Recent Prog Horm Res* 1999, **54**:315-342; discussion 342-313.
14. Smalera I, Williamson JM, Baginsky W, Leiting B, Mazur P: **Expression and characterization of protein geranylgeranyltransferase type I from the pathogenic yeast *Candida albicans* and identification of yeast selective enzyme inhibitors.** *Biochim Biophys Acta* 2000, **1480**:132-144.
15. Long SB, Casey PJ, Beese LS: **Reaction path of protein farnesyltransferase at atomic resolution.** *Nature* 2002, **419**:645-650.
16. Taylor JS, Reid TS, Terry KL, Casey PJ, Beese LS: **Structure of mammalian protein geranylgeranyltransferase type-I.** *Embo J* 2003, **22**:5963-5974.
17. Hast MA, Nichols CB, Armstrong SM, Kelly SM, Hellinga HW, Alspaugh JA, Beese LS: **Structures of *Cryptococcus neoformans* Protein Farnesyltransferase Reveal Strategies for Developing Inhibitors That Target Fungal Pathogens.** *Journal of Biological Chemistry* 2011, **286**:35149-35162.

18. Hast MA, Beese LS: **Structure of Protein Geranylgeranyltransferase-I from the Human Pathogen *Candida albicans* Complexed with a Lipid Substrate.** *Journal of Biological Chemistry* 2008, **283**:31933-31940.
19. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA: **Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities.** *J Mol Biol* 2010, **395**:176-190.
20. Mathis JR, Poulter CD: **Yeast protein farnesyltransferase: a pre-steady-state kinetic analysis.** *Biochemistry* 1997, **36**:6367-6376.
21. Furfine ES, Leban JJ, Landavazo A, Moomaw JF, Casey PJ: **Protein farnesyltransferase: kinetics of farnesyl pyrophosphate binding and product release.** *Biochemistry* 1995, **34**:6857-6862.
22. Maurer-Stroh S, Eisenhaber F: **Refinement and prediction of protein prenylation motifs.** *Genome Biol* 2005, **6**:R55.
23. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F: **Towards complete sets of farnesylated and geranylgeranylated proteins.** *PLoS Comput Biol* 2007, **3**:e66.
24. Ellman GL: **Tissue sulfhydryl groups.** *Arch Biochem Biophys* 1959, **82**:70-77.
25. Cassidy PB, Dolence JM, Poulter CD: **Continuous fluorescence assay for protein prenyltransferases.** *Methods Enzymol* 1995, **250**:30-43.
26. Reid TS, Terry KL, Casey PJ, Beese LS: **Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity.** *J Mol Biol* 2004, **343**:417-433.
27. Karplus PA: **Hydrophobicity regained.** *Protein Sci* 1997, **6**:1302-1307.
28. Hougland JL, Lamphear CL, Scott SA, Gibbs RA, Fierke CA: **Context-dependent substrate recognition by protein farnesyltransferase.** *Biochemistry* 2009, **48**:1691-1701.

## CHAPTER THREE

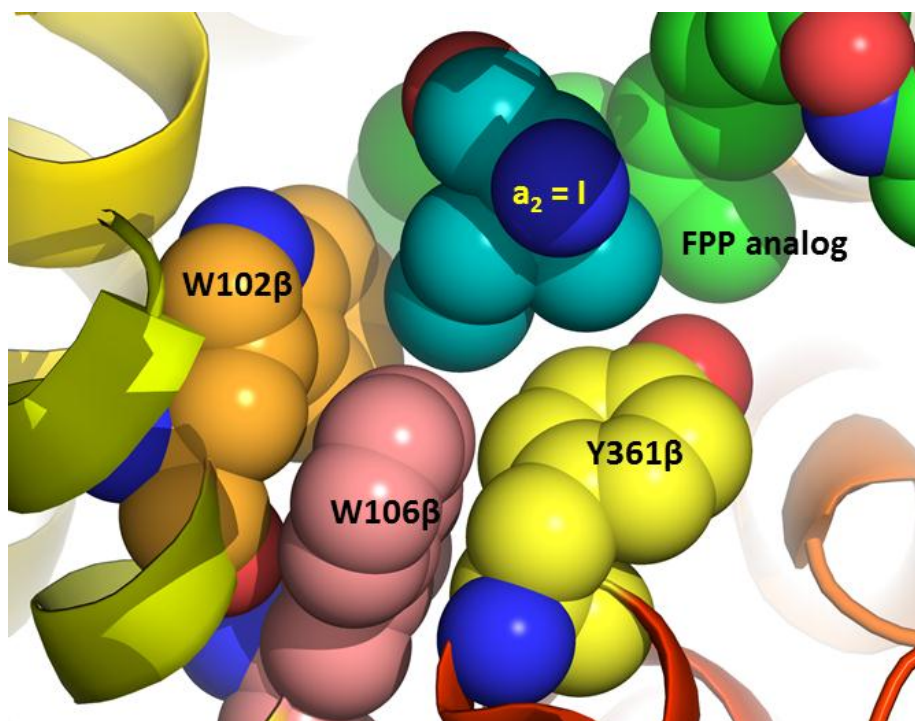
### MOLECULAR BASIS OF Ca<sub>1</sub>a<sub>2</sub>X RECOGNITION IN MAMMALIAN AND *CANDIDA ALBICANS* PRENYLTRANSFERASES

#### INTRODUCTION

FTase and GGTase-I catalyze prenylation of a range of protein substrates, including those implicated in various human diseases [1,2]. These enzymes are proposed to recognize their substrate proteins by the C-terminal “CaaX” motif, where “C” is the modified cysteine residue, “a” is an aliphatic amino acid, and “X” represents an amino acid that guides specificity for either FTase or GGTase-I [3]. Although this model describes a subset of FTase and GGTase-I substrates, recent studies have shown that it is not adequate to encompass all CaaX sequences that can undergo prenylation [4-6]. In Chapter 2, peptide libraries were used to determine substrate recognition parameters of *C. albicans* FTase and GGTase-I. These data demonstrated that at the a<sub>2</sub> position amino acid preferences of FTases are similar, while at the X position the yeast enzyme has a slight preference for hydrophilic residues compared to the rat FTase. Unexpectedly, *C. albicans* GGTase-I showed a significant preference for large (F, W) residues at the a<sub>2</sub> position, while maintaining a similar profile to the rat enzyme at the X position.

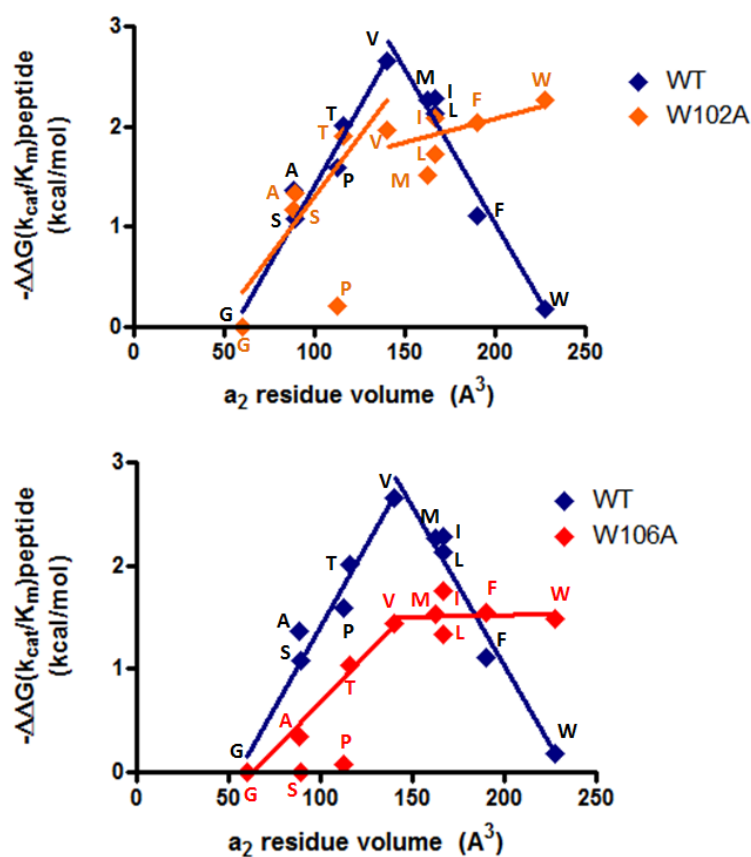
Crystal structures of mammalian FTase and GGTase-I complexed with numerous peptide substrates, lipid donors, products, and inhibitors have been obtained [7,8]. With these crystal structures, the Ca<sub>1</sub>a<sub>2</sub>X substrate recognition model has been refined to incorporate potential

interactions between substrates and the enzyme active site. Given the  $Ca_1a_2X$  sequences of known mammalian FTase substrates, selectivity at the  $a_1$  position appears to be quite relaxed, a finding that is consistent with structural studies which indicate that the  $a_1$  residue of the peptide substrate is exposed to solvent within the interface of the FTase  $\alpha$  and  $\beta$  subunits [9,10]. In contrast to the  $a_1$  position, naturally prenylated proteins appear to favor a subset of moderately sized hydrophobic amino acids (Val, Ile, Leu, Met and Thr) at the  $a_2$  position [11]. The structural basis for this selectivity at the  $a_2$  position is proposed to be due to the hydrophobic nature of the active site surrounding the  $a_2$  residue, largely composed of two Trp residues (W102 $\beta$  and W106 $\beta$ ), a Tyr residue (Y361 $\beta$ ) and the third isoprenoid unit of the FPP prenyl donor, as pictured in Figure 3.1.



**Figure 3.1. Structure of a peptide substrate bound to FTase illustrating the  $a_2$  residue binding site.** The  $a_2$  residue of the peptide substrate KKKSKTKCVIM (teal) is surrounded by residues W102 $\beta$  (orange), W106 $\beta$  (pink) and Y361 $\beta$  (yellow) within the active site of FTase. PDB ID 1D8D.

A rigorous study of the  $a_2$  residue recognition was carried out for mammalian FTase [6]. Reactivity of mammalian FTase was measured with several panels of peptides where the  $a_2$  residue was substituted with all 20 amino acids while keeping the remainder of the  $Ca_1a_2X$  sequence constant. This study showed a positive correlation between peptide reactivity and  $a_2$  residue volume as the amino acid volume increases from 60 to 150 Å<sup>3</sup>, after which a negative correlation was observed. This was interpreted to mean that a positive hydrophobic interaction between the  $a_2$  binding pocket and the  $a_2$  residue exists that enhances substrate reactivity, but there is also a steric clash that prevents larger  $a_2$  side chains from fitting into the  $a_2$  binding pocket. W102βA and W106βA mutations alleviated the steric clash for large  $a_2$  residues (Elaina Zverina, James Hougland, Carol A. Fierke, unpublished data, Figure 3.2).



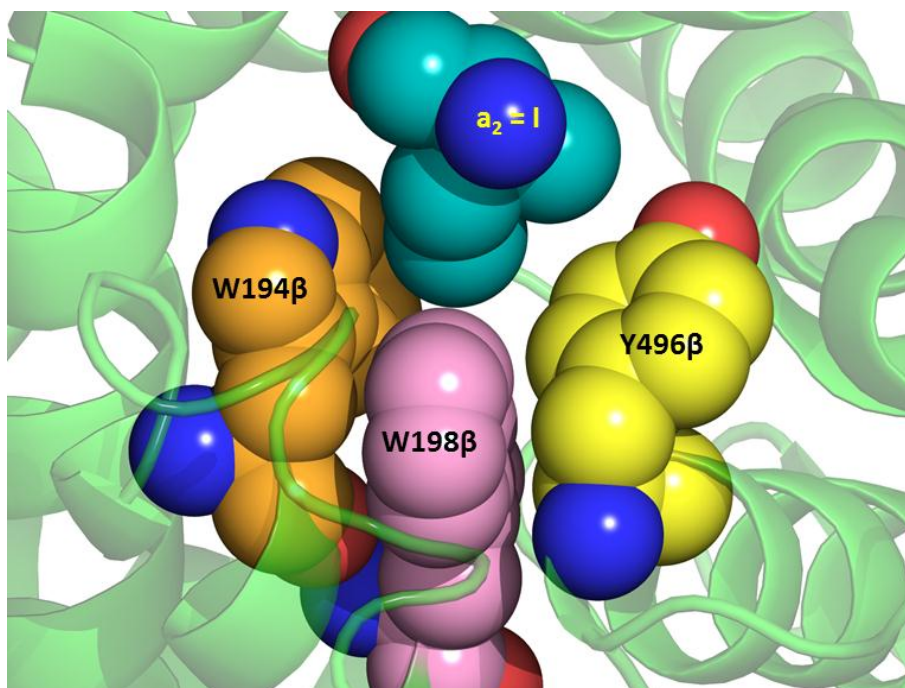
**Figure 3.2.  $a_2$  reactivity of mammalian FTase W102Aβ and W106βA mutants.** Removal of a bulky W102β and W106β residues in the  $a_2$  binding pocket relieves a steric clash and allows peptides with large amino acids at the  $a_2$  position to be prenylated by the enzyme.

In addition to discovering that substrate selectivity in mammalian FTase is based on steric discrimination at the  $a_2$  position, studies have shown that the dependence of reactivity on the volume of the  $a_2$  residue depends on the identity of the X group [6]. When the X residue is Ala or Ser, for FTase the correlation between the  $a_2$  residue volume and peptide reactivity follows the trend shown for wild-type enzyme in Figure 3.2. However, when the X residue is Gln or Met, the steric discrimination for larger side chains ( $> 150 \text{ \AA}^3$ ) disappears. This interplay between the identity of the  $a_2$  and X residues indicates that mammalian FTase recognizes the  $a_2X$ , and possibly  $Ca_1a_2X$ , sequence as a whole rather than separate recognition of each amino acid. Therefore, each position contributes to but does not define substrate recognition.

Analysis of substrate selectivity of *C. albicans* FTase using peptide library approach demonstrated significant similarity between mammalian and *C. albicans* FTases, especially at the  $a_2$  position. The majority of mammalian and *C. albicans* FTase MTO substrates contain canonical (I, L, M, T or V) residues at the  $a_2$  position (84 and 83%, respectively). Although currently there is no crystal structure of *C. albicans* FTase, the structure of *Cryptococcus neoformans* FTase was recently published [12]. Sequence alignment of mammalian, *C. albicans*, and *C. neoformans* FTases shows high sequence conservation in the  $a_2$  binding pocket, with mammalian and *C. albicans* enzymes having exactly the same residues, and the only non-conservative substitution found in *C. neoformans* FTase, where Y365 $\beta$  in rat FTase is replaced by N413 $\beta$  in *C. neoformans* FTase. Based on the mammalian and *C. neoformans* FTase structures, a model of *C. albicans* FTase was developed using Protein Homology/analogy Recognition Engine (PHYRE) [13], and the active site contacts with the  $a_2$  side chain (without FPP) are shown in Figure 3.3. As expected, contacts between the  $a_2$  side chain and the active site

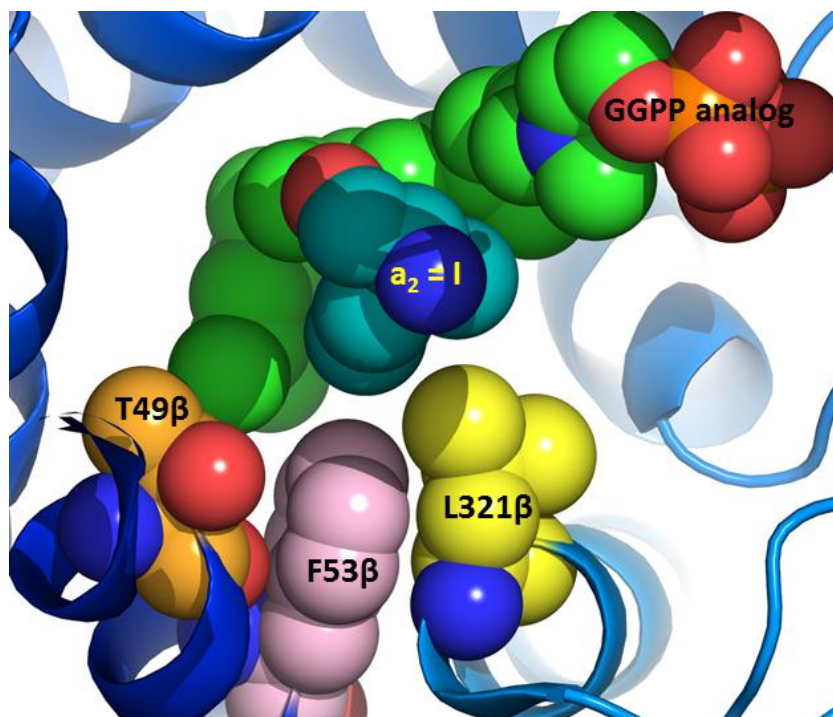


of *C. albicans* FTase are identical to the mammalian enzyme, providing structural support for similar recognition of the  $a_2$  residue by these enzymes.



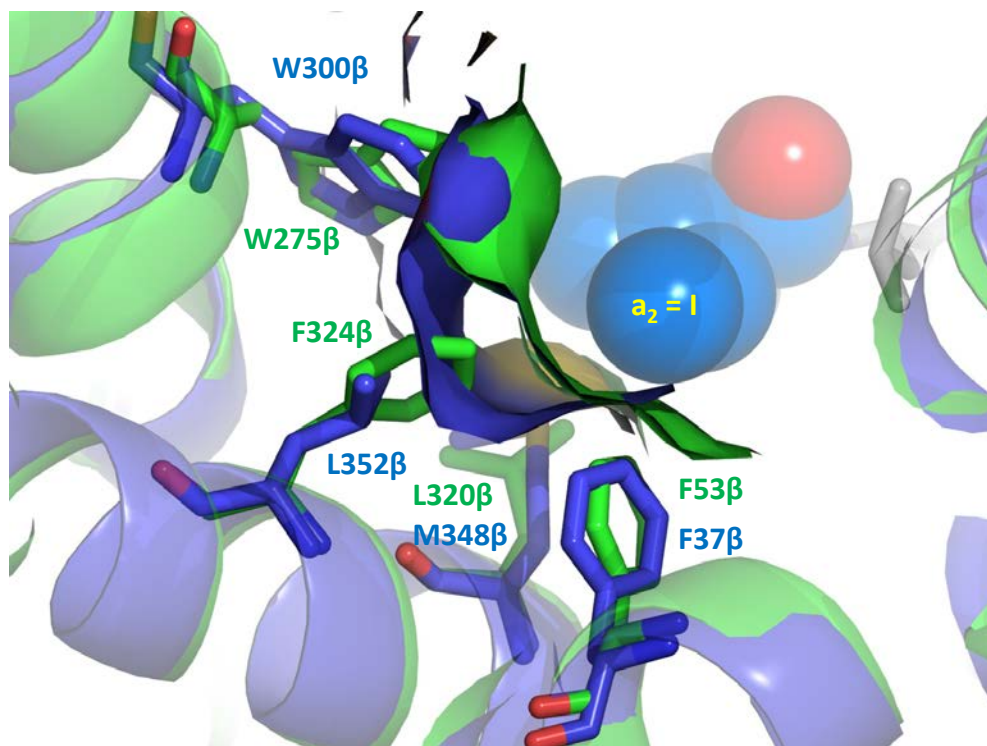
**Figure 3.3. Model of *C. albicans* FTase with a bound peptide substrate illustrating the proposed  $a_2$  residue binding site.** The  $a_2$  residue of the peptide substrate TKCVIM (teal) is surrounded by residues W194 $\beta$  (orange), W198 $\beta$  (pink) and Y496 $\beta$  (yellow) within the active site of FTase. This model was generated using PHYRE (Protein Homology/analogY Recognition Engine) [13].

Structural data indicate that the  $a_2$  binding site in mammalian GGTase-I is somewhat larger and less aromatic in character than the corresponding binding site in FTase. As pictured in Figure 3.4, residues that line the  $a_2$  binding pocket in GGTase-I are T49 $\beta$ , F53 $\beta$ , and L321 $\beta$ , as well as isoprene units 3 and 4 of the GGPP prenyl donor. The effect of these structural changes on the recognition of the  $a_2$  side chain by GGTase-I has not yet been rigorously investigated.



**Figure 3.4. Structure of a peptide substrate bound to GGTase-I illustrating the  $\alpha_2$  residue binding site.** The  $\alpha_2$  residue of the peptide substrate KKKSKTKCVIL (teal) is surrounded by residues T49 $\beta$  (orange), F53 $\beta$  (pink) and L321 $\beta$  (yellow) within the active site of GGTase-I. PDB ID 1N4Q.

A comparison of the crystal structures of mammalian GGTase-I and *C. albicans* GGTase-I with a bound TKCVIL peptide model [14] indicates that the  $\alpha_2$  binding pocket in *C. albicans* GGTase-I is  $\sim 25 \text{ \AA}^3$  larger than rat GGTase-I (Figure 3.5). Consistent with this, a key difference between mammalian and *C. albicans* GGTase-I substrate preferences was that substrates that contain large residues Phe and Trp at the  $\alpha_2$  position are enriched in the MTO substrate pool of *C. albicans* GGTase-I. These data are consistent with the structural studies and support the hypothesis that *C. albicans* GGTase-I has a larger  $\alpha_2$  binding pocket that favors recognition of large residues at the  $\alpha_2$  position of the substrate.



**Figure 3.5. Mammalian and *C. albicans* GGTase-I  $a_2$  binding pocket.** Mammalian (green) and *C. albicans* (blue) GGTase-I were aligned in PyMOL. *C. albicans* GGTase-I  $a_2$  binding pocket is  $\sim 25 \text{ \AA}^3$  larger than the mammalian site due to substitution of L352 $\beta$  and M348 $\beta$  for F324 $\beta$  and L320 $\beta$ . Mammalian PDB ID 1N4Q, *C. albicans* PDB ID 3DRA [8,14].

The aim of this work is to elucidate the specific selectivity criteria used by *C. albicans* FTase and GGTase-I, as well as mammalian GGTase-I, by performing a detailed structure-activity study of these enzymes with panels of peptides with different amino acids at the  $a_2$  and X positions. Reactivity trends that emerge can reveal important features of the  $a_2$  and X residues that contribute to their molecular recognition, as well as any cooperativity between the  $a_2$  and X residues that may provide evidence of a functional interconnection in recognition of these two residues. This study provides insight into the differential substrate recognition of mammalian and *C. albicans* prenyltransferases, as observed in the peptide library analysis (Chapter 2). In combination with structural information, these data may identify important interactions between peptide substrates and the enzymes.

## EXPERIMENTAL PROCEDURES

### *Multiple Turnover Prenylation Activity Assay*

*C. albicans* FTase and GGTase-I were expressed and purified according to the protocols described in Chapter 2. Peptide substrates were assayed for activity with FTase and GGTase-I using a continuous fluorescence-based assay [15]. Peptides were of the form dansyl-GCV<sub>a</sub><sub>2</sub>X, where the residue immediately upstream of the Cys was G instead of TK used in the initial peptide screening libraries. This was done to directly compare the data with the *C. albicans* enzymes to the rat enzymes [6]. A glycine residue upstream of the conserved cysteine had previously been incorporated into peptide substrates to avoid inhibitory interactions between the peptide N-terminal amino group and the prenyl donor cosubstrate [16]. Peptides were stored at -80°C in absolute ethanol containing 10% (v/v) DMSO. Peptide concentrations were measured spectrophotometrically at 412 nm by the reaction of the thiol with 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) using an extinction coefficient of 14,150 M<sup>-1</sup> cm<sup>-1</sup> [17].

For all assays, peptides were incubated in assay buffer (50 mM Hepes pH 7.8, 5 mM TCEP, and 5 mM MgCl<sub>2</sub>) for 30 min. Reactions were initiated by addition of enzyme pre-incubated with prenyl donor in assay buffer. Prenylation of a dansyl peptide enhances the fluorescence intensity and the initial rate of fluorescence change was monitored continuously ( $\lambda_{\text{ex}} = 340 \text{ nm}$ ,  $\lambda_{\text{em}} = 520 \text{ nm}$ ) in a 96-well plate (Costar, black, non-binding surface) using a POLARstar Galaxy plate reader (BMG Technologies). Activity was assayed at 25°C for 1.5 - 2 hours, and all peptides were assayed with no-enzyme controls to correct for background changes in fluorescence. The peptide concentration was varied from 0.2 to 15  $\mu\text{M}$ . The upper limit was determined by initial fluorescence signal intensity. Prenyl donor concentrations were varied from

5 to 15  $\mu\text{M}$ . Enzyme concentrations were adjusted to obtain sufficient signal and linear time courses, usually between 10 and 150 nM.

### *Data analysis*

The initial rate for prenylation at each peptide concentration was fit by a linear regression model. The steady-state kinetic parameters were determined from a fit of the Michaelis-Menten equation (Eq. 1) to the dependence of the initial rate on the peptide concentration:

$$v = \frac{\frac{k_{\text{cat}}}{K_M} [S]}{1 + \frac{[S]}{K_M}} \quad (\text{Eq. 1})$$

For some peptides, this equation was modified to include a term for substrate inhibition (Eq. 2):

$$v = \frac{\frac{k_{\text{cat}}}{K_M} [S]}{1 + \left(\frac{[S]}{K_M}\right) \left(1 + \frac{[S]}{K_i}\right)} \quad (\text{Eq. 2})$$

The  $k_{\text{cat}}/K_M$  values were plotted on a log scale for ease of comparison between peptides.

## **RESULTS**

### *C. albicans FTase reactivity with GCVa<sub>2</sub>X where X=A, S, Q, M, L and F*

To examine substrate selectivity at the a<sub>2</sub> and X residues directly, the steady state kinetic parameters for farnesylation catalyzed by *C. albicans* FTase were measured with dns-GCVa<sub>2</sub>X peptides where the a<sub>2</sub> residue was Gly, Ala, Ser, Pro, Thr, Val, Met, Ile, Leu, Phe, Tyr or Trp, and X was Ala, Ser, Gln, Met, Leu or Phe (Table 3.1). Over half of the peptides had measurable  $k_{\text{cat}}$  and  $K_M$  values under assay conditions described in the protocol. However, for some peptides accurate  $k_{\text{cat}}$  and  $K_M$  values could not be determined because  $K_M$  values were too large, and lower

limits for both  $k_{\text{cat}}$  and  $K_M$  values are given. Some peptides had very small  $K_M$  values, and for these peptides an upper limit of the  $K_M$  value is provided. The  $k_{\text{cat}}/K_M$  steady-state kinetic parameter was used for analysis as it best reflects the selectivity of an enzyme for different substrates [18].

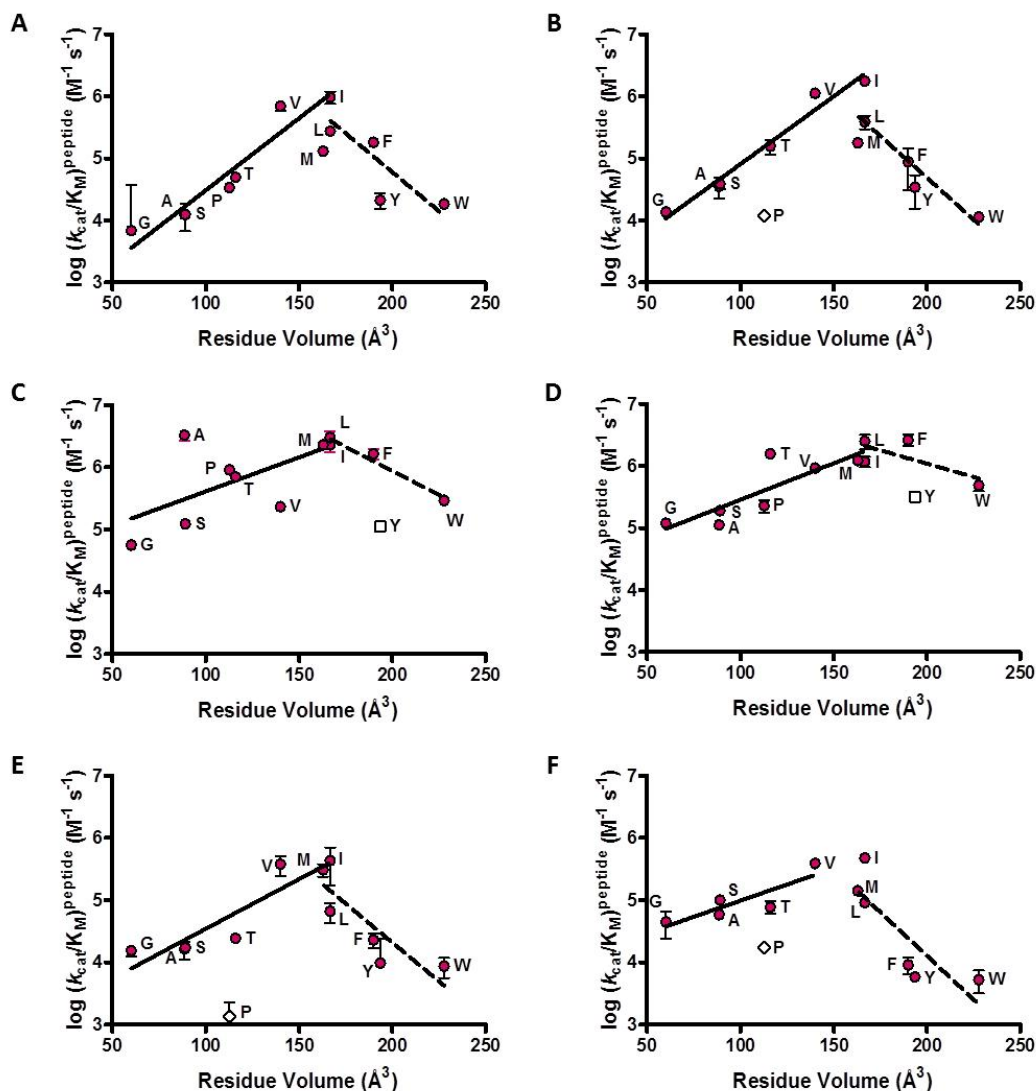
Peptide	$k_{\text{cat}}/K_M$ ( $\text{M}^{-1} \text{s}^{-1}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_M$ ( $\mu\text{M}$ )
dns-GCVAA	$1.3 \times 10^4 \pm 6.8 \times 10^2$	>0.15	>15
dns-GCVFA	$1.8 \times 10^5 \pm 1.9 \times 10^4$	$1.6 \pm 0.2$	$8.8 \pm 1.9$
dns-GCVGA	$6.9 \times 10^3 \pm 3.0 \times 10^4$	~0.01	~1.5
dns-GCVIA	$9.8 \times 10^5 \pm 2.3 \times 10^5$	$4.3 \pm 0.6$	$4.3 \pm 1.6$
dns-GCVLA	$2.7 \times 10^5 \pm 2.5 \times 10^4$	$4.2 \pm 0.6$	$16 \pm 4$
dns-GCVMA	$1.3 \times 10^5 \pm 1.0 \times 10^4$	$1.8 \pm 0.2$	$14 \pm 3$
dns-GCVPA	$3.4 \times 10^4 \pm 1.3 \times 10^3$	>0.5	>15
dns-GCVSA	$1.3 \times 10^4 \pm 6.0 \times 10^3$	>0.2	>15
dns-GCVTA	$5.0 \times 10^4 \pm 4.4 \times 10^3$	$0.8 \pm 0.1$	$16 \pm 4$
dns-GCVVA	$7.0 \times 10^5 \pm 1.2 \times 10^5$	$5.1 \pm 0.8$	$7.2 \pm 3.2$
dns-GCVWA	$1.9 \times 10^4 \pm 1.9 \times 10^3$	>0.2	>15
dns-GCVYA	$2.2 \times 10^4 \pm 6.1 \times 10^3$	>0.7	>15
dns-GCVAS	$3.6 \times 10^4 \pm 1.3 \times 10^4$	>0.7	>15
dns-GCVFS	$9.0 \times 10^4 \pm 5.8 \times 10^4$	>0.6	~6.8
dns-GCVGS	$1.4 \times 10^4 \pm 2.0 \times 10^3$	>0.2	ND
dns-GCVIS	$1.8 \times 10^6 \pm 2.6 \times 10^5$	$3.3 \pm 0.2$	$1.8 \pm 0.3$
dns-GCVLS	$3.9 \times 10^5 \pm 9.6 \times 10^4$	$3.6 \pm 0.9$	$9.1 \pm 4.5$
dns-GCVMS	$1.8 \times 10^5 \pm 1.3 \times 10^4$	$3.8 \pm 0.6$	$21 \pm 5$
dns-GCVPS	$1.2 \times 10^4 \pm 9.1 \times 10^2$	>2	>15
dns-GCVSS	$3.9 \times 10^4 \pm 6.9 \times 10^3$	>2	>15
dns-GCVTS	$1.6 \times 10^5 \pm 4.1 \times 10^4$	>4	>15
dns-GCVVS	$1.1 \times 10^6 \pm 1.2 \times 10^5$	$5.4 \pm 0.4$	$4.8 \pm 0.8$
dns-GCVWS	$1.1 \times 10^4 \pm 9.4 \times 10^2$	>0.1	>15
dns-GCVYS	$3.4 \times 10^4 \pm 1.9 \times 10^4$	>0.4	>15
dns-GCVAQ	$3.3 \times 10^6 \pm 5.9 \times 10^5$	$2.3 \pm 0.1$	$0.7 \pm 0.1$
dns-GCVFQ	$1.7 \times 10^6 \pm 3.4 \times 10^5$	$1.8 \pm 0.1$	$1.1 \pm 0.3$
dns-GCVGQ	$5.7 \times 10^4 \pm 7.5 \times 10^3$	>6	>15
dns-GCVIQ	$2.3 \times 10^6 \pm 5.7 \times 10^5$	$0.48 \pm 0.03$	$0.21 \pm 0.06$
dns-GCVLQ	$3.0 \times 10^6 \pm 7.8 \times 10^5$	$1.05 \pm 0.05$	$0.3 \pm 0.1$
dns-GCVMQ	$2.3 \times 10^6 \pm 3.2 \times 10^5$	$2.39 \pm 0.09$	$1.0 \pm 0.2$
dns-GCVPQ	$9.0 \times 10^5 \pm 1.2 \times 10^5$	$1.81 \pm 0.09$	$2.0 \pm 0.4$
dns-GCVSQ	$1.2 \times 10^5 \pm 5.9 \times 10^3$	$2.3 \pm 0.2$	$18.7 \pm 2.6$
dns-GCVTQ	$7.1 \times 10^5 \pm 8.0 \times 10^4$	$2.2 \pm 0.1$	$3.1 \pm 0.5$
dns-GCVVQ	$2.4 \times 10^5 \pm 1.3 \times 10^4$	$0.78 \pm 0.02$	$3.3 \pm 0.3$
dns-GCVWQ	$2.9 \times 10^5 \pm 3.6 \times 10^4$	$0.46 \pm 0.02$	$1.6 \pm 0.2$
dns-GCVYQ	$1.1 \times 10^5 \pm 4.6 \times 10^3$	$1.01 \pm 0.04$	$9.1 \pm 0.7$
dns-GCVAM	$7.7 \times 10^4 \pm 1.1 \times 10^4$	>8	>15
dns-GCVFM	$1.1 \times 10^6 \pm 2.3 \times 10^5$	$1.67 \pm 0.08$	$0.6 \pm 0.1$
dns-GCVGM	$1.2 \times 10^5 \pm 8.4 \times 10^3$	$0.43 \pm 0.03$	$3.5 \pm 0.5$
dns-GCVIM	$1.1 \times 10^6 \pm 2.1 \times 10^5$	$1.8 \pm 0.1$	$1.6 \pm 0.4$
dns-GCVLM	$1.3 \times 10^6 \pm 3.4 \times 10^5$	$3.9 \pm 0.3$	$1.6 \pm 0.5$

dns-GCVMM	$9.6 \times 10^5 \pm 1.3 \times 10^5$	$4.5 \pm 0.3$	$3.8 \pm 0.7$
dns-GCVPM	$1.8 \times 10^5 \pm 4.3 \times 10^4$	$0.8 \pm 0.1$	$3.5 \pm 1.2$
dns-GCVSM	$1.2 \times 10^5 \pm 1.6 \times 10^4$	$1.8 \pm 0.2$	$9.2 \pm 2.4$
dns-GCVTM	$6.4 \times 10^5 \pm 9.3 \times 10^4$	$3.4 \pm 0.2$	$2.1 \pm 0.4$
dns-GCVVM	$8.9 \times 10^5 \pm 1.4 \times 10^5$	$2.7 \pm 0.2$	$2.8 \pm 0.6$
dns-GCVWM	$2.9 \times 10^5 \pm 5.3 \times 10^4$	$1.22 \pm 0.09$	$2.5 \pm 0.6$
dns-GCVYM	$2.0 \times 10^5 \pm 5.3 \times 10^4$	$2.2 \pm 0.5$	$6.8 \pm 3.2$
dns-GCVAL	$1.7 \times 10^4 \pm 5.3 \times 10^3$	$0.15 \pm 0.05$	$8.8 \pm 5.5$
dns-GCVFL	$2.3 \times 10^4 \pm 6.2 \times 10^3$	$0.11 \pm 0.02$	$4.7 \pm 2.0$
dns-GCVGL	$1.6 \times 10^4 \pm 3.0 \times 10^3$	$>0.2$	$>15$
dns-GCVIL	$4.4 \times 10^5 \pm 2.6 \times 10^5$	$1.2 \pm 0.3$	$2.8 \pm 2.4$
dns-GCVLL	$6.6 \times 10^4 \pm 2.3 \times 10^4$	$0.48 \pm 0.15$	$7.3 \pm 4.7$
dns-GCVML	$3.1 \times 10^5 \pm 6.8 \times 10^4$	$0.98 \pm 0.10$	$3.2 \pm 1.0$
dns-GCVPL	$1.4 \times 10^3 \pm 9.5 \times 10^2$	$>0.02$	$>15$
dns-GCVSL	$1.7 \times 10^4 \pm 2.3 \times 10^3$	$>0.8$	$>15$
dns-GCVTL	$2.5 \times 10^4 \pm 3.6 \times 10^3$	$>0.6$	$>15$
dns-GCVVL	$3.8 \times 10^5 \pm 1.4 \times 10^5$	$1.0 \pm 0.2$	$2.8 \pm 1.4$
dns-GCVWL	$8.7 \times 10^3 \pm 3.2 \times 10^3$	$0.037 \pm 0.008$	$4.2 \pm 2.3$
dns-GCVYL	$9.7 \times 10^3 \pm 1.5 \times 10^4$	$0.024 \pm 0.009$	$\approx 1.5$
dns-GCVAF	$5.9 \times 10^4 \pm 5.4 \times 10^3$	$0.36 \pm 0.03$	$6.1 \pm 1.0$
dns-GCVFF	$9.2 \times 10^3 \pm 2.7 \times 10^3$	$0.16 \pm 0.11$	$>15$
dns-GCVGF	$4.4 \times 10^4 \pm 2.1 \times 10^4$	$0.23 \pm 0.07$	$5.1 \pm 3.9$
dns-GCVIF	$4.8 \times 10^5 \pm 5.5 \times 10^4$	$2.98 \pm 0.26$	$6.2 \pm 1.2$
dns-GCVLF	$9.1 \times 10^4 \pm 1.1 \times 10^4$	$0.65 \pm 0.07$	$7.1 \pm 1.5$
dns-GCVMF	$1.4 \times 10^5 \pm 1.4 \times 10^4$	$1.06 \pm 0.09$	$7.5 \pm 1.3$
dns-GCVPF	$1.8 \times 10^4 \pm 1.1 \times 10^4$	$0.07 \pm 0.02$	$4.2 \pm 3.9$
dns-GCVSF	$1.0 \times 10^5 \pm 1.3 \times 10^4$	$0.64 \pm 0.06$	$6.3 \pm 1.4$
dns-GCVTF	$7.8 \times 10^4 \pm 1.7 \times 10^4$	$0.44 \pm 0.07$	$5.6 \pm 2.1$
dns-GCVVF	$3.9 \times 10^5 \pm 5.7 \times 10^4$	$1.56 \pm 0.13$	$4.0 \pm 0.9$
dns-GCVWF	$5.2 \times 10^3 \pm 2.1 \times 10^3$	$>0.06$	$>15$
dns-GCVYF	$5.9 \times 10^3 \pm 5.6 \times 10^2$	$>0.07$	$>15$

**Table 3.1. Steady-state kinetic parameters measured for *C. albicans* FTase.** The steady-state kinetic parameters were determined at saturating FPP (10  $\mu$ M) and varying peptide concentrations (0.2 – 15  $\mu$ M) and 10 nM FTase. Peptides are grouped by the X residue, and ordered alphabetically by the  $a_2$  residue.

To facilitate comparison across different  $a_2$  and X groups, the data were analyzed using  $\log k_{\text{cat}}/K_M^{\text{peptide}}$ , as shown in Figure 3.6. Peptide reactivity was plotted against the  $a_2$  residue volume (Table 3.2) as previous studies have found the steric volume of the  $a_2$  residue to be a driver of selectivity [6]. All of the peptide panels varying  $a_2$  position show a pyramidal shape with a maximum  $k_{\text{cat}}/K_M$  value at  $\sim 150 \text{ \AA}^3$  ( $a_2 = \text{I, L, M}$ ), independent of the identity of the X group. However, the dependence of the  $k_{\text{cat}}/K_M$  values on the volume of the  $a_2$  side chain is greater when X = Ala or Ser compared to X = Gln or Met. This dependence manifests as both a

greater enhancement in reactivity for the smaller amino acids (slope  $\sim 0.023$  versus slope  $\sim 0.012$ ) and a larger steric discrimination against the larger amino acids (slope  $\sim -0.026$  versus slope  $\sim -0.012$ ).



**Figure 3.6. *C. albicans* FTase-I reactivity with GCVa<sub>2</sub>X peptides.** (A) Dns-GCVa<sub>2</sub>A peptide panel. The solid line has a slope of  $0.023 \pm 0.003$  ( $R^2 = 0.91$ ) and the dashed line has a slope of  $-0.025 \pm 0.009$  ( $R^2 = 0.71$ ) (B) Dns-GCVa<sub>2</sub>S peptide panel. The solid line has a slope of  $0.022 \pm 0.002$  ( $R^2 = 0.97$ ) and the dashed line has a slope of  $-0.027 \pm 0.008$  ( $R^2 = 0.75$ ) (C) Dns-GCVa<sub>2</sub>Q peptide panel. The solid line has a slope of  $0.011 \pm 0.005$  ( $R^2 = 0.43$ ) and the dashed line has a slope of  $-0.015 \pm 0.002$  ( $R^2 = 0.96$ ) (D) Dns-GCVa<sub>2</sub>M peptide panel. The solid line has a slope of  $0.012 \pm 0.002$  ( $R^2 = 0.76$ ) and the dashed line has a slope of  $-0.009 \pm 0.006$  ( $R^2 = 0.52$ ). (E) Dns-GCVa<sub>2</sub>L peptide panel. The solid line has a slope of  $0.016 \pm 0.003$  ( $R^2 = 0.85$ ) and the dashed line has a slope of  $-0.025 \pm 0.008$  ( $R^2 = 0.71$ ) (F) Dns-GCVa<sub>2</sub>F peptide panel. The solid line has a slope of  $0.010 \pm 0.004$  ( $R^2 = 0.74$ ) and the dashed line has a slope of  $-0.028 \pm 0.009$  ( $R^2 = 0.70$ )



Amino acid	Residue volume ( $\text{\AA}^3$ )
A	88.6
F	189.9
G	60.1
I	166.7
L	166.7
M	162.9
P	112.7
S	89
T	116.1
V	140
W	227.8
Y	193.6

**Table 3.2. Residue volumes used in analysis of peptide reactivity.** Values were obtained from reference [19].

These data demonstrate that *C. albicans* FTase recognizes the  $a_2X$  as a cooperative unit and not as individual amino acids. Therefore, there is a functional interconnection in recognition of these two side chains by *C. albicans* FTase. Peptides that terminate in Ala or Ser show a stronger effect of the  $a_2$  residue volume on reactivity than those that have  $X = \text{Gln}$  or  $\text{Met}$ . Glutamine and methionine residues are larger than alanine or serine, and therefore these side chains could potentially participate in more interactions with the enzyme (i.e. contribute more binding energy) to exert more impact on substrate recognition. Similar trends were observed for mammalian FTase where Met or Gln are proposed to form hydrogen bonds with W102 $\beta$  and S99 $\beta$  side chains while Ala or Ser are unable to form these interactions [6]. However, in rat FTase removing these hydrogen bond partners via mutagenesis (W102 $\beta$ A and S99 $\beta$ A) did not alter the peptide reactivity profiles as peptides with  $X = \text{Met}$  did not show the same dependence on the  $a_2$  residue volume with these mutant enzymes as the wild-type enzyme with small X residues. Thus, these interactions either did not occur in the enzyme or their contribution was not large enough to be the driving force behind differences in X residue interaction.

To further investigate the contribution of the size of the X residue to molecular recognition, reactivity of peptides with varying  $a_2$  residue volumes that have  $X = \text{L}$  and  $\text{F}$  was

measured (Figure 3.6). Although peptides that have Leu and Phe amino acids at the X position are generally thought to be GGTase-I substrates, a recent report showed that prenylation of a significant number of peptides from a -CaaL library was catalyzed by mammalian FTase [5]. Thus, we also wanted to determine whether *C. albicans* FTase can react with canonical GGTase-I substrates. Somewhat surprisingly, *C. albicans* FTase readily catalyzes farnesylation of all these peptides. Furthermore, peptide reactivity as a function of the  $a_2$  residue volume is qualitatively similar for peptides terminating in L and F amino acids, showing a pyramidal correlation. In both cases, the highest  $k_{cat}/K_M$  value occurs for a side chain volume of  $\sim 150 \text{ \AA}^3$ ; however, the maximum value of  $k_{cat}/K_M$  for X = L or F is  $\sim 3.5$ -fold lower than for peptides where X = M or Q ( $4.6 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  versus  $1.6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ).

Dependence of *C. albicans* FTase reactivity on the  $a_2$  residue volume for peptides with X = L or F appears to be a hybrid between trends observed with the X = Ala/Ser and Gln/Met peptides. For peptides that have L or F at the X position, the increase in reactivity as the side chain volume increases (slope  $\sim 0.013$ ) is closer in magnitude to those measured for X = Gln or Met (slope  $\sim 0.012$ ) than for X = Ala or Ser. After peak peptide reactivity is reached, steric hindrance decreases peptide reactivity with both X = L and F amino acids at the X position. However, in this case the anti-correlation coefficient (slope  $\sim -0.026$ ) is more similar to that measured for X = A or S (slope  $\sim -0.026$ ) than for X = Q or M. Together these data suggest that when the X residue is large, the  $a_2$  residue volume is less important for reactivity when the  $a_2$  residue is small, but as the  $a_2$  residue volume increases, this residue becomes a more significant contributor to substrate recognition.

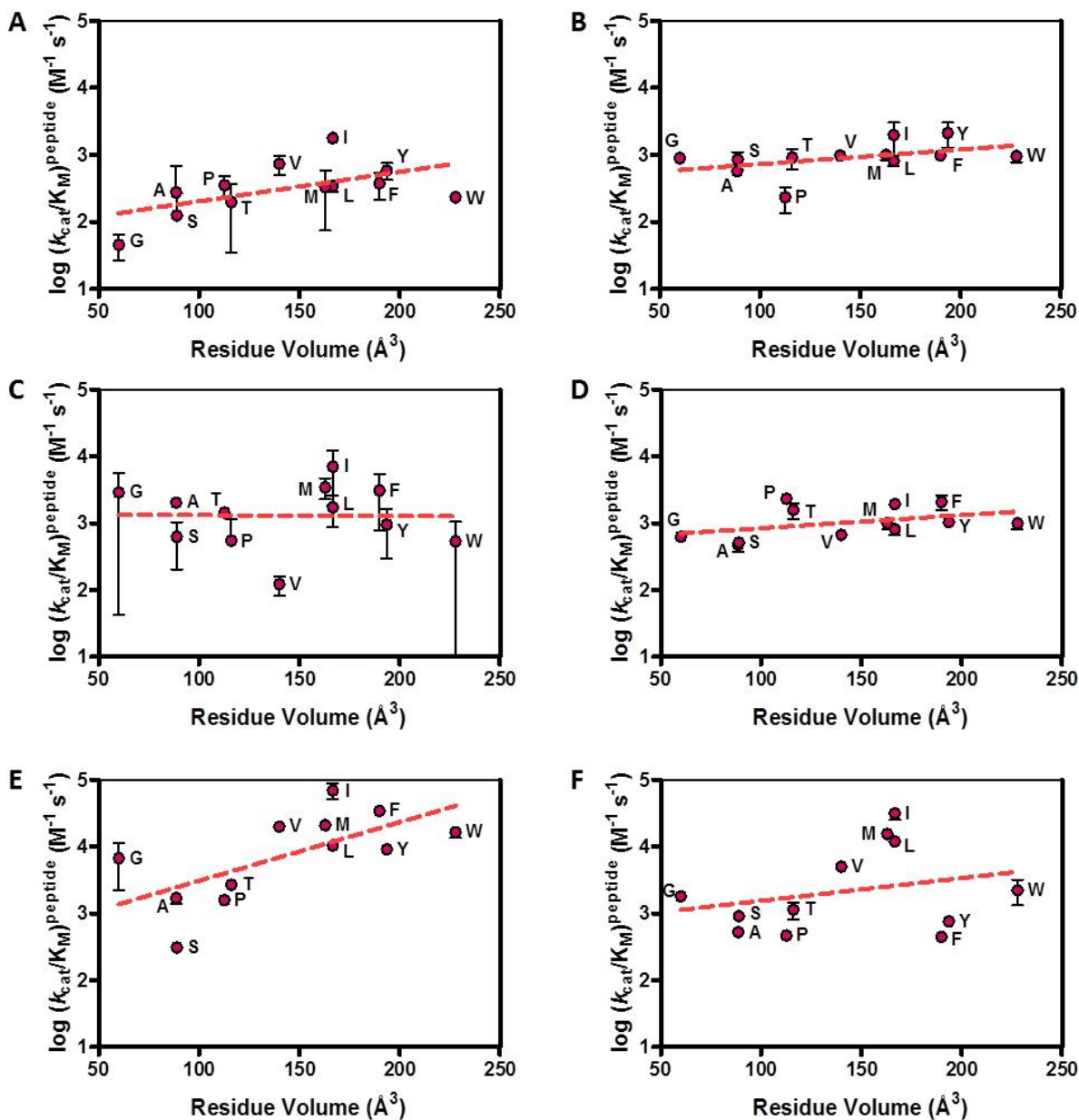
*C. albicans* GGTase-I reactivity with GCVa<sub>2</sub>X where X=A, S, Q, M, L and F

Reactivity of *C. albicans* GGTase-I was measured with dns-GCVa<sub>2</sub>X peptides where a<sub>2</sub> residue was Gly, Ala, Ser, Pro, Thr, Val, Met, Ile, Leu, Phe, Tyr or Trp, and X was Ala, Ser, Gln, Met, Leu or Phe. The steady-state kinetic parameters for GGTase-I catalyzed geranylgeranylation of these peptides were obtained using a fluorescence-based assay and are shown in Table 3.3. Overall, *C. albicans* GGTase-I had higher K<sub>M</sub> values for peptides than FTase, leading to more error in determination of the k<sub>cat</sub>/K<sub>M</sub> value. The values for k<sub>cat</sub>/K<sub>M</sub><sup>peptide</sup> were converted to log scale in order to facilitate comparison across different a<sub>2</sub> and X groups, and the results are presented in Figure 3.7.

Peptide	k <sub>cat</sub> /K <sub>M</sub> (M <sup>-1</sup> s <sup>-1</sup> )	k <sub>cat</sub> (s <sup>-1</sup> )	K <sub>M</sub> (μM)
dns-GCVAA	2.8 x 10 <sup>2</sup> ± 4.2 x 10 <sup>2</sup>	>0.0006	~2
dns-GCVFA	3.8 x 10 <sup>2</sup> ± 1.7 x 10 <sup>2</sup>	>0.003	~8
dns-GCVGA	4.6 x 10 <sup>1</sup> ± 1.9 x 10 <sup>1</sup>	>0.0006	>15
dns-GCVIA	1.8 x 10 <sup>3</sup> ± 2.4 x 10 <sup>2</sup>	0.0055 ± 0.0004	3.1 ± 0.6
dns-GCVLA	3.5 x 10 <sup>2</sup> ± 6.3 x 10 <sup>1</sup>	>0.004	>15
dns-GCVMA	3.3 x 10 <sup>2</sup> ± 2.6 x 10 <sup>2</sup>	>0.004	>15
dns-GCVPA	3.6 x 10 <sup>2</sup> ± 1.2 x 10 <sup>2</sup>	>0.003	~8
dns-GCVSA	1.3 x 10 <sup>2</sup> ± 1.4 x 10 <sup>2</sup>	>0.0006	~5
dns-GCVTA	2.0 x 10 <sup>2</sup> ± 1.7 x 10 <sup>2</sup>	>0.002	~9
dns-GCVVA	7.5 x 10 <sup>2</sup> ± 2.4 x 10 <sup>2</sup>	0.0028 ± 0.0005	3.7 ± 1.8
dns-GCVWA	2.3 x 10 <sup>2</sup> ± 1.8 x 10 <sup>1</sup>	>0.003	>15
dns-GCVYA	5.8 x 10 <sup>2</sup> ± 1.6 x 10 <sup>2</sup>	0.0018 ± 0.0002	3.1 ± 1.2
dns-GCVAS	5.8 x 10 <sup>2</sup> ± 7.5 x 10 <sup>1</sup>	>0.007	>15
dns-GCVFS	9.9 x 10 <sup>2</sup> ± 1.2 x 10 <sup>2</sup>	>0.01	>15
dns-GCVGS	8.9 x 10 <sup>2</sup> ± 9.6 x 10 <sup>1</sup>	>0.01	>15
dns-GCVIS	2.0 x 10 <sup>3</sup> ± 1.0 x 10 <sup>3</sup>	>0.02	>15
dns-GCVLS	8.1 x 10 <sup>2</sup> ± 1.4 x 10 <sup>2</sup>	>0.01	>15
dns-GCVMS	1.0 x 10 <sup>3</sup> ± 1.7 x 10 <sup>2</sup>	>0.01	>15
dns-GCVPS	2.4 x 10 <sup>2</sup> ± 9.9 x 10 <sup>1</sup>	>0.003	>15
dns-GCVSS	8.4 x 10 <sup>2</sup> ± 2.4 x 10 <sup>2</sup>	>0.01	>15
dns-GCVTS	9.0 x 10 <sup>2</sup> ± 3.0 x 10 <sup>2</sup>	>0.01	>15
dns-GCVVS	9.9 x 10 <sup>2</sup> ± 1.3 x 10 <sup>2</sup>	>0.01	>15
dns-GCVWS	9.4 x 10 <sup>2</sup> ± 1.7 x 10 <sup>2</sup>	>0.01	>15
dns-GCVYS	2.1 x 10 <sup>3</sup> ± 8.4 x 10 <sup>2</sup>	>0.03	>15
dns-GCVAQ	2.1 x 10 <sup>3</sup> ± 3.5 x 10 <sup>2</sup>	>0.03	>15
dns-GCVFQ	3.1 x 10 <sup>3</sup> ± 2.3 x 10 <sup>3</sup>	0.007 ± 0.002	~2.5
dns-GCVGQ	2.9 x 10 <sup>3</sup> ± 2.8 x 10 <sup>3</sup>	0.007 ± 0.003	~2.5
dns-GCVIQ	7.1 x 10 <sup>3</sup> ± 5.2 x 10 <sup>3</sup>	0.034 ± 0.017	~4
dns-GCVLQ	1.7 x 10 <sup>3</sup> ± 8.3 x 10 <sup>2</sup>	>0.02	>15
dns-GCVMQ	3.4 x 10 <sup>3</sup> ± 1.2 x 10 <sup>3</sup>	0.009 ± 0.003	2.7 ± 1.6

dns-GCVPQ	$1.5 \times 10^3 \pm 1.6 \times 10^2$	>0.02	>15
dns-GCVSQ	$6.4 \times 10^2 \pm 4.3 \times 10^2$	$0.0011 \pm 0.0003$	$1.8 \pm 1.5$
dns-GCVTQ	$5.5 \times 10^2 \pm 6.3 \times 10^2$	>0.007	>15
dns-GCVVQ	$1.2 \times 10^2 \pm 4.0 \times 10^1$	>0.001	>15
dns-GCVWQ	$5.4 \times 10^2 \pm 5.4 \times 10^2$	>0.006	>15
dns-GCVYQ	$9.6 \times 10^2 \pm 6.7 \times 10^2$	>0.01	>15
dns-GCVAM	$4.8 \times 10^2 \pm 1.0 \times 10^2$	$0.005 \pm 0.001$	$10.9 \pm 5.1$
dns-GCVFM	$2.1 \times 10^3 \pm 5.0 \times 10^2$	$0.0037 \pm 0.0004$	$1.8 \pm 0.6$
dns-GCVGM	$6.3 \times 10^2 \pm 3.3 \times 10^1$	>0.008	>15
dns-GCVIM	$1.9 \times 10^3 \pm 2.0 \times 10^2$	>0.05	>15
dns-GCVLM	$8.2 \times 10^2 \pm 1.4 \times 10^2$	>0.02	>15
dns-GCVMM	$1.0 \times 10^3 \pm 1.8 \times 10^2$	>0.02	>15
dns-GCVPM	$2.3 \times 10^3 \pm 3.1 \times 10^2$	$0.0064 \pm 0.0004$	$2.7 \pm 0.5$
dns-GCVSM	$5.1 \times 10^2 \pm 6.9 \times 10^1$	$0.006 \pm 0.001$	$11.5 \pm 3.5$
dns-GCVTM	$1.6 \times 10^3 \pm 4.3 \times 10^2$	$0.021 \pm 0.008$	$13.3 \pm 8.6$
dns-GCVVM	$6.7 \times 10^2 \pm 7.4 \times 10^1$	$0.009 \pm 0.001$	$12.8 \pm 3.3$
dns-GCVWM	$1.0 \times 10^3 \pm 1.8 \times 10^2$	$0.015 \pm 0.004$	$15 \pm 7$
dns-GCVYM	$1.0 \times 10^3 \pm 1.4 \times 10^2$	$0.014 \pm 0.003$	$13.7 \pm 4.5$
dns-GCVAL	$1.7 \times 10^3 \pm 3.0 \times 10^2$	$0.014 \pm 0.002$	$8.3 \pm 2.9$
dns-GCVFL	$3.4 \times 10^4 \pm 4.2 \times 10^3$	$0.0278 \pm 0.0009$	$0.81 \pm 0.11$
dns-GCVGL	$6.8 \times 10^3 \pm 4.6 \times 10^3$	$0.03 \pm 0.01$	~4.5
dns-GCVIL	$6.9 \times 10^4 \pm 1.8 \times 10^4$	$0.22 \pm 0.03$	$3.2 \pm 1.2$
dns-GCVLL	$1.1 \times 10^4 \pm 1.1 \times 10^3$	$0.027 \pm 0.001$	$2.5 \pm 0.4$
dns-GCVML	$2.1 \times 10^4 \pm 2.5 \times 10^3$	$0.088 \pm 0.006$	$4.2 \pm 0.8$
dns-GCVPL	$1.6 \times 10^3 \pm 1.3 \times 10^2$	>0.02	>15
dns-GCVSL	$3.1 \times 10^2 \pm 2.1 \times 10^1$	>0.004	>15
dns-GCVTL	$2.7 \times 10^3 \pm 4.2 \times 10^2$	$0.032 \pm 0.007$	$12.0 \pm 4.3$
dns-GCVVL	$2.0 \times 10^4 \pm 1.9 \times 10^3$	$0.13 \pm 0.01$	$6.5 \pm 1.1$
dns-GCVWL	$1.6 \times 10^4 \pm 2.6 \times 10^3$	$0.0168 \pm 0.0008$	$1.0 \pm 0.2$
dns-GCVYL	$9.1 \times 10^3 \pm 9.3 \times 10^2$	$0.044 \pm 0.003$	$4.8 \pm 0.8$
dns-GCVAF	$5.2 \times 10^2 \pm 3.1 \times 10^1$	>0.006	>15
dns-GCVFF	$4.5 \times 10^2 \pm 2.8 \times 10^1$	>0.005	>15
dns-GCVGF	$1.8 \times 10^3 \pm 1.5 \times 10^2$	>0.02	>15
dns-GCVIF	$3.1 \times 10^4 \pm 6.4 \times 10^3$	$0.14 \pm 0.02$	$4.6 \pm 1.4$
dns-GCVLF	$1.2 \times 10^4 \pm 1.4 \times 10^3$	$0.036 \pm 0.002$	$3.0 \pm 0.5$
dns-GCVMF	$1.5 \times 10^4 \pm 1.7 \times 10^3$	$0.065 \pm 0.004$	$4.2 \pm 0.7$
dns-GCVPF	$4.6 \times 10^2 \pm 2.7 \times 10^1$	>0.006	>15
dns-GCVSF	$9.0 \times 10^2 \pm 6.7 \times 10^1$	>0.01	>15
dns-GCVTF	$1.1 \times 10^3 \pm 3.3 \times 10^2$	$0.009 \pm 0.003$	$8.1 \pm 4.5$
dns-GCVVF	$5.1 \times 10^3 \pm 5.0 \times 10^2$	$0.049 \pm 0.005$	$9.8 \pm 2.0$
dns-GCVWF	$2.2 \times 10^3 \pm 8.9 \times 10^2$	$0.011 \pm 0.003$	$5.1 \pm 3.2$
dns-GCVYF	$7.5 \times 10^2 \pm 4.5 \times 10^1$	>0.009	>15

**Table 3.3. Steady-state kinetic parameters measured for *C. albicans* GGTase-I.** The steady-state kinetic parameters were determined at saturating GGPP (10  $\mu$ M) and varying peptide concentrations (0.2 – 15  $\mu$ M) and 100 nM GGTase-I. Peptides are grouped by the X residue, and ordered alphabetically by the a<sub>2</sub> residue.



**Figure 3.7. *C. albicans* GGTase-I reactivity with GCVa<sub>2</sub>X peptides.** (A) Dns-GCVa<sub>2</sub>A peptide panel. The dashed line has a slope of  $0.0044 \pm 0.0020$  ( $R^2 = 0.32$ ) (B) Dns-GCVa<sub>2</sub>S peptide panel. The dashed line has a slope of  $0.0022 \pm 0.0014$  ( $R^2 = 0.20$ ) (C) Dns-GCVa<sub>2</sub>Q peptide panel. The dashed line has a slope of  $-0.0001 \pm 0.003$  ( $R^2 = 0.0002$ ) (D) Dns-GCVa<sub>2</sub>M peptide panel. The dashed line has a slope of  $0.0020 \pm 0.0014$  ( $R^2 = 0.17$ ) (E) Dns-GCVa<sub>2</sub>L peptide panel. The dashed line has a slope of  $0.0088 \pm 0.0031$  ( $R^2 = 0.44$ ) (F) Dns-GCVa<sub>2</sub>F peptide panel. The dashed line has a slope of  $0.0034 \pm 0.0039$  ( $R^2 = 0.07$ )

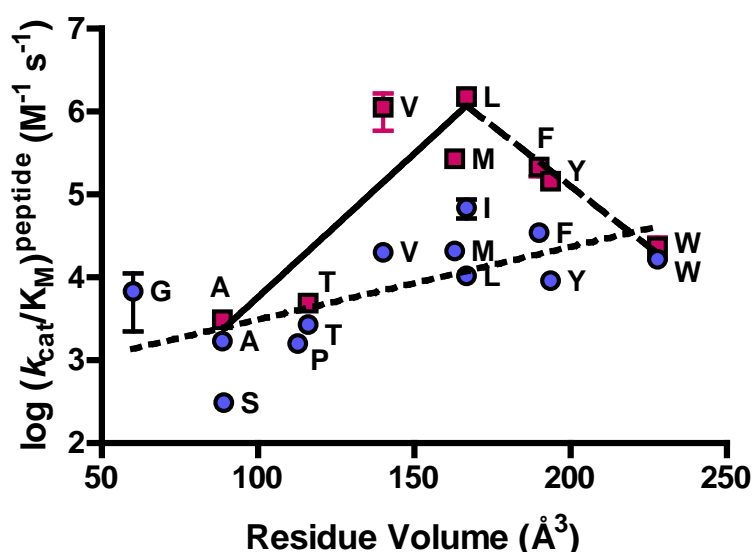
Unlike mammalian and *C. albicans* FTase, *C. albicans* GGTase-I reactivity has little dependence on the identity of the amino acid at the  $a_2$  position. In particular, there is no discrimination against peptides containing Trp and Phe at the  $a_2$  position. This result is consistent with the peptide library data (Chapter 2) where these residues were significantly enriched at the  $a_2$  position. A comparison of the impact of the  $a_2$  residue volume on peptide reactivity for each X group shows that there are some minor differences based on the nature of the X group. When X is Gln (panel C in Figure 3.7), the  $a_2$  residue volume does not affect peptide reactivity (slope  $\approx 0$ ). When X is either Met or Ser (panels D and B, respectively), there is a small positive correlation between  $a_2$  volume and reactivity (slopes  $\approx 0.0020$  and  $0.0022$ , respectively). When X is Phe, Ala, and Leu, there is a stronger correlation between  $a_2$  volume and reactivity (slopes  $\approx 0.0034$ ,  $0.044$ , and  $0.0083$ , respectively). Thus, it appears that as the X residue becomes larger and more hydrophobic, the reactivity with peptides containing larger  $a_2$  residues is enhanced.

#### *Comparison of mammalian and C. albicans GGTase-I reactivity with GCV $a_2$ L*

Biochemical reactivity and structural studies suggest that mammalian GGTase-I preferentially recognizes Leu, Ile and Val amino acids at the  $a_2$  position, although it can also accept larger and less flexible amino acids Phe, Tyr, Pro, Thr and Met, as shown by both biochemical reactivity and structural studies [10]. To rigorously compare the dependence of the reactivity of rat GGTase-I on the  $a_2$  position, the steady-state kinetics for geranylgeranylation of  $-\text{Ca}_1a_2\text{L}$  catalyzed by GGTase-I were measured (Table 3.4, Figure 3.8).

Peptide	$k_{cat}/K_M$ ( $M^{-1} s^{-1}$ )	$k_{cat}$ ( $s^{-1}$ )	$K_M$ ( $\mu M$ )
dns-GCV <sub>2</sub> AL	$3.1 \times 10^3 \pm 1.9 \times 10^2$	$0.0112 \pm 0.0004$	$3.6 \pm 0.3$
dns-GCV <sub>2</sub> FL	$2.1 \times 10^5 \pm 4.6 \times 10^4$	$0.0135 \pm 0.0004$	<0.2
dns-GCV <sub>2</sub> GL	ND		
dns-GCV <sub>2</sub> IL	ND		
dns-GCV <sub>2</sub> LL	$1.5 \times 10^6 \pm 5.8 \times 10^5$	$0.092 \pm 0.005$	<0.2
dns-GCV <sub>2</sub> ML	$2.7 \times 10^5 \pm 3.8 \times 10^4$	$0.112 \pm 0.005$	$0.42 \pm 0.07$
dns-GCV <sub>2</sub> PL	ND		
dns-GCV <sub>2</sub> SL	ND		
dns-GCV <sub>2</sub> TL	$4.9 \times 10^3 \pm 4.6 \times 10^2$	$0.026 \pm 0.002$	$5.2 \pm 0.8$
dns-GCV <sub>2</sub> VL	$1.1 \times 10^6 \pm 5.3 \times 10^5$	$0.098 \pm 0.008$	<0.2
dns-GCV <sub>2</sub> WL	$2.4 \times 10^4 \pm 6.9 \times 10^3$	$0.0016 \pm 0.0001$	<0.2
dns-GCV <sub>2</sub> YL	$1.5 \times 10^5 \pm 1.1 \times 10^4$	$0.079 \pm 0.003$	$0.54 \pm 0.06$

**Table 3.4. Steady-state kinetic parameters measured for mammalian GGTase-I.** The steady-state kinetic parameters were determined at saturating GGPP (10  $\mu M$ ) and varying peptide concentrations (0.2 – 15  $\mu M$ ) and 100 nM GGTase-I.



**Figure 3.8. Comparison of mammalian and *C. albicans* GGTase-I reactivity with dns-GCV<sub>2</sub>L peptides.** Mammalian GGTase-I data points are red-filled squares. The solid line has a slope of  $0.035 \pm 0.011$  ( $R^2 = 0.78$ ) and the dashed line has a slope of  $-0.029 \pm 0.003$  ( $R^2 = 0.98$ ). *C. albicans* GGTase-I data points are blue-filled circles. The dotted line has a slope of  $0.0088 \pm 0.0031$  ( $R^2 = 0.44$ )

The dependence of the  $k_{cat}/K_M$  value on the volume of the  $a_2$  residue in dns-GCV<sub>2</sub>L is similar to mammalian FTase when X is small. For small  $a_2$  residues  $\log k_{cat}/K_M$  increases with volume (slope = 0.035) to a maximum at  $\sim 150 \text{ \AA}^3$ . The value of  $\log k_{cat}/K_M$  then decreases as the  $a_2$  volume increases (slope = -0.029). This result is in sharp contrast to the dependence of the

reactivity of *C. albicans* GGTase-I. Thus *C. albicans* GGTase-I has unique molecular recognition properties in that it does not pay a penalty in reactivity for either small (G, A, S) or large (F, Y, W) amino acids at the a<sub>2</sub> position (Figure 3.7). Structural data indicate that *C. albicans* GGTase-I a<sub>2</sub> pocket is larger than the mammalian GGTase-I pocket (Figure 3.5), consistent with the loss of steric discrimination against peptides with large a<sub>2</sub> groups.

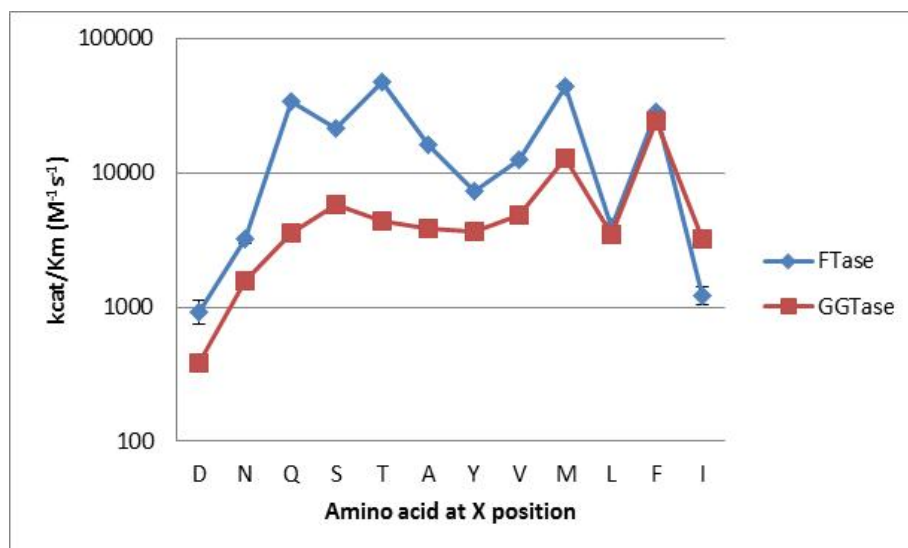
#### *Recognition of the X group by mammalian and C. albicans prenyltransferases*

To analyze recognition of the X group of the CaaX motif by *C. albicans* FTase and GGTase-I, the steady-state kinetics were measured for this reaction with dns-TKCVIX peptides (Table 3.5, Figure 3.9).

Peptide	$\Delta G_{\text{transfer}}$ (kcal/mol)	FTase $k_{\text{cat}}/K_M$ ( $M^{-1} s^{-1}$ )	GGTase-I $k_{\text{cat}}/K_M$ ( $M^{-1} s^{-1}$ )
dns-GCVIA	0.42	$1.6 \times 10^4 \pm 4.1 \times 10^2$	$3.9 \times 10^3 \pm 4.6 \times 10^1$
dns-GCVID	-1.05	$9.3 \times 10^2 \pm 1.9 \times 10^2$	$3.8 \times 10^2 \pm 1.0 \times 10^1$
dns-GCVIF	2.43	$2.9 \times 10^4 \pm 1.2 \times 10^3$	$2.4 \times 10^4 \pm 7.2 \times 10^2$
dns-GCVII	2.45	$1.2 \times 10^3 \pm 1.9 \times 10^2$	$3.2 \times 10^3 \pm 4.2 \times 10^1$
dns-GCVIL	2.31	$4.0 \times 10^3 \pm 1.9 \times 10^2$	$3.5 \times 10^3 \pm 3.8 \times 10^1$
dns-GCVIM	1.67	$4.4 \times 10^4 \pm 2.5 \times 10^3$	$1.3 \times 10^4 \pm 1.2 \times 10^3$
dns-GCVIN	-0.82	$3.2 \times 10^3 \pm 2.0 \times 10^2$	$1.6 \times 10^3 \pm 2.1 \times 10^1$
dns-GCVIQ	-0.3	$3.4 \times 10^4 \pm 4.7 \times 10^2$	$3.6 \times 10^3 \pm 5.8 \times 10^1$
dns-GCVIS	-0.05	$2.2 \times 10^4 \pm 5.8 \times 10^2$	$5.8 \times 10^3 \pm 1.0 \times 10^1$
dns-GCVIT	0.35	$4.7 \times 10^4 \pm 1.0 \times 10^3$	$4.4 \times 10^3 \pm 5.8 \times 10^1$
dns-GCVIV	1.66	$1.2 \times 10^4 \pm 5.5 \times 10^2$	$4.9 \times 10^3 \pm 3.2 \times 10^1$
dns-GCVIY	1.31	$7.4 \times 10^3 \pm 2.6 \times 10^2$	$3.6 \times 10^3 \pm 4.0 \times 10^1$

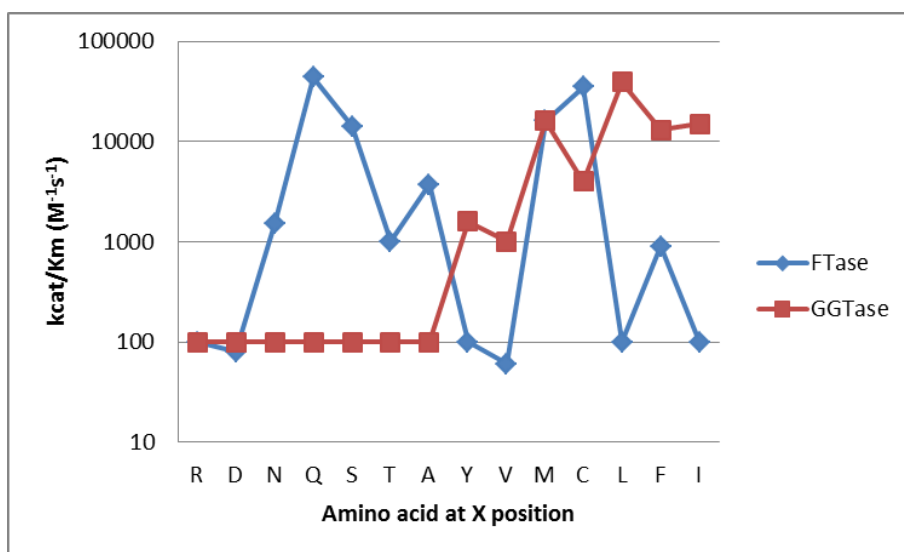
**Table 3.5. Steady-state kinetic parameters measured for *C. albicans* FTase and GGTase-I.** The steady-state kinetic parameters were determined at saturating FPP/GGPP (10  $\mu$ M) and varying peptide concentrations (0.2 – 15  $\mu$ M) with 10 or 100 nM FTase or GGTase-I, respectively.





**Figure 3.9. Differential reactivities of *C. albicans* FTase and GGTase-I as a function of hydrophobicity of the X residue.** Multiple turnover rate constant  $k_{cat}/K_M$  was measured with dns-TKCVIX peptide.

Remarkably, both *C. albicans* FTase and GGTase-I recognize a wide variety of X groups. In particular, *C. albicans* GGTase-I readily catalyzes prenylation of peptides with hydrophilic residues (Asp, Asn, Gln, Ser, and Thr) and small amino acids (Ala) at the X position. This is in contrast to the mammalian GGTase-I which does not catalyze multiple turnover reactions with peptides that contain these residues (Figure 3.10).



**Figure 3.10. Differential reactivities of mammalian FTase and GGTase-I as a function of hydrophobicity of the X residue.** Multiple turnover rate constant  $k_{cat}/K_M$  was measured with dns-TKCVIX peptide. Figure adapted from [20].

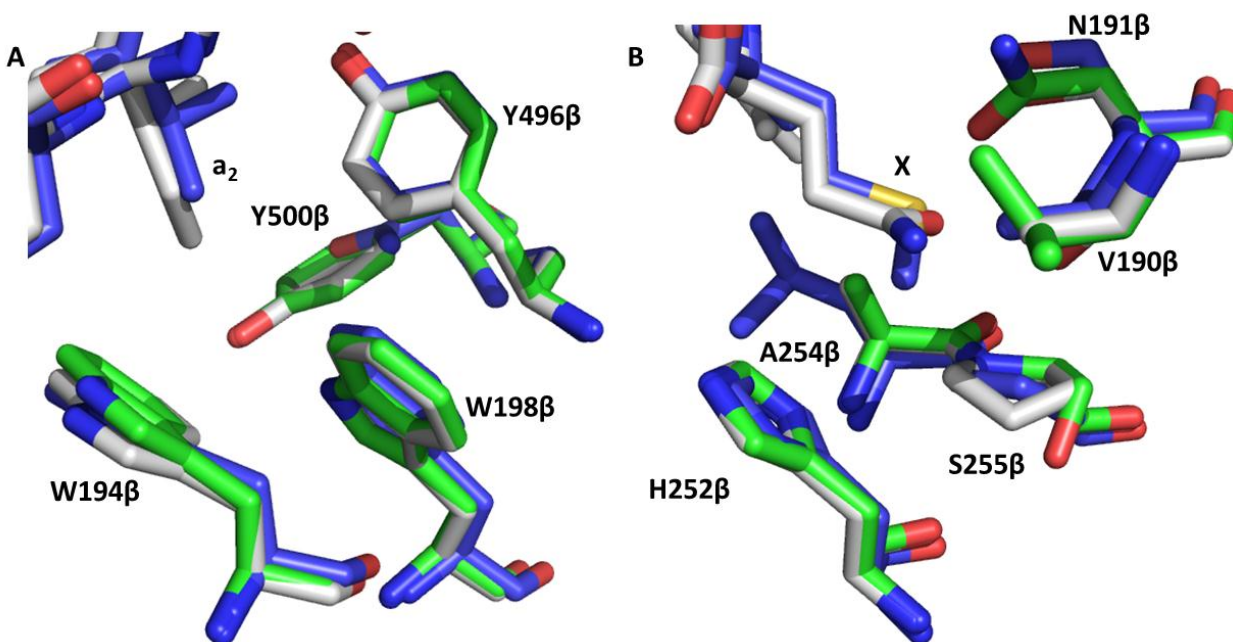
Previously, Hartman and coworkers showed that the MTO reactivity of peptides with mammalian FTase and GGTase-I in general correlates with the hydrophobicity and/or volume of the X group of the peptide [20]. Specifically, it was shown that  $k_{cat}/K_M$  values decrease for FTase and increase for GGTase-I as the hydrophobicity or volume of the C-terminal X residue increases. Peptides that contain X residues that are less hydrophobic than tyrosine become single turnover substrates for mammalian GGTase-I. Since these peptides are capable of MTO catalysis with *C. albicans* GGTase-I, and in general *C. albicans* GGTase-I has very few substrates with STO reactivity, it suggests that it is the rate of product release that differentiates the two GGTases, with the exit groove possibly playing a role in faster product release in the yeast enzyme (discussed in detail in Chapter 2).

## DISCUSSION

### *Mammalian and C. albicans FTase recognition of $-a_2X$*

The  $a_2$  binding site in mammalian FTase is formed by residues W102 $\beta$ , W106 $\beta$ , Y361 $\beta$ , and Y365 $\beta$ , as well as the third isoprenoid unit of the FPP co-substrate. These residues are conserved in FTase from the *C. neoformans* and *C. albicans* fungal pathogens, with the sole exception of Asn substitution at Y365 in the *C. neoformans* enzyme. The X binding pocket of mammalian and *C. neoformans* FTase is mainly conserved, with the exception of an A151 $\beta$  to L141 $\beta$  substitution in the *C. neoformans* enzyme. However, the  $C_\alpha$  positions of the residues lining the X residue pocket diverge slightly in *C. neoformans* relative to mammalian FTase, with an aggregate effect of altering the shape of the X binding cavity, possibly leading to differences in substrate recognition [12]. The differences between the sequence of the mammalian and *C. albicans* FTase X residue binding pockets are more pronounced, with A98 $\beta$  being replaced by

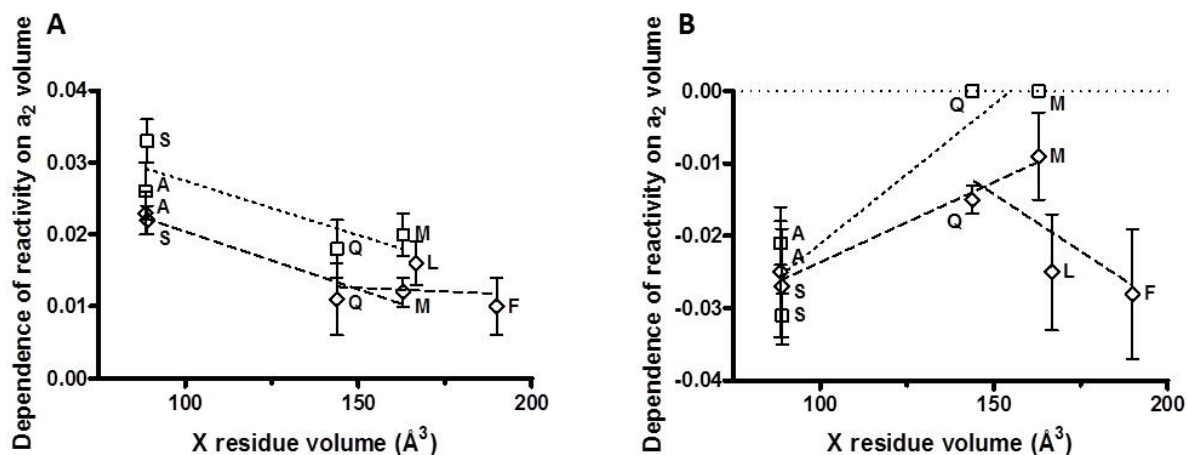
V190 $\beta$ , S99 $\beta$  by N191 $\beta$ , and P152 $\beta$  by S255 $\beta$  in *C. albicans* relative to mammalian FTase. Overlays of the a<sub>2</sub> and X binding pockets of FTase from these three different organisms are shown in Figure 3.11.



**Figure 3.11. Structure of mammalian, *C. albicans*, and *C. neoformans* FTase a<sub>2</sub> and X binding pockets.** *C. neoformans* FTase (blue, 3Q75), mammalian FTase (grey, 1TN6), *C. albicans* FTase (green, homology model). Residue numbers are for *C. albicans* FTase. Panel A shows the structure of the a<sub>2</sub> binding pocket. Panel B shows the structure of the X binding pocket.

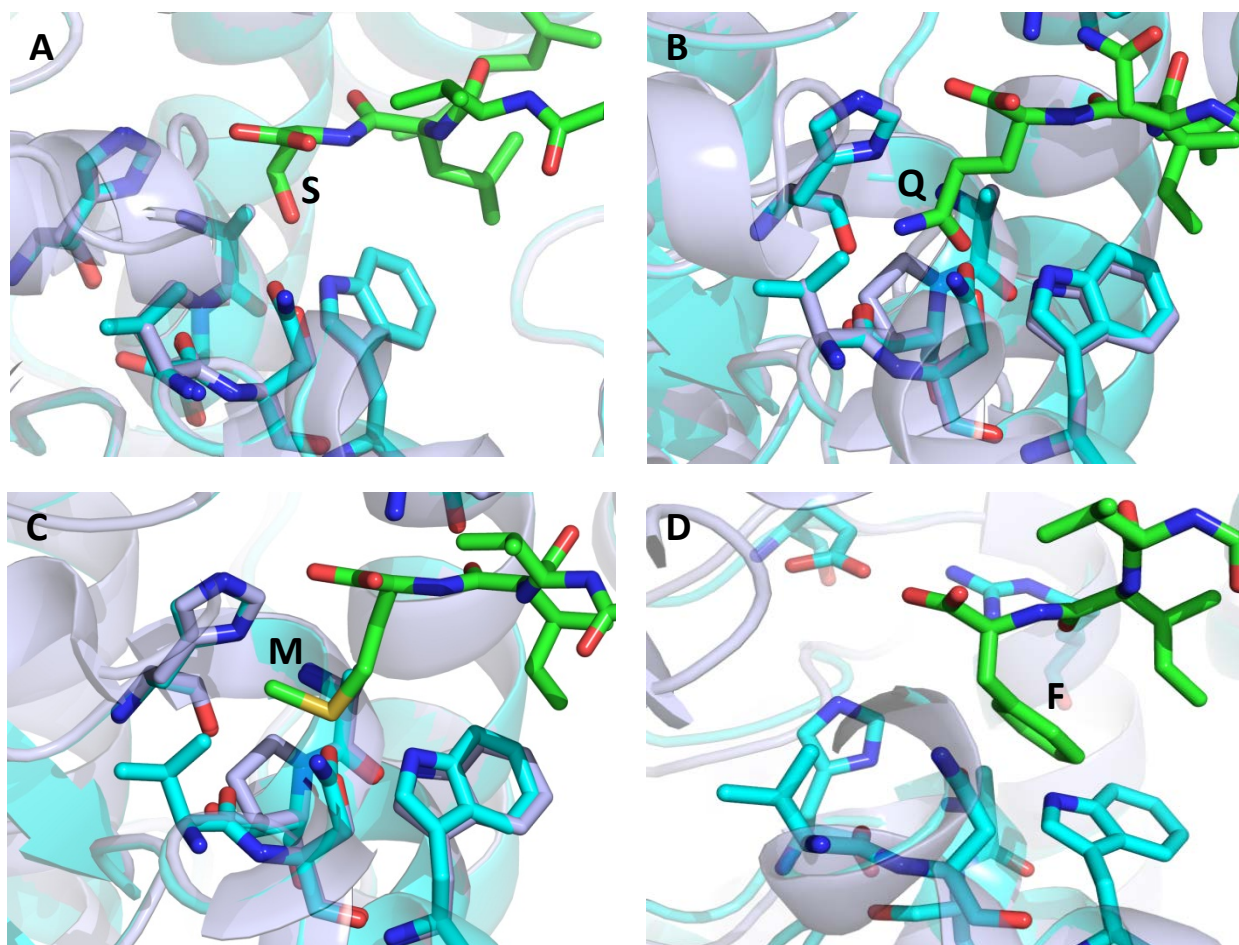
Previous studies carried out in our lab showed a positive correlation of peptide reactivity and a<sub>2</sub> residue volume as the a<sub>2</sub> residue volume increases from ~60 to 150 Å<sup>3</sup>, after which point a negative correlation was observed [6]. In addition, the dependence on the volume of the a<sub>2</sub> residue differed based on the identity of the amino acid at the X position, indicating cooperativity between these two residues. This trend was also observed for *C. albicans* FTase (Figure 3.6). Correlations between the a<sub>2</sub> residue volume and peptide reactivity were expressed as slopes calculated from these reactivity plots (Figure 3.6 and Figure 3.7). The slopes describing both positive and negative interactions with the a<sub>2</sub> residue are the same within error for mammalian and *C. albicans* FTase. This suggests that when the X residue is small, both enzymes gain the

same amount of relative reactivity as the  $a_2$  residue volume increases up to  $\sim 150 \text{ \AA}^3$ , and then lose the same amount of reactivity as the  $a_2$  residue volume increases beyond this size.



**Figure 3.12. Mammalian and *C. albicans* FTase linear correlations between  $a_2$  residue volume and peptide reactivity for peptides with varying X residues.** A) Positive interactions for mammalian (square) and *C. albicans* (diamond) FTase; B) Negative interactions for mammalian (square) and *C. albicans* (diamond) FTase. Slopes for mammalian FTase were taken from [6].

When the X residue = Q or M, mammalian and *C. albicans* FTase both show a weaker dependence of reactivity on the  $a_2$  residue volume. As shown in Figure 3.12, increase in the side chain volume of the X residue correlates with a decrease in enzyme sensitivity to the  $a_2$  residue volume. Larger peptides could form contacts within the FTase•peptide complex that are absent when the X residue is small, and thus drive peptide reactivity. Crystal structures of peptide substrates terminating in Ser, Glu and Met bound in the mammalian FTase active site are shown in Figure 3.13 (panels A-C). When X = Ser, there is a significant portion of the X binding pocket in FTase that is unoccupied when compared to peptides that contain X = Gln or Met. It is quite clear that larger residues are able to extend farther into the peptide binding site of FTase and form positive interactions that could dominate substrate selectivity.



**Figure 3.13. Comparison of FTase interactions with peptides that have X = S, Q, M, and F.** A) FTase with CVLS peptide shows unoccupied space between Ser and FTase (PDB ID 1TN8) B) FTase with CNIQ peptide shows Gln extends into the X binding pocket (PDB ID 1TN6) C) FTase with CVIM peptide shows Met extends into the X binding pocket (PDB ID 1D8D) D) FTase with CVIF peptide shows Phe side chain shifts towards the  $a_2$  binding pocket (PDB ID 1TNB) Mammalian FTase is shown in grey, *C. albicans* FTase in cyan, and peptide substrate is green.

As the residue volume of X is increased further (X = L or F), when the  $a_2$  residue volume is small, the X residue continues to dominate peptide reactivity (Figure 3.12, panel A). It is noteworthy to point out that the positive correlation slope when X = L is somewhat larger than those when X = Q, M, or F, shifting the interaction toward more  $a_2$  influence. Although Leu has a larger residue volume than Met or Gln ( $166.7 \text{ \AA}^3$  versus  $162.9 \text{ \AA}^3$  versus  $143.8 \text{ \AA}^3$ ), Leu is a branched amino acid while Met and Gln are not, and thus Met and Gln can reach deeper into the X binding pocket and potentially contribute more favorable interactions than Leu.

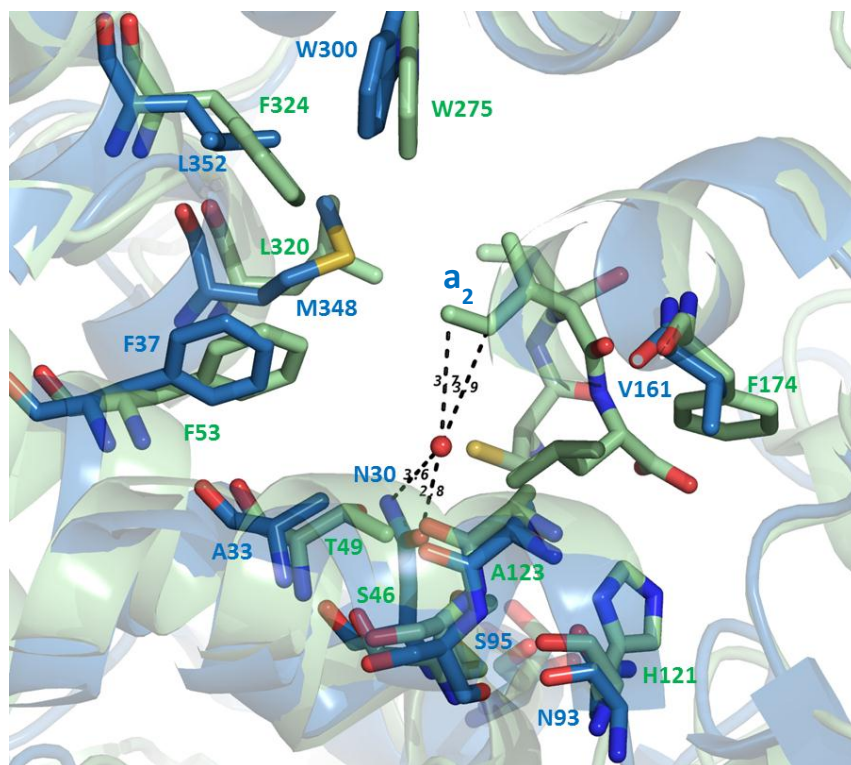
The combination of large side chains at both the  $a_2$  and X positions is unfavorable (Figure 3.12, panel B), possibly causing movement of the X residue such that it is no longer able to maintain all of the favorable contacts with the X binding pocket. A structure of mammalian FTase with CVIF peptide (Figure 3.13, panel D) shows that this peptide binds to FTase with the Phe side chain in an alternative hydrophobic binding site close to the  $a_2$  binding pocket while two solvent molecules move in to occupy the “normal” X binding site [10]. This could mean that while small  $a_2$  residues are able shift to accommodate large X groups, larger  $a_2$  residues are unable to do so and result in a reactivity penalty. Although this analysis does not directly explain how the interplay between the  $a_2$  and X residue recognition results in differences in contributions to substrate reactivity by various interactions, it does suggest that these enzymes have multiple modes for binding a wide variety of  $a_2X$  combinations.

#### *Mammalian and C. albicans GGTase-I recognition of $-a_2X$*

Peptide library studies with *C. albicans* GGTase-I revealed that this enzyme has a preference for large aromatic amino acids Phe and Trp at the  $a_2$  position. This was an unexpected observation given that studies of mammalian prenyltransferases have shown that typically  $a_2$  residues are medium-sized hydrophobic amino acids such as Ile, Leu and Val. Although structural evidence indicates that the  $a_2$  binding site can also accommodate other amino acids, including Phe, Tyr, Pro, Thr, and Met, it has been shown that depending on the X group, there is a steric discrimination against large amino acids at the  $a_2$  position [8,16,21]. However, comparison of mammalian and *C. albicans* GGTase-I  $a_2$  binding pockets (Figure 3.5) showed that the yeast enzyme  $a_2$  binding site is  $\sim 25 \text{ \AA}^3$  larger than the mammalian as a result of two amino acid substitutions in the *C. albicans* enzyme: M384 $\beta$  for L320 $\beta$  and L352 $\beta$  for F324 $\beta$ . Reactivity with GCV $a_2X$  peptide series where  $a_2$  residues were hydrophobic amino acids of

varying sizes showed either no dependence of peptide reactivity on the  $a_2$  residue size, or a slight positive correlation between size and reactivity (Figure 3.7). It appears that as the X residue becomes larger and more hydrophobic, the reactivity with peptides containing larger  $a_2$  residues is enhanced. In contrast, mammalian GGTase-I showed both a positive and a negative correlation with  $a_2$  residue volume (Figure 3.8), a pattern similar to FTases. Thus, the contribution of a larger  $a_2$  binding cavity in *C. albicans* GGTase-I to CaaX sequence recognition was confirmed by the lack of a discrimination against peptides with large  $a_2$  residues.

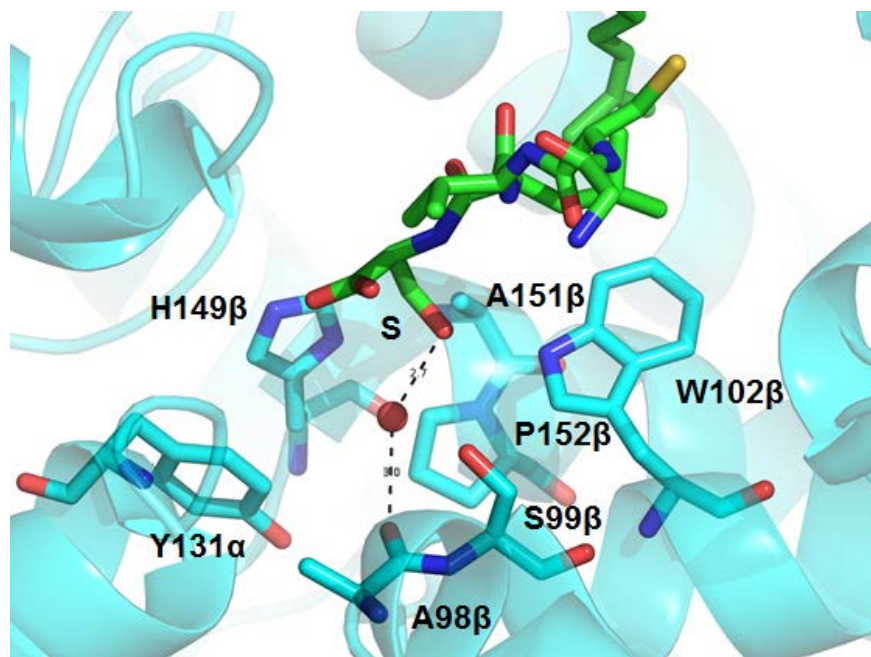
In addition to accepting large  $a_2$  residues, *C. albicans* GGTase-I also does not discriminate against small  $a_2$  side chains. The  $a_2$  binding pocket of mammalian GGTase-I is exclusively hydrophobic, completely buried in the active site, and the  $a_2$  residue forms hydrophobic interactions and van der Waals contact with the enzyme. In contrast, *C. albicans* GGTase-I  $a_2$  binding pocket, while still primarily hydrophobic, contains a buried water molecule that is located 3.7 Å away from the  $a_2$  residue in the substrate, and 2.8 and 3.6 Å away from oxygen and nitrogen atoms of N30 $\beta$  side chain amide, as shown in Figure 3.14. Thus, when the  $a_2$  residue is small, this water molecule can potentially fill unoccupied space and continue to make hydrogen bonds in the  $a_2$  pocket and thus help the enzyme to overcome the penalty for insufficient contacts with smaller substrates. Conversely, it is also possible that the  $a_2$  residue does not form specific contacts with *C. albicans* GGTase-I, and remains mainly hydrated, thus contributing to the lack of variability in reactivity of peptides with different  $a_2$  residues.



**Figure 3.14. Comparison of  $a_2$  binding pockets of mammalian and *C. albicans* GGTase-I.** Mammalian (green) and *C. albicans* (blue) GGTase-I with CVIL peptide, where mammalian  $a_2$  pocket is exclusively hydrophobic and buried deep in the enzyme, while *C. albicans* enzyme has a buried water molecule (red) that can continue to make hydrogen bonds in this pocket when  $a_2$  residue is small, and thus no reactivity penalty is observed with small  $a_2$  residues.

Unfortunately this analysis could be somewhat misleading as *C. albicans* GGTase-I structure only contains GGPP prenyl donor and no peptide, and this water molecule may be absent once a peptide is bound to the enzyme. However, a similar strategy is utilized by mammalian FTase in its X binding pocket; FTase harbors a buried water molecule in the X binding site, as shown in Figure 3.15, and thus it can accommodate small and polar residues while GGTase-I favors hydrophobic X groups. These data strengthen the hypothesis that a buried water molecule in *C. albicans* GGTase-I could contribute to its ability to efficiently prenylate peptides with small  $a_2$  residues.



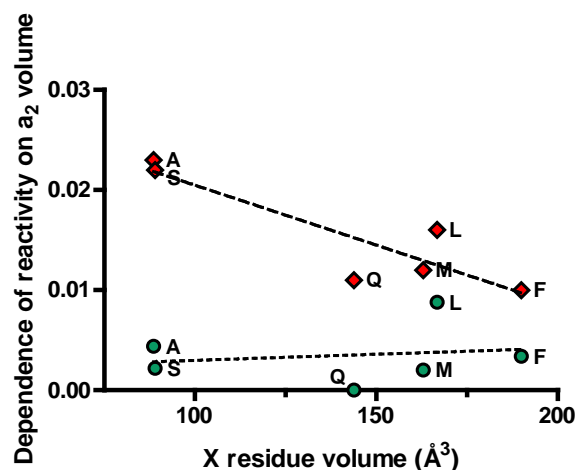


**Figure 3.15. FTase X binding pocket.** When the substrate X residue is Ser, it is accompanied by a water molecule hydrogen bonded to the Ser hydroxyl group and the adjacent carbonyl oxygen atom of A98 $\beta$ . (PDB ID 1TN8)

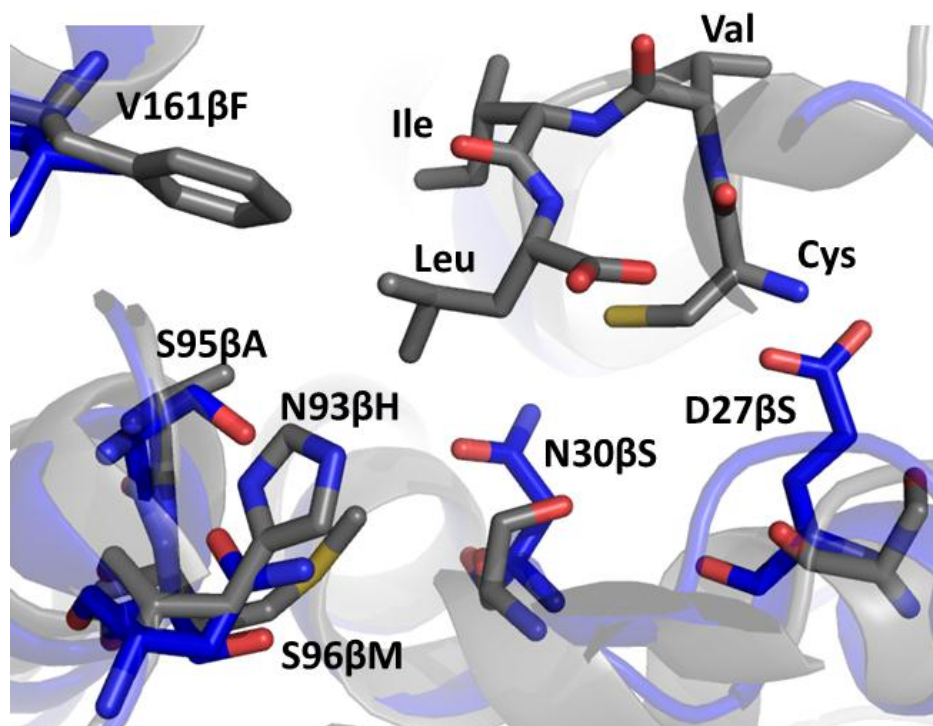
*Recognition of X group by mammalian and C. albicans prenyltransferases*

*C. albicans* GGTase-I catalyzes MTO reactions with a wider range of X residues than mammalian enzyme (Figure 3.10 versus Figure 3.9), in particular those containing hydrophilic and small side chains. In contrast to *C. albicans* FTase and mammalian GGTase-I, *C. albicans* GGTase-I is also largely insensitive to the  $a_2$  and X residue volumes, as shown in Figure 3.16. One possibility that can account for this lack of specificity is the apparent lack of an exit groove in the yeast enzyme. However, comparison of residues that form the X residue binding site in mammalian and *C. albicans* GGTase-I also reveals that the yeast enzyme pocket is more hydrophilic in nature, as can be seen in their superposition in Figure 3.17. The X pocket of mammalian GGTase-I consists of S42 $\beta$ , S46 $\beta$ , H121 $\beta$ , A123 $\beta$ , M124 $\beta$ , and F174 $\beta$ , while the corresponding residues in *C. albicans* GGTase-I are D27 $\beta$ , N30 $\beta$ , N93 $\beta$ , S95 $\beta$ , S96 $\beta$ , and

V161 $\beta$ , where every amino acid in the yeast enzyme is more hydrophilic than the corresponding one in the mammalian GGTase-I.

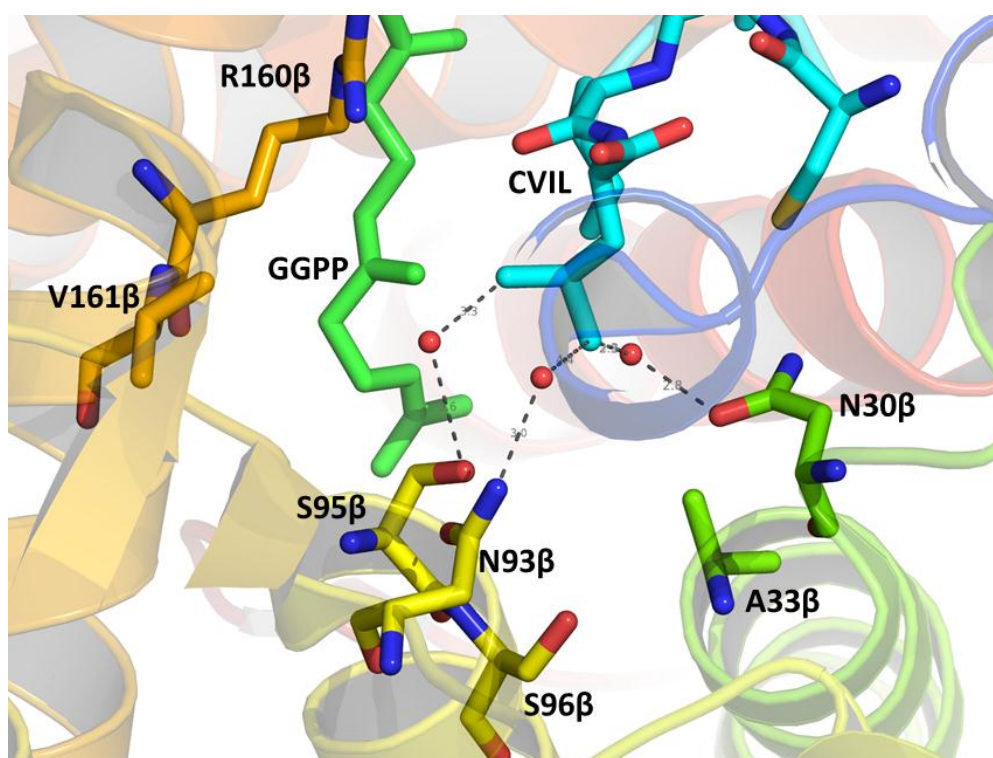


**Figure 3.16.** *C. albicans* FTase and GGTase-I reactivity as a function of  $a_2$  and X residue volumes. FTase reactivity is dependent on both  $a_2$  and X residue volume. GGTase-I reactivity is largely insensitive to  $a_2$  and X residue volume. FTase is shown with red diamonds, and GGTase-I is shown with green circles.



**Figure 3.17.** Mammalian and *C. albicans* GGTase-I X residue binding pocket. *C. albicans* GGTase-I (blue, 3DRA) residues are labeled and numbered, mammalian GGTase-I complexed with CVIL peptide (grey, 1N4Q).

In addition to differences in hydrophobicity of the X residue pockets, solvent molecules can be seen in the *C. albicans* GGTase-I X binding pocket, while mammalian GGTase-I X binding pocket is not accessible to solvent. In particular, three water molecules are found in *C. albicans* GGTase-I that are within hydrogen bonding distance of the N30 $\beta$ , N93 $\beta$ , and S95 $\beta$  amino acid side chains and the a<sub>2</sub> residue of the substrate (Figure 3.18). Their presence serves both to fill unoccupied space for small X groups, and facilitates hydrophilic interactions for polar residues, thus providing *C. albicans* GGTase-I the ability to accept these amino acids at the X position. Presumably these water molecules are displaced when substrates with larger and more hydrophobic X groups bind to the enzyme.



**Figure 3.18. X group contacts in *C. albicans* GGTase-I.** *C. albicans* GGTase-I is modeled with CVIL peptide substrate. X group binding pocket consists of N30 $\beta$ , A33 $\beta$ , N93 $\beta$ , S95 $\beta$ , S96 $\beta$ , R160 $\beta$ , V161 $\beta$  and GGPP isoprenoid. Water molecules (red spheres) are found in the *C. albicans* GGTase-I binding pocket, and can hydrogen bond with side chains of N30 $\beta$ , N93 $\beta$ , and S95 $\beta$  amino acid. Presence of water molecules can accommodate small amino acids by filling unoccupied space, and polar amino acids by creating a hydrophilic environment.

While solvent accessibility of the peptide binding site could be a valid explanation for its better tolerance of hydrophilic substrates, it should be noted that the *C. albicans* GGTase-I structure contains no bound peptide, and thus could be misleading. In an attempt to alleviate this concern, the binary complex of mammalian GGTase-I with GGPP (PDB ID 1N4P) was compared to both *C. albicans* GGTase-I (PDB ID 3DRA) as well as the ternary complex with 3-aza-GGPP and CaaX peptide (PDB ID 1N4Q). Additional water molecules were not observed in the structure of mammalian GGTase-I without bound peptide. Furthermore, *C. albicans* GGTase-I contained more water molecules in the putative peptide binding site than the mammalian enzyme, both in the absence of bound peptide. This analysis suggests that higher water content of *C. albicans* GGTase-I peptide binding site is not due to the absence of a bound peptide substrate but rather an intrinsic property of this enzyme. Therefore, the additional water molecules may play an important role in the broadened selectivity of *C. albicans* GGTase-I.

In summary, these data demonstrate both that *C. albicans* FTase and GGTase-I have broader substrate selectivity with more overlap between the two prenyltransferases in comparison to their mammalian homologs. This overlap could help explain the apparent resistance by *C. albicans* organism to FTase-only inhibitors, and only morphological abnormalities with no growth inhibition that results from GGTase-I inhibition. These biochemical data are in agreement with the *in vivo* findings of GGTase-I being the more indispensable enzyme, as *C. albicans* GGTase-I is able to prenylate a wider range of CaaX sequences than FTase. These findings support the hypothesis of a dual-enzyme inhibitor for antimicrobial agents against *C. albicans* pathogen.

## REFERENCES

1. Zhang FL, Casey PJ: **Protein prenylation: molecular mechanisms and functional consequences.** *Annu Rev Biochem* 1996, **65**:241-269.
2. Benetka W, Koranda M, Eisenhaber F: **Protein prenylation: An (almost) comprehensive overview on discovery history, enzymology, and significance in physiology and disease.** *Monatshefte fur Chemie* 2006, **137**:1241-1281.
3. Fu HW, Casey PJ: **Enzymology and biology of CaaX protein prenylation.** *Recent Prog Horm Res* 1999, **54**:315-342; discussion 342-313.
4. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA: **Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities.** *J Mol Biol* 2010, **395**:176-190.
5. Krzysiak AJ, Aditya AV, Hougland JL, Fierke CA, Gibbs RA: **Synthesis and screening of a CaaL peptide library versus FTase reveals a surprising number of substrates.** *Bioorg Med Chem Lett* 2010, **20**:767-770.
6. Hougland JL, Lamphear CL, Scott SA, Gibbs RA, Fierke CA: **Context-dependent substrate recognition by protein farnesyltransferase.** *Biochemistry* 2009, **48**:1691-1701.
7. Long SB, Casey PJ, Beese LS: **Reaction path of protein farnesyltransferase at atomic resolution.** *Nature* 2002, **419**:645-650.
8. Taylor JS, Reid TS, Terry KL, Casey PJ, Beese LS: **Structure of mammalian protein geranylgeranyltransferase type-I.** *Embo J* 2003, **22**:5963-5974.
9. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F: **Towards complete sets of farnesylated and geranylgeranylated proteins.** *PLoS Comput Biol* 2007, **3**:e66.
10. Reid TS, Terry KL, Casey PJ, Beese LS: **Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity.** *J Mol Biol* 2004, **343**:417-433.
11. Maurer-Stroh S, Eisenhaber F: **Refinement and prediction of protein prenylation motifs.** *Genome Biol* 2005, **6**:R55.
12. Hast MA, Nichols CB, Armstrong SM, Kelly SM, Hellinga HW, Alspaugh JA, Beese LS: **Structures of Cryptococcus neoformans Protein Farnesyltransferase Reveal Strategies for Developing Inhibitors That Target Fungal Pathogens.** *Journal of Biological Chemistry* 2011, **286**:35149-35162.
13. Kelley LA, Sternberg MJ: **Protein structure prediction on the Web: a case study using the Phyre server.** *Nat Protoc* 2009, **4**:363-371.
14. Hast MA, Beese LS: **Structure of Protein Geranylgeranyltransferase-I from the Human Pathogen Candida albicans Complexed with a Lipid Substrate.** *Journal of Biological Chemistry* 2008, **283**:31933-31940.
15. Cassidy PB, Dolence JM, Poulter CD: **Continuous fluorescence assay for protein prenyltransferases.** *Methods Enzymol* 1995, **250**:30-43.
16. Long SB, Hancock PJ, Kral AM, Hellinga HW, Beese LS: **The crystal structure of human protein farnesyltransferase reveals the basis for inhibition by CaaX tetrapeptides and their mimetics.** *Proc Natl Acad Sci U S A* 2001, **98**:12948-12953.
17. Riddles PW, Blakeley RL, Zerner B: **Reassessment of Ellman's reagent.** *Methods Enzymol* 1983, **91**:49-60.

18. Fersht A: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*: W. H. Freeman; 1998.
19. Zamyatnin AA: **Protein volume in solution**. *Prog Biophys Mol Biol* 1972, **24**:107-123.
20. Hartman HL, Hicks KA, Fierke CA: **Peptide specificity of protein prenyltransferases is determined mainly by reactivity rather than binding affinity**. *Biochemistry* 2005, **44**:15314-15324.
21. Dolence JM, Steward LE, Dolence EK, Wong DH, Poulter CD: **Studies with recombinant *Saccharomyces cerevisiae* CaaX prenyl protease Rce1p**. *Biochemistry* 2000, **39**:4096-4104.

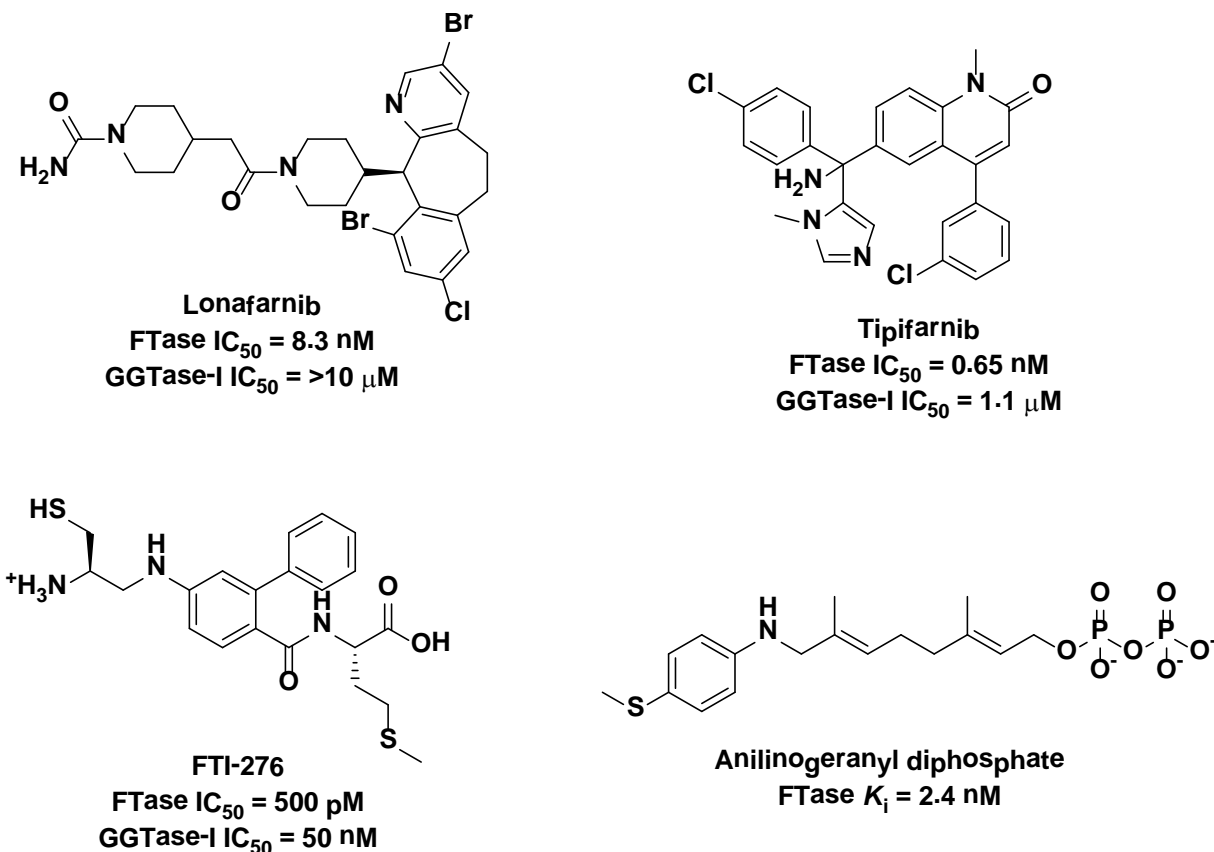
## CHAPTER FOUR

### INVESTIGATION OF SMALL MOLECULE INHIBITORS OF MAMMALIAN AND *CANDIDA ALBICANS* PRENYLTRANSFERASES

#### INTRODUCTION

In the last decade, a number of FTase inhibitors (FTIs) have been evaluated in clinical trials as cancer therapeutics [1], and a selection of the most advanced compounds is shown in Figure 4.1. Based on the biology of Ras farnesylation and structural studies of FTase-ligand complexes, rational development of FTIs was undertaken. These agents can be classified as: 1) peptidomimetics that resemble and compete with the CaaX motif of natural protein substrates for binding to FTase, 2) FPP analogs, 3) bisubstrate analogs that incorporate the structural motif of both FPP and CaaX, and 4) non-peptidomimetic small molecules [2,3]. FTI-276 is an example of a peptidomimetic FTI incorporating a methionine mimic at the C terminus, and it is ~100-fold more potent against FTase than GGTase-I *in vitro* [4]. FPP analogs anilino geranyl diphosphates potently inhibit FTase with  $K_i$  values in the 2.4 – 18 nM range [5], and 3,7-disubstituted FPP analogs can be both inhibitors and alternative substrates for FTase [6]. Bisubstrate inhibitors that are proposed to mimic the transition state and include an imidazole central scaffold moiety to interact with the zinc atom showed modest activity against *S. cerevisiae* FTase, with  $IC_{50}$  values in the 35 – 400  $\mu$ M range [7]. The most potent non-peptidomimetic inhibitors, lonafarnib,

tipifarnib, and BMS-214662 were identified by screening compound libraries and are peptide competitive inhibitors.

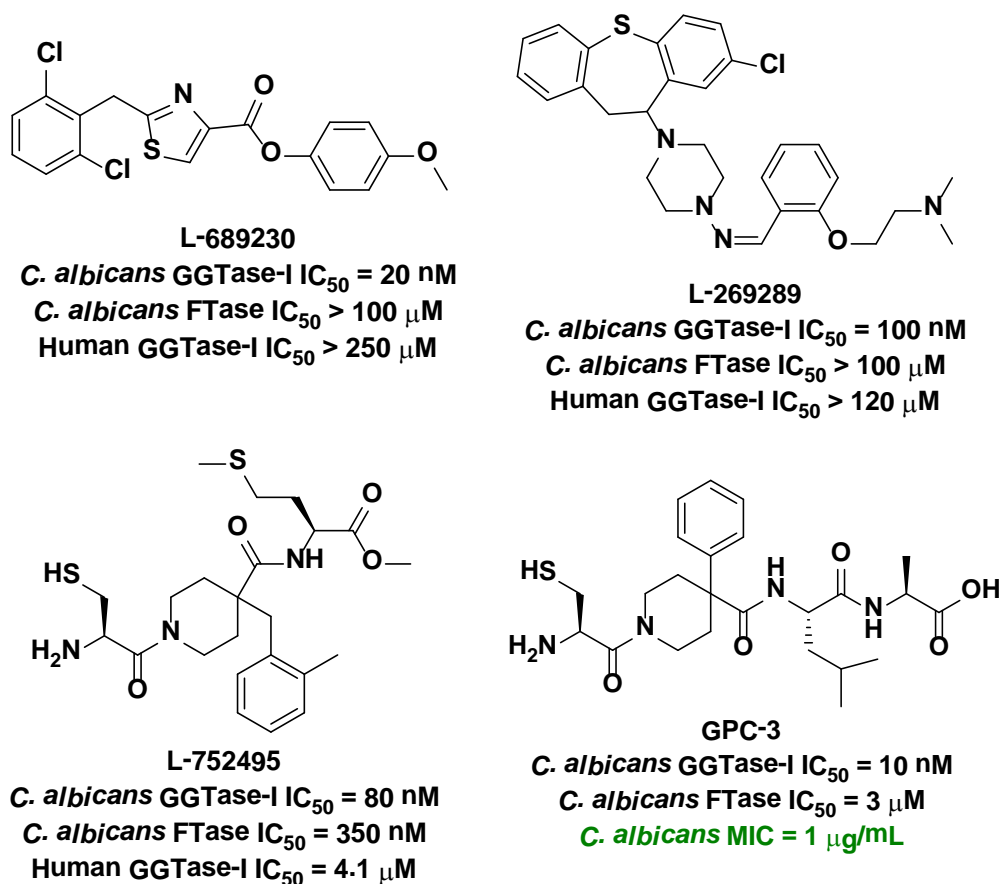


**Figure 4.1. Selected FTase and GGTase-I inhibitors.** Lonafarnib (Sarasar<sup>®</sup>) has been evaluated in numerous clinical trials for cancer, and is currently being evaluated as a potential treatment for progeria. Tipifarnib (Zarnestra<sup>®</sup>) is being evaluated in patients with acute myeloid leukemia. FTI-276 is a peptidomimetic ~100-fold more selective for FTase. Anilinoogeranyl diphosphate is an FPP analog and a potent FTase inhibitor.

Although the majority of FTIs and GGTIs developed to date have been targeted against mammalian prenyltransferases, several reports of *C. albicans*-specific inhibitors have been published. Smalera and coworkers report identification of eight structurally diverse *C. albicans* GGTase-I inhibitors using a scintillation proximity assay-based high-throughput screen [8]. These compounds have IC<sub>50</sub> values ranging from 20 to 260 nM against *C. albicans* GGTase-I, and the selectivity for the yeast enzyme compared to human GGTase-I ranged from a modest 7-fold to a dramatic > 12500-fold (compound L-689230 in Figure 4.2). However, these potent *C.*



*albicans* GGTase-I inhibitors do not inhibit *C. albicans* growth, consistent with the determination that GGTase-I is not essential for *C. albicans* viability [9]. Inhibition of both FTase and GGTase-I appears to be necessary to achieve *in vivo* antifungal activity. L-752495 was the only compound to have dual specificity, but it also inhibited human GGTase-I, making it a poor candidate for further development [8]. A similar lack of antifungal activity was observed with another potent inhibitor of *C. albicans* GGTase-I ( $IC_{50} = 120$  nM) that was 2500-fold less potent against *C. albicans* FTase, supporting the hypothesis that a dual FTase/GGTase-I inhibition is necessary for antimycotic therapy [9-11].



**Figure 4.2. *C. albicans* FTase and GGTase-I inhibitors.** L-689230 and L-269289 are selective for *C. albicans* GGTase-I but lack activity against *C. albicans* FTase. L-752495 is active against both *C. albicans* prenyltransferases, but it also inhibits mammalian FTase. GPC-3 is the only compound that shows *in vivo* anti-fungal activity.

To clarify the need for dual FTase/GGTase-I inhibition, Murthi and coworkers synthesized a peptidomimetic compound GPC-3 that was active against both *C. albicans* prenyltransferases, with  $IC_{50} = 10$  nM and  $IC_{50} = 3$   $\mu$ M for GGTase-I and FTase, respectively [12]. This compound had *in vivo* activity against *C. albicans* with a minimal inhibitory concentration (MIC) of 1  $\mu$ g/mL. This result suggests that a dual inhibitor is required to achieve *in vivo* efficacy. However, complete inhibition of FTase by this compound was likely not achieved as its *in vitro* activity against FTase is only 3  $\mu$ M, and FTase substrate Ras1p localized to the membrane fraction. Thus, the question of dual FTase/GGTase-I inhibition currently remains unresolved.

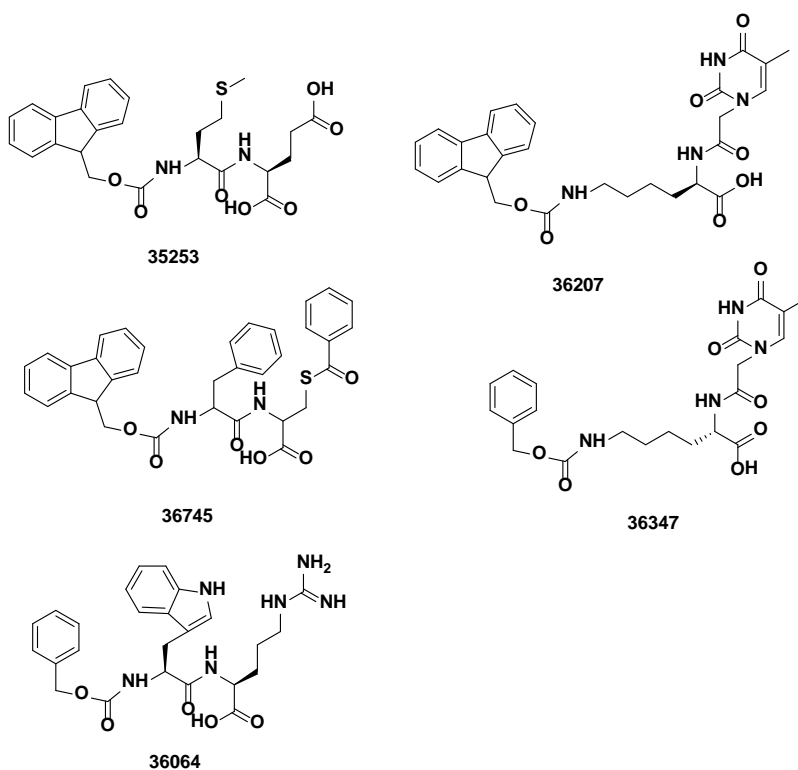
The overall aim of this work was to identify new small molecules that inhibit mammalian and/or *C. albicans* prenyltransferases, and to determine if any of the active molecules had selectivity for FTase *versus* GGTase-I or human *versus* *C. albicans* enzymes. The goal was to find potential starting scaffolds or chemical features that could be used to generate new lead compounds for further inhibitor development. The approach undertaken was to perform docking studies of compounds with mammalian FTase to predict inhibitory activity (in collaboration with the Merz group). These small molecules were then tested for inhibition of prenyltransferases in an *in vitro* assay. Initial studies identified an Fmoc-protected dipeptide with excellent activity and selectivity against *C. albicans* FTase. Further kinetic studies revealed a mixed mode of inhibition with these compounds binding to both free enzyme, E•FPP, and a ternary complex of the enzyme. Given these initial results, Fmoc-capped dipeptides consisting of all 20 natural amino acids were scored using a docking algorithm (in collaboration with the Merz group) based on their complementarity to the prenyltransferase peptide binding site, and the top twenty dipeptides were tested *in vitro*. Based on these studies, metal-coordinating ligands such as

carboxylates, thiolates, and imidazoles were found to enhance *in vitro* potency of these inhibitors.

## EXPERIMENTAL PROCEDURES

### *Selection of small molecule prenyltransferase inhibitors*

Prof. Kenneth Merz and Dr. Mark Benson (University of Florida) performed a virtual screen [13] of compounds found in the Florida Center for Heterocyclic Compounds (maintained and curated by Prof. Alan Katritzky, University of Florida) using the rat FTase structure. Five compounds (Figure 4.3) with the best docking scores were chosen for further evaluation. These compounds were a gift from Prof. Katritzky. All compounds were dissolved in dimethyl sulfoxide (DMSO) to a stock concentration of 20 mM and stored at -20°C.



**Figure 4.3. Small molecules selected in the initial virtual screen against mammalian FTase.** These five compounds were selected for evaluation based on their *in silico* docking scores. Carboxylate functional group was designed to coordinate zinc metal ion in the FTase active site.

### *Prenyltransferase inhibition assays*

*C. albicans* prenyltransferases were purified as described in Chapter 2. Mammalian prenyltransferases were purified by Corissa L. Lamphear. Inhibition of rat and *C. albicans* FTase and GGTase-I was measured using a continuous assay with dansylated peptide substrate monitoring the increase in dansyl group fluorescence (same as in Chapters 2 and 3). The standard assay buffer included: 50 mM HEPES pH 7.8, 5 mM TCEP and 5 mM MgCl<sub>2</sub>. Inhibitors were assayed at 10 concentrations (obtained by 2-fold serial dilutions) with 400 μM as the highest concentration. At first reactions were initiated with enzyme; subsequently the order of addition was switched and the compound was incubated with the enzyme in the presence of the prenyl donor for ~20 minutes, and reaction was initiated by adding peptide. FTase was assayed with 10 μM FPP and 5 μM dns-GCVLS; GGTase-I was assayed with 10 μM GGPP and 5 μM dns-GCVLL. The final concentration of DMSO was maintained at 2% for all inhibitor concentrations and negative and positive controls. The MAX and MIN initial rates for each enzyme were obtained using a DMSO blank and no enzyme, respectively, and the percent inhibition was calculated using these values according to equation (1) where Act is the initial rate at a given inhibitor concentration:

$$\% inhibition = \frac{MAX - Act}{MAX - MIN} \quad (\text{Eq. 1})$$

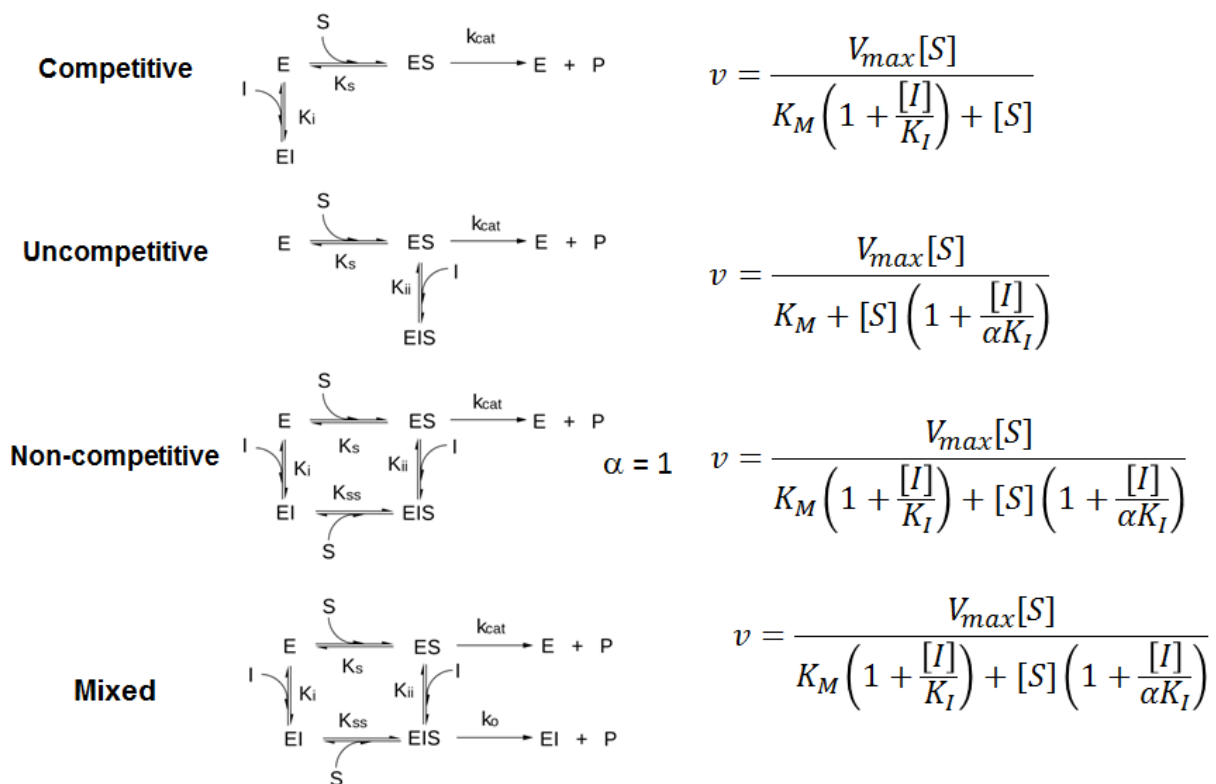
IC<sub>50</sub> values were calculated using a two parameter dose response model where n is the hill coefficient:

$$\% inhibition = \frac{100}{1 + \left(\frac{IC_{50}}{[I]}\right)^n} \quad (\text{Eq. 2})$$

All data analysis was performed using Prism GraphPad software.

### *Mode of inhibition of C. albicans FTase by compound 36745*

To determine the mode of inhibition of 36745, the concentration of FPP, peptide and inhibitor were varied. [dns-GCVLS] = 1, 2, 4, 6, 8, 10, 15, 20, 25, and 30  $\mu\text{M}$  ( $K_{\text{M,dns-GCVLS}} = 8.2 \mu\text{M}$ ); [FPP] = 0.075, 0.1, 0.15, 0.2, 0.3, 0.4, 0.6, 0.8, 1.2, 1.6, 2.4, and 3.2  $\mu\text{M}$  ( $K_{\text{M,FPP}} = 1.5 \mu\text{M}$ ); and [36745] = 50, 100, 200, 400, 600, 800, and 1000 nM. To simplify experimental setup and analysis, only one substrate was varied at a time, while the other substrate was held constant at a saturating concentration ([FPP] = 10  $\mu\text{M}$  or [dns-GCVLS] = 5  $\mu\text{M}$ ), thus resulting in two 2-dimensional matrices instead of a 3-dimensional matrix. The standard assay buffer was used and the concentration of *C. albicans* FTase was 1 nM. Four different types of reversible inhibition are possible: 1) competitive, where I competes directly with S for binding to E, 2) uncompetitive, where I only binds to E once S is bound, or E•S complex, 3) non-competitive, where I has the same affinity for E and E•S, and thus binds to both forms of E equally well, and 4) mixed inhibition, where I binds to both E and E•S, but with different affinities. Schematic representations of these different types of reversible inhibition are given in Figure 4.4, along with equations that are used to describe these different modes of inhibition.



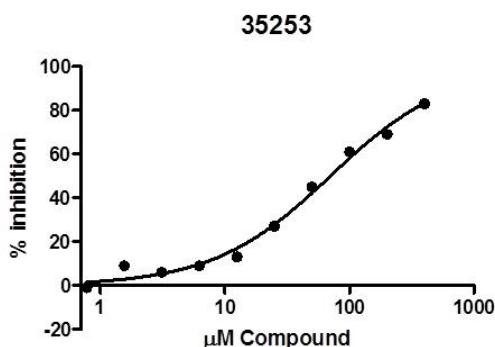
**Figure 4.4. Types of reversible inhibition.** In the competitive inhibition model, I only binds to E; in uncompetitive inhibition, I only binds to E•S; the non-competitive inhibition model is a special case of mixed inhibition where I binds to both E and E•S with the same affinity, whereas in the mixed inhibition model, I has different affinities for E and E•S. Equations on the right are used to fit the experimentally determined initial velocities as a function of S and I concentration.

## RESULTS

### *FTase and GGTase-I inhibition*

Dr. Mark Benson in the Merz group used computational methods [13] to rank the likelihood that compounds in the Florida Center for Heterocyclic Compounds database [14] would bind to the active site of rat FTase. Prof. Katritzky supplied us with five compounds that had the highest docking scores. IC<sub>50</sub> values were determined for these compounds (Figure 4.3) for inhibition of rat and *C. albicans* FTase and GGTase-I. All compounds appeared to be soluble in assay buffer at all concentrations tested. Additionally, all of the compounds behaved well and

did not appear to aggregate under the assay conditions, indicated by the hill coefficients of ~1 [15,16]. A representative IC<sub>50</sub> curve is shown in Figure 4.5, and a summary of IC<sub>50</sub> values and hill slopes is shown in Table 4.1. All IC<sub>50</sub> values > 400 μM were extrapolated based on the assumption of 100% inhibition.



**Figure 4.5.** IC<sub>50</sub> curve of compound 35253 with mammalian FTase. IC<sub>50</sub> = 70 ± 5 μM, Hill coefficient = 0.9.

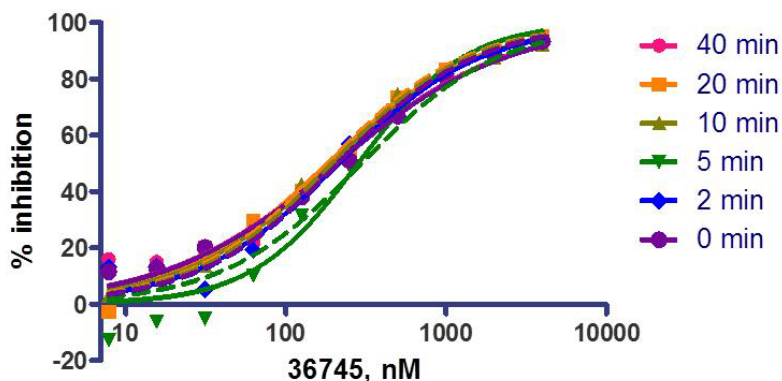
Enzyme	Rat FTase		Rat GGTase-I		<i>C. alb</i> FTase		<i>C. alb</i> GGTase-I	
Compound	IC <sub>50</sub> (μM)	n	IC <sub>50</sub> (μM)	n	IC <sub>50</sub> (μM)	n	IC <sub>50</sub> (μM)	n
<b>35253</b>	70 ± 5	0.9	>400 <sup>a</sup>		469 ± 44 <sup>c</sup>	1.3	553 ± 85 <sup>d</sup>	1.0
<b>36064</b>	>400 <sup>a</sup>		>400 <sup>a</sup>		>400 <sup>a</sup>		>400 <sup>a</sup>	
<b>36207</b>	595 ± 67 <sup>b</sup>	1.6	>400 <sup>a</sup>		199 ± 17	1.0	200 ± 21	1.0
<b>36347</b>	>400 <sup>a</sup>		>400 <sup>a</sup>		>400 <sup>a</sup>		>400 <sup>a</sup>	
<b>36745</b>	245 ± 35	1.1	>400 <sup>a</sup>		0.20 ± 0.02	1.0	73 ± 15	1.0

**Table 4.1.** Summary of mammalian and *C. albicans* FTase and GGTase-I inhibition. <sup>a</sup><15% inhibition was observed at 400 μM inhibitor, <sup>b</sup>34% inhibition at 400 μM inhibitor, <sup>c</sup>43% inhibition at 400 μM inhibitor, <sup>d</sup>40% inhibition at 400 μM inhibitor.

Compounds 35253, 36207, and 36745 have activity against rat FTase, *C. albicans* FTase and *C. albicans* GGTase-I, whereas compounds 36064 and 36347 do not inhibit any of the prenyltransferase enzymes. All of the active compounds contain a large aromatic fluorene substituent attached to the core molecule via an ester linkage. In contrast, the inactive compounds contain a benzyl group, and this moiety may not be large enough to make sufficient

hydrophobic interactions with the enzyme to enhance the binding affinity. Compounds 35253 and 35207 have modest activity against rat FTase and *C. albicans* prenyltransferases. However, compound 36745 has potent ( $IC_{50} = 200$  nM) and selective activity against *C. albicans* FTase. None of the compounds inhibit mammalian GGTase-I under these conditions.

Since compound 36745 is a potent inhibitor of *C. albicans* FTase, time dependence of inhibition was evaluated. Compound 36745 was incubated with *C. albicans* FTase for varying lengths of time (0 to 40 min) and then FTase activity was measured. An overlay of the  $IC_{50}$  curves obtained for different pre-incubation times is shown in Figure 4.6. While some variability ( $\leq 20\%$ ) in the initial rates was observed, the  $IC_{50}$  value for compound 36745 does not change upon pre-incubation ( $IC_{50} = 206 \pm 13$  nM for 0 min and  $IC_{50} = 201 \pm 21$  nM for 40 min). These data indicate the compound 36745 is not a time-dependent inhibitor but is a rapidly equilibrating, reversible inhibitor.



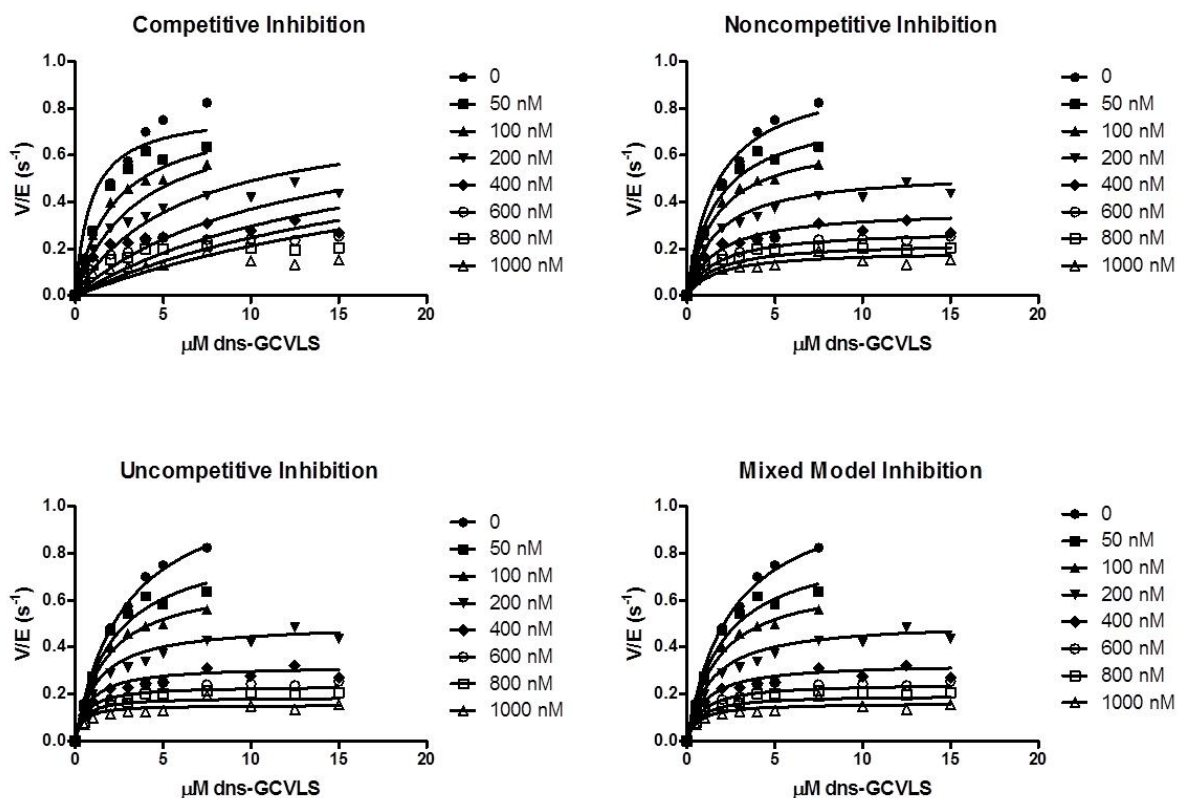
**Figure 4.6. Time dependence of *C. albicans* FTase inhibition by compound 36745.** Compound 36745 was incubated with *C. albicans* FTase for varying lengths of time, and then the enzyme activity was measured.  $IC_{50}$  value for compound 36745 does not change with the pre-incubation time indicating that inhibition is not time dependent.

#### *Mode of inhibition of C. albicans FTase by 36745*

Initial velocities for catalysis of farnesylation by *C. albicans* FTase at each substrate and inhibitor concentration were analyzed by different inhibition models using GraphPad Prism



curve fitting software, where shared constraints were applied to the  $K_M$ ,  $K_i$  and  $V_{max}$  parameters to perform a global fit analysis for all of the data sets. First, the mode of inhibition by compound 36745 with respect to the peptide substrate was determined. The data were fit to the four different inhibition models, as shown in Figure 4.7. Just from visually inspecting these plots, it is apparent that the model for competitive inhibition does not fit the data at high peptide concentrations as well as the other models, ruling out a purely competitive mode of inhibition. This deviation from the fit is also reflected in the lowest  $R^2$  value of 0.84 for the competitive inhibition model as seen in Table 4.2. However, the other models appear to fit equally well, with very similar values for  $R^2$  of 0.98 and 0.99, and thus it is impossible to distinguish between them just from looking at the graphs and the global fits.



**Figure 4.7. Mode of inhibition of 36745 with respect to dns-GCVLS peptide substrate.** Initial velocities were fit to competitive, non-competitive, uncompetitive, and mixed models of inhibition.

Global parameter	Competitive inhibition	Noncompetitive inhibition	Uncompetitive inhibition	Mixed model inhibition
$k_{\text{cat}}$ , s <sup>-1</sup>	0.79 ± 0.06	0.96 ± 0.03	1.15 ± 0.04	1.10 ± 0.04
$K_M$ , μM	0.89 ± 0.28	1.69 ± 0.12	2.83 ± 0.21	2.54 ± 0.21
$K_i$ , nM	34.0 ± 8.9	250 ± 10	155 ± 7.3	1362 ± 625
$\alpha$				0.13 ± 0.06
$R^2$	0.84	0.98	0.99	0.99

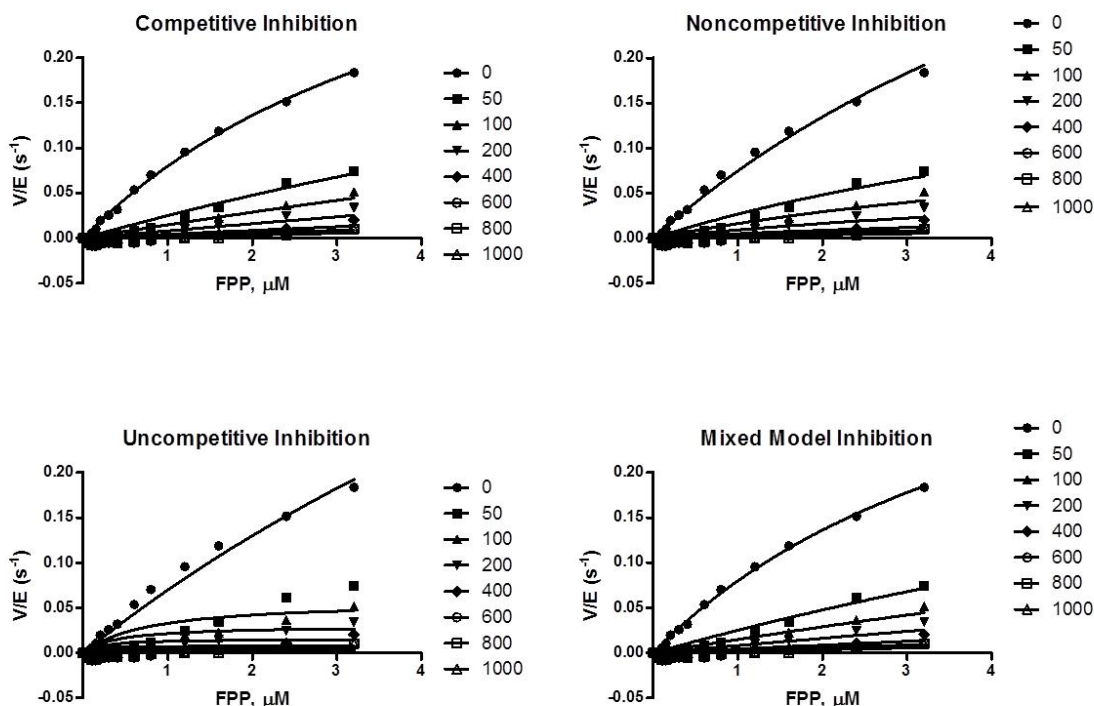
**Table 4.2. Global fit parameters for inhibition of *C. albicans* FTase by 36745 with respect to dns-GCVLS.**  $k_{\text{cat}}$ ,  $K_M$ , and  $K_i$  values were optimized for all the different substrate and inhibitor concentrations using global fit analysis. These values are intrinsic to the system and therefore common to all the curves.  $\alpha$  indicates the relative values of  $K_i$  and  $K_{iS}$ .

To distinguish between noncompetitive, uncompetitive and mixed modes of inhibition, the apparent steady-state kinetic parameters for each inhibitor concentration were calculated by fitting the Michaelis-Menten equation to these data, shown in Table 4.3. These data demonstrate that all three parameters,  $k_{\text{cat}}$ ,  $k_{\text{cat}}/K_M$  and  $K_M$  decrease as the inhibitor concentration increases. These data rule out: 1) competitive inhibition since the value of  $k_{\text{cat}}$  decreases, 2) uncompetitive inhibition since the value of  $k_{\text{cat}}/K_M$  decreases and 3) non-competitive inhibition since the value of  $K_M$  decreases. Therefore, these data are most consistent with a mixed mode of inhibition, where compound 36745 binds to and inhibits both E•FPP and a ternary complex containing both substrates such as E•FPP•Pep or E•Product. A fit of the equations describing the dependence of the kinetic parameters on the inhibitor concentration for mixed inhibition indicates that this inhibitor binds to E•S with 10-fold higher affinity than E (150 nM *versus* 1350 nM, respectively).

[36745], nM	$K_M$ , $\mu\text{M}$	$k_{\text{cat}}$ , $\text{s}^{-1}$	$k_{\text{cat}}/K_M$ , $\text{M}^{-1}\text{s}^{-1}$
0	$3.02 \pm 0.27$	$1.18 \pm 0.05$	$3.92 \times 10^5 \pm 2.1 \times 10^4$
50	$1.70 \pm 0.32$	$0.82 \pm 0.05$	$4.81 \times 10^5 \pm 6.4 \times 10^4$
100	$1.54 \pm 0.14$	$0.67 \pm 0.02$	$4.38 \times 10^5 \pm 2.9 \times 10^4$
200	$1.64 \pm 0.19$	$0.50 \pm 0.02$	$3.08 \times 10^5 \pm 2.9 \times 10^4$
400	$0.95 \pm 0.18$	$0.31 \pm 0.01$	$3.33 \times 10^5 \pm 5.4 \times 10^4$
600	$0.86 \pm 0.14$	$0.27 \pm 0.01$	$3.12 \times 10^5 \pm 4.4 \times 10^4$
800	$0.81 \pm 0.13$	$0.22 \pm 0.01$	$2.75 \times 10^5 \pm 3.8 \times 10^4$
1000	$0.73 \pm 0.26$	$0.16 \pm 0.01$	$2.21 \times 10^5 \pm 6.8 \times 10^4$

**Table 4.3. Apparent  $K_M$ ,  $k_{\text{cat}}$ , and  $k_{\text{cat}}/K_M$  values for inhibition of *C. albicans* FTase by 36745 with respect to dns-GCVLS.** Michaelis-Menten kinetics were calculated for each inhibitor concentration. As [36745] increases,  $K_M$ ,  $k_{\text{cat}}$ , and  $k_{\text{cat}}/K_M$  values all decrease.

An analogous analysis of 36745 mode of inhibition with respect to FPP was carried out. Equations describing the four different models of inhibition were fit to the initial velocity data, as shown in Figure 4.8. Visual inspection of these fits suggests that uncompetitive model does not fit the data as well as the others.



**Figure 4.8. Mode of inhibition of 36745 with respect to FPP prenyl donor.** Competitive, non-competitive, uncompetitive, and mixed models of inhibition were fit to the initial velocity data.

Global parameter	Competitive inhibition	Noncompetitive inhibition	Uncompetitive inhibition	Mixed model inhibition
$k_{cat}$ , s <sup>-1</sup>	0.46 ± 0.05	0.66 ± 0.12	0.96 ± 0.52	0.46 ± 0.05
$K_M$ , μM	4.74 ± 1.45	7.88 ± 1.89	12.7 ± 8.18	4.74 ± 1.24
$K_i$ , nM	18.9 ± 0.80	27.8 ± 1.80	3.23 ± 1.76	18.9 ± 3.56
$\alpha$				>1000
$R^2$	0.97	0.97	0.90	0.97

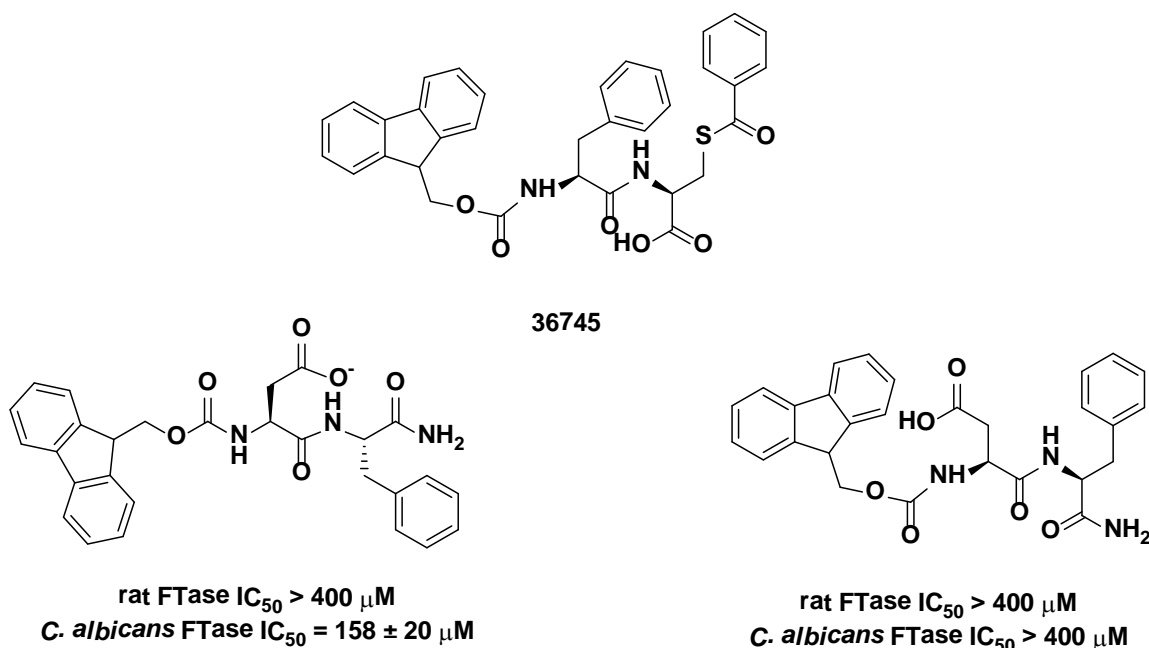
**Table 4.4. Global fit parameters for 36745 inhibition of *C. albicans* FTase wrt FPP.**  $k_{cat}$ ,  $K_M$ , and  $K_i$  values were optimized for all the different substrate and inhibitor concentrations using global fit analysis. These values are intrinsic to the system and therefore common to all the curves.

As in the previous analysis, steady-state kinetic parameters obtained from the global fit reflect that the uncompetitive inhibition model is not a good fit to the data, as it has the lowest  $R^2$  value of 0.90 (Table 4.4). The other models appear to fit equally well, as indicated by the  $R^2$  values of 0.97, but the mixed mode of inhibition can be ignored, since the large  $\alpha$  value means that the inhibitor has much greater affinity for E than E•S, and therefore this model basically reflects a competitive model. To distinguish between noncompetitive and competitive inhibition, the apparent  $K_M$  values were calculated and they increase as a function of inhibitor concentration. This is a hallmark of competitive inhibition, and thus compound 36745 directly competes with FPP prenyl donor for binding to *C. albicans* FTase.

#### *Analogs of 36745 identified in ZINC database*

To explore structure-activity relationships, a search for compounds similar to 36745 was performed using the ZINC database. This database contains over 20 million commercially available compounds that can be used for structure based virtual screening [17,18]. It can be queried based on both protein target and ligand chemotype, as well as a number of physicochemical properties that bin available small molecules into three categories: (1) “fragment-like” molecules that have low molecular weights and low affinities; (2) “lead-like” molecules which have higher molecular weights and are more likely to have higher affinities,

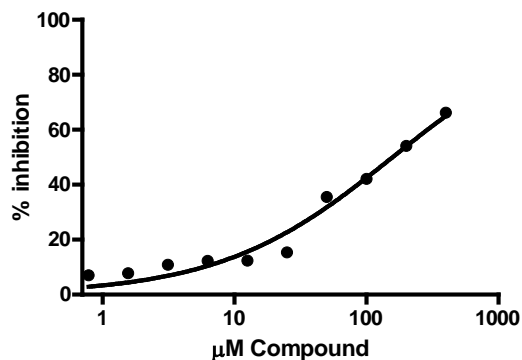
and (3) “drug-like” structures which have been curated to remove or flag potentially problematic functional groups, such as reactive trifluoromethanesulfonates, alkyl halides, perchlorates, Michael acceptors, aldehydes and thiols. A ZINC database search restricted to compounds that are 90% similar to compound 36745 yielded a set of 93 compounds, where approximately half of them were Fmoc-protected dipeptides. Of these compounds, two were selected to be tested for inhibition of mammalian and *C. albicans* FTase, and the results are shown in Figure 4.9.



**Figure 4.9. Activity of analogs of compound 36745.**  $IC_{50}$  values against mammalian and *C. albicans* FTase. Fmoc-Asp-Phe analog has weak activity against *C. albicans* FTase with  $IC_{50} = 158 \pm 20 \mu M$ , but no measurable activity against mammalian enzyme. Fmoc-Phe derivative has no measurable activity against either mammalian or *C. albicans* FTase.

Of these two compounds, only the Fmoc-Asp-Phe analog had any inhibitory activity against FTase, with an  $IC_{50}$  value of  $158 \pm 20 \mu M$  for *C. albicans* FTase (Figure 4.10). It had no measurable activity against mammalian FTase, and the second compound had no measurable activity against either FTase. These results suggest that having only one amino acid, even if it contains both an Fmoc group on the amine terminal and a large phenyl group on the carboxy terminal, is not sufficient to inhibit FTase. On the other hand, having a negatively charged

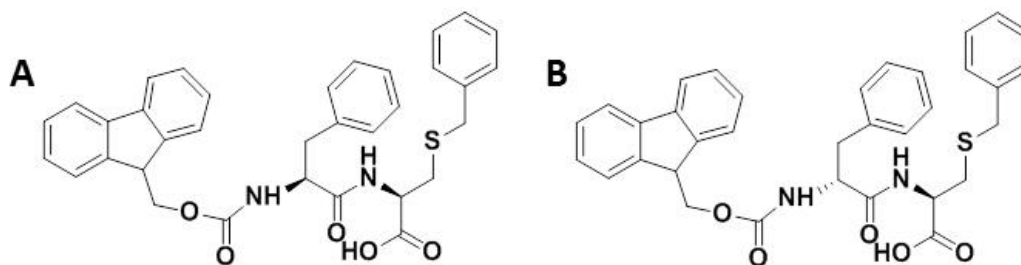
aspartate in place of a phenylalanine found in compound 36745 at the amine terminal of the dipeptide significantly affects activity against *C. albicans* FTase, but does not render the compound completely inactive. Additionally, having a C terminal amide instead of the carboxylic acid that is proposed to coordinate the catalytic zinc ion is also a tolerated replacement.



**Figure 4.10.** IC<sub>50</sub> curve of Fmoc-Asp-Phe analog with *C. albicans* FTase. IC<sub>50</sub> = 158 ± 20 μM, Hill coefficient = 0.66 ± 0.06

#### *C. albicans* FTase inhibition by close analogs of compound 36745

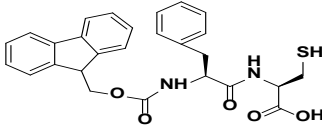
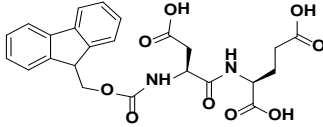
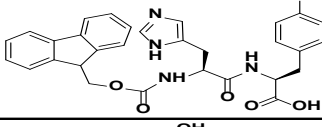
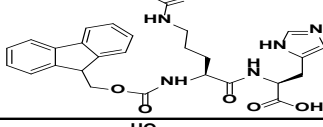
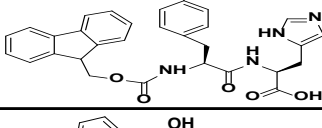
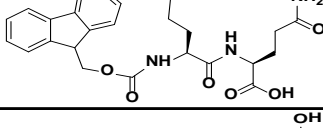
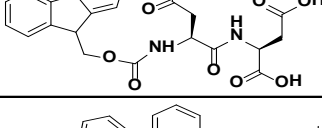
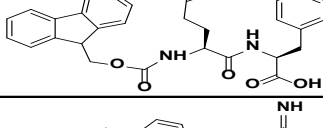
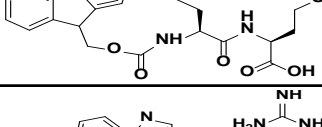
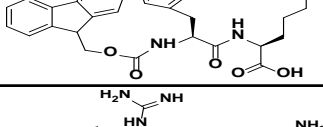
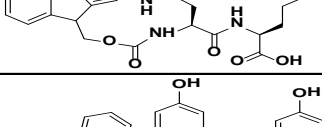
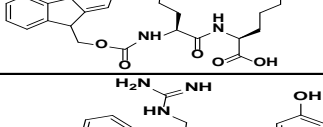
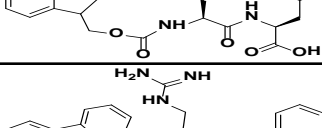
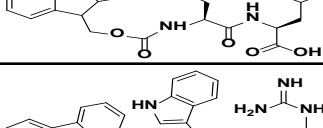
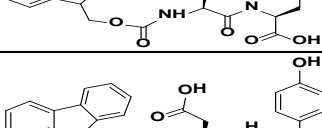
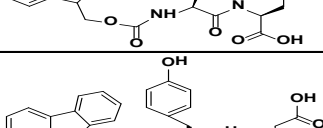
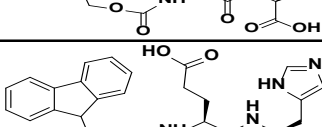
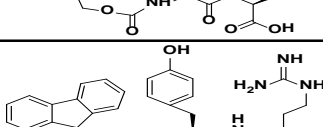
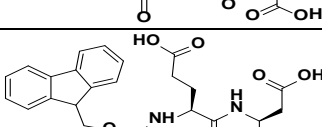
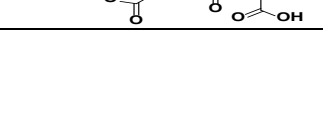
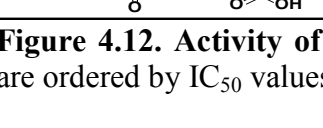
Review of the structure of compound 36745 reveals that it is a decorated dipeptide molecule consisting of Phe and Cys amino acids. The N terminal amine of Phe is modified with fluorenylmethoxycarbonyl (Fmoc), a protecting group that is often used in solid-phase peptide synthesis. Additionally, the C terminal Cys residue is capped with a benzoyl moiety attached to the sulfur atom, and it contains a free carboxylate that could function as a zinc-binding warhead. The thioester linkage in 36745 is not very stable, especially at high pH (> 8.0). Thus, two analogs of 36745 that contain a more stable thioether bond (Figure 4.11) were purchased from Sigma and tested for inhibition of *C. albicans* FTase. Both compounds inhibit this enzyme, where the natural L-Phe-containing compound is more potent, at IC<sub>50</sub> = 1.5 μM, than the unnatural D-Phe-containing compound, with IC<sub>50</sub> = 66 μM; however, the IC<sub>50</sub> values of both are decreased compared to 36745 (IC<sub>50</sub> = 0.2 μM).



**Figure 4.11. Benzylthioether analogs of compound 36745.** A) Fmoc protected L-Phe S-benzyl-L-Cys B) Fmoc protected D-Phe S-benzyl-L-Cys

#### *Inhibition of C. albicans FTase by Fmoc-substituted dipeptides*

Given that 36745 is an Fmoc protected dipeptide, a docking screen of natural amino acids with an Fmoc protecting group was performed by Prof. Kenneth Merz's group. The top 20 best scoring compounds were purchased from Sigma and inhibition of *C. albicans* FTase was measured. The results along with chemical structures are shown in Figure 4.12. It is interesting to note that seven out of twenty of these compounds contain a negatively charged aspartate or glutamate at the N terminal position, and four out of twenty have a negatively charged residue at the C terminal position, whereas the initial hit compound has a phenylalanine and a benzoyl group at these positions, respectively. In addition, three out of these seven compounds have negatively charged residue at both positions, suggesting that the binding site for these dipeptides can accommodate both hydrophobic and electrostatic interactions depending on the relative position and orientation of these diverse side chains. Furthermore, four out of the twenty amino acids contain a positively charged arginine at the N terminal position, and five out of the twenty contain a positively charged residue at the C terminal position. Eight dipeptides contain an aromatic residue at the N terminal and nine contain an aromatic residue at the C terminal position. Thus, these top scoring peptides include a mix of charged and aromatic amino acids, but there are no small amino acids (e.g. glycine, alanine, serine, or threonine) or aliphatic amino acids (e.g. valine, leucine, or isoleucine).

Structure	IC <sub>50</sub> (μM)	Hill Slope	Structure	IC <sub>50</sub> (μM)	Hill Slope
	0.28 ± 0.02	1.1 ± 0.1		448 ± 205	0.6 ± 0.2
	7.2 ± 0.2	0.8 ± 0.0		522 ± 108	0.6 ± 0.1
	50 ± 4	0.7 ± 0.0		610 ± 77	0.9 ± 0.1
	139 ± 13	0.7 ± 0.1		632 ± 96	0.9 ± 0.1
	156 ± 17	1.0 ± 0.1		752 ± 97	0.9 ± 0.1
	192 ± 20	0.7 ± 0.1		>1000	
	213 ± 22	0.6 ± 0.0		>1000	
	238 ± 26	0.9 ± 0.1		>1000	
	248 ± 59	1.3 ± 0.4		>1000	
	327 ± 38	0.8 ± 0.1		>1000	
	427 ± 82	0.5 ± 0.1			

**Figure 4.12.** Activity of Fmoc-protected dipeptides against *C. albicans* FTase. Compounds are ordered by IC<sub>50</sub> values.



The most active dipeptide inhibitor of *C. albicans* FTase is Fmoc-Phe-Cys, with an  $IC_{50} = 0.28 \mu\text{M}$ . This molecule closely resembles the initial hit, but it is lacking the benzoyl modification on the cysteine sulfur. This result suggests that while the benzoyl group does contribute to the overall potency, it is not a strictly required moiety as there is a less than 2-fold difference in the  $IC_{50}$  values. However, replacement of cysteine with methionine in Fmoc-Phe-Met increases  $IC_{50}$  to  $156 \pm 17 \mu\text{M}$ , a dramatic decrease in potency (over 550-fold).

The next most potent dipeptides after Fmoc-Phe-Cys contain two aromatic amino acids, histidine and tyrosine, where histidine at the N terminal position is ~7-fold more potent than tyrosine at the N terminal position, with  $IC_{50}$  of  $7.2 \pm 0.2 \mu\text{M}$  and  $50 \pm 4 \mu\text{M}$ , respectively. These two dipeptides present two of the three dipeptide combinations that contain aromatic amino acids at both positions. The other dipeptide containing tyrosine residues at both positions ( $IC_{50} = 213 \pm 22 \mu\text{M}$ ) is about 30-fold and 4-fold less active than Fmoc-His-Tyr and Fmoc-Tyr-His, respectively. This result suggests that having a histidine at the N terminal position is important for maintaining potency, possibly due to coordination with the active site zinc ion.

The next most potent combination is a dipeptide containing two aspartic acid residues, with  $IC_{50} = 139 \pm 13 \mu\text{M}$ . Replacing aspartate with a larger glutamate at either the N terminal or C terminal position leads to > 3-fold drop in activity, with  $IC_{50} = 448 \pm 205 \mu\text{M}$  and  $427 \pm 82 \mu\text{M}$ , respectively. The position of glutamate does not appear to make a difference for activity, especially since the errors on the  $IC_{50}$  values are large. Replacing Asp at the C terminal position with an aromatic amino acid tyrosine leads to a less than 2-fold decrease in potency, with  $IC_{50}$  for Fmoc-Asp-Tyr being  $248 \pm 59 \mu\text{M}$ . It is interesting to note that this dipeptide closely resembles the dipeptide identified using ZINC database (see Figure 4.9), where the ZINC compound is ~1.5-fold more active than the Fmoc-Asp-Tyr molecule. The Fmoc-Glu-Tyr

combination has an  $IC_{50} = 632 \pm 96 \mu\text{M}$ , which is  $\sim 2.5$ -fold less potent than the Fmoc-Asp-Tyr combination. This difference is consistent with the 3-fold drop in activity observed for the Asp to Glu substitution, suggesting that the same interactions between Asp or Glu in the inhibitor and the enzyme are present whether the other residue is negatively charged or aromatic. Replacement of tyrosine with histidine to make Fmoc-Glu-His restores almost 2-fold potency with  $IC_{50} = 327 \pm 38 \mu\text{M}$ . Although this “rescue” of potency is not as large as when C terminal tyrosine in a Fmoc-Tyr-Tyr dipeptide is replaced with histidine, which results in 4-fold potency increase, these results are consistent with the overall trend where having a histidine at either position, all other things being equal, results in a more potent compound. Replacement of tyrosine with glutamine to make Fmoc-Glu-Gln does not change the potency of the dipeptide within error, with the  $IC_{50} = 610 \pm 77 \mu\text{M}$ . Finally, changing the order of Glu and Tyr in the dipeptide renders the compound inactive, with Fmoc-Tyr-Glu  $IC_{50} > 1000 \mu\text{M}$ .

Moving on to dipeptides that contain positively charged amino acids, Fmoc-His-Arg is the most potent inhibitor with an  $IC_{50}$  of  $192 \pm 20 \mu\text{M}$ . Comparing this inhibitor to Fmoc-His-Tyr that contains an aromatic tyrosine residue instead of a positively charged arginine, the arginine-containing inhibitor is  $\sim 26$ -fold less potent than the tyrosine-containing one, suggesting that an aromatic moiety makes a significant contribution to enzyme inhibition. Replacement of histidine with phenylalanine to make Fmoc-Phe-Arg further decreases potency by almost 4-fold, giving an  $IC_{50} = 752 \pm 97 \mu\text{M}$ . This decrease in activity is consistent with the trend observed for replacing histidine with tyrosine at the N terminal position, although the change from histidine to phenylalanine appears to be less detrimental to overall activity. However, this conclusion may be somewhat premature, given that histidine to tyrosine replacement data is for a dipeptide where C terminal amino acid is tyrosine, and the histidine to phenylalanine replacement data is for a

dipeptide with arginine at the C terminus, and thus the context of these replacements is different and may not be interpretable with such a simplistic approach. However, replacement of histidine with tryptophan or tyrosine results in complete loss of inhibitor activity, where Fmoc-Trp-Arg and Fmoc-Tyr-Arg  $IC_{50}$ 's are both  $> 1000 \mu\text{M}$ . The fact that Fmoc-Tyr-Arg lacks measurable activity against *C. albicans* FTase further supports the hypothesis that a histidine to phenylalanine change is less disruptive to inhibitor-enzyme interaction than a histidine to tyrosine substitution.

The last four inhibitors in this study contain arginine at the N terminal position. Of these, the Fmoc-Arg-Phe dipeptide is the most potent inhibitor with an  $IC_{50} = 238 \pm 26 \mu\text{M}$ . This compound is  $>3$ -fold more potent than Fmoc-Phe-Arg, where the placement of charged and aromatic amino acids is reversed. This trend is consistent with the results for Fmoc-His-Tyr and Fmoc-Tyr-His, where the former compound is  $\sim 7$ -fold more active than the latter. The next most potent dipeptide with arginine at the N terminal position is Fmoc-Arg-His, with an  $IC_{50} = 522 \pm 108 \mu\text{M}$ , which is a  $\sim 2$ -fold drop in potency compared to Fmoc-Arg-Phe. This result represents a reversal of the trend observed with Fmoc-Tyr-His and Fmoc-Tyr-Tyr, where the histidine-containing dipeptide was over 4-fold more active than the tyrosine one. This suggests that there may be some context-dependence of recognition of individual residues in the dipeptide inhibitor. Finally, replacement of C terminal phenylalanine with lysine or tyrosine results in complete loss of potency as both Fmoc-Arg-Lys and Fmoc-Arg-Tyr have  $IC_{50}$ 's  $> 1000 \mu\text{M}$ . Here once again the trend that tyrosine at the C terminus results in lower potency than histidine at this position is confirmed.

## DISCUSSION

### *Prenyltransferase inhibition*

Virtual inhibitor screening by the Merz group identified a ranked list of potential FTase inhibitors. Five of the top scoring compounds were tested for inhibition of rat and *C. albicans* prenyltransferases. Of these, three compounds (35253, 36207 and 36745) inhibited rat and *C. albicans* FTase and *C. albicans* GGTase-I. This shows a >50% overall success rate of this type of *in silico* approach, which is quite good, especially compared to *in vitro* high-throughput screening strategies of random libraries of small molecules, where true hit rates are well below 1%. It is somewhat puzzling that even though the virtual screen was performed using rat FTase, the most best *in vitro* activity was observed for *C. albicans* FTase. Compound 36745 is the most potent inhibitor of *C. albicans* FTase with an  $IC_{50} = 200$  nM, and 73  $\mu$ M against *C. albicans* GGTase-I. It has an  $IC_{50} = 250$   $\mu$ M against mammalian FTase, making it greater than 1000-fold selective against *C. albicans* than mammalian FTase. Compound 35253 is the most potent rat FTase inhibitor with  $IC_{50} = 70$   $\mu$ M, with modest (7-8-fold) selectivity for this enzyme compared to *C. albicans* prenyltransferases. Compound 36207 has a 2-3-fold selectivity for *C. albicans* FTase and GGTase-I ( $IC_{50} = 200$   $\mu$ M) compared to rat FTase. Further investigation of the most potent compound, 36745, for inhibition of FTase revealed that this compound is time-independent and competitive with FPP,  $K_i = 18.9 \pm 0.8$  nM, and uncompetitive with the peptide substrate,  $K_i = 155 \pm 7$  nM. Further analysis of inhibition pattern of this compound is warranted.

Given excellent *in vitro* potency and selectivity of compound 36745, measurement of its *in vivo* activity against *C. albicans* pathogen would be the next logical step. However, compound 36745 has two readily identifiable metabolic liabilities, the ester bond that connects the fluorene moiety to the dipeptide, and a thioester linkage of the benzoyl group to the sulfur atom of the

cysteine. These functional groups render this compound unusable in whole-cell applications, and thus *in vivo* activity was not measured. Substitution of the thioester linkage with a thioether (Figure 4.11) was a first attempt to overcome this issue. This compound was still an inhibitor although with decreased potency. However, the remaining ester group still precluded further *in vivo* inhibition measurement. In the future, the ester group could be replaced by a carbonyl or an amide group (oxygen replaced by methylene or nitrogen) to create a more metabolically stable acetamide substitution.

#### *SAR of Fmoc-modified natural dipeptides*

Based on the structure of compound 36745, Fmoc protected natural dipeptides were docked into the rat FTase active site to select the dipeptide combinations most likely to inhibit the enzyme. The top twenty Fmoc-substituted dipeptide molecules were evaluated in an *in vitro* assay and their analysis revealed insights into structure-activity relationships for *C. albicans* FTase inhibitors. A grid of all the tested amino acid combinations and their IC<sub>50</sub> values is shown in Figure 4.13.

		Position 1						
		D	E	F	H	R	W	Y
Position 2	C			0.28				
	D	139	427					
	E	448						>1000
	F					238		
	H		327			522		50
	K					>1000		
	M			156				
	Q		610					
	R			752	192		>1000	>1000
	Y	248	632		7.2	>1000		213

**Figure 4.13. *C. albicans* FTase inhibition results for Fmoc-substituted dipeptide compounds.** Numbers in this grid are IC<sub>50</sub> values against *C. albicans* FTase expressed in μM. Position 1 (across) refers to the N terminal amino acid, and position 2 (down) refers to the C terminal amino acid. Color coding is based on a red-green color scale, with red being the most active compound, and green being the least. Inactive compounds where IC<sub>50</sub>'s are > 1000 μM are excluded from the color scale.

Fmoc-Phe-Cys is the most active dipeptide, followed by Fmoc-His-Tyr and Fmoc-Tyr-His. Aromatic amino acids are thus beneficial for activity, with even Fmoc-Tyr-Tyr exhibiting modest activity. In general, having a His at either position is beneficial for activity, as all dipeptides that contain histidine show some inhibition of the enzyme. On the other hand, positively charged amino acids are not optimal, as having arginine is generally not good for activity, with four out of the five molecules that lack measurable activity containing an arginine at either position. The dipeptide combinations that contain arginine that retain inhibitory activity are those that also have either histidine or phenylalanine, while arginine combinations with

tyrosine, tryptophan, or lysine all lead to loss of measurable inhibition. Negative charge on the dipeptide appears to be tolerated much better than positive charge, as all but one (Fmoc-Tyr-Glu) dipeptide combinations that contain aspartate or glutamate retain some activity, and even dipeptides where both residues are charged are active against *C. albicans* FTase. This preference for aromatic and negatively charged amino acids is consistent with mode of inhibition results for compound 36745, as it showed that this compound is competitive with FPP substrate, which contains both hydrophobic and negatively charged moieties. Although it is not guaranteed that these Fmoc-dipeptides bind to *C. albicans* FTase in the same way as compound 36745, the ‘anchoring’ moieties fluorene and carboxylic acid are present in all of these molecules, and thus similar binding mode for these compounds is a reasonable proposition.

It should be noted that activity against *C. albicans* FTase has been measured for only a subset of possible 400 dipeptide combinations; however, these were chosen based on the *in silico* screening. The results of this analysis should only serve as preliminary guides for further development of *C. albicans* FTase inhibitors. A more exhaustive investigation of dipeptide combinations may uncover additional or different activity trends and be able to better guide rational inhibitor design.

## REFERENCES

1. Berndt N, Hamilton AD, Sebti SM: **Targeting protein prenylation for cancer therapy.** *Nature Reviews Cancer* 2011, **11**:775-791.
2. Wong NS, Morse MA: **Lonafarnib for cancer and progeria.** *Expert Opin Investig Drugs* 2012, **21**:1043-1055.
3. Charron G, Tsou LK, Maguire W, Yount JS, Hang HC: **Alkynyl-farnesol reporters for detection of protein S-prenylation in cells.** *Mol Biosyst* 2011, **7**:67-73.
4. Lerner EC, Qian Y, Blaskovich MA, Fossum RD, Vogt A, Sun J, Cox AD, Der CJ, Hamilton AD, Sebti SM: **Ras CAAX peptidomimetic FTI-277 selectively blocks oncogenic Ras signaling by inducing cytoplasmic accumulation of inactive Ras-Raf complexes.** *J Biol Chem* 1995, **270**:26802-26806.
5. Troutman JM, Subramanian T, Andres DA, Spielmann HP: **Selective Modification of CaaX Peptides with ortho-Substituted Anilino geranyl Lipids by Protein Farnesyl Transferase: Competitive Substrates and Potent Inhibitors from a Library of Farnesyl Diphosphate Analogues.** *Biochemistry* 2007, **46**:11310-11321.
6. Rawat DS, Krzysiak AJ, Gibbs RA: **Synthesis and Biochemical Evaluation of 3,7-Disubstituted Farnesyl Diphosphate Analogues.** *The Journal of Organic Chemistry* 2008, **73**:1881-1887.
7. de Figueiredo RM, Coudray L, Dubois J: **Synthesis and biological evaluation of potential bisubstrate inhibitors of protein farnesyltransferase. Design and synthesis of functionalized imidazoles.** *Organic & Biomolecular Chemistry* 2007, **5**:3299.
8. Smalera I, Williamson JM, Baginsky W, Leiting B, Mazur P: **Expression and characterization of protein geranylgeranyltransferase type I from the pathogenic yeast *Candida albicans* and identification of yeast selective enzyme inhibitors.** *Biochim Biophys Acta* 2000, **1480**:132-144.
9. Kelly R, Card D, Register E, Mazur P, Kelly T, Tanaka KI, Onishi J, Williamson JM, Fan H, Satoh T, et al.: **Geranylgeranyltransferase I of *Candida albicans*: null mutants or enzyme inhibitors produce unexpected phenotypes.** *J Bacteriol* 2000, **182**:704-713.
10. Nishimura S, Matsunaga S, Shibasaki M, Suzuki K, Furihata K, van Soest RW, Fusetani N: **Massadine, a novel geranylgeranyltransferase type I inhibitor from the marine sponge *Stylissa aff. massa*.** *Org Lett* 2003, **5**:2255-2257.
11. Sunami S, Ohkubo M, Sagara T, Ono J, Asahi S, Koito S, Morishima H: **A new class of type I protein geranylgeranyltransferase (GGTase I) inhibitor.** *Bioorg Med Chem Lett* 2002, **12**:629-632.
12. Murthi KK, Smith SE, Kluge AF, Bergnes G, Bureau P, Berlin V: **Antifungal activity of a *Candida albicans* GGTase I inhibitor-Alanine conjugate. inhibition of Rho1p prenylation in *C. albicans*.** *Bioorganic & Medicinal Chemistry Letters* 2003, **13**:1935-1937.
13. Zheng Z, Merz KM, Jr.: **Ligand Identification Scoring Algorithm (LISA).** *J Chem Inf Model* 2011, **51**:1296-1306.
14. Katritzky AR, Tala SR, Abo-Dya NE, Ibrahim TS, El-Feky SA, Gyanda K, Pandya KM: **Chemical ligation of S-scylated cysteine peptides to form native peptides via 5-, 11-, and 14-membered cyclic transition states.** *J Org Chem* 2011, **76**:85-96.



15. Feng BY, Simeonov A, Jadhav A, Babaoglu K, Inglese J, Shoichet BK, Austin CP: **A high-throughput screen for aggregation-based inhibition in a large compound library.** *J Med Chem* 2007, **50**:2385-2390.
16. Shoichet BK: **Interpreting steep dose-response curves in early inhibitor discovery.** *J Med Chem* 2006, **49**:7274-7277.
17. Irwin JJ, Shoichet BK: **ZINC--a free database of commercially available compounds for virtual screening.** *J Chem Inf Model* 2005, **45**:177-182.
18. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG: **ZINC: A Free Tool to Discover Chemistry for Biology.** *J Chem Inf Model* 2012.

## CHAPTER FIVE

### SUMMARY, CONCLUSIONS, AND FUTURE DIRECTIONS

#### SUMMARY AND CONCLUSIONS

In this work, the substrate specificity and molecular recognition of *C. albicans* FTase and GGTase-I were investigated using peptide library approaches. The proteins modified by the prenyltransferases are involved in a variety of important pathways both in yeast and in mammalian cells, and inhibitors that are selective for the yeast enzymes could be developed into anti-fungal therapies. *C. albicans* is the major opportunistic human fungal pathogen and is the cause of serious systemic disease in immunocompromised patients and topical infections in healthy individuals [1]. Since prenyltransferases are not unique to pathogenic organisms, it is important to find selective inhibitors that will only inhibit growth of pathogenic species, and will not affect human cells. Mammalian FTase inhibitors (FTIs) have been investigated for the treatment of cancer and other diseases [2-4], but unfortunately they have not met with much clinical success due to lack of efficacy. Although this is not a desirable outcome, it presents an opportunity for development of FTase and/or GGTase-I-targeted anti-fungal agents as they are likely to be well tolerated. This lowers the bar on having to discover an exquisitely selective compound, potentially removing some of the inherent risk that is involved in developing therapeutics against novel targets. The challenge then is to determine how much overlap *C. albicans* FTase and GGTase-I have in their substrate specificity, i.e. how well they can

compensate for each other and cross-prenylate substrates. This will dictate whether a single specificity inhibitor is sufficient for anti-fungal activity, or if a dual inhibitor or a combination is required to achieve *in vivo* efficacy.

#### *Substrate recognition by C. albicans FTase and GGTase-I*

In this study, first a methodology for cloning, expression, and purification of *C. albicans* FTase and GGTase-I was established. Then reactivity of a library of 328 peptides of the form dns-TKCxxx was measured with these enzymes under MTO and STO conditions. *C. albicans* GGTase-I differed from mammalian and *C. albicans* FTase in that it was able to catalyze prenylation of 63% of all peptides in the library under MTO conditions, a number that was almost twice that for the other enzymes. Structural data on this enzyme showed that it lacked a proper “exit groove”, a tunnel where the isoprenoid moiety of the prenylated product is found before it dissociates from the enzyme [5,6]. Since product release is often the rate-limiting step for mammalian prenyltransferases, it is plausible that an enzyme that lacks such product binding site is more likely to exhibit faster product dissociation kinetics. Distribution of amino acids at the a<sub>1</sub>, a<sub>2</sub> and X positions of the CaaX sequence suggested a broader specificity of *C. albicans* GGTase-I as it showed more variability of residues at all positions that were efficient MTO substrates.

Statistical analysis was used to determine amino acids that were over-represented or under-represented in each pool of substrates as compared to its presence in the overall library. In this analysis, *C. albicans* GGTase-I showed a significant enrichment in large amino acids, F and W, at the a<sub>2</sub> position, an unexpected finding given that mammalian FTase and GGTase-I, as well as *C. albicans* FTase all show a preference for I, L and V at the a<sub>2</sub> position [7]. Structural

examination of the *C. albicans* GGTase-I  $a_2$  binding pocket revealed that it is  $\sim 25 \text{ \AA}^3$  larger than mammalian GGTase-I, primarily due to a L352 $\beta$  substitution for F324 $\beta$ .

Sequence preferences of *C. albicans* FTase and GGTase-I showed that both of these enzymes have strong preferences at the  $a_1$  position, with FTase showing a preference for I, V and T and GGTase-I preferring I, L, W and C. Although historically  $a_1$  has been considered a highly variable position, recent studies with mammalian enzymes identified sequence preferences at the  $a_1$  position [7]. Charged residues (D, E, R and K) are selected against by *C. albicans* FTase and GGTase-I at all positions, a trend that was also observed with mammalian enzymes.

Based on the results of the statistical analysis, a simple algorithm was derived to predict *C. albicans* FTase and GGTase-I substrates. 25 unique Cxxx> sequences found in the *C. albicans* genome for which reactivity with *C. albicans* FTase and GGTase-I had been measured were used as a test set, and the new algorithm was compared to the PrePS algorithm [8] for the ability to correctly predict substrates. While PrePS had a similar false positive rate to the new algorithm, it had a higher false negative rate, especially for *C. albicans* GGTase-I (48% versus 12%). As PrePS was designed to predict mammalian FTase and GGTase-I substrates, it is not surprising that it is less accurate in predicting yeast substrates. However, studies with mammalian FTase have also found that PrePS is too narrow in its definition of substrates; this algorithm is good at predicting “canonical” substrates but misses substrates that are less well described by the traditional CaaX sequences [7].

#### *Molecular basis of Ca<sub>1</sub>a<sub>2</sub>X recognition by C. albicans FTase and GGTase-I*

Based on the insights gained about substrate recognition from the peptide library studies, this study further explored the determinants of substrate recognition in *C. albicans* FTase and GGTase-I. Systematic variation of the side chain volume of the  $a_2$  residue with a number of

different X residues was used to probe substrate recognition determinants as well as the role of cooperativity between the  $a_2$  and X residues. Previous studies carried out in our lab with mammalian FTase showed a positive correlation between peptide reactivity and  $a_2$  residue volume as the amino acid volume increases from  $\sim 60 \text{ \AA}^3$  to  $150 \text{ \AA}^3$ , after which point a negative correlation was observed [9]. *C. albicans* FTase showed a similar dependence on the  $a_2$  residue volume as that for mammalian FTase. When the X residue is small (A, S), the size of the  $a_2$  residue exerts a larger influence on the peptide reactivity, and when the X residue is large (Q, M), the impact of the  $a_2$  residue volume is diminished. However, when X is too large (L, F), the contribution of a positive interaction between the  $a_2$  residue and the enzyme is diminished, while steric discrimination is enhanced. Therefore, *C. albicans* FTase has a nuanced and cooperative recognition of the CaaX sequence. In contrast, *C. albicans* GGTase-I shows virtually no dependence on  $a_2$  residue volume, and little influence on peptide reactivity by the X residue. Mammalian GGTase-I showed a similar  $a_2$  residue volume dependence as FTase.

Closer investigation of mammalian and *C. albicans* GGTase-I reactivity with varying X residues revealed that while the mammalian enzyme is able to prenylate a wide variety of X residues, it is only able to do so under STO conditions, as product release is prohibitively slow for efficient progression through the entire catalytic cycle. On the contrary, all of these peptides are MTO substrates for *C. albicans* GGTase-I. These data provide support for the faster product release hypothesis, and the role of the exit groove (or lack thereof) in the GGTase-I catalytic cycle. In addition to the potential role of product release in broadening MTO substrate specificity of *C. albicans* GGTase-I, structural data also provide a possible explanation for this phenomenon. In addition to the increase in the size of the  $a_2$  binding cavity, *C. albicans* GGTase-

I contains several solvent molecules in its putative peptide binding site, enabling hydrogen bond formation.

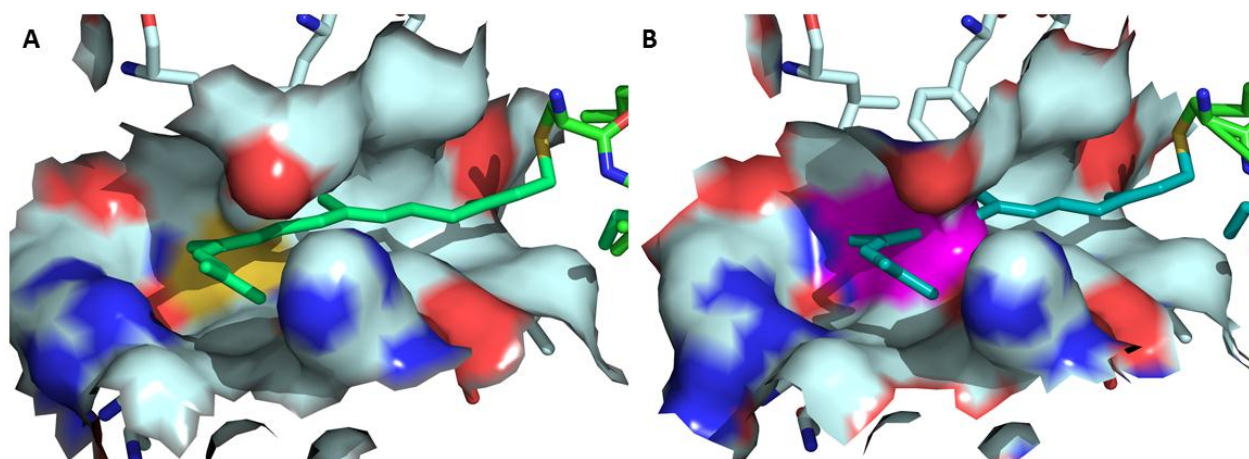
### *Inhibition of C. albicans FTase and GGTase-I*

This work was a collaboration between Dr. Kenneth Merz and Dr. Mark Benson, who identified prenyltransferase inhibitors by performing a docking study with the rat FTase structure. Three out of the top five compounds showed measurable inhibition of mammalian FTase and *C. albicans* FTase and GGTase-I in an *in vitro* assay. The most potent compound, a fluorene-substituted dipeptide, had an  $IC_{50} = 200$  nM against *C. albicans* FTase, and showed >1000-fold selectivity for the yeast over the rat enzyme. The mode of action of this compound is most consistent with competitive binding with FPP and uncompetitive binding with peptide substrate. The top twenty scoring compounds from a second round of docking studies with fluorene-substituted dipeptides were assayed for inhibition of *C. albicans* FTase. Although the docking protocol placed the C-terminal carboxylate as the zinc-binding moiety, the SAR that emerged from these data identified several other potential zinc ligands. C-terminal thiolate of cysteine, carboxylate side chains of Asp and Glu amino acids, and the histidine nitrogen could all potentially coordinate the zinc ion; dipeptides that contain these amino acids show good to moderate *in vitro* potency. In addition, negatively charged amino acids could bind in the diphosphate binding pocket and form favorable electrostatic interactions. Positively charged amino acids Arg and Lys were generally not well-tolerated, unless they were paired with Phe or His, but not Tyr or Trp. This context-dependent inhibitor recognition mimics the substrate recognition of  $a_2X$  by *C. albicans* FTase, where the CaaX amino acids are not recognized individually, but rather seen as an ensemble of di- or even tripeptides.

## FUTURE DIRECTIONS

### *Role of the exit groove in C. albicans GGTase-I product release*

Based on the evidence presented in this work, it is tempting to hypothesize that the large number of *C. albicans* GGTase-I MTO substrates is due at least in part to an increase in the rate constant for product release, and thus determination of this kinetic parameter will provide insight into the mechanism of *C. albicans* GGTase-I broadened substrate specificity. If this enzyme shows a faster rate of product release, then the role of the exit groove in product release can be investigated. As it is difficult to widen the exit groove without disrupting the integrity of the enzyme, a steric block can be engineered into rat GGTase-I via mutagenesis of a small side chain that points into the exit groove cavity to a large amino acid. Figure 5.1 shows one such approach where C32 $\beta$  in mammalian GGTase-I (panel A, yellow) is replaced with W32 $\beta$  (panel B, magenta) that creates a narrowing of the exit groove that should not permit the isoprenoid unit on the product to reside there. Y40 $\beta$  and H316 $\beta$  close around the isoprenoid moiety on the opposite side from C32 $\beta$ , so the geranylgeranyl group should not be able to simply shift slightly out of the tunnel to avoid the clash with W32 $\beta$ . If the exit groove slows down product release, then C32 $\beta$ W mutant GGTase-I should have a faster rate of product release than the wild-type enzyme, and potentially STO substrates can be converted into MTO substrates by introduction of this mutation. This work could provide further insight into the GGTase-I kinetic mechanism.



**Figure 5.1. Structure of the mammalian GGTase-I exit groove.** A) Mammalian exit groove with geranylgeranyl group of the product bound (teal). It is able to fit well in the exit tunnel. B) Proposed mutation in mammalian GGTase-I at C32 $\beta$ W (shown in magenta) that will block entry of the prenyl group. (PDB ID 1N4S)

#### *Prediction of *C. albicans* FTase and GGTase-I substrates*

As there are currently no computational models that can predict yeast FTase and GGTase-I substrates, a significant contribution to the field would be an algorithm that can perform this function. In this study, a very simple algorithm was proposed based on statistically significant enrichment and depletion of amino acids at the  $a_1$ ,  $a_2$  and X positions. Based on a test set of 25 peptides, it showed reasonable predictive ability, but more CaaX sequences should be tested (especially the ones found in the *C. albicans* genome) to validate this prediction methodology. It is highly likely that further testing will provide new insights about substrate recognition, and those can be incorporated into the substrate prediction model. Having a pool of CaaX sequences that are only FTase or GGTase-I substrates will help in developing an inhibitor that has antimycotic activity against *C. albicans* pathogen. On the other hand, if it is found that there is so much overlap between substrate specificity of these enzymes that inhibition of one enzyme is not sufficient, this will definitively show that either a dual-specificity inhibitor or a combination of an FTI and a GGTI is necessary to achieve *in vivo* efficacy.



### *C. albicans* inhibitor design

Based on the SAR data obtained with Fmoc-substituted dipeptides, other dipeptides can be synthesized and tested for inhibitory activity against *C. albicans* FTase. Since His correlated favorably with inhibition of *C. albicans* FTase, Fmoc-His-His, Fmoc-His-Cys, and Fmoc-His-Phe could be tested next. Additionally, Fmoc-His-Asp and Fmoc-Asp-His combinations may also be active, where the carboxylate on the Asp could coordinate the Zn<sup>2+</sup> metal ion, or it could interact with the charged diphosphate binding pocket, depending on its binding mode. In addition, Fmoc-dipeptides should be tested against *C. albicans* GGTase-I and mammalian FTase and GGTase-I to see if any of them are dual-specificity inhibitors, and also determine if they have any selectivity for *C. albicans* over rat enzymes. Finally, a move towards less metabolically problematic compounds involving an isosteric replacement of the ester functionality with an amide or substitution of the oxygen with a methylene group should prevent esterase catalyzed hydrolysis. The peptide bond is also a liability, as it is susceptible to proteolysis, and it will also need to be replaced with a more stable linkage [10].

Overall, this work explored the substrate selectivity and inhibition of *C. albicans* FTase and GGTase-I. It showed that for the most part *C. albicans* FTase has similar substrate specificity to mammalian FTase, while *C. albicans* GGTase-I has a more relaxed substrate specificity, including recognition of substrates with large amino acids. In addition, *C. albicans* GGTase-I catalyzes prenylation of a substantially larger pool of peptides under multiple turnover conditions than other prenyltransferases. Further investigation of *C. albicans* GGTase-I can enhance our understanding of prenyltransferase substrate recognition in general, as well as gain insight into using prenyltransferases as anti-fungal drug targets.

## REFERENCES

1. Kelly R, Card D, Register E, Mazur P, Kelly T, Tanaka KI, Onishi J, Williamson JM, Fan H, Satoh T, et al.: **Geranylgeranyltransferase I of *Candida albicans*: null mutants or enzyme inhibitors produce unexpected phenotypes.** *J Bacteriol* 2000, **182**:704-713.
2. Berndt N, Hamilton AD, Sebti SM: **Targeting protein prenylation for cancer therapy.** *Nature Reviews Cancer* 2011, **11**:775-791.
3. Wong NS, Morse MA: **Lonafarnib for cancer and progeria.** *Expert Opin Investig Drugs* 2012, **21**:1043-1055.
4. Garcia-Ruiz C, Morales A, Fernandez-Checa JC: **Statins and protein prenylation in cancer cell biology and therapy.** *Anticancer Agents Med Chem* 2012, **12**:303-315.
5. Long SB, Casey PJ, Beese LS: **Reaction path of protein farnesyltransferase at atomic resolution.** *Nature* 2002, **419**:645-650.
6. Hast MA, Beese LS: **Structure of Protein Geranylgeranyltransferase-I from the Human Pathogen *Candida albicans* Complexed with a Lipid Substrate.** *Journal of Biological Chemistry* 2008, **283**:31933-31940.
7. Hougland JL, Hicks KA, Hartman HL, Kelly RA, Watt TJ, Fierke CA: **Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities.** *J Mol Biol* 2010, **395**:176-190.
8. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F: **Towards complete sets of farnesylated and geranylgeranylated proteins.** *PLoS Comput Biol* 2007, **3**:e66.
9. Hougland JL, Lamphear CL, Scott SA, Gibbs RA, Fierke CA: **Context-dependent substrate recognition by protein farnesyltransferase.** *Biochemistry* 2009, **48**:1691-1701.
10. Meanwell NA: **Synopsis of some recent tactical application of bioisosteres in drug design.** *J Med Chem* 2011, **54**:2529-2591.